A critical review of 10 years of Privacy Technology

George Danezis* Seda Gürses†

August 12, 2010

Abstract

This paper is still being revised. It will eventually appear in the proceedings of the 2010 surveillance studies conference. It can then be cited as follows: George Danezis and Seda Gürses, A critical review of 10 years of Privacy Technology, In the Proceedings of Surveillance Cultures: A Global Surveillance Society?, April 2010, UK.

Since 2000 there has been a renewed interest amongst computer scientists in the field of "privacy technology". This includes mechanisms for "anonymous" communications, censorship resistance, selective disclosure credentials, as well as privacy in databases - all of which are meant to shield the user from some aspects of on-line surveillance. Beyond the lab, some of those systems have been deployed and are widely used today.

Yet, the type of surveillance against which privacy technologies are supposed to offer protection is often ill-defined, and widely varying between works: from an individual who wishes "to hide an occasional purchase from his spouse", to "groups coordinating political dissent under totalitarian regimes". While privacy is seen as the key unifying theme of these works only one aspect of it is systematically represented, namely "confidentiality". Privacy as self-definition, informational self-determination or as a public good that needs to be negotiated is often neglected. Further, the increasing omni-presence of surveillance technologies, the informatisation of every day life, as well as active resistance to on-line surveillance are used as justifying departure points for privacy technologies but they have so far not been explored in depth in the privacy research field.

In this paper, we explore the development of contemporary privacy technologies, its key results and methodologies. At its heart our argument is that the field of privacy technology was seeded by computer security and cryptography experts that rushed to apply their tools to new problems, yielding mixed results. Additional pressures from different stakeholders to devise technology that will make large IT systems acceptable to the public, has led to further confusion about the goals and methods most appropriate to embed privacy friendly values into computer systems. Using concrete examples, we seek to explain why some paradigms came to dominate the field, their advantages, but also their blind spots, and unfulfilled promises. From the results of the analysis we expect to infer new requirements for future privacy research.

1 Introduction

The technical field of privacy technology can be traced back to the late 70s to 1981s, when David Chaum, proposed in his seminal paper the first method for supporting anonymous communications, over networks where all communications are under surveillance [13]. This early work was technically enabled by the then recent invention of public key cryptography. Over the following decade, the work was influential but the community researching privacy technologies remained small (prominently counting Andreas Pfitzmann, Birgit Pfitzmann and Michael Waidner), and focused around the topic of confidentiality in communications within the wider blooming cryptograhic community. By confidentiality in communications they aimed for both confidentiality of the communication content, but also keeping the participants and the time of a communication secret from unauthorized others. Chaum himself was at the time integral to the community, and in setting up the International Association for Cryptologic Research (IARC) that manages the most highly regarded conferences in the field (CRYPTO and EUROCRYPT).

In the mid 80s the idea of combining authenticity and privacy was proposed again by David Chaum [15]. Over the 90s, alongside the internet boom, this was to become the second dominant paradigm in the small, but growing privacy technology community. David Chaum developed blind signatures to support anonymous electronic cash [14], Stefan Brands developed efficient schemes for single show selective

*Microsoft Research, Cambridge, U.K. gdane@microsoft.com

†ESAT/COSIC and IBBT, K.U. Leuven, Heverlee, Belgium. sguerses@esat.kuleuven.be

disclosure credentials [7], and Camenisch significantly extended these ideas to build multi-show credentials [11]. These selective disclosure credentials allowed users to prove that they fulfill certain conditions for accessing systems while revealing minimal information, e.g., only revealing that the user is of a given age without revealing her birthday or any other information about her. These techniques were at their very heart cryptographic: they made use of cutting edge cryptographic primitives, such as zero knowledge proofs, and pushed the cryptographic state-of-the-art to accommodate seemingly contradictory security properties, namely anonymity without the ability to forge credentials.

The years 2000 to 2010 were instrumental in privacy technology becoming a larger established field within computer security. Two key venues were established to publish work exclusively on privacy technology, the Privacy Enhancing Technologies Workshop (PET, later to become a Symposium) and the ACM Workshop on Privacy in the Electronic Society (WPES). The community considerably widened and changed to include a number of people from computer systems, networking, theory and usability background. The dominance of cryptography, which was by then well developed, lessened, and privacy technology work adopted a more general computer security character, considering the privacy properties of whole system beyond their mathematical core.

This work examines those last tens years of privacy technology developed by the PETs community. We aim to provide a road-map to these privacy technologies, an overview of their maturity, and their caveats. Yet our aim is to explore the extent to which influences other than the actual end-user requirements were responsible for shaping them as they are today, as well as the resulting privacy properties. These include the established methodologies of cryptographers and computer security researchers, the attempt to integrate them within government mandated data protection frameworks, or the will to use them to make technologies that are intrinsically intrusive, from both privacy and surveillance perspectives, acceptable to the public.

Throughout this work we present a wide selection of technologies, but we dive deeper into two of them to illustrate our arguments. These are current proposals and models for *anonymous communications* (that we define later – as the elusiveness of definition itself is part of our argument), and *identity management systems* underpinning the current thinking on how to combine authentication and respect for privacy, as it is defined by data protection. Within those families of systems a variety of solutions exist, making different assumptions and allowing different functionality, and the choice of one over the other has significant political and social consequences, as we shall see.

Finally, while our review is critical, and attempts to uncover the blind spots of privacy technology, it does not place us above the rest of the privacy technology community and many of our criticisms apply to our own work as much as the work of others. The conceptual problems and biases we describe embody a level of genuine intellectual difficulty – such as the concept of power and identity, accountability and limited identification, models of social cohesion for trusted infrastructures, the rules that govern the state's legitimacy to conduct surveillance, etc.

No systematic attempt has been made so far to confront privacy technology with the fact that these problems lie at the core of the proposed technologies, beyond attempts to argue that the whole field is bound to failure [60]. Yet, the observed success of even imperfect privacy technology solutions, largely invisible to the public, have made such arguments mildly redundant. Such arguments tend to be human-centric, emphasizing how humans (or organizations) make sense of and interact with technology in various circumstances, minimizing the role of technology itself. We agree with [52] that discussions of privacy technologies that dismiss any agency to the proposed solutions disregard how surveillance and privacy debates are constitutively entangled with the underlying technologies, and vice versa. In this work in progress, we map out examples of this entanglement, in the hope to open an academic debate on the future directions of privacy technology, the assumptions to be made, and the privacy properties of systems that are and will be actually developed by researchers and valued by technology users. The debate needs to be robust and ambitious, and not shy away from uneasy questions, including "is privacy even the right point of departure to think about developing counter-surveillance technologies?"

2 A brief review of current PETs

We start our discussion with a brief overview of Privacy Technologies developed or refined over the last 10 years.

- **Privacy Enhancing Technologies (PETs):**

  Communications Anonymity and Anonymizers: These systems "delink" the identity of an actor from the traces of their activities in information systems. Anonymity is achieved when

an individual is not identifiable within a limited set of users, called the anonymity set [54]. On the Internet, this consists of developing systems that protect the IP address or other network addresses of a user from the communication partners using proxies, or, in further varieties, even from the anonymity service provider itself. Latter systems are variations on Chaum's Mixes [13]. Research in the field focuses on anonymous communication at the communication layer (e.g., [24, 62]), traffic analysis (e.g., [38, 59, 23]), and anonymous publishing (sometimes also called 'censorship resistance', Freenet being one of the more popular applications [16]). The current state-of-the-art system, Tor, counts thousands of volunteer relays and hundreds of thousands of users, and focuses a lot of research effort on how to scale the system beyond this. This success is a clear illustration that the "end of privacy" was overstated [60], and PETS, even with their weaknesses can be popular technologies.

Identity Management Systems (IDMS): allow individuals to establish and secure identities, describe those identities using attributes, follow the activities of their identities, and delete identities. The research in this field is concerned with developing cryptographic schemes and protocols used to enable anonymous or pseudonymous credentials [15, 10] that are ideally used in combination with anonymity services discussed earlier. Further, they depend on policies that define access control rules with respect to revealed information (e.g.,[3, 18]). In practice, developing such user-centric systems demands the definition and evaluation of appropriate design heuristics for such systems [5, 33, 35] and in general studies on conceptions of (digital) identity (e.g., [2]). These are mature technologies, and some of their technical core is currently being included in commercial projects.

Privacy Policy Languages and Policy Negotiation: the objective of this research is to develop platform independent privacy policy languages that allow users and organizations to express the privacy controls that they desire. Using such policy languages service providers can encode their data-collection and data-use practices in a machine-readable XML format (e.g., using the Platform for Privacy Preferences (P3P) Privacy Policies [67]). These are then compared and matched with user policies which state the preferences of a user in a set of preference-rules, interpreted by their user agents to make automated or semi-automated decisions regarding the acceptability of a given privacy policy [41]. Later privacy policy models have enabled the creation of access control policies that are then enforced by the service providers (e.g., Enterprise Privacy Authorization Language (EPAL) [58] for representing authorization and entitlement policies). Such technologies aim to enable distributed systems to enforce a common privacy policy, and users being able to discover and reason about the information needed to access a service or leaked as part of accessing a service. Some approaches go a bit further to allow users to negotiate their preferences privately to access those services. Such approaches have seen only a limited deployment, with P3P being present in some browsers, but hardly used by on-line services.

- **Privacy Preserving Data Publishing and Mining:**

Tabular Data Analysis: the objective of Privacy Preserving Data Publishing (PPDP) methods is to provide the analysts of personal information databases (microdata) with the ability to analyze and infer certain information from the database, while forestalling the inference of certain other information (information that could lead to privacy breaches). Privacy is guaranteed through suppressing or generalizing some of the attributes (called quasi identifiers), so that they cannot be used to identify individuals uniquely or in a small set of data records, or link individuals to a subset of their attributes (e.g.,[61, 39, 43, 57]). A similar line of research concentrates on avoiding unwanted inferences from statistical data, known as statistical data control [46]. Privacy Preserving Data Mining (PPDM) is concerned with randomization and perturbation methods, as well as the use of cryptography, allowing data mining on a modified version of the data that contains no sensitive information [65]. PPDM in distributed models focuses on secure mutliparty computation models, so that multiple parties can carry out distributed computing tasks without revealing their data sets to the other parties [69]. PPDP focuses on (micro)data, while PPDM focuses on the data mining results.

Differential Privacy: An alternative to PPDP is possible in interactive systems with a secure query interface. The objective of differential privacy is to publish data such that the probability of a privacy breach occurring is similar whether or not that person's information is contained in the data. Hence, no additional harm is done by releasing additional records in response to

a query [27]. Intuitively, differential privacy guarantees that a querying observer cannot tell if a given individual is in the set of microdata that is being analyzed or not.

Social Network Data Analysis: network or graph data cannot be anonymized using the same methods for anoymizing microdata, as the structure of the data reveals further information and some information is multifaceted (a single record in the network may affect the privacy of many entities in the network [19]). Hence, naive graph anonymization techniques are subject to local [36], global [49] and injection attacks [26]. Techniques have been developed to avoid re-identification and edge disclosure. In parallel to microdata publishing algorithms, these include alteration based approaches [45] achieved through adding edges to the graph to reach k-anonymity among nodes or subgraphs, generalization based approaches [36] that obscure local details while keeping global properties; ramdomization based approaches that change the graph randomly [56]; and, publishing the network interactively using differential privacy [26]. Further methods have been concerned with constructing the network privately using a mix of access control models and cryptography [12, 30], protecting the graph from the service provider or data publisher itself as well as potential inferences made by other (colluding) users.

Privacy Preserving Recommender systems: Looks at ways of providing privacy from centralized recommender system providers, as well as the privacy of individual users from other users of the system. Recently, new directions of research consider integrating recommender systems and differential privacy [47].

3 Privacy as Confidentiality

In one of its historical moments, privacy has been defined as "the right to be let alone" [66]. Although originally formulated by legal scholars as a right that protects individuals against gossip and slander, this construct has since then acquired a wider meaning. Namely, it refers to an individualistic liberal tradition in which an intrinsic pre-existing self is granted a sphere of autonomy free from intrusions from both an overbearing state and the pressure of social norms [55].

This definition has also been popularly used by some of the privacy researchers in computer science and has been interpreted as an autonomous (digital) sphere in which the data about persons is protected so that unauthorized others cannot access it, also known as data confidentiality. Privacy is hence defined as avoiding making personal information accessible to a greater public. If the personal data becomes public, privacy is lost.

As we shortly introduced above, numerous privacy technologies are concerned with data confidentiality. The main objective of such technologies is to enable the use of information based services while either minimizing the collected information, anonymizing the collected information, or securing the collected information from unauthorized access.

Data confidentiality:

Once data about a person exists in a digital form, it is very difficult to provide individuals with any guarantees on the control of that data. In order to keep data private, in other words confidential from a greater public, various cryptographic building blocks can be utilized to achieve a number of properties. These building blocks can be used to achieve system properties like unlinkability, undetectability, unobservability, and communications content confidentiality. Various formal definitions of these properties exist. According to [54] unlinkability between two information items holds when an observer of the system cannot distinguish if the two information items (in a system) are related or not. Undetectability of an information item of interest is guaranteed when the attacker cannot sufficiently distinguish whether the information item exists or not. Unobservability of an information item of interest is guaranteed when both the undetectability of the item against all subjects uninvolved, and the anonymity of the subjects, even against other subjects involved in the information item of interest, hold. Different metrics can be used to quantify the degree of linkability, undetectability and unobservability that an observer identifies after her observation of the system given her a-priori knowledge. Communications content confidentiality is guaranteed through encryption of the content, where the guarantees are based on computational metrics.

With respect to the conditions of the communication further confidentiality properties can be achieved which can be guaranteed through anonymity. Unlinkability can also be seen as a generalization of anonymity e.g., sender anonymity is when a message and the sender cannot be linked. We will see later that most of the building blocks that we categorize under data confidentiality are also used in identity management systems.

The past few year have seen a radical paradigm shift in data anonymization research. First of all a number of very powerful and general de-anonymization attacks [48] have been demonstrated on real world datasets, that cast serious doubt on whether rich data-sets containing relational data between users (such as ratings of movies) can ever be made safe to release. This, in effect shatters the promises of previous work (such as k-anonymity) that technical solutions will allow both privacy and useful public data-sets. The second line of work on differential privacy sets the problem of privacy for published statistics on a firm theoretical basis, and shows how much perturbation is necessary to prevent inferences about individuals. While this is a positive result the mechanisms that are needed to provide these guarantees differ significantly from current (un-safe) practice.

Anonymity in communications: In communications, anonymity is achieved when an individual is not identifiable within a limited set of users, called the anonymity set [54]. An individual carries out a transaction anonymously if she cannot be distinguished by an observer from others in that set. The observer, often also called the adversary, may obtain some additional information [22]. This means that the observer captures probabilistic information about the likelihood of different subjects having carried out a given transaction. The observing party may be the service provider or some other party with observation capabilities or with the ability to actively manipulate messages. Depending on the observer's capabilities, different models can be constructed with varying degrees of anonymity for the given anonymity set. The degrees of anonymity is calculated using metrics, the most popular of which are entropy based metrics. Exactly what degree of anonymity is sufficient in a given context is dependent on legal and social consequences of a data breach and is an open question [22].

Communications anonymity keeps the identity of the persons in information systems confidential but is not necessarily concerned with how public the traces subsequently become. This is also reflected in data protection legislation which by definition cannot and does not protect anonymous data [32]. Technically, applying data protection to anonymous communications would paradoxically require that an identifier is left behind. Leaving identifiers behind are likely to conflict with the desire to communicate anonymously. This is a fundamental tension between the requirements of Data Protection regimes, and the goals of most common anonymizers as well as the desires of users, that is not resolved and hardly discussed: in order to implement an effective and comprehensive data protection regime, we fist have to implement the most extensive surveillance and tracking infrastructure.

As for data anonymization some very important impossibility results have also been shown for communications anonymity. Any systems that only protects users within anonymity sets that are changing will in the long run leak the long term communication partners of users. In other words communications anonymity is at best a tactical protection, as long term profiles will be extracted. The full impact of this result has not yet been fully digested by the research community: many systems prefer to ignore this reality in their security analysis, while other systems further weaken their security models on the basis that in the long run all anonymity is lost. It is a multi-disciplinary problem what types of communications anonymity are needed for different activities, and whether tactical short term anonymity is a useful property at all. Conversely, it is questionable (but hardly questioned) whether long term profiles or identities even exist: could it be the case that a users preferences, profiles and identities change faster than the rate at which they can be uncovered?

Distrbuted Architecture, data minimization and confidentiality: Another confidentiality approach depends on the underlying architecture of the system-to-be to guarantee information confidentiality. In a distributed system, where the personal data is collected through distributed clients (or devices) various steps can be taken to minimize the collection of information and to obtain data confidentiality. For example, in a distributed system like a car toll system, devices can be embedded in vehicles that track the vehicles' use of toll roads. This can be done such that the personal data collected by the embedded devices are minimized [4]; the surveillance information is aggregated before it is sent to a central server so that detailed location information collection is avoided; and, the communication between the devices and centralized system are secured in order to avoid unauthorized access and unwanted inferences from communication patterns.

Anonymity in databases and networks: One difference to communications anonymity is that PPDP methods aspire to protect the utility of the anonymized surveillance information for data analysts. Hence, in PPDP models the database or service provider, for example an SNS, is trusted with all the data. Guaranteeing anonymity is a requirement when the (SNS) database has to be analyzed (e.g. data-mined), especially so when this is done by third parties.

PPDP research on relational databases as well as network data has shown that not only simple anonymization techniques but that most existing anonymization techniques do not work, because ill-meaning or even unsuspecting analysts (also called "adversaries") may, through additional information,

like the network structure in a social network, recover the supposedly-unlinked identities and/or find more information about these data subjects [61, 39, 43, 57, 25]. Differential privacy may offer some alternatives, although it is a very specific way of publishing data i.e., interactively and the usability and feasibility of applying this new method is a topic of future research [27].

3.1 Potentials and limitations of confidentiality mechanisms:

We have discussed three types of privacy solutions that rely on the assumption that keeping data confidential, minimizing the collection of data, and de-identifying collected data protects individual privacy. By arguing so, these solutions assume that certain information practices, i.e., confidentiality and data collection minimization, are better for privacy than others. For example, communications anonymity keeps the identity of the persons in information systems confidential but is not necessarily concerned with how public the traces subsequently become. Hence, the solutions assume that unlinking the (unidentifiable) traces an individual leaves behind is a desirable and sufficient protection of the communications of that individual. Hence, communications anonymity privileges protection through the unlinkability of individuals to their traces, over the visibility of that link protected through additional technical and legal measures. There are legal, technical and social problems with this position.

The legal problems arise from the fact that data protection applies only to a subset of collected information, namely 'personal information'. At first sight, data protection legislation echoes the assumption that anonymity is sufficient to protect individual privacy: by definition data protection cannot and does not protect anonymous data [32]. It only applies to personal data; or, in other words, data that can be identifiable, hence linked back to the individual. As a result, applying data protection to anonymous communications would paradoxically require that an identifier is left behind. Leaving identifiers behind are likely to conflict with the desire to communicate anonymously. This is a fundamental tension between the requirements of Data Protection regimes, the objectives of anonymizers, as well as the desires of users. This tension even manifests itself beyond applications like anonymizers: in order to implement an effective and comprehensive data protection regime, we first implement the most extensive surveillance and tracking infrastructure.

Technical problems have to do with weaknesses of anonymous communication systems with respect to long term communication, as well as the results from PPDP showing the impossibility of anonymizing databases. Systems that only protect users within anonymity sets that are changing will in the long run leak the long term communication partners of users. In other words communications anonymity is at best a tactical protection, as long term profiles will be extracted. Further technical problems arise due to the impossibility results from both PPDP and PPDM. With the increasing application of data mining to content analysis and given the ability to link information from multiple sources the likelihood of re-identification of users or the linking of messages from a single user using anonymous communications will further increase. The full impact of these results has not yet been fully digested by the research community: many systems prefer to ignore this reality in their security analysis, while other systems further weaken their security models on the basis that in the long run all anonymity is lost. It is a multi-disciplinary problem what types of communications anonymity are needed for different activities, and whether tactical short term anonymity is a useful property at all. Conversely, it is questionable (but hardly questioned) whether long term profiles or identities even exist: could it be the case that a users preferences, profiles and identities change faster than the rate at which they can be uncovered?

Privacy as confidentiality solutions are also made more vulnerable through the increasing popularity of data intensive applications like SNS. Yet, the recent popularity of The Onion Router (TOR) [63] shows that anonymity services are also indispensable. The increase in the collection of massive amounts of personal data as well as in the use of anonymity services underline the challenges to and the importance of guaranteeing privacy through confidentiality solutions. In the least, these challenges demand that researchers advance new solutions that are robust enough against the vulnerabilities that occur as a result of advances in data mining. For requirements engineering, these results mean that both PPDP results and existing anonymity metrics have to be considered when deciding on the necessary anonymity models to implement.

Moreover, from different surveillance studies perspectives, it can be questioned in the privileging of confidentiality of the link between a user and her traces is always desirable (also for privacy). [60] argues that our societies are increasingly organized as networks underpinned by digital information and communication technologies. He claims that in such networks the characteristics are determined primarily by its connections, rather than its intrinsic properties, Hence isolation, as would be argued for in the privacy as confidentiality paradigm, is an undesirable option [60]. [71] claim that PETs offer "customers"

only a false perception of autonomy. The consumer self is ontologically not distinct from its representation in the electronic market-space. The consumer categories cannot be manipulated, because the consumer is constituted by language and the language governing the electronic market space is constituted by (marketing) databases. [55] argues for the importance of anonymity for the negotiation of the public and private boundaries for social issues, but urges not to argue for conservative privacy absolutes that reinforce normative boundaries.

4 Privacy as Control

A wider notion of privacy, appearing in many legal codifications, defines privacy not only as a matter of concealment of personal information, but also as the ability to control what happens with it. One reason for this notion, which does not call for strict data parsimony, is that the revelation of data is necessary and beneficial under many circumstances – and that control may help to prevent abuses of data thus collected.

This idea is expressed in Westin's [68] definition of *(data) privacy*: "the right of the individual to decide what information about himself should be communicated to others and under what circumstances" and in the term *informational self-determination* [9]. Informational self-determination is also expressed in international guidelines for data protection such as the OECD's Guidelines on the Protection of Privacy and Transborder Flows of Personal Data [51], the Fair Information Practices (FIP) *notice, choice, access, and security* [64, 31], or the principles of the EU Data Protection Directives [28, 29]. As an example, consider the principles set up in the OECD Guidelines: collection limitation, data quality, purpose specification, use limitation, security safeguards, openness, individual participation, and accountability.

This concept of control underpins the intended privacy provided by current *identity management* (IM) designs. Identity management systems emerged in the 90s, initially without considering privacy as a core feature. The first widely deployed system was Microsoft Passport, a service which allowed users to sign-up and log in to an account once and then use this account with other cooperating services on-line. The passport system was also the earliest failure of an ambitions general purpose IM system. The system was shunned by third-parties, as it locked them in a relationship with a single identity provider, namely Microsoft. The lack of privacy, and the ability of the identity provider to observe the browsing habits of users, was only a marginal issue in this debate.

The industry's response to Passport, and its failure was to launch the Liberty Alliance [40] – a group of initially 30 companies (now about 150) that promoted open standards for *federated* identity management. The emphasis was placed on the fact that different *identity providers* can inter-operate and make claims about a data subject. These claims are relayed by the data subject to relying parties that accept those claims and grant the subject some privileges. This 3-party architecture, with the user in the middle, has since become the dominant patterns for building IM systems. Microsoft cardspace, that shipped with Vista, adopts it to implement a multi-provider IM system that has again not been a wide success. The liberty protocols, while technically influential, have also failed to establish themselves as a de-facto standard. While Liberty Alliance claims over a billion devises support their protocols even technical people would struggle to name a singe on-line credentials they have using this protocol.

Under the hood IM systems implement privacy using a varying quality of technology. The Passport system allowed the single identity provider, Microsoft, to observe all the interactions between users and relying parties. The Liberty protocols allow multiple identity providers and put the user in the middle of the interaction between them. Still, a cooperating identity provider and relying party can put their information together and infer all of a user's activities. Thus the Liberty protocols rely on these two parties not cooperating to violate privacy, hence being different, a third party not being able to gain this information, and presumably a market system that allows users to chose service providers that would respect their privacy in this way. The Microsoft Cardspace technology made similar assumptions when it was shipped in 2007, and we can see that this is the prevailing architecture for IM in the last decade.

Yet, it turns out, this is a weak privacy mechanism, and the same functionality as in Liberty and Cardspace can be implemented using *selective disclosure credentials*. Those are cryptographic constructions that allow a credential to be issued by an identity provider to a data subject, that then shows it to a relying party to gain some privileges. This is equivalent to the Liberty architecture, except that the cryptographic construction guarantees that the issuing and the showing of the credentials cannot be linked to each other – effectively guaranteeing privacy against them colluding. From a cryptographic and security engineering point of view this is the proper and most secure way of implementing identity management systems that follow the user-centric paradigm – it is simply magical that credentials can be unlinkable and yet not forgeable. They technically allow users to create and use pseudonyms that are

unlinkable to their real identity, or to only disclose some certified attributes to relying parties. Systems using selective disclosure credentials have been implemented in large scale prototypes (PRIME [34]) and are now being rolled out as part of new IM frameworks by Microsoft in Windows Identity Foundation.

But is it the case that this common IM architecture can really guarantee privacy in terms of control?

4.1 The illusion of control

The dominant IM architecture places the user in the middle, mediating the flows of certified attributes from the identity provider to the relying party. Arguably, this position is the ultimate privacy safeguard, as it puts the user firmly in control of those information flows. This is of course an illusion, and being in the middle of a sticky situation does not automatically put one in control of anything.

The inspiration for the name of the first widely deployed IM system, Passport, illustrates this insight. A citizen or subject is issued a passport that certifies their status vis-a-vis the issuing country, certifies some of their attributes, and binds them to a biometric. The passport, or identity card, is then held by a citizen and presented at the borders to cross, at police checks in the case of identity cards in some jurisdiction, or to access services in others. In all these interactions the subject is in the middle, yet none of them are particularly empowering. Instead, one may argue, one has to suffer the double humiliation of not only being subject to registration and further controls, but also being used as part of the identity machinery as a mere information channel, i.e. a carrier of the passport or ID documents.

Current implementations of IM systems are much less coercive, and give a greater illusion of control. Special interfaces are drafted that allow users to select from a variety of credentials which ones to show a relying party, and of course allow the user to not proceed with a transaction if they feel that the relying party demands too strong a form of identification. Yet, choice at this level as well as the option to abstain from a service or on-line space is only a caricature of control. To pursue our parallel with national identification schemes, it is equivalent to allowing a choice between showing a driver's licence or a passport as ID or abstaining from travelling or opening a bank account. This is often not possible.

In practice the concrete balance of power between a relying party and a data subject will dictate how much information about the one or the other is exchanged. This is hardly ever the subject of a individual negotiation in real-life (although often at the core of collective demands of social movements), and has hardly even been in the on-line world beyond well documented attempts at market segmentation or academic exercises. The whole field of *trust negotiation* has struggled to find a single application that gives the user genuine choice about the amount of information to be disclosed to a service provider. This is again interpreting privacy through the lens of the familiar bias: an individualistic approach – choice here – to privacy.

4.2 The tyranny of certified attributes

We have argued that traditionally the balance of power between the data subject and the service provider (or relying party) has dictated the amount of information the one requires from the other. Thus putting the subject "in control" of tokens certifying attributes of their identity, does not in practice give them any "informational self-determination", as they are subject to coercion to provide this information or be excluded from services (or worse liable for non-compliance). Yet, the IM architectures hardly ever fail due to this apparent lack of privacy – they fail because they ignore this original power dynamic by introducing a third party that certifies attributes, i.e. the identity provider.

The concept of an identity provider is a strange one to start with, and one has to wonder what are the concrete tasks they are entrusted with. First of all, it is questionable what elements of one's identity and what attributes of a person are amenable to certification. An identity provider may be require to certify a variety of things: that an identity is fresh, and identity is unique, that an identity is expensive to change, that a single real physical person is in control of the identity, that the person controlling the identity can be found (physically), that the person behind this identity is the same as the person behind a previous interaction, etc. All of those have some use in different security protocols, and different security mechanisms exist today for checking different properties. Having a separate party certifying an identity would have to be accompanied with a very precise description of the actual property that the certification guarantees, for a relying party to be able to use it at all, let alone rely on it to secure anything.

The fact that the identity provider is certifying attributes on which the relying party makes decisions requires them to share a common ontology of attributes. This is in practice very difficult: when looking closely there are few 'facts' of a person's identity that are self-explanatory and universally understood. The name of a person is an obvious attribute: it can change, it can be the same as many other people's names, and is difficult to match for equality due to potentially different equivalent spellings. Age, or

date of birth, is another obvious general purpose attribute that IM systems refer to: yet out of context it is of little use. It has to be compared with a complex set of rules to act as an attribute that are also often determined by the identity provider: is the person allowed to drink, drive, have sex? Age by itself cannot answer that question. Beyond those attributes IM systems can certify either further demographic information or certify system specific attributes. Demographic attributes – gender, place of birth, religion – are of questionable use for security decisions beyond discrimination, that is often illegal. System level attributes – unique passport ID, login name, account balance on a date, class of service – are so system specific that they deny the utility of a general purpose identity infrastructure. There are intimately tied in with the system of the relying party, and have no meaning outside it. Once the ontology is only used by a single relying party, the triangle becomes meaningless, and the identity provider might as well not be there at all and folded into the service provider.

Finally, the separation of the relying party and the identity provider fundamentally shifts balances of power. In the two party relationship between the user and the service provider, we already argued, the provider is effectively in control. With the introduction of the identity provider, this balance of power shifts, and rests with the identity provider or providers. This can be understood in two ways: the relying party is in effect making use of the certified attributes to make security decisions. If the identity provider makes false claims about an identity, then the security of the relying party can be compromised – in technical terms the identity provider is withing the trusted computing base of the relying party. Secondly, the identity provider makes judgements that cannot be ultimately tested by the relying party – if there was an objective truth about the certified attributes there would be no need for a third party certifying them. By entrusting the identity provider to make these judgements the relying parties accept the ontologies and semantics the identity providers use to interpret attributes – which is itself an exercise in power.

4.3 The temptation of escrow

As we have seen the triangle of service provider, subject and identity provider is by far not a "natural" one, even if the user is in the middle of it. So why is this architecture so favoured within e-government circles? This uneasy triangle introduces mediation: tokens provided by the identity provider, mediate all transactions between subject and service providers – and this mediation as we have seen is great for surveillance.

Privacy friendly identity management systems are built using selective disclosure credentials that make "identity provision" and "identity use" unlinkable. Researchers in this field are very attached to this notion and have proven that schemes can be build that are information theoretically secure, preventing an adversary with even infinite computational power to break the privacy of the schemes. Yet, over the past 10 year, and particularly within the European research agenda it has become accepted that such scheme should be made intentionally weaker to allow a "trusted third party" to de-anonymise transactions if there is ever such a need (and against dire warning against this practice from earlier work by Brands [8]). This practice is veiled under civilised and attractive names such as "accountability" or more honestly "traceability", while its opponents denounce it as "escrow" (alluding to "key escrow" that fuelled the crypto-wars in the 1990') or simply "surveillance by design".

Two aspects of turning IMS systems into surveillance infrastructures are noticeable: the fact that they are outsourced and the fact that their architecture and rationale makes escrow fatally attractive.

First of all it is interesting that proposed mechanisms for identity escrow do not allow the identity provider to initiate the tracing of a transaction of link them to uses of the identity, but instead only allows, usually unspecified, third parties. Ideally those third parties are judicial authorities, potentially they are law-enforcement, and in practice they will be intelligence agencies that are commonly entrusted with holding cryptographic material on behalf of governments. So, if only those actors can convince service providers that independent identity provision is economically preferable, and if only identity providers can be sustained by a business model, they will find themselves with a distributed surveillance infrastructure that can be turned on-and-off at will, under the veil of the highest standard of privacy technology. Furthermore, this surveillance infrastructure is self-funded and ubiquitous.

Escrow mechanisms are orthogonal to selective disclosure credentials, and one could envisage an IM infrastructure free from surveillance by design. So, why is it that the latest projects include an escrow component? The first explanation might be simple: their funding comes from e-govenment projects, and the funders can see the value of surveillance. Yet, this explanation is only partial, as there is a genuine feeling, even amongst some privacy researchers that a general purpose escrow mechanism is needed for large IM systems. Full identity escrow is inevitable due to the general and open ended nature of the

IMS systems themselves: by definition the identity providers are separated from service provision, and therefore do not know and cannot possibly predict what the credentials are going to be used for. This open-endedness makes it impossible to predict the effectiveness of any custom-built abuse prevention measures: for every abuse prevention measure short of revealing the full identity of the user, a fictitious abuse scenario within a fictitious service can be constructed to assert it would not be sufficient to avert a drama.

Practical custom built privacy preserving mechanisms have been successfully proposed for specific applications of credentials: double-spending prevention for e-cash, n-periodic spending for tickets, black-listing of users (without revealing their identities) for forum abuse, and reputation systems to prevent spam. Given a specific security goal of a system the privacy research community has systematically provided specific abuse prevention mechanisms that preserve privacy. Yet, it is impossible to build those for the open ended scenarios for which IM systems are destined, making an on-off privacy switch as the only acceptable mechanism.

4.4 Breaking up the identity management menage-a-trois

It is within that IM context that selective disclosure credentials are meant to provide privacy: a user that is coerced into participating into a system for lack of choice or even by dictate – for e-governement architectures; a relying party that is losing control over part of its security infrastructure and outsources security decisions to an identity provider that may not share the same ontology or even concept of identity with it; finally, an identity provider that certifies attributes that are either meaningless for most security decisions (demographic) or so tied into a system that would be better checked by the relying party; over all this looms the spectre of a surveillance entity that can, at the flick of a switch, de-anonymize any or all transactions. While perfect privacy and unlinkability can be proved beyond doubt for the cryptographic constructions the idea of privacy as informational self-emancipation is far from being realised. How can privacy technology disentangle itself from this situation?

The first hopeful direction is to accept that the three parties will in practice have to be reduced to two, by leaving out the separate identity provider. This can be done in two ways, depending on the application:

1. The subject or user of a service can self-certify their identity and attributes. This is in practice the current model in many "free" on-line services (Hotmail, gmail, facebook, flickr, ...). The service provider just has to rely on the information provided, even though it might be potentially false. For some important aspects of identity this is fully sufficient: the OpenID mechanism, as well as self-issued credentials, can ensure for example that two otherwise anonymous transactions are performed by the same person.

2. In the second case, the service provider acts as the identity provider. This is a common scenario in workplaces or educational establishments that also run services. Those places use off-line methods to establish who is who, and then distribute credentials to authorised users to access services. Currently these credentials are privacy invasive and link transactions to a single identity. A promising avenue for identity management would look at how selective disclosure credentials can be used to make this process more privacy friendly. This solutions avoid the ontological gap that is introduced by separating service and identity provider at the detriment of supposed cost as each provider has to replicate an identity infrastructure – this has so far been the only systematically successful strategy, and for good reason as we have seen.

A third strategy is just emerging from the field of PETS about how to establish identity and certify attributes, potentially in as privacy friendly manner. It consists of crowd sourcing both identity provision as well as attribute certification: this involves users certifying each others identities, and potentially attributes. This has been deployed by google that customarily requires an invitation to create accounts on services, as well as a way to bootstrap "reputation" in on-line services such as `couchsurfing.com`. More theoretical research shows how a friend-of-a-friend approach can be used to solve key problems in identity: how to ensure identities are not fake and aiming to flood a system [70, 21], or how to recover lost or stolen passwords [6]. Making those distributed approaches to identity privacy friendly could be the key research challenge of the next decade.

When it comes to giving users real control solutions might escape the realm of identity management, something that will be hard for researchers to accept, as well as a realisation that goes counter to established business models. Real user control is probably best served, not by allowing multiple providers, but by allowing user data to be fully under the control of the user: this allows easy and cheap migration

between services, as well as local processing. The data protection regimes, as well as many PETS that aim to support them, are poor substitute for this simple architecture: users, and software under their control, should be able at all times to access fully, process, copy, and delete information that is held on their behalf to provide a service.

Finally, the limitations of IM systems have to be recognised when it comes to privacy, particularly when the academic community engages in conversations with policy makers. IM Systems cannot offer any protection to the data once it is in the possession of third party service providers. Control in this case is reduced to users being able to chose which entity can violate their privacy though malice or carelessness. The proliferation of service providers, and the ephemeral nature of the interactions with many of them, makes it a management nightmare for users to keep track of where their information is being processed, and IM systems offer no solution to this. The inherent difficulties in anonymizing fully complex interactions makes information help in user profiles sensitive. While IM systems might help keep track of authentication relationships actual user data stored in services is not mediated through them, and cannot be easily audited. Thus, even when it comes to enforcing strict data protection requirements, IM systems are only a small part of the solution.

Finally, IM systems do not recognise how people actually construct and play with their fluid identities when interacting with each other and private or government services. The current, monolithic, approach to identity management presents inherent limitations in that respect.

5 Privacy as Practice?

The privacy as practice paradigm mechanisms are an even more recent development in the research field of privacy technologies. These mechanisms are about making it possible to intervene in the flows of existing data and the re-negotiating of boundaries with respect to collected data by making transparent the way in which information is collected, aggregated into data sets, analyzed and used for decision making. These two activities rest on, but extend the idea of privacy as informational self-determination in that they demand transparency with respect to aggregated data sets and the analysis methods and decisions applied to them. In this sense, these approaches define privacy not only as a right, but also as a public good [37]. They also make use of the definition of privacy as the "the freedom from unreasonable constraints on the construction of one's own identity" [17]. Since these mechanisms are relatively new, it is difficult to say exactly the principles that they apply. Nevertheless, they are distinct enough to be categorized independently of the two other types of privacy mechanisms.

Palen and Dourish argue that "privacy management in everyday life involves combinations of social and technical arrangements that reflect, reproduce and engender social expectations, guide the interpretability of action, and evolve as both technologies and social practices change" [53]. Boyd and Ellison state that privacy in social networks sites is also implicated in users' ability to control impressions and manage social contexts [20].

These definitions emphasize that confidentiality and individual control are part of privacy, but not all. Privacy includes strategic concealment, but also revelation of information in different contexts, and these decisions are based on – and part of – a process of collective negotiation. Tools that support data concealment and revelation individually and collectively through feedback are hence typical for privacy as practice mechanisms. They can also be distinguished from mechanisms in the other two paradigms, since they are very much based on experience and hence incorporate changes with respect to privacy concerns over time.

As an example of feedback and awareness tools, Lederer et al.[42] suggest improving privacy sensitivity in systems through feedback that enhances users' understanding of the privacy implications of their system use. This can be coupled with control mechanisms that allow users to conduct socially meaningful actions through them. These ideas have led to suggestions like the identityMirror [44] which learn and visualize a dynamic model of user's identity and tastes.

A similar approach is suggested in the concept of privacy mirrors [50]. The authors criticize purely technical privacy preservation solutions that do not take the social and physical environments in which the technical systems are embedded into consideration. Making the data visible would make the underlying systems more understandable, enabling users to better shape those socio-technical systems, not only technically, but also socially and physically. A first implementation of a "privacy mirror" exists in Facebook through which users can set controls on their profile information and then check how their profile is seen by their friends.

6 Conclusion

We have traced the recent history of the privacy technology field, provided a necessarily abbreviated survey of key technologies, and looked in depth at anonymous communications and identity management systems and their context. In both cases we see that the otherwise broad concepts of privacy are narrowed down to very specific interpretations when translated to technology. Privacy in communications is interpreted as an individualistic need for hiding a network address, and privacy in identity management systems is interpreted as the technical ability to issue and use credentials in an unlinkable manner. Within their wider contexts privacy friendly communications only support some specific privacy needs. Privacy friendly IM systems on the other hand are embedded in a much more complex social reality, that often negates their privacy benefits or worse has the potential to turn them into a surveillance infrastructure in disguise.

When building technical systems, it is necessary to make choices about what properties they will implement, and making correct choices is part of providing value to users. We do not have a fundamental objection to this approach. Yet, narrow interpretations of privacy sometimes come to redefine what privacy means, and that can lead a whole academic community to solve specific problems while ignoring the larger picture.

Anonymous communications and selective disclosure credentials are the oldest and most mature privacy technologies. A battery of new technologies are currently being developed to address other aspects of privacy. They include privacy friendly data mining, data privacy, election systems, location privacy systems, and complex privacy preserving computations. As these become more prominent we aim to widen our analysis of how the properties they support match privacy aspirations of users in different contexts, and expect to find similar biases and narrow interpretations of privacy there.

Along with newer privacy technologies newer approaches are emerging from within the privacy technology field: transparency tools that illustrate the potential for surveillance, break the glass policies that provide auditing capabilities, the use of social networking platforms to bootstap identity, and privacy as practice. None of these is a silver bullet, and they come with their own open problems, contradictions and necessarily narrow interpretations of what privacy means – likely to fuel the field for many years to come. Yet, these approaches show that despite the dominant paradigms in privacy technology there exists an internal capacity to innovate technically as well as conceptually.

References

[1] *2008 IEEE Symposium on Security and Privacy (S&P 2008), 18-21 May 2008, Oakland, California, USA*, 2008. IEEE Computer Society.

[2] Phillip Agre. The architecture of identity: Embedding privacy in market institutions. *Information, Communication and Society*, 2(1), 1999.

[3] Claudio A. Ardagna, Jan Camenisch, Markulf Kohlweiss, Ronald Leenes, Gregory Neven, Bart Priem, Pierangela Samarati, Dieter Sommer, and Mario Verdicchio. Exploiting cryptography for privacy-enhanced access control. *Journal of Computer Security*, 18(1), 2009.

[4] Josep Balasch, Alfredo Rial, Carmela Troncoso, Christophe Geuens, Bart Preneel, and Ingrid Verbauwhede. Pretp: Privacy-preserving electronic toll pricing. Cosic internal report, 2010.

[5] Katrin Borcea-Pftizmann, Marit Hansen, K liesebach, Andreas Pfitzmann, and Sandra Steinbrecher. What user-controlled identity management should learn from communities. In *Information Security Technical Report*, volume 11, pages 119–128, 2006.

[6] John G. Brainard, Ari Juels, Ronald L. Rivest, Michael Szydlo, and Moti Yung. Fourth-factor authentication: somebody you know. In Ari Juels, Rebecca N. Wright, and Sabrina De Capitani di Vimercati, editors, *ACM Conference on Computer and Communications Security*, pages 168–178. ACM, 2006.

[7] Stefan Brands. Rapid demonstration of linear relations connected by boolean operators. In *EUROCRYPT*, pages 318–333, 1997.

[8] Stefan A. Brands. *Rethinking Public Key Infrastructures and Digital Certificates: Building in Privacy.* MIT Press, Cambridge, MA, USA, 2000. ISBN 0262024918.

[9] Bundesverfassungsgericht. BVerfGE 65, 1 – Volkszählung. Urteil des Ersten Senats vom 15. Dezember 1983 auf die mündliche Verhandlung vom 18. und 19. Oktober 1983 – 1 BvR 209, 269, 362, 420, 440, 484/83 in den Verfahren über die Verfassungsbeschwerden, 1983.

[10] Jan Camenisch and Anna Lystanskaya. Efficient non-transferable anonymous multi-show credential system with optional anonymity revocation. Technical Report RZ 3295, IBM Research, 2000 (Nr93341).

[11] Jan Camenisch and Anna Lysyanskaya. An efficient system for non-transferable anonymous credentials with optional anonymity revocation. In Birgit Pfitzmann, editor, *EUROCRYPT*, volume 2045 of *Lecture Notes in Computer Science*, pages 93–118. Springer, 2001. ISBN 3-540-42070-3.

[12] Barbara Carminati and Elena Ferrari. Privacy issues in web-based social networks. In Elena Ferrari and Francesco Bonchi (Eds.), editors, *Privacy-Aware Knowledge Discovery: Novel Applications and New Techniques,*. Chapman and Hall/CRC Data Mining and Knowledge Discovery book series, Florida, U.S.A. (In print), 2010.

[13] David Chaum. Intraceable electronic mail, return adresses and digital pseudonyms. In *Communications of the ACM*, 1981.

[14] David Chaum. Blind signature system. In *CRYPTO*, page 153, 1983.

[15] David Chaum. Showing credentials without identitfication: Signatures transferred between unconditionally unlinkable pseudonyms. In *Advances in Cryptology - EUROCRYPT 1985 Workshop on the Theory and Application of Cryptographic Techniques*, 1985.

[16] Ian Clarke, Oskar Sandberg, Brandon Wiley, and Teodore W. Hong. Freenet: A distributed anonymous information storage and retrieval system. In *Designing Privacy Enhancing Technologies*, pages 46 – 66, 2000.

[17] Richard Clarke. The digital persona and its application to data surveillance. In Phillip Agre and Marc Rotenberg, editors, *Technology and Privacy: The New Landscape*. MIT: Cambridge, Massachusetts, 1994.

[18] Sebastian Clauß, Dogan Kesdogan, Tobias Kölsch, Lexi Pimenidis, Stefan Schiffner, and Sandra Steinbrecher. Privacy enhanced identity management: Design considerations and open problems. In *ACM CCS2005 Workshop on Digital Identity Management*, 2005.

[19] Scott E. Coull, Fabian Monrose, Michael Reiter, and Michael Bailey. The challenges of effectively anonymizing network data. *Conference For Homeland Security, Cybersecurity Applications & Technology*, 0:230–236, 2009. doi: 10.1109/CATCH.2009.27. URL `http://dx.doi.org/10.1109/CATCH.2009.27`.

[20] danah boyd and Nicole Ellison. Social network sites: Definition, history and scholarship. *Journal of Computer-Mediated Communication*, 2007.

[21] George Danezis and Prateek Mittal. Sybilinfer: Detecting sybil nodes using social networks. In *NDSS*. The Internet Society, 2009.

[22] Claudia Diaz. *Anonymity and Privacy in Electronic Services*. Katholieke Universiteit Leuven, 2005.

[23] Claudia Diaz, Stefaan Seys, Joris Claessens, and Bart Preneel. Towards measuring anonymity. In *Designing Privacy Enhancing Technologies*, pages 54 – 68, 2002.

[24] Roger Dingledine, Nick Mathewson, and Paul Syverson. Tor: the second generation onion router. In *13th USENIX Security Symposium*, pages 303 – 320, 2004.

[25] J. Domingo-Ferrer and V. Torra. A critique of k-anonymity and some of its enhancements. In *Third International Conference on Availability, Reliability and Security, 2008. ARES 08.*, 2008.

[26] Cynthia Dwork:. Differential privacy. In *ICALP (2)*, pages 1–12, 2006.

[27] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *TCC*, 2006.

[28] EU. 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. *Official Journal of the European Communities*, (L. 281), November 1995.

[29] EU. Eu, directive 2002/58/ec of the european parliament and of the council concerning the processing of personal data and the protection of privacy in the electronic communications sector., 2002.

[30] Keith B. Frikken and Philippe Golle. Private social network analysis: how to assemble pieces of a graph privately. In *WPES '06: Proceedings of the 5th ACM workshop on Privacy in electronic society*, pages 89–98, New York, NY, USA, 2006. ACM. ISBN 1-59593-556-8. doi: http://doi.acm.org/10.1145/1179601.1179619.

[31] Federal Trade Commission (FTC). Privacy online: Fair information practices in the electronic marketplace: A federal trade commission report to congress, May 2000. `http://www.ftc.gov/reports/privacy2000/privacy2000.pdf`.

[32] Paolo Guarda and Nicola Zannone. Towards the development of privacy-aware systems. *Information and Software Technology*, 51(2):337 – 350, 2009.

[33] Marit Hansen. Linkage control - integrating the essence of privacy protection into identity management. In *eChallenges*, 2008.

[34] Marit Hansen, Ronald Leenes, and Jan Schallabck. Prime white paper. Technical report, European Community, May 2008. URL `https://www.prime-project.eu/prime_products/whitepaper/index_html`.

[35] Marit Hansen, Andreas Pfitzmann, and Sandra Steinbrecher. Identity management throughout one's whole life. In *Information Security Technical Report*, number 13. Elsevier, 2008.

[36] Michael Hay, Gerome Miklau, David Jensen, Don Towsley, and Philipp Weis. Resisting structural re-identification in anonymized social net- resisting structural re-identification in anonymized social networks. In *VLDB*, 2008.

[37] Mireille Hildebrandt. Profiling and the identity of the european citizen. In Mireille Hildebrandt and Serge Gutwirth, editors, *Profiling the European Citizen: Cross Disciplinary Perspectives*. Springer Science and Business Media B. V., 2008.

[38] Dogan Kesdogan and Lexi Pimenidis. The hitting set attack on anonymity protocols. In *6th Information Hiding Workshop*, pages 326 – 339, 2004.

[39] Daniel Kifer and Johannes Gehrke. l-diversity: Privacy beyond k-anonymity. In *IEEE 22nd International Conference on Data Engineering (ICDE'07*, 2006.

[40] Susan Landau, Hubert Gong, and Robin Wilton. Achieving privacy in a federated identity management system. pages 51–70, 2009. doi: http://dx.doi.org/10.1007/978-3-642-03549-4_4.

[41] M. Langheinrich. A P3P Preference Exchange Language (APPEL), 2001. W3C Working Draft. 26 February 2001, `http://www.w3.org/TR/P3P-preferences`.

[42] Scott Lederer, Jason I. Hong, Anind K. Dey, and James A. Landay. Personal privacy through understanding and personal privacy through understanding and action: Five pitfalls for designers. *Personal Ubiquitous Computing*, 8(6):440–454, 2004.

[43] Ninghui Li and Tiancheng Li. t-closeness: Privacy beyond k-anonymity and -diversity. In *IEEE 23rd International Conference on Data Engineering (ICDE'07*, 2007.

[44] Hugo Liu, Pattie Maes, and Glorianna Davenport. Unraveling the taste fabric of social networks. *International Journal on Semantic Web and Information Systems*, 2(1):42–71, 2006.

[45] Kun Liu and Evimaria Terzi. Towards identity anonymization on graphs. In *SIGMOD*, 2008.

[46] Paul Massell. Statistical disclosure control for tables: Determining which method to use. In *Statistics Canada International Symposium Series*, 2003.

[47] Frank McSherry and Ilya Mironov. Differentially private recommender systems: building privacy into the net. In *KDD '09: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 627–636, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-495-9. doi: http://doi.acm.org/10.1145/1557019.1557090.

[48] Arvind Narayanan and Vitaly Shmatikov. Robust de-anonymization of large sparse datasets. In *IEEE Symposium on Security and Privacy* DBL [1], pages 111–125.

[49] Arvind Narayanan and Vitaly Shmatikov. De-anonymizing social networks. In *30th IEEE Symposium on Security and Privacy, Oakland, CA*, pages 173 – 187, 2009.

[50] David H. Nguyen. Privacy mirrors: Understanding and shaping socio-technical ubiquitous computing. Technical Report, 2002.

[51] OECD. Guidelines on the protection of privacy and transborder flows of personal data., 1980.

[52] Wanda J. Orlikowski. Sociomaterial practices: Exploring technology at work. *Organization Studies*, 28, 2007.

[53] Leysia Palen and Paul Dourish. Unpacking "privacy" for a networked world. In *CHI '03*, pages 129 – 136, 2003.

[54] Andreas Pfitzmann and Marit Hansen. Anonymity, unobservability, and pseudonymity: A consolidated proposal for terminology. Technical report, Technical University, Dresden, 2008. URL `http://dud.inf.tu-dresden.de/Anon_Terminology.shtml`.

[55] David J. Phillips. Privacy policy and PETs. *New Media and Society*, 6(6):691–706, 2004.

[56] Vibhor Rastogi, Sungho Hong, and Dan Suciu. The boundary between privacy and utility in data publishing. In *VLDB*, 2007.

[57] David Rebollo-Monedero, Jordi Forné, and Josep Domingo-Ferrer. From t-closeness to pram and noise addition via information theory. In *PSD '08: Proceedings of the UNESCO Chair in data privacy international conference on Privacy in Statistical Databases*, 2008.

[58] Matthias Schunter. Enterprise privacy authorization language (epal), version 1.2. IBM, 2003. URL `http://www.w3.org/Submission/2003/SUBM-EPAL-20031110/`.

[59] Andrei Serjantov and George Danezis. Towards an information theoretic metric for anonymity. In *Designing Privacy Enhancing Technologies*, pages 41 – 53, 2002.

[60] Felix Stalder. The failure of privacy enhancing technologies (pets) and the voiding of privacy. *Sociological Research Online*, 7, 2002.

[61] Latanya Sweeney. k-anonymity: a model for protecting privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10(5):557–570, 2002.

[62] Paul Syverson, Gene Tsudik, Michael Reed, and Carl Landwehr. Towards an analysis of onion routing security. In *Designing Privacy Enhancing Technologies*, pages 96 – 114, 2000.

[63] TOR. Tor: Anonymity online, 2010. URL `http://www.torproject.org/`.

[64] Education U.S. Department of Health and Welfare (HEW). Secretary's advisory committee on automated personal data systems, records, computers, and the rights of citizens viii, 1973.

[65] Vassilios S. Verykios, Elisa Bertino, Igor Nai Fovino, Loredana Parasiliti Provenza, Yucel Saygin, and Yannis Theodoridis. State-of-the-art in privacy preserving data mining. *SIGMOD Rec*, 2004.

[66] S. Warren and L. Brandeis. The right to privacy. *Harward Law Review*, 4:193–220, 1890.

[67] Rigo Wenning and Matthias Schunter. The Platform for Privacy Preferences 1.1 (P3P1.1) Specification, 2006. W3C Working Group Note 13 November 2006, `http://www.w3.org/TR/P3P11/`.

[68] A. F. Westin. *Privacy and freedom*. Atheneum, New York, 1970.

[69] A. C. Yao. How to generate and exchange secrets. In *27. Annual IEEE Symposium on Foundations of Computer Science*, pages 162 – 167, 1986.

[70] Haifeng Yu, Phillip B. Gibbons, Michael Kaminsky, and Feng Xiao. Sybillimit: A near-optimal social network defense against sybil attacks. In *IEEE Symposium on Security and Privacy* DBL [1], pages 3–17.

[71] Detlev Zwick and Nikhilesh Dholakia. Whose identity is it anyway? comsumer representation in the age of database marketing. *Journal of MacroMarketing*, 2003.