

# Technical Aspects of Online Privacy\*

Marco Ghiglieri<sup>1</sup>, Hervais Simo<sup>1, 2</sup>, and Michael Waidner<sup>1, 2</sup>

<sup>1</sup> Technische Universität Darmstadt, FB Informatik/FG SIT

<sup>2</sup> Fraunhofer Institute for Secure Information Technology (SIT), Darmstadt

{FirstName.LastName}@sit.fraunhofer.de

**Abstract.** We discuss the privacy problems introduced by online services, in particular online social networks like *Facebook* (Section 1), summarize the state of the art in privacy technology (Section 2) and conclude with recommendations and challenges for research and development (Section 3).

## 1 The Privacy Problem

More than 80% of all Germans use the Internet and more than 30% do so over a wireless broadband connection [I11]. In 2011, roughly half of the German population actively published personal data on the Internet, predominantly through online social networks (OSN) [B11]. Arguably, all 80% who used the Internet at all left some digital traces like the IP numbers of their devices, the web pages they looked at and the terms they searched for.

The online social network *Facebook* estimates that 20-30% of all Germans are subscribers<sup>1</sup> [E12]. Worldwide *Facebook* has more than 480 million active subscribers per day, out of a total of 845 million [E12]. The company *Google* offers an even broader range of online services. Recently *Google* started linking subscriber data across all of its services, which turns them into an integrated offering.<sup>2</sup> *Google* has approx. 90% market share for Internet search in Germany.<sup>3</sup> Worldwide *Google* has over 350 million active *Gmail/Google Mail* subscribers<sup>4</sup> and over 800 million active *YouTube* users<sup>5</sup>. Worldwide and in Germany *Google* owns the by far most frequently visited web sites.<sup>6</sup>

More than half of all Germans – 55% overall, 65% for age group 14-29 – do *not* believe that their personal data are adequately protected once handed over to some online service. Consequently only 40-50% deem their online service providers trustworthy [B11].

This very moderate level of trust in the operations of online services does not impact all service providers uniformly. The online retailer and cloud services provider *Amazon* controls one of the most valuable customer databases, yet it is trusted by 77% of the German population, according to a

---

\* This work was supported by projects ECSPRIDE (BMBF) and CASED (LOEWE).

<sup>1</sup> We distinguish between *service users* and *service subscribers*. A *subscriber* is a *user* who has registered with a service, i.e., has a personal account and typically logs on to the service before using it. Many providers serve also unsubscribed users, i.e., not every *user* is a *subscriber*.

<sup>2</sup> <http://googleblog.blogspot.com/2012/01/updating-our-privacy-policies-and-terms.html>

<sup>3</sup> <http://www.spiegel.de/netzwelt/netzpolitik/0,1518,775833,00.html>

<sup>4</sup> <http://thenextweb.com/google/2012/01/19/gmail-closes-in-on-hotmail-with-350-mm-active-users/>

<sup>5</sup> [http://www.youtube.com/t/press\\_statistics](http://www.youtube.com/t/press_statistics)

<sup>6</sup> <http://www.alexa.com/>

2011 survey among 102 top brands. This actually turns *Amazon* into the most trusted brand in Germany.<sup>7,8</sup> *Facebook* is trusted by only 10%, according to the same survey.

Overall, there is agreement that online privacy is important, and also that it is at serious risk in certain contexts.

Society values privacy because it is considered a prerequisite for personal freedom, self-determination, and the protection of minorities. People need protected spaces where they can act and communicate freely and without fear of discrimination and repression. They need confidence that sensitive information does not flow in an uncontrolled fashion between social contexts [N04]. People have the fundamental right to determine their own faith and develop their own personality (GG Art. 1&2). In an online world this means they must be able to determine and control, directly or through trusted intermediaries, their online identity [B83, BPB11, W67]. The digital identity (or rather: the multiple digital identities) of a person is defined as the totality of all digital data that relates to that person [HSC08].

From an engineering perspective, there are three major problems when realizing online privacy, which we will discuss in the following three subsections.

## 1.1 Information Security

Online privacy strongly depends on information security: information must be adequately protected from criminals who want to break into the IT systems of users and online service providers, or into the communication and processing systems connecting them. Without online security, there is no online privacy.

Securing information technology is a very difficult task, as one may easily see from the seemingly never ending list of stories about identity theft on the Internet, data security incidents, and the rising number of software vulnerabilities, computer viruses, Trojan horses and other malware. In general there is no actual IT system that is absolutely and unconditionally secure. This is neither technically nor economically feasible. Securing online social networks seems to be particularly difficult, as a recent study of popular networks demonstrated [P08].

Our focus is on privacy and therefore we will not elaborate in detail on the problem of information security. We just point out that online social networks have certain characteristics which make them particularly vulnerable:

- Many online social networks use a centralized IT architecture, i.e., they have a central database of all subscribers, or they even store all data centrally (e.g., *Facebook* stores and processes all data in a datacenter in Prineville, Oregon<sup>9</sup>). This creates single points of failures, i.e., criminals just need to break into one single IT system and gain access to the complete network.
- Most online social networks offer their services globally and have their main operations outside the European Union. Almost necessarily, they store personal data of European citizens abroad. This is certainly not a security problem by itself. However, it limits the power of European privacy and data security regulations, and it may give foreign law enforcement and intelligence agencies privileged access to personal data of EU citizens [H11].

---

<sup>7</sup> Studie von *Musiol Munzinger and Sasserath*, zitiert nach <http://blog.mediaroute.de/2011/12/ranking-deutsche-vertrauen-amazon-und-nivea-und-misstrauen-olgesellschaften/>

<sup>8</sup> Similar statistics exist for the US: The *2011 Temkin Trust Rating* puts *Amazon* at #2, the *2010 Ponemon Privacy Rating* at #8.

<sup>9</sup> <http://www.wired.com/wiredenterprise/2011/12/facebook-data-center/all/1>

- Most online social networks are actually software platforms: third parties can write “apps,” i.e., programs, and offer these apps to all subscribers, paid directly or through delivering advertisements “on the side.” Typical apps are games, calendar add-ons, but also apps that provide some statistics for what a subscriber’s contacts are interested in. If a subscriber uses such an app then, in current systems, the app has either no access at all or full access to everything the subscriber has access to, e.g., the subscriber’s personal profile and whatever that subscriber may see of the profiles of his contacts. An app may also act offline on a subscriber’s behalf (e.g., automatically send a birthday card). Apps are typically server-based and thus necessarily communicate back to the third party (e.g., apps can and often also do send back the complete contact list of subscribers). There is no reason to believe that these apps are of better security quality than other software, and actually there have been many examples of poorly written or even malicious apps [EMKK11].

## 1.2 Paradigm Change: Sharing is better than Non-Sharing

Classical privacy technologies (often called Privacy Enhancing Technologies, PETs) are based on the assumption that a priori people do *not* want to share data with others. Sharing always creates a privacy risk, and accordingly sharing needs to be minimized and very tightly controlled. This applies to the people (I should not give away more data about myself than absolutely necessary) but also to the organizations processing data (they should not share more data about me than needed, and agreed by me a priori). If there are no personal data then nothing needs to be controlled and protected. This is a perfectly logical perspective, which is even encoded in European privacy legislation.<sup>10</sup>

Today’s Internet economy clearly works on the completely opposite paradigm: most people *want* to share data. Otherwise, they won’t subscribe to services that are predominantly about sharing information. Certainly most people prefer *controlled* sharing over *uncontrolled* sharing, but actually a large portion of all online data is created by individuals and released to the world without a very clear purpose.

Sharing comes with benefits (e.g., service access, gain in quality and efficiency) and costs (e.g., loss of control over personal data). Each subscriber will have to decide individually on how to balance these benefits and costs. Different subscribers with different backgrounds will decide differently: what appears to be a severe privacy invasion and breach of trust to some will be seen as a valuable service and great communication tool to others.

More than 80% of all Germans say they care about their online privacy, roughly 15% say they don’t [B11]. Similar numbers show up across all Western countries and have been fairly stable over the years [C08]. Those concerned expect adequate and convincing privacy controls. Most are unsure what to do (47% don’t know how to protect themselves [B11]). A majority expects the government to take care of the problem for them (55% want government-endorsed privacy seals, 72% want stronger regulations), or the service providers (81% want better controls). They live with the dilemma of considering the risk as too high but accepting it anyway. (All numbers are for Germany and taken from [B11].)

Technology must support the complete range of possible perspectives on privacy. Offering many choices can easily result in overly complex technical solutions, very hard to use interfaces and incomprehensible policy documents. The challenge is to keep the range of choices small and easy to understand, while at the same time accommodate the needs of the vast majority of users.

---

<sup>10</sup> Portal of the European Commission on the Protection of Personal Data: [http://ec.europa.eu/justice/data-protection/index\\_en.htm](http://ec.europa.eu/justice/data-protection/index_en.htm).

The tradeoff between privacy and sharing shows up in two versions, vis-à-vis the online service provider (Section 1.2.1) and vis-à-vis the other subscribers (Section 1.2.2).

### 1.2.1 Know and Respect Your Subscriber: The Purpose of Data

Most online service providers build up subscriber profiles and use them for customizing their service, providing personalized recommendations and, a special case, targeted advertisements. Here we assume that this is done with the explicit consent of the subscribers (just to keep the discussion focused on the technical challenges).

Profiles include essentially everything a subscriber provides explicitly when subscribing (e.g., name, payment details) and everything the service provider observes and traces back to that subscriber later on. Online retailers are likely to record shipping addresses, purchasing history, what products a subscriber searched for and how they reviewed and rated products. Search engines may store search queries and which of the results were selected. Online social networks almost necessarily store the social relationships of their subscribers, and maybe also each single step they take, including time and location. News aggregators may store what someone reads, maybe also when, where and for how long.

Subscribers can be clustered into sets of subscribers with similar profiles, which can be used to predict what a specific subscriber may search for next, read next, buy next, and also what social relationship a subscriber may want to establish next. Being able to predict “what’s next” is the key to targeted advertising, which is the main source of revenue for many online services [B08, KR11].

For instance, *Facebook* is free of charge for its 845 million subscribers, but in 2011 it generated more than 3.1B\$ revenue from advertising (plus 0.5B\$ from other services), or roughly 3.70\$ per subscriber [E12]. In the same year *Google* generated 36.5B\$ revenue from advertising (plus 2.4B\$ from other services).<sup>11</sup> The more a service knows about its subscribers the better it can predict their needs and interests, and the higher the chances are an online ad shown to a specific subscriber turns into business.

Typical benefits for subscribers are:

- A personalized service might be more convenient and efficient than a standard service.
- Personalized recommendations may accelerate identifying and getting at what one is interested in, and may avoid being overburdened with irrelevant marketing information.
- Personalized services may be free of charge for subscribers.

Typical costs for subscribers are:

- Subscribers give up control over the personal data collected by the service. Obviously the “amount” of privacy a subscriber loses depends on how they use the service – what information they provide, how complete, accurate and truthful that information is – and also on what the service provider is allowed to do with these data.
- Subscribers take the risk that sometime in the future these data may be exploited to their disadvantage. Service *providers* may change their business model or may be acquired by some other company, and as a result may want to do more with the data they have. Service *subscribers* may change their mind (e.g., because they get older, change their political perspective, change jobs, etc.) and may consider certain previously accepted uses of their data as disadvantageous. Finally, data may be stolen by someone who uses it for criminal purposes, like identity theft, facilitating social engineering, or identifying suitable targets for terrorist attacks, blackmailing or other acts of crime.

---

<sup>11</sup> <http://investor.google.com/financial/tables.html>

In between costs and benefits is discrimination, the flip side of personalization: Subscribers who are particularly lucrative for the service provider may receive better service and lower prices [A08, AV05]. This may range from special offers at online travel agencies for members of some frequent flyer programs, shorter response times from online social networks for people who click very easily on online advertisements, or lower health insurance premiums for people who appear to live a happy, healthy life.

### 1.2.2 Know and Respect Your Neighbor: The Context of Data

Most online services are “social,” which means they let their subscribers interact with each other. Interaction may mean about anything: sending electronic mail, instant messaging, discussion groups, audio and video conferences, file sharing (text, audio, video), online publication of files, broadcasting “tweets”, collaborative authoring of documents and collaborative development of programs, scheduling and planning of meetings and events. Some services specialize in just one of these services, but increasingly providers turn into wholesale providers (we already mentioned *Google's* move to link all their online services<sup>2</sup>).

Online social networks like *Facebook* assume more specifically that subscribers can describe themselves towards other subscribers, can create contact lists of subscribers with whom they want to share certain data, and can inspect their own contact lists and to a certain degree also the contact lists created by other subscribers [BE07].

This social nature of online services is the key to their success, but it also and necessarily creates the potential for misuse: There is always the risk that data released in one social context is moved – accidentally or intentionally – into another social context where it hurts the originator. Nobody wants their family secrets being broadcast on national TV. And many may feel very creepy when they read their e-mail and see online ads popping up that strongly relate to the content of their very personal messages.

These problems are not new, and like in face-to-face, physical interactions online users have to continuously balance the benefit of sharing data with the costs (risk of being betrayed or exploited, risk of hurting someone else). But the effects of making a mistake are amplified online, through the large scale and global reach of online services, the potential to link profiles across services and across social contexts, the inability of the Internet to “forget”, and the lack of reliable and scalable means for establishing trust between subscribers.

There are many examples of unintended consequences of sharing data. The obvious ones are ill-considered postings, which happen online just as in the offline world [WKG+11]. For instance, two people were denied access to the US because of inappropriate although supposedly funny remarks they posted on *Twitter* about the United State.<sup>12</sup>

A majority of HR managers – 59% in Germany, 89% in the US [C10] – search for candidates online and check what prospective new employees may have posted online, and to a lesser degree what others may have to say about them [C10, DMB11]. 16% of German and 70% of US HR managers have rejected candidates because of what they found [C10]. Presumably the best known case of such a rejection due to bad online reputation is Stacy Snyder's, a student teacher who got denied a teaching degree following the disclosure of a “drunken pirate” photo on some online social network [U08]. Of course, candidates may also turn this into an advantage by consciously designing their online representation [K11].

Online social networks are also the perfect platform for “spear phishing”, i.e., social engineering attacks targeted at specific groups (“all employees of xyz”) or individuals (“the CEO of xyz”), as well

---

<sup>12</sup> <http://www.thesun.co.uk/sol/homepage/news/4095372/Twitter-news-US-bars-friends-over-Twitter-joke.html>

as for fishing for sensitive information of enterprises and individuals, automating social engineering [LPBK10] and automating identity theft [BSBK09]. Security vendors and analysts generally agree that organized cybercrime is transitioning from mounting mass attacks (massive spam, stealing masses of credit card data) to targeted attacks using spear phishing (e.g., [B11b, C11]).

## 2 State of the Art

Figure 1 shows the standard set-up for an online social network: On the left we have users and subscribers (i.e., logged-in users) who connect to the network using a variety of devices (e.g., PC, laptop, tablet, smart phone, smart TV, car entertainment system). The network connects them with other users (top; this is the “social” aspect) and with other services (bottom; e.g., targeted advertising, business optimization, opinion research). Users access online services through some Internet Service Provider (ISP).

The ISP may pose a privacy risk just like the online social network. But it may also act as a trusted intermediary between users and service providers, helping both to improve security and privacy. Here we will not discuss these risks and potential roles of ISPs any further.

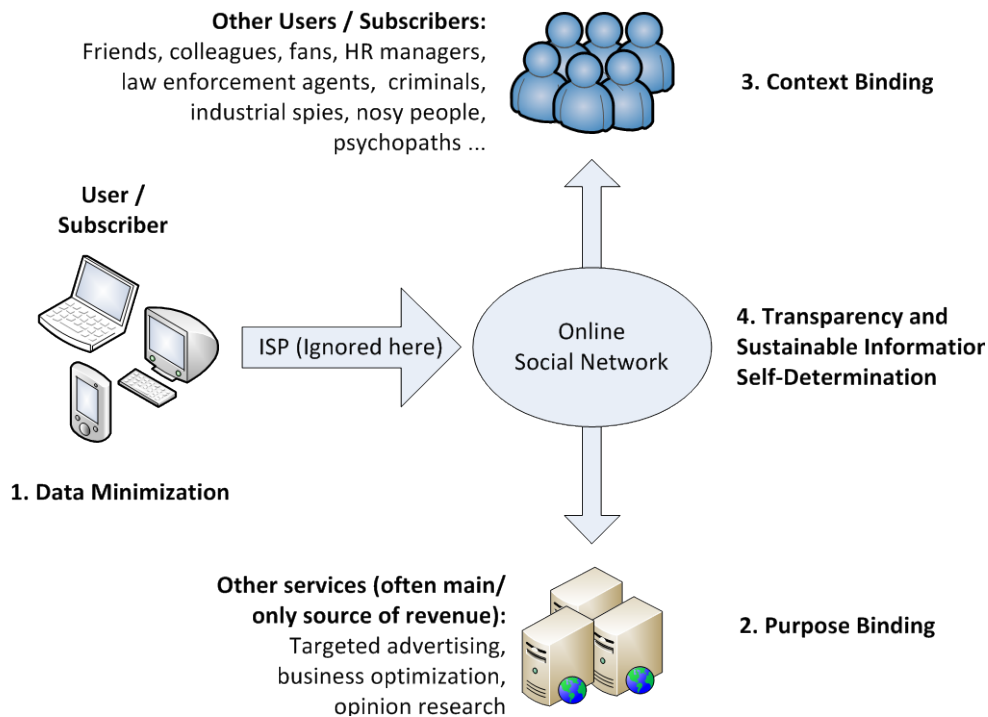


Figure 1: Four Aspects of Online Privacy

The following subsections discuss the four aspects shown in Figure 1.

### 2.1 Data Minimization

Data minimization is the most fundamental principle of privacy, as already explained at the beginning of Section 1.2. The idea is to minimize what is shared to the bare minimum deemed necessary for a specific purpose and in a specific context.

### 2.1.1 Pseudonyms

Online services frequently ask users to identify themselves, e.g., give their name and contact details. Very often a user could delay or completely avoid identifying or could act under a pseudonym instead of a real name. For instance, reading a blog should not require identification, and commenting on a blog posting of someone else could be done anonymously or under a pseudonym.

Many online services support pseudonyms. Some do so directly, often using a two-step approach: they ask for a real name during subscription (e.g., because they feel legally obliged, or they need to charge fees to a credit card), but let the subscriber choose a pseudonym for all interactions with other subscribers. Interestingly, a recent study showed that specifically for discussion groups this yields contributions of higher quality than if subscribers use their real name towards the other subscribers.<sup>13</sup>

Most online services support pseudonyms indirectly, by allowing for arbitrary user names, or by using very weak identity verification methods. Such weak identity verification techniques just aim at preventing that one person registers more than once, which is likely sufficient for achieving a level of data quality good enough for targeted advertising and for preventing the automated creation of fake accounts [D02].

Few online services prohibit the use of pseudonyms, at least for some services. For instance, *Facebook* prohibits pseudonyms. *Amazon* enables customers to write product reviews using a pseudonym but supports also verified real names as an option.

A technically somewhat more complex but also more powerful approach for avoiding identification is replacing names in access control decisions by other attributes: If you have a paid subscription to a newspaper, you are probably required to log into the newspaper website with your name and password, i.e., you identify yourself. Logically this is too much, it would be sufficient to just prove somehow that you have a paid subscription – you could stay anonymous among all other online subscribers. It would even be sufficient to just pay electronically for the one newspaper you want to read, using some kind of anonymous electronic cash, and you could stay anonymous among all users of the electronic cash service.

Technically, this is perfectly feasible. Most modern access control systems support attribute-based access control (also called claims-based access control), based on the *XACML* language for expressing access control policies [M05]. It is also possible for a subscriber to just prove that he or she has certain attributes, without revealing anything more than that. Assuming trust in the third parties vouching for attributes this can be done using the *SAML* industry standard and commercial products [R08], or the very popular *OpenID*<sup>14</sup> standard.

Using some sophisticated but perfectly practical crypto techniques it is even possible to do this without revealing the link between two transactions to anybody: I can prove twice that I have a newspaper subscription, and the publisher won't be able to realize that it is me again, nor will I be able to hand over my subscription to someone else without being identified – then with my real name – as a cheater [B00, CL01, CFR11 Ch. 5]. Based on these technologies one can essentially create a full-blown electronic commerce system where people can buy and sell electronic goods, and nobody needs to ever identify themselves or reveal more than the attributes necessary [C85, PWP00]. Unfortunately, none of the major technology vendors are supporting this approach in their mainline products. *Microsoft* and *IBM* both developed such technologies (based on [B00] and [CL01], respectively) but neither offers it as supported products.

Usability is one of the major challenges of any pseudonym system. There is some agreement across research and industry that the best approach for managing pseudonyms is through information

---

<sup>13</sup> <http://disqus.com/research/pseudonyms>

<sup>14</sup> <http://openid.net/developers/specs/>

cards, which is the technical concept behind user-centric identity management.<sup>15</sup> Such a card is a graphical representation of a certain set of attributes, and technology can help users to understand which card may possibly be used in a certain situation and which cards have been used previously in the same or a similar context. There are several implementations of this concept (e.g., *Cardspace* by Microsoft [C06], which has been discontinued in February 2011<sup>16</sup>; *Project Higgins' Identity Selector*<sup>17</sup>), but neither is widely used.

### 2.1.2 Client-controlled Selection of Data

To stick with our newspaper example: if you read a paid newspaper online you will probably click on the articles of interest, giving the publisher the chance of profiling exactly what you are reading and for how long. This may be good feedback for the newspaper and so at the end you might benefit from this. But you may also prefer “minimizing away” that data by downloading the full newspaper and browsing through it locally at your laptop. In essence that is the idea behind client-controlled selection of data.

Another example is online advertisements: usually the service provider selects what ads to show to a specific subscriber based on that subscriber’s profile and current activity. Doing the ad selection locally is not that easy, but in combination with some crypto and trusted third parties it could be done [FVH09, TNB+10, BKMP12]. None of these approaches is used in practice though.

### 2.1.3 Data Minimization and Business Processes

Individuals and organizations interact with each other, and the rules according to which these interactions happen are called business processes. For instance, subscribing to a social network, placing an order for a newspaper, or opening an account with a bank are all specific business processes.

In practice business processes determine the limits for data minimization, and in particular where identification is needed and which transactions must be linked. They are often old, complex and hard to change. New processes may be designed with data minimization in mind. But the business purpose of many online social networks is selling targeted ads, and from that perspective they have a business need to know as much as possible about their subscribers. In general minimizing data adds to the design complexity, and tools for designing business processes do not focus on supporting data minimization concept (e.g., pseudonyms, or lowering the fidelity of data).

### 2.1.4 Limitations of Data Minimization: Unintended Traces

So far we talked about the data consciously disclosed to other parties, and how to minimize them. There are also unconsciously generated digital traces, and typically those are even more difficult to deal with.

Devices like PCs and mobile phone, and web browsers can be “fingerprinted” with a variety of methods. The most trivial “fingerprint” is the IP number of a device, which typically stays constant for a couple of hours and which can be seen by each communication partner. IP numbers can be hidden through so-called anonymous proxies (or “mixes”), or sequences of such proxies [C81, AGLD07 Ch. 1 – 2, DD08]. Commercial and free-of-charge offerings exist (e.g., Hotspot Shield<sup>18</sup>,

---

<sup>15</sup> <http://www.incontextblog.com/?p=728>

<sup>16</sup> <http://gcn.com/articles/2011/02/23/ecg-microsoft-kills-cardspace.aspx>

<sup>17</sup> <http://www.eclipse.org/higgins/>.

<sup>18</sup> <http://hotspotshield.com/>



JonDoNym<sup>19</sup> and JAP<sup>20</sup>, Tor<sup>21</sup>) but are not used very widely. A certain loss in performance and reliability is certainly one reason, but a more critical limitation might be the technical understanding assumed by all these tools.

Web browsers can be fingerprinted by collecting their configuration and status data [E10]. “Private browsing” modes somewhat reduce the problem, but do not completely avoid it. Even devices can be fingerprinted at the physical level [DZC11]. There are projects which try to prevent this fingerprinting through filtering and proxies, but none is at production level.

A more official way of generating traces is through browser cookies and through elements which are specifically put on web pages to enable tracking (this can be invisible elements, but also plug-ins like the “I like buttons” recently introduced by some social networks). Users can avoid this kind of tracking by consciously managing their browser cookies (e.g., restrict which sites can set cookies and for how long) and their sessions (e.g., explicitly terminate sessions when not needed anymore). There are also browser plug-ins that support this kind of cookie management.

Currently various groups, in particular the World Wide Web Consortium (W3C)<sup>22</sup>, are working on web standards for a “do not track” option. Such a standard would allow users to signal that they do not want to be tracked. Some services already support such a feature [W11], and some technologies exist that can help users and third parties to check whether online services observe this option.<sup>23</sup>

Traces are also generated through personal characteristics, including a user’s face, voice or writing style. Experiments have shown that it is very easy to decide whether two even relatively short emails or postings have been written by the same author [RR00]. Personal characteristics also allow to link data about a person from different sources.

Traces even exist in supposedly anonymized data sets, like anonymized health records, or anonymized lists of your most favorite movies. Data sets are anonymized by removing obvious identifiers (e.g., name and address) and by generalizing some data (e.g., “Darmstadt” may be generalized to “Germany”, or age “12” may be generalized into an age group “10-19”) [S02]. In [S97] it was shown that one can often identify the records of specific persons in anonymized databases if one knows just a few attributes of that person (like sex, city, date of birth). Using the same principle, although with more sophisticated correlation algorithms, it was also shown how to deanonymize users of social networks, or how to deanonymize large parts of databases that were published for testing purposes [NS08, WHKK10]. Measuring the quality of anonymization technologies is still an open problem in practice; popular definitions are  $k$ -anonymity (everybody is hidden in a group of at least  $k$  individuals) [S02] and differential privacy (being included in a statistical database gives an adversary with regular access to the database no significant advantage in guessing some personal attribute) [D06].

## 2.2 Purpose Binding

Service providers must use their subscribers’ data only for the agreed purposes. This purpose needs to fit with the service, i.e., must not be unnecessarily broad. Service providers are also obliged to secure these data properly while they store and use them, and to delete them as soon as they are not needed anymore.

---

<sup>19</sup> <http://anonymous-proxy-servers.net/>

<sup>20</sup> [http://anon.inf.tu-dresden.de/index\\_en.html](http://anon.inf.tu-dresden.de/index_en.html)

<sup>21</sup> <https://www.torproject.org/>

<sup>22</sup> <http://www.w3.org/2011/track-privacy/>, <http://www.w3.org/2011/tracking-protection/>

<sup>23</sup> E.g., PrimeLife/W3C Privacy Dashboard, <http://www.primelife.eu/ttt>; Privacy Violation Detector <http://www.aisee.fraunhofer.de/en/fields-of-expertise/projects/prividor.html>

Purposes are formalized in privacy policies. Often such policies offer some choices, and users may opt-in (i.e., explicitly accept, otherwise reject) or opt-out (i.e., explicitly reject, otherwise accept) certain choices. For instance, one may agree or disagree to authorize using an address for marketing purposes. Opt-in approaches are generally safer for privacy, as they protect privacy by default.

Purpose may be very narrow or very broad, depending on the service. Online social networks offer a very broad range of data sharing services, which necessarily results in a very broad purpose. Their business model is often targeted advertising, suggesting that data can be used for any marketing purpose.

Privacy policies may exist in two versions, as human readable text – written by lawyers and displayed to users and subscribers – and in some machine-readable format. The correspondence between both is typically maintained manually (and due to inherent ambiguities in human-readable policies it has been proven difficult to automate this step).

Human readable privacy policies are notorious difficult to understand. It is well known that users rarely read such policies, and research has shown that users even have a hard time distinguishing between good and bad policies [BGS05]. They easily accept disadvantageous policies if they are presented in a seemingly trustworthy and convincing style. The development of standardized privacy policies and standardized interfaces for presenting and opting-in/opting-out is still an open challenge.

Machine-readable privacy policies are often expressed in *XACML* [M05], mostly as part of a larger security-focused access control policy. Machine-readable privacy policies are relatively well understood [KSW03, PHB06, AGLD07 Ch. 7, CFR11 Ch. 16-18]. Data management – or in this context: privacy management – technology exists to monitor and enforce such policies, to generate all necessary reports and to give users access to their data.

Doing so across organizations is still a challenge [KSW03]. There are also many practical problems within a single organization: e.g., management roles and responsibilities are not always clear; there might be technical integration problems that let some IT components bypass the privacy management system; sometimes not all locations of personal data are known; sometimes it is not clear what policies govern a specific data set; sometimes data that should be deleted cannot be deleted for purely technical reasons. Therefore, privacy policies are only partially enforced through technology.

Strictly speaking, a privacy policy is just a promise, and the subscriber has little or no actual control over whether this promise is met or not. Even if a service provider is willing to enforce a policy end-to-end through technology, that enforcement is done under the service provider's control, not under the subscriber's or some third party's control. There is also still the problem of information security, see Section 1.1.

Giving the subscriber actual control over the data processing at the service provider's side is still a matter of research. Promising concepts are Trusted Computing<sup>24</sup> and computing on encrypted data, e.g., using fully homomorphic encryption [G09]. Both are still subject to research.

Overall purpose binding is the most mature part of privacy technology.

### **2.3 Contextual Integrity**

People act in a specific context when they disclose data to someone else. Telling an anecdote to a few personal friends, even in a chat room, happens in a different social context than outlining one's CV to a potential employer in an email. Each society has explicit and implicit norms that govern how

---

<sup>24</sup> <http://www.trustedcomputinggroup.org/>

much information and what type of information is fitting a certain context. Contextual integrity means a state where these norms are respected [N98, N04].

Someone who receives personal information in one context and discloses it in a completely different context is easily committing a violation of contextual integrity. This may happen completely outside any technical system, and thus technology cannot prevent it from happening.

Purpose binding and privacy policies support contextual integrity, as they codify some aspects of context and contextual integrity. But “contextual integrity” is a much richer and fuzzier concept than “purpose.” Purpose binding is not sufficient for protecting contextual integrity.

One may “stick” meta-data to each piece of data, describing aspects of the context under which they have been disclosed, and use those meta-data to prevent or at least detect violations of contextual integrity [KSW03, BPB11]. One may also enrich the usual vocabulary of privacy policies to capture more components of “context.” Arguably this will always be incomplete, and implementing such “sticky policies” has been proven very difficult.

Technology in support of contextual integrity mostly focuses on trust and reputation and on user centric identity management. The former helps estimating whether it is safe and appropriate to disclose data in a certain situation. The latter helps defining communities of trust and keeping track of what each community knows.

Technology for building reputation and establishing trust between parties exists but is still at a very experimental stage [J09]. In particular there is no agreement on the precise meaning of these terms. Early work focused on trust in public-key infrastructures (as needed for digital signatures and public-key encryption) and more generally trust in the validity of digitally signed statements [M96, BFK98]. Since then the focus moved to reputation of parties in electronic commerce (e.g., sellers in auctions or retail platforms) [RZFK00] and more generally members of online social networks [J1B07, J09, S08]. An attempt at developing an industry standard for the exchange of reputation data stalled.<sup>25</sup>

Pseudonyms have already been discussed in Section 2.1.1. Through the use of different pseudonyms for different communities one can somewhat control what information may flow between communities. Of course, the isolation between communities is never perfect since data can be correlated between two communities. Context is also still a much richer concept than community.

## **2.4 Transparency and Sustainable Informational Self-Determination**

Most subscribers of online services actively want to share their personal data with the service and with other users. Data minimization helps avoiding sharing too much, the concepts of purpose and context help reducing the risk of misuse of the data shared. The concepts of transparency and sustainable informational self-determination help keeping control over the data once they have been shared.

### **2.4.1 Transparency and Personal Data Stores**

Transparency enhancing technologies (TET) support users in their awareness of what service providers and other users know about them, and what these other parties may be able to do with these data.

Essentially all subscription-based online services provide some account management functions. Such functions let subscribers check and change their account data, e.g., their user name and privacy settings. This bare minimum hardly counts as transparency *enhancing*. Some service providers go

---

<sup>25</sup> <http://www.oasis-open.org/committees/orms/charter.php>

one step further and offer “dashboards,” i.e., a function where a subscriber can see and check all data the service provider stores about that subscriber, and can modify account and privacy settings accordingly.<sup>26</sup>

Such provider-specific dashboards have, almost necessarily, two shortcomings: Firstly, they cover only what a service provider knows *directly* about a subscriber. Service providers may also collect a lot of indirect knowledge about subscribers, e.g., their analysis of a subscriber’s interest profile, or the actual revenue they derived from this subscriber through targeted advertisements. Another example is all the data a provider collects unknowingly about a subscriber, e.g., when that subscriber uses the service anonymously or under another subscriber identity. Providers may be able to link that unknowingly collected data to a subscriber, through clever correlation of usage patterns. The second shortcoming is that such provider-specific dashboards cover only the data of one provider. They may point to other service providers, and may list what was shared with these other service providers and for what purpose. But today they do not support a subscriber in understanding the full picture, i.e., in understanding what different service providers know about them and how those different knowledge sets relate.

Technically it would be relatively simple to integrate the knowledge and account management functions of different service providers. Ideally all service providers would agree on a common technical standard and ontology for personal data and account management. Such an integrated privacy dashboard could be implemented as an online service or, better for privacy but more complicated for users, as a client-side tool. Such a tool could also support users in assessing risks and deciding what information to reveal to whom and under what pseudonym in what context, and take care of the actual disclosure of data.

Several research and development projects have been working on this idea. Presumably the biggest issue with this type of tools is usability, mental models and interface standards. This aspect has been the focus of *PrimeLife DataTrack* [CFR11 Ch. 13] and to a certain degree of *ProjectVRM* (VRM = Vendor Relationship Management)<sup>27</sup>.

Many projects have been working on the more technical aspects. The key components are always a personal data store, which stores and manages a user’s data, and protocols for attribute exchange, which let other services access the user’s data under the user’s control. All this can be implemented as an online service or as a client-side tool. The same idea is known under a variety of names: personal data management [BLK+01], federated identity management [PW03], distributed identity management (mostly *OpenID*<sup>28</sup>), and personal data store [M10]. The *Kantara Initiative* (formerly known as *Liberty Alliance*) is currently working on this concept under the title User Managed Access.<sup>29</sup> We already mentioned *OASIS SAML* and *OpenID* in Section 2.1.1. Both support this concept, but in practice they are just used for single sign on (i.e., an identity established at one specific service provider can be recognized by several others, sparing the user the effort to register and log in again and again). In theory they could also be used for empowering the users to actually control their own data.

Today online social networks are, conceptually, centrally controlled databases which store all data of all their users. Based on the approach of personal data stores one can implement an online social network in a completely distributed fashion, replacing the one database by a network of subscriber-controlled databases (e.g., *Diaspora*<sup>30</sup>). Alternatively one could use a centrally structured online social network but store all relevant content in encrypted format only, and create an overlay

---

<sup>26</sup> A good example is *Google Dashboard*, <https://www.google.com/dashboard/>.

<sup>27</sup> [http://cyber.law.harvard.edu/projectvrm/Main\\_Page](http://cyber.law.harvard.edu/projectvrm/Main_Page)

<sup>28</sup> <http://openid.net/>

<sup>29</sup> <http://kantarainitiative.org/confluence/display/uma/Home>

<sup>30</sup> <http://diasporaproject.org/>

network for exchanging keys (e.g., *PrimeLife's Clique*<sup>31</sup> and *Scramble!*<sup>32</sup>). All these approaches are at prototype stage. Naturally the usual ad-based business model would not work for them.

## **2.4.2 Sustainable Informational Self-Determination**

The main purpose of online social networks is the sharing and publishing of information, which means that almost necessarily personal data of one user will end up in some places outside his or her direct control. Today this necessarily includes the service provider's database – which we covered in Sections 2.2 and 2.4.1 – but also databases of other service providers and devices of other users.

Some data may be ignored and never reach any other database or device. Some may just stay within a small group, and some may spread across the world. Informational self-determination suggests that users should be able to maintain control over their data even if it spreads widely. Control includes at least the ability to find data, correct and withdraw or delete it.

### **2.4.2.1 Digital Rights Management**

Digital Rights Management (DRM) is a similar problem: Users acquire digital content – books, music, movies, software, etc. – and store it locally. Content owners want to be able to control the use of their content despite the fact that the users have full physical control over the content. Some DRM technologies aim at preventing illegal use by requiring a specific hardware token or a secret registration key. Most DRM technologies just deter illegal copying: the content includes an invisible watermark which identifies the copyright holder or even the user who acquired the content legally, and if such a watermark is found in some inappropriate place it will identify its source.

Despite the obvious analogy it turns out that DRM technology does not work overly well for privacy [BP08]. One general reason is that content owners are primarily concerned with mass fraud, but a few illegal copies, or even a large number of low-quality copies typically do not matter. In contrast one copy of a diary, an image or even just a name showing up in the wrong context might be sufficient to change one's life. In all of these cases it is completely impossible to prevent low-quality copying (remembering a diary, taking a photo of a computer monitor, etc.). With current client technology it is even very easy to make perfect copies of everything one sees or hears online.

A more technical problem is that watermarking, the main technique for deterring illegal copying, does not work for small data. For instance, there is no way that one could watermark a name.

Watermarking does work for larger data sets, e.g., an address book, or for personal images. Still this would not help much with the protection of a single person's privacy. But it would help detecting mass fraud, like a service provider who illegally uses multimedia content produced by its subscribers.

### **2.4.2.2 Restricting Access to Data After the Fact**

DRM is actually a simpler problem than informational self-determination. For DRM it is enough to enforce a given usage policy remotely. Informational self-determination also requires that users may change this policy, i.e., restrict access to or even delete data after it has been shared.

Data might move from a subscriber via some first service provider to a second, third, fourth, etc., service provider, and any change in the policy with the first service provider may have to trickle down to the second, third, fourth, etc. provider. This is similar to purpose binding across multiple organizations, which means it is a challenge but at least doable, as long as one trusts all service providers that they won't ignore policy changes [SW07].

---

<sup>31</sup> <http://www.primelife.eu/results/opensource/64-clique>; <http://clique.primelife.eu/>

<sup>32</sup> <http://www.primelife.eu/results/opensource/65-scramble>

Online social networks encourage its users to spend as much time as possible within the network, and as a side effect discourage making local copies. As long as data stay within the service any changes in policy will automatically also impact all the other users.

As already explained, technically there is not much that prevents users from making local copies. Arguably most of the content on the Internet has no value beyond the moment, and only few if any will copy them locally. But content that is deemed interesting in whatever sense will be copied, and then might get out of control. One must even expect that any attempt at restricting access to some moderately interesting content will cause it to be spread even further.<sup>33</sup>

The easiest, and therefore most recommended, approach for keeping control over data is to introduce expiration dates, after which data is deleted from the service. By default content handed over to an online social network could expire after some predefined time, and would stay online only if the originator actively extends the period. Certain content could be marked explicitly as “permanently published” in order to keep it forever.

Using encryption and a third-party provided key management infrastructure one could implement expiration dates for some media files even without the service provider’s cooperation [BBD11].

### 3 Summary and Outlook

In Section 2 we discussed four aspects of online privacy:

Data minimization (Section 2.1) enables users to control and limit what and how much they share. On the one hand research has produced many excellent technical results, and at least the basics are also available on the market. Usability and tool support (for users, providers and developers) is one of the main open practical problems. Developing meaningful privacy metrics and improving the efficiency of advanced cryptographic concepts (like fully homomorphic encryption) are the main open theoretical challenges. On the other hand most users seem to accept that personal data has become the main currency on the Internet. In general industry is moving towards collecting more information and applying sophisticated search and data mining algorithms (aka “big data”) in order to maximize the value of those data.

Purpose binding (Section 2.2) is the basis for privacy-respecting processing of data. This is the most mature part of privacy technology, both in research and in practice. There are still many open problems, like creating a map of all relevant data within an enterprise, or supporting purpose binding across organizations, or actually enforcing purpose binding through cryptography or trusted computing concepts. One should also keep in mind that purpose binding assumes there is an agreed purpose. Often this agreement is very fragile, considering the frequent changes of privacy policies of major online social networks, and also considering the generally low confidence of consumers in online services.

Contextual integrity (Section 2.3) is the concept that information is released in a certain social context, and moving information between social contexts can easily destroy privacy. This is hard to grasp and handle in a technical sense, and accordingly not much technical support exists yet. Notable exception is user-centric identity management, which supports the isolation of contexts from each other. Despite a lot of research and very mature technologies industry has not seriously taken up the idea of user centric identity management, and sometimes it has even slowed down in its adoption.

---

<sup>33</sup> This is often called the Streisand Effect, [http://en.wikipedia.org/wiki/Streisand\\_effect](http://en.wikipedia.org/wiki/Streisand_effect). Barbara Streisand tried to delete rather unspectacular images of her residence from the Internet, with the effect that suddenly these images caught attention and spread beyond any chance to remove them.

Transparency and sustainable informational self-determination (Section 2.4) are well understood problems, but only very basic technologies exist in practice. Some services offer privacy dashboards where subscribers can check what the service knows about them directly. Tools for deeper transparency, or transparency across multiple service providers, or quantifying privacy, are all still matters of research. Tools for pseudonym management also support informational self-determination, but more powerful personal data stores are just emerging and arguably are still at prototype stage. Recently the problem of “forgetting,” i.e., withdrawing information from the Internet, has received a lot of attention. This is an essentially unsolvable problem. The best approximation would be mandatory expiration dates for user-generated content.

Figure 4 summarizes this assessment (more black means technically more mature, better understood in R&D, more effective in practice).





<b>Data minimization</b>	
<b>Purpose binding</b>	
<b>Contextual Integrity</b>	
<b>Transparency and Sustainable Informational Self-determination</b>	

Figure 2: Maturity of the Four Aspects

There is clearly the need and the potential for more research and development in privacy. Several research challenges have been mentioned above and throughout the paper. See also [FHK+11] for a more complete discussion of research challenges.

Short-term the focus of R&D in privacy technology should be on the development of simple but reasonably effective tools. Instead of striving for perfection we recommend applying the 80/20-principle: looking for the really simple ideas where a small (“20%”) effort yields a big (“80%”) impact. Ideally such technology is so intuitive and simple that it can be mandated, which would take it out of the normal cost considerations of IT vendors and service providers.

Good candidates for such 80/20-solutions are the development of technical and interface standards. Such standards are needed for identity, attributes and relationships; for privacy options and for expressing sticky policies across organizations; and for data management practices like expiration dates. In particular introducing mandatory expiration dates for all user generated content would be an easy win.

Similarly easy wins are standards for accessing one’s own personal data, i.e., for transparency. Such a standard would enable the creation of personal privacy dashboards that would span multiple service providers. Transparency could and should also be extended to show the financial value of personal information for the service provider. Actually measuring this value would be very difficult, in particular without cooperation of the service providers. But requiring service providers to always show the current value of personal data to them would be easy.

Ultimately such standards need to be global, reflecting the global nature of most online services. But standards of this type could be introduced regionally, e.g., in the European Union, and arguably the market forces would ensure that global service providers would follow the lead.

## Acknowledgements

We thank Andreas Poller, Martin Steinebach and the members of the acatech project “Internet Privacy” for interesting discussions. We thank Martin Steinebach in particular for pointing out the potential value of watermarking for detecting mass fraud in online social networks.

## 4 References

- [A08] Alessandro Acquisti: Identity Management, Privacy, and Price Discrimination; IEEE Security & Privacy 6/2 (2008) 45-50.
- [AGLD07] Alessandro Acquisti, Stefanos Gritzalis, Costos Lambrinouidakis, Sabrina De Capitani di Vimercati: Digital Privacy: Theory, Technologies, and Practices; Auerbach Publications, Boca Raton 2007.
- [AV05] Alessandro Acquisti, Hal R. Varian: Conditioning Prices on Purchase History; Marketing Science 24 (2005) 367-381.
- [B00] Stefan Brands: Rethinking Public Key Infrastructures and Digital Certificates; MIT Press, 2000.
- [B08] Stephen Baker: The Numerati: How They'll Get My Number and Yours; Jonathan Cape Ltd, 2008.
- [B11] Datenschutz im Internet; Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e.V. (BITKOM), Berlin 2011.
- [B11b] Wade Baker et. al.: 2011 Data Breach Investigations Report; Verizon, 2011 ([http://www.verizonbusiness.com/resources/reports/rp\\_data-breach-investigations-report-2011\\_en\\_xg.pdf](http://www.verizonbusiness.com/resources/reports/rp_data-breach-investigations-report-2011_en_xg.pdf)).
- [B83] Bundesverfassungsgericht: Urteil vom 15. Dezember 1983, Az. 1 BvR 209, 269, 362, 420, 440, 484/83 (“Volkszählungsurteil”).
- [BBD11] Julian Backes, Michael Backes, Markus Dürmuth, Sebastian Gerling, Stefan Lorenz: X-pire! A Digital Expiration Date for Images in Social Networks; CoRR abs/1112.2649 (2011).
- [BE07] Danah M. Boyd, Nicole B. Eklisonb: Social Network Sites: Definition, History, and Scholarship; Journal of Computer-Mediated Communication 13/1 (2007).
- [BFK98] Matt Blaze, Joan Feigenbaum, Angelos D. Keromytis: KeyNote: Trust Management for Public-Key Infrastructures; Security Protocols Workshop, Cambridge 1998; Springer-Verlag, LNCS 1550, 59 – 63.
- [BGS05] Bettina Berendt, Oliver Günther, Sarah Spiekermann: Privacy in E-Commerce: Stated Preferences vs. Actual Behavior; Communications of the ACM 48/3 (2005)
- [BKMP12] Michael Backes, Aniket Kate, Matteo Maffei, Kim Pecina: ObliviAd: Provably Secure and Practical Online Behavioral Advertising; 2012 IEEE Symposium on Security and Privacy, IEEE, Oakland 2012.
- [BLK+01] Kathy Bohrer, Xuan Liu, Dogan Kesdogan, Edith Schonberg, Moninder Singh, Susan L. Spraragen: Personal Information Management and Distribution; International Conference on Electronic Commerce Research (ICECR-4) Dallas, 2001.
- [BP08] Rainer Böhme, Andreas Pfitzmann: Digital Rights Management zum Schutz personenbezogener Daten?; Datenschutz und Datensicherheit – DuD 32/5 (2008) 342-347.
- [BPB11] Katrin Borcea-Pfitzmann, Andreas Pfitzmann, Manuela Berg: Privacy 3.0 := Data Minimization + User Control + Contextual Integrity; Information Technology 53/1 (2011) 34 – 40.
- [BSBK09] Leyla Bilge, Thorsten Strufe, Davide Balzarotti, Engin Kirda: All your contacts are belong to us: automated identity theft attacks on social networks; 18th International Conference on World Wide Web, ACM, Madrid 2009.
- [C05] Kim Cameron: The Laws of Identity; Microsoft Corporation, Redmond 2005 (<http://msdn.microsoft.com/en-us/library/ms996456>).
- [C06] David Chappell: Introducing Windows CardSpace; Microsoft, Redmond 2006 (<http://msdn.microsoft.com/en-us/library/aa480189.aspx>).
- [C08] Roger Clarke: Reference List: Surveys of Privacy Attitudes; 1996 – 2008 (<http://www.rogerclarke.com/DV/Surveys.html>).
- [C10] Online Reputation in a Connected World; Cross Tab, 2010 (Research done for Microsoft) (<http://go.microsoft.com/?linkid=9709510>).
- [C11] Email Attacks: This Time It's Personal; Cisco Security White Paper, San Jose 2011.
- [C81] David Chaum: Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms. Communications of the ACM 24/2 (1981) 84 – 88.
- [C85] David Chaum: Security Without Identification: Transaction Systems to Make Big Brother Obsolete; Communications of the ACM 28/10 (1985) 1030 – 1044.
- [CFR11] Jan Camenisch, Simone Fischer-Hübner, Kai Rannenberg (eds.): Privacy and Identity Management for Life; Springer-Verlag, Heidelberg 2011.



- [CL01] Jan Camenisch, Anna Lysyanskaya: An Efficient System for Non-transferable Anonymous Credentials with Optional Anonymity Revocation; Eurocrypt 2001, Innsbruck 2001, Springer-Verlag, LNCS 2045, 93 – 118.
- [D02] John R. Douceur: The Sybil Attack; International Workshop on Peer-To-Peer Systems, Cambridge 2002.
- [D06] Cynthia Dwork: Differential Privacy; ICALP 2006, Venice 2006; Springer-Verlag, LNCS 4052, 1 – 12.
- [DD08] George Danezis, Claudia Diaz: A Survey of Anonymous Communication Channels; Microsoft Research MSR-TR-2008-35, Cambridge 2008 (<ftp://ftp.research.microsoft.com/pub/tr/TR-2008-35.pdf>).
- [DMB11] H. Kristl Davison, Catherine Maraist, Mark N Bing: Friend or Foe? The Promise and Pitfalls of Using Social Networking Sites for HR Decisions; Journal of Business and Psychology 26 (2011) 153 – 159.
- [DZC11] Boris Danev, Davide Zanetti, Srdjan Capkun: On Physical-layer Identification of Wireless Devices; ETH Zürich, 2011 (<http://www.syssec.ethz.ch/research/OnPhysId.pdf>).
- [E10] Peter Eckersley: How Unique is Your Browser? A Report on the Panoptick Experiment; Electronic Frontier Foundation, 2010 (<http://panoptick.eff.org/browser-uniqueness.pdf>).
- [E12] David A. Ebersman, CFO of Facebook: S-1 Registration Statement prepared for the United States Securities and Exchange Commission; Menlo Park, 2012.
- [EMKK11] Manuel Egele, Andreas Moser, Christopher Kruegel, Engin Kirda: PoX: Protecting users from malicious Facebook applications; IEEE PERCOM Workshops, Seattle 2011.
- [FHK+11] Simone Fischer-Hübner, Chris Jay Hoofnagle, Ioannis Krontiris, Kai Rannenberg, Michael Waidner: Online Privacy: Towards Informational Self-Determination on the Internet (Dagstuhl Perspectives Workshop 11061); Dagstuhl Report 1/2 and Dagstuhl Manifesto 1/1, Schloss Dagstuhl 2011.
- [FVH09] Julien Freudiger, Nevena Vratonjic, Jean-Pierre Hubaux: Towards Privacy-Friendly Online Advertising; Web 2.0 Security and Privacy, Oakland 2009.
- [G09] Craig Gentry: Fully Homomorphic Encryption Using Ideal Lattices; ACM Symposium on Theory of Computing (STOC), ACM, Bethesda 2009.
- [H11] Joerg Heidrich: Datenschutz in der Wolke; Spiegel-Online, Sept. 3, 2011 (<http://www.spiegel.de/netzwelt/web/0,1518,783446,00.html>).
- [HSC08] Marit Hansen, Ari Schwartz, Alissa Cooper: Privacy and Identity Management; IEEE Security & Privacy 6/2 (2008) 38 – 45.
- [HZN09] Kevin Hoffman, David Zage, Cristina Nita-Rotaru: A Survey of Attack and Defense Techniques for Reputation Systems; ACM Computing Surveys 42/1 (2009) 1-31.
- [I11] International Telecommunication Union: Measuring the Information Society; Geneva 2011.
- [J09] Audun Jøsang: Trust and Reputation Systems; Tutorial at IFIPTM 2009, Purdue University, June 2009 (<http://folk.uio.no/josang/tr/IFIPTM2009-TrustRepSys.pdf>).
- [JIB07] Audun Jøsang, Roslan Ismail, Colin Boyd: A Survey of Trust and Reputation Systems for Online Service Provision; (Elsevier) Journal on Decision Support Systems 43/2 (2007) 618 – 644.
- [K11] Dave Kerpen: Likeable Social Media: How to Delight Your Customers, Create an Irresistible Brand, and Be Generally Amazing on Facebook (And Other Social Networks); McGraw-Hill, 2011.
- [KR11] Constanze Kurz, Frank Rieger: Die Datenfresser: Wie Internetfirmen und Staat sich unsere persönlichen Daten einverleiben und wie wir die Kontrolle darüber zurückerlangen; S. Fischer Verlag, Frankfurt 2011.
- [KSW03] Günter Karjoth, Matthias Schunter, Michael Waidner: The Platform for Enterprise Privacy Practices: Privacy-enabled Management of Customer Data; Privacy Enhancing Technologies, San Francisco 2002, 69 – 84.
- [LPBK10] Tobias Lauinger, Veikko Pankakoski, Davide Balzarotti, Engin Kirda: Honeybot, Your Man in the Middle for Automated Social Engineering; 3rd USENIX Conference on Large-scale Exploits and Emergent Threats; USENIX Association, 2010.
- [M05] Tim Moses (ed.): eXtensible Access Control Markup Language (XACML) Version 2.0; OASIS, Feb 2005.
- [M10] The Case for Personal Information Empowerment: The rise of the personal data store; Mydex CIC, Bath 2010.
- [M96] Ueli Maurer: Modelling a Public-Key Infrastructure; ESORICS, Rome 1996; Springer-Verlag, LNCS 1146, 325 – 350.
- [N04] Helen Nissenbaum: Privacy as Contextual Integrity; Washington Law Review 79/1 (2004) 119 – 158.
- [N98] Helen Nissenbaum: Protecting Privacy in an Information Age: The Problem of Privacy in Public; Law and Philosophy 17 (1998) 559 – 596.
- [NS08] Arvind Narayanan, Vitaly Shmatikov: Robust De-anonymization of Large Sparse Datasets; IEEE Symposium on Security and Privacy, Oakland 2008, 111 – 125.
- [P08] Andreas Poller: Privatsphärenschutz in Soziale-Netzwerke-Plattformen; Fraunhofer-Institut für Sichere Informationstechnologie; Darmstadt 2008. (<http://sit.sit.fraunhofer.de/studies/de/studie-socnet-de.pdf>).
- [PH10] Andreas Pfitzmann, Marit Hansen: A Terminology for Talking about Privacy by Data Minimization (v0.34); TU Dresden and ULD Kiel, 2010 ([http://dud.inf.tu-dresden.de/Anon\\_Terminology.shtml](http://dud.inf.tu-dresden.de/Anon_Terminology.shtml)).
- [PHB06] Alexander Pretschner, Manuel Hilty, David Basin: Distributed Usage Control; Communications of the ACM 49/9 (2006) 39 – 44.
- [PW03] Birgit Pfitzmann, Michael Waidner: Privacy in Browser-Based Attribute Exchange; ACM Workshop on Privacy in the Electronic Society (WPES) 2002, ACM Press 2003, 52-62.

- [PWP00] Birgit Pfizmann, Michael Waidner, Andreas Pfizmann: Secure and Anonymous Electronic Commerce: Providing Legal Certainty in Open Digital Systems Without Compromising Anonymity; IBM Research Report RZ 3232 (#93278), Zurich 2000.
- [R08] Nick Ragouzis et. al. : Security Assertion Markup Language (SAML) V2.0 Technical Overview; OASIS, March 2008.
- [RR00] Josyula R. Rao, Pankaj Rohatgi: Can Pseudonymity Really Guarantee Privacy?; 9th USENIX Security Symposium, USENIX Association, Denver 2000.
- [RZFK00] Paul Resnick, Richard Zeckhauser, Eric Friedman, Ko Kuwabara: Reputation Systems; Communications of the ACM 43/12 (2000) 45 – 48.
- [S02] Latanya Sweeney: *k*-anonymity: A Model for Protecting Privacy; International Journal on Uncertainty, Fuzziness and Knowledge-based Systems 10/5 (2002) 557 – 570.
- [S08] Neel Sundaresan: Online Trust and Reputation Systems; ACM conference on Electronic commerce, San Diego 2007.
- [S97] Latanya Sweeney: Weaving Technology and Policy Together to Maintain Confidentiality; Journal of Law, Medicine and Ethics 25 (1997) 98 – 110.
- [SW07] Matthias Schunter, Michael Waidner: Simplified Privacy Controls for Aggregated Services – Suspend and Resume of Personal Data; Privacy Enhancing Technologies, Ottawa 2007, 218 – 232.
- [TNB+10] Vincent Toubiana, Arvind Narayanan, Dan Boneh, Helen Nissenbaum, Solon Barocas: Adnostic: Privacy Preserving Targeted Advertising; NDSS 2010, Internet Society, San Diego 2010.
- [U08] Stacey Snyder v. Millersville University et al.; US District Court for the Eastern District of Pennsylvania, 07-1660, December 3, 2008 (<http://www.paed.uscourts.gov/documents/opinions/08d1410p.pdf>).
- [W11] W3C Workshop on Web Tracking and User Privacy, Princeton, 2011 (<http://www.w3.org/2011/track-privacy/report.html>).
- [W67] Alan F. Westin: Privacy and Freedom; Atheneum, New York 1967.
- [WHKK10] Gilbert Wondracek, Thorsten Holz, Engin Kirda, Christopher Kruegel: A Practical Attack to De-anonymize Social Network Users; IEEE Symposium on Security and Privacy, Oakland 2010, 223 – 238.
- [WKG+11] Yang Wang, Saranga Komanduri, Pedro Giovanni Leon, Gregory Norcie, Alessandro Acquisti, Lorrie Faith Cranor : “I regretted the minute I pressed share”: A Qualitative Study of Regrets on Facebook; Symposium On Usable Privacy and Security, Pittsburgh, PA, 2011.