enisa
European Network and Information Security Agency

Privacy, Accountability and Trust – Challenges and Opportunities

About ENISA

The European Network and Information Security Agency (ENISA) is an EU agency created as a response to security issues of the European Union. The Agency's mission is essential to achieving a high and effective level of network and information security within the European Union. Together with the EUinstitutions and the Member States, ENISA seeks to develop a culture of network and information security for the benefit of citizens, consumers, business and public sector organisations in the European Union. ENISA is a centre of competence for the European Member States and European institutions in network and information security, giving advice and recommendations and acting as a switchboard of information for good practices. Moreover, the agency facilitates contacts between European institutions, the Member States and industry players.

Acknowledgments

This study relies on the contributions and the expertise of ENISA's expert group on Privacy, Accountability and Trust. ENISA would like to thank its members for their support in integrating this study. The following external experts participated in the Privacy, Accountability and Trust expert group:

- Claude Castelluccia (INRIA, France)
- Peter Druschel (Max Planck Institute for Software Systems, Germany)
- Simone Fischer Hübner (Karlstad University, Sweden)
- Aljosa Pasic (Athos Origin, Spain)
- Bart Preneel (K.U.Leuven, Belgium)
- Hannes Tschofenig (NSN, Finland)

Without their work and support this could not have been achieved.

We would like to acknowledge useful comments received on the Trusted Framework Section from David Simonsen, Alan DeKok, John Levine, Stefan Winter, Klaas Wierenga, Josh Howlett, Heather West, Jim Fenton, Dave Crocker, Don Thibeau, Mary Rundle; on the Economics of Privacy Section from Nicola Jentzsch; on the Architecture Side Section from Claudia Diaz and Seda Gürses.

Contact details

For enquiries about this study, please use the following contact details:
European Network and Information Security Agency, Technical Competence Department
Email: sta@enisa.europa.eu
Internet: http://www.enisa.europa.eu/act/it/
Supervisor of the project: Rodica Tirtea – ENISA.
ENISA staff involved in the project: Demosthenes Ikonomou, Slawomir Gorniak.

Executive summary 5

I. Introduction 6

1.1. EU data protection regulatory framework 6

1.2. Context of this work 7

1.3. Structure of the study 7

2. Business and user perspective 8

2.1. Service value chain in the Future Internet 8

2.1.1. Stakeholder analysis 9

2.1.2. Accountability principle and compliance 12

2.1.3. Other issues 14

2.1.4. Conclusions 14

2.1.5. Recommendations 15

2.2. Behavioural tracking and profiling on the Internet 16

2.2.1. Behavioural profiling 16

2.2.1.1. Definition 16

2.2.1.2. Motivations: why are we being profiled? 17

2.2.1.3. Profiling and privacy 17

2.2.2. Web tracking and profiling 18

2.2.2.1. Cookies 18

2.2.2.2. Supercookies and Evercookies 19

2.2.2.3. Browser fingerprinting 19

2.2.3. Location tracking and profiling 19

2.2.3.1. Location privacy 19

2.2.3.2. Location based services 20

2.2.4. Social network tracking and profiling 21

2.2.4.1. Online Social Networks 21

2.2.4.2. Mobile Online Social Networks 22

2.2.5. Some existing solutions 22

2.2.6. Conclusions and recommendations 22

2.3. On monetising privacy 25

2.4. Transparency-enhancing technologies 26

2.4.1. Background 26

2.4.2. Example technologies 26

2.4.3. Conclusions 29

2.5. HCI (Human Computer Interaction) for policy display and informed consent 30

2.5.1. Multi-layered privacy policy display 30

2.5.2. Policy icons for displaying policy components 31

2.5.3. Icons for displaying policy (mis-) match information 33

2.5.4. Consent via click-through agreements 34

2.5.5. Consent via menu choices 35

2.5.6. Consent via DaDAs 35

2.5.7. Conclusions 37

3. Architecture side 38

3.1. Architecture and privacy. Terminology and initiatives 38
Privacy by design and data minimisation principle 39
Privacy impact assessment 40

3.2. Identity management and privacy 40

3.3. Accountability 42
3.3.1. Definitions 42
3.3.2. Types of accountability 42
3.3.3. Threats against accountability 43
3.3.4. Existing reputation systems 44
3.3.5. Existing work on integrity accountability 44
3.3.6. Existing work on information accountability 45
3.3.7. Conclusions 45

3.4. Trust frameworks 45
3.4.1. Introduction 45
3.4.2. Email architecture 47
3.4.3. Voice over IP and Instant Messaging architectures 49
3.4.4. Web-based federations 50
inCommon 50
WAYF 50
EDUGAIN 51
3.4.5. Authentication, Authorisation and Accounting (AAA) frameworks 51
3.4.6. Payment card industry 52
3.4.7. Conclusions 53
3.4.8. Further investigations 53

3.5. Examples of privacy preserving architectures and trust frameworks 54
3.5.1. Privacy-preserving advertising 54
3.5.2. Location privacy 55
3.5.3. Smart metering 56

3.6. Privacy at lower layers 56
3.6.1. Definitions and basic notions 57
3.6.2. High Latency communication: remailers and mixes 58
3.6.3. Low latency communication: onion routing 58

4. Concluding remarks 60

List of recommendations 62

5. References 64

Executive summary

Due to the new challenges for personal data protection, in particular in the light of new technologies and globalisation, a review of EU data protection regulatory frameworks has been initiated, with a view to enhancing individuals' confidence and strengthening their privacy rights.

According to the EU legal framework, citizens of the European Union enjoy, a series of rights in the digital environment, such as protection of personal data and privacy, freedom of expression and information. However certain aspects of protection of personal data are difficult to address and implement completely.

In the study, we focus on some of the available technologies and research results addressing privacy and data protection and topics related to, or influencing privacy, such as consent, accountability, trust, tracking and profiling. The objective is to provide a comprehensive and realistic view of both limitations generated and possibilities provided by technologies in the case of personal data protection rights.

The study covers three perspectives: user side, business side and architecture side. State-of-the-art Human-Computer Interaction for policy display and informed consent, as well as transparency-enhancing technologies, are covered on the user side part of the study, next to an analysis of behavioural tracking and profiling on the Internet. In the business side perspective, the service value chain in the Future Internet is addressed and the current status of studies related to monetizing privacy is briefly introduced. The architecture side focuses on privacy aspects related to identity management, accountability and trust frameworks.

Each part of the study identifies and proposes clear means to better privacy protection and enhancement that are useful for several stakeholders. As will be seen in the concluding remarks section, many more efforts by the technical community will, however, be necessary to reach a good level of understanding, a common language and clear principles recognised by designers.

Based on the findings of this study, ENISA and other stakeholders could support this area by being active in the effort to

- Raise the level of awareness and education regarding privacy. The risks associated with profiling and tracking, i.e. from an economic perspective, should be assessed (dissemination of such studies should be supported).

- Promote technologies and initiatives addressing privacy. As in the online environment information stays around forever, techniques should be promoted to support the 'right to be forgotten'. Concepts such as *privacy certification* could be supported; this would allow labelling sites and services according to their privacy policy claims and to evidence they provide for compliance with these policies. *Data minimisation, privacy enhancing technologies* and *privacy by design concepts* should be well understood and promoted by ENISA in an effort to prevent rather than cure. Research on information accountability technology should be promoted, aimed at the technical ability to hold information processors accountable for their storage, use and dissemination of third-party data. Informed user consent should be supported in a transparent and user friendly manner, i.e. using transparent privacy policies with icons

- Support policy initiatives in the field and the revision process of the data protection directive. Clear legal provisions limiting behaviour tracking and profiling should be promoted, as should clear definitions and guidelines in the field, by raise awareness of data mining techniques and their possibilities to de-anonymize data and profiles (linking in this way information that initially is not considered personal data)

1. Introduction

Individuals and enterprises are increasingly using the Internet for their interaction with public authorities, for finding information on goods and services and for purchasing or providing goods and services online.

According to EUROSTAT publications, about 30% of individuals aged 16 to 74 (average for EU 27 Member States) have used the Internet, in the last 3 months, for interaction with public authorities (i.e. have used the Internet for one or more of the following activities: obtaining information from public authorities' web sites, downloading official forms, sending filled in forms) [1] while 72% of enterprises have used the Internet to interact with public authorities (i.e. have used the Internet for one or more of the following activities: obtaining information, downloading forms, filling-in web-forms, full electronic case handling) [2].

More than 50% of individuals used the Internet for finding information about goods or services for private purposes (in the last 3 months, data from 2009), while nearly 40% of individuals in EU-27 shopped online (in the last 12 months, in 2009) and about 20% of e-shoppers bought from other EU countries [3].

Between the individuals who did not buy or order over the Internet in the last 12 months (data for 2009), the following reasons have been mentioned: for 35% of individuals, *payment security concerns* have been pointed out, for more than 25%, *privacy concerns*, e.g. giving personal details and, for almost 25% of the individuals identified, *trust concerns* about receiving, returning goods, complaint or redress concerns [3].

In the study we address topics such as privacy, security and trust. The purpose is to identify to what extent the technologies currently used or proposed by research communities offer sufficient solutions to support the legal framework regarding protection of personal data and privacy and, at the same time, sustain an increase in trust in the online environment.

1.1. EU data protection regulatory framework

In the Digital Agenda [4], Europe's strategy for a digital economy by 2020, policies and actions are identified and outlined to maximise the benefit of the Information and Communication Technologies (ICT). Among the key actions of the Digital Agenda is the review of the EU data protection regulatory framework, with a view to enhancing individuals' confidence and strengthening their rights.

The Treaty of Lisbon [5] assures the protection of personal data. The Treaty on the Functioning of the European Union states, at Article 16 "*Everyone has the **right to the protection of personal data** concerning them*". In the Charter of Fundamental Rights of the European Union at Article 8, addressing Protection of personal data, it is stated "*Everyone has the right to the **protection of personal data** concerning him or her*"; furthermore, in the same article "*Such **data must be processed fairly for specified purposes** and on the basis of the **consent of the person concerned** or some other legitimate basis laid down by law. Everyone has the **right of access** to data which has been collected concerning him or her, and the **right to have it rectified.***" Besides the Treaty of Lisbon, The European Convention on Human Rights [6], adopted by states member of The Council of Europe, specifies at Article 8 the right to respect for private and family life "*Everyone has the **right to respect for his private and family life, his home and his correspondence.***"

The main building block of data protection law within the EU is Directive 95/46/EC [7]. Other EU legislative instruments for data protection are Regulation (EC) Nr. 45/2001 [8], applicable to data processing by EU institutions and bodies, Directive 2002/58/EC [9] on privacy and electronic communications, Framework Decision 2008/977/JHA [10] on data protection in the area of police and judicial cooperation in criminal matters and Directive 2009/136/EC [11] (amending Directive 2002/22/EC on universal service and users' rights relating to electronic communications networks and services, Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector and Regulation (EC) No 2006/2004 on cooperation between national authorities responsible for the enforcement of consumer protection laws).

On 9 July 2009, the Commission launched a consultation on the legal framework regarding the fundamental right to protection of personal data, with Directive 95/46/EC as its main object. The consultation focuses on the new challenges for personal data protection, in particular in the light of new technologies and globalisation.

1.2. Context of this work

In the ENISA Working Programme [12] for 2010, a *Preparatory Action* entitled *Trust and privacy in the Future Internet* covers in one work package *Identity, accountability and trust in the Future Internet*. The objective is to study security models of electronic services and their performance in highly distributed environments, such as today's Internet. Furthermore, ENISA investigates various ways of assuring privacy and accountability on the Internet, reviewing the most prominent methods used, studying their mapping to the underlying architectures and assessing their level of effectiveness and performance. ENISA also works towards the development of recommendations on the use of specific service models in given environments and architectures. This entails the development of guidelines for necessary actions with regard to privacy and trust.

Two studies are addressing these objectives. A survey of privacy, security, trust, reputation, tracking, consent accountability mechanisms deployed in online services became available at the end of 2010 [13]. Complementary to the survey, the current study focuses available technologies and research results addressing privacy, consent, tracking, accountability, trust, etc. and how well they are implementing or supporting the EU individual's data protection rights.

Some limitations. This study does not cover all aspects related to privacy and personal data protection and their supporting technologies. For instance, regarding consent, in this study we do not consider aspects related to the actual user being in a position to give consent (children, adults with diminished responsibility, etc.).

1.3. Structure of the study

The current study covers three perspectives: business side, user side and architecture side.

In the business perspective the service value chain in the Future Internet is addressed. Some terminology is provided to enable the reader to understand the current legal framework for data protection. Some new trends and topics related to privacy are briefly introduced and the stakeholders are identified. The current status of practices related to profiling and tracking is presented. It is obvious that such practices are encouraged by the added value generated by the use of personal information and personal data. Monetising privacy is briefly introduced and some recommendations are presented.

We will shift our view towards the means that the *user* has at his disposal to protect his privacy. State-of-the-art Human-Computer interaction for policy display and informed consent, as well as transparency-enhancing technologies, are covered on the user side of the study.

The *architecture part* focuses on privacy aspects related to identity management, accountability and trust frameworks. Available solutions are presented and topics that require more research are identified.

Each section/part contains conclusions for the topic that it covers and the concluding remarks section summarises the findings. Topics for further research, as well as recommendations for different stakeholders, can be found both in the conclusions to sections and in the final considerations section. The main recommendations are listed at the end of conclusions section.

2. Business and user perspective

Section 2.1. introduces both the legal terminology of Data Protection Directives [7] and the solutions to address privacy requirements. In Section 2.2. we take a realistic view of the current context offered by surveillance, tracking and profiling mechanisms used mostly for gaining economic advantages i.e. by the advertising industry at the expense of users interested to get services oncline. In Section 2.3. economics of privacy is introduced and points for consideration are provided regarding the current understanding and quantification of economics in the context of personal information.

In Section 2.4. and 2.5. we follow a user perspective and cover issues regarding consent and privacy policies display as well as transparency-enhancing technologies.

2.1. Service value chain in the Future Internet

In the Future Internet, almost everything will become a service. On-demand delivery and charge per usage are already replacing ownership or licensing models, but this is only the beginning. Existing services will be not only delivered in a new way, but also designed in a different way, for example by taking privacy requirements into account, or by direct service co-creation between providers and users. Physical products (e.g vehicles) will contain features that enable their use as a service (e.g. rent-a-car per minute) and a new breed of services, such as derivates and added-value services will emerge. For example, there could be a new e-service called 'tax reduction' that is offered to those who travel outside commuter hours (physical fact verified through combination of e-services), share their car with the other passengers (verified through another combination of e-services), or carry out an even more simple physical act that can be verified through e-service, such as performing some sort of energy-saving or environmentally-friendly deed.

Traditional classification of online services and business models will not be applicable and flexibility and personalisation attributes will become predominant features. Profiling and customisation will increase service appeal, but will introduce change in attitudes, involving trust and privacy considerations. These considerations should be taken into account in the design of user interfaces as well as architectures supporting these service ecosystems. The role of intermediation or brokerage services will also change as the service value chain evolves, and the shift from the distribution role of intermediaries towards the 'experience enablers' role will be more evident. Service customisation, iterative personalisation or proactive maintenance will bring changes in what we today find acceptable for data collection, use or data disclosure. Shared service will bring changes in how and where our data has been stored. Service discovery and composition will use shared contents where users' consent, as we know it today, becomes inefficient. Finally, accountability principles will bring changes not only to service composition and delivery, but also to service design. Here, concepts such as Privacy by Design (PbD, see Section 3.1.) will play a major role. In a typical scenario involving the composition of services owned or operated by multiple service providers, roles and responsibilities will be shared in order to provide the optimal end-to-end experience. Joint controllership of personal data will increase and it might fall to one service chain stakeholder alone to fulfil all data controllers' obligations, as stipulated in the current Data Protection directive [7]. The increasing use of shared services, such as cloud computing, is probably the fact with the strongest impact on the privacy principles and future business models. The reliance on external resources introduces new risks for which the service providers do not necessarily have an appropriate answer today. Similarly, traceability of the resources used to provide a service might become an architectural challenge and, as a result, the accountability for security and privacy management might become unfeasible. Last, but not least, redefinition of identity and "identifiability" is an open discussion that needs to be expanded and to include various public and private stakeholders in service lifecycles and provision chains (users, shared service resources, service providers, service delivery).

2.1.1. Stakeholder analysis

The most comprehensive piece of European data protection legislation is Data Protection Directive 95/46/EC [7], which has been complemented by Directive 2002/56/EC [9], also known as E-Privacy Directive. It mainly targets publicly available electronic communications, such as telecom operators, Internet access providers, digital TV providers and similar communication providers, although the amendment from November 2009 also targets devices connected to publicly available electronic communications networks [11].

In the context of the processing of personal data, three distinct categories of parties are recognised:

- Data subject: the individual who is the subject of the personal data
- Data controller: a person (natural or legal) who alone or jointly with others "determines the purposes and means of the processing of personal data" (Art. 2(d) in [7])
- Data processor: a third party who simply processes personal data on behalf of the data controller without controlling the contents or making use of the data (Art. 2(e) in [7])

However, in the emerging service composition and provision models, the main issue is to identify where processing of data finds its place and what actually can be considered as 'data processing', whether or not using automated means. A set of operations might include collection, organisation, storage, transmission, erasure, destruction, alteration etc.

The Article 29 Working Party (WP29) is an independent advisory body on data protection composed of representatives of national data protection authorities; in 2010 it gave its opinion on the concepts of 'controller' and 'processor' [14]. WP 29 recognises organisational differentiation in the public and in the private sector, the importance of emerging and future development of ICT as well as the globalisation of data processing. It concludes that the concept of controller is autonomous, in the sense that it should be interpreted mainly according to Community data protection law, and functional, in the sense that it is intended to allocate responsibilities where the factual influence is, and thus based on a factual rather than a formal analysis. The existence of processor depends on a decision taken by the controller, who can decide either to process data within his organisation or to delegate all or part of the processing activities to an external organisation. In the Future Internet, many scenarios will involve controllers and processors, alone or jointly, with different degrees of autonomy and responsibility, so there is a clear need to allocate responsibility in such a way that compliance with accountability principles can be ensured in practice. WP29 recognises that in the future we will face examples of different kinds and degrees of joint control that may give rise to different degrees of responsibility and liability. It concludes that access to personal data and the exercising of data subjects' other rights can also be ensured at different levels by different stakeholders.

If we look at the service provision chain, we increasingly see many types of service chain stakeholders with their corresponding roles and objective (see Table 2.1.):

Stakeholders	Business Objective
Primary service supplier	To sell products or services to customers.
Identity provider	To assure identity of user and enable authentication.
Credentials provider	To provide credentials needed for authorisation related to specific service.
Data collection/aggregator	To collect and aggregate data and to transmute basic products/services into a full package offering and an end-to-end experience.
Intermediary	To offer information and matchmaking of services/products provided by others.
Shared service supplier	To provide scalable and low cost services concerning issues like storage or computational resources.
Managed service provider	To provide outsourcing for specialised tasks, such as technical (e.g. security management) or non-technical (financial management, planning, administration, marketing, branding etc).
Third-party service supplier	To provide complementary, derived or added-value services.
Personalised service provider	To supply personalised, location-based services/products and/or manage inter-organisational processes.
Infrastructure provider	To own and manage the physical network infrastructure.
Auditing and accountability tracking	To perform assessment related to regulatory compliance.
User	To buy products and services

Table 2.1.: Business objectives of service provision chain stakeholders (a non-exhaustive list)

The relationship between stakeholders will typically be encoded in some sort of policy, agreement or contact. *Service policy* represents the constraints or the conditions on the use, deployment or description of a service while a *contract* is a measurable assertion that governs the requirements and expectations of one or more parties. Policies potentially apply to various aspects of SOA (Service Oriented Architecture), such as security, manageability, privacy, etc. but they could also be applied to business-oriented aspects, e.g. hours of business. In their turn, contracts can also cover a wide range of service aspects: quality of services agreements, interface and choreography agreements, commercial agreements, etc.

The primary service provider is also a 'front-end' for the user, and therefore is also responsible for notice, transparency, and collection of service data, such as name or credit card details. This service provider should consider examples and take into account recommendations presented later in this report (user side). By default, all data related to users should be private and not accessible by any other service provider unless the user decides to make them (partially or completely) public. The client application should allow the customisation of the privacy rules to enable partial sharing of data between service providers. In PICOS **[15]** project, for example, users can create multiple partial identities attached to the same user and they can have a presence attribute attached to the primary identity, or to a partial identity. Privacy rules allow the definition of rules that may request user B authorisation when a user requests presence or subscribes to presence. The presence attributes contain not only status (online/offline), but also user status (mood, situation...) and timestamp (last update of the presence). Creating a partial identity also creates a specific context for the user profile, the presence, the privacy rules and the reputation. Other features in the PICOS approach include a flexible policy engine to manage privacy and privilege rules, as well as a privacy advisor, which warns the user of any action that would endanger his or her privacy.

The identity provider could be a public authority (e.g. electronic ID card) or private organisation. Depending on the strength of authentication (ID card, certified signature, passwords, PIN etc.) the level of authentication assurance should be assigned (such as QAA (Quality Authentication Assurance levels) in the STORK project [16]). From a user's perspective, it is not desirable to have too many levels of assurance, as some research results suggest that a user can handle at most three levels of granularity/complexity [17]. The user may be confused and lose confidence (trust) in the authentication framework and the applications or services using this framework. If national registry numbers are used as unique identifiers, one should keep in mind that they have been originally intended to be used for public sector applications only. The other authentication and authorisation schemes are presented at the end of this report, in the chapter dedicated to architecture and trust frameworks.

The provisioning of persistent user identifiers to service providers and relying parties, however, is not an option since identifiers that are used by the identity provider and service provider are directly linked to each other without any obfuscation. Some solutions that have been suggested include the use of opaque and transient identifiers, privacy enhancing technologies, and explicit user consent via user-centric identity management solutions, treated in more detail later in this report. User's identity and actions are harder to track as the user uses different service providers and in this case only the identity provider is able to map the different identities on to each other via the opaque handles.

Credentials providers or attribute providers are often separate from identity providers and in cases where particular attributes are requested for authorisation purposes, user consent is required before the attributes are provided to the service provider. The integrity and authenticity of the attributes needs to be guaranteed, while services must be able to specify and communicate the attributes they require for authorisation (e.g. age, gender, profession…) and/or further personalisation of the service (e.g. language). From a privacy perspective, only a minimal set of necessary attributes should be communicated and with the consent of the user, so what should be considered is an assertion expressing the authentication status of the user and relevant attributes for authorisation. Once again, this represents both a service design and an architectural challenge.

As the amount of data, both structured and unstructured, increases exponentially, a new type of stakeholder, namely the data collector/aggregator, is emerging. Sensors, cameras, satellites and many other data sources are producing vast quantities of potentially valuable information that needs to be extracted, aggregated, correlated, etc. Public sector data, user generated content, and many other publicly available sources are increasing privacy threats. Disclosed and entrusted data (blog entries, photographs, messages, comments...) might reveal personal information. More often, these types of stakeholder will be interested in behavioural data which aggregates various sources in order to perform profiling (described later in this report, in Section 2.2.). Situation based profiling is another example of an approach that allows mobile service providers to apply information about the mobile identity of the customer and that enables service providers to identify high value customers and to sponsor their data transmission costs. The service provider will be able to decide how much he or she is willing to invest in any customer, or to assign a "business value" to each customer in each situation. This, however, will be an automated process based on some sort of profiling algorithm. The placement of profiling components is an important architectural decision, since it involves the handling of personal information. Information flow and information abstraction are very important for architectural decisions, especially when more operators (e.g. roaming, virtual operator) are involved. Both third-party services and mobile network services may interact with the network operator's authentication and billing infrastructure in the back-end (e.g., to settle payment). Ideally, neither the mobile network operator, nor the third-party service provider, should be able to trace the identity of the user and should not be able to link multiple transactions as originating from the same source. Other technological approaches that are focusing on the analysis of user preferences or recommendations include usage analysis.

Intermediary service providers might be used for support, such as service matching or adding third party content. One of the main issues here is trust, especially in situations where service providers had no previous interaction. Trust among service providers might rely on reputation management designed to increase trust, or on some other rating scheme introduced by an intermediary. This intermediary might also take care of data provenance, assuring the user that he can rely on the provenance of information. Services that support the establishment of secure transactions are also fundamental features that need to be provided and need to protect the privacy and integrity of sensible data exchanged in the course of business transactions.

Shared Service Data storage is another important issue to consider, especially if a cross-country situation applies, since there might be a conflict between jurisdictions. A similar situation holds for managed operations, where data is transferred across borders (e.g. in the case of a Security Operations Centre (SOC) where SOC is based in a different country). The advent of cloud service providers, for example, may now allow specialised business process outsourcing (BPO) providers to focus on knowledge and still provide compelling scale in the infrastructure by using shared services online. In a 'classic' business process outsourcing context, the outsourcing partner not only takes over the ICT but also takes responsibility for the accountability-related tasks, such as evidence collection. Business Process (BP) execution is often assisted by software tools that should be able to collect specific evidence and to generate respective execution events. Although business process outsourcing involves a re-allocation of responsibility for performing all or part of a business process, the regulatory compliance responsibility and obligations remain with the process owner, which puts a lot of emphasis on trust relationships between data controller and data processing organisation.

A third-party service provider receives user data from another stakeholder. In this situation, a one-way hash function of identifier can perhaps be used, or even some sort of indirect linking to implement pseudonymity. Transient linking does not provide an identifier, but a temporary anonymous handle that is valid for a single session or a part session. The key issue for service composition designers is to enable privacy-friendly yet efficient service discovery. In addition, a third party provider might use personalisation services, which involve data mining, profiling or location providers. Data disclosure, such as locations disclosed by network operators, is an example of a process that needs some sort of accountability. The "communications data" referred to in the European Directive context are all traffic and location data held by Internet service providers and landline and mobile telephone companies about their customers. It therefore includes people's browsing patterns, phone and e-mail details (geographic location of mobile phone users, call time and duration, number dialled, callers' and recipients' names, e-mail addresses), chatroom user IDs, credit cards, etc. Disclosure of any of this information to the third party could, for example, be subject to 'accountability principle'.

Finally, in regard to the user's role and responsibilities it is also worth mentioning that the user's attitude toward privacy is rather dynamic. It changes if the user is at the airport (we allow our luggage to be searched) and shops (where we can have stronger anonymity). The established, or at least generally accepted, perception of privacy in public might change in the future. The ability to transmit and share large amounts of information, the ability to aggregate disparate sets of information into large databases, the increase in processing power to ease the processing and analysis of data etc. will eventually also contribute to shifting norms of privacy in public places. Norms of privacy appropriateness and, even more importantly, provisions for a dynamicity of these norms, will have to be incorporated in architectural principles of the future Internet. What is considered private today might not be tomorrow and vice versa. The distribution of personal information across service chain stakeholders might be dynamically governed by the norms of the moment in any given context.

2.1.2. Accountability principle and compliance

When it comes to auditing and accountability services, the placement of logging or monitoring components and access to evidences plays an essential role. It is likely that in a typical scenario of the Future Internet, end-to-end service would include several 'primary service suppliers', together with various other stakeholders that might qualify as 'data controllers'. This complicates implementation of the so-called accountability principle, recently proposed by Article 29 Working Party (WP29) **[18]**. In **[18]** there is also certain vagueness about metrics and criteria used by data controllers in the implementation of policies and procedures, especially when it comes to the way they respond to the request, or the time they take to do so. In the complex outsourcing scenario, where legal responsibility remains with the data controller, but processing, deployment of effective measures and evidence gathering is done elsewhere, it is essential to establish an agreement between parties that includes some sort of metric on privacy policy enforcement and assurance. The basic idea behind the cross-organisational privacy metrics or indicators is that they give a meaningful evaluation of how well the actual processes comply with requirements derived from data protection and privacy directive. The set of metrics or indicators should not only cover effectiveness of privacy preserving or privacy protection mechanisms, but also correctness (has the service provider applied the correct measures to protect privacy?) and coverage (has the service provider covered the whole scope of privacy principles and requirements, as described in EU directive?). The outsourcing client (or service requester) and outsourcing provider (or service provider) need to agree on how much control and visibility are provided on the one hand and how much responsibility the provider holds on the other hand. The challenge in this case is an appropriate disaggregation of the accountability-related requirements and the corresponding evidence collection and assessment process – parts of it will be ensured on the provider side and other parts on the requester side.

We should also consider design issues relevant to the proposed 'accountability principle'. What requirements do we have to take into account during the design phase in order to enable posterior (automated) evidence collection and data protection/privacy compliance assessment? Some challenges include advertising of 'privacy audit-ready' services, service discovery based on privacy-labels, event/traces semantics, separation of log traces in a case of shared services, authenticated encryption during event/evidence transfer, treatment of untrusted events/traces etc.

The concepts of automated event collection and accountability only make sense if the raw data (e.g. log events) that is used to compute them is correct and trustworthy. Therefore, the privacy by design principles (PbD, see section 3.1) should also consider the authentication of events as they are emitted by the service provider infrastructure and collected by the auditor or third party assessment. Events generated by different components should contain correct, i.e. truthful and complete, information about the observed actions/events related to privacy policy. In addition, they should be secured according to the necessary security and trust requirements, such as message origin authentication, message integrity, data privacy and so on.

Since the complex scenario with many stakeholders, presented in the previous section, involves many events coming from various distributed sources, we have to consider privacy requirements together with operational requirements, such as scalability and efficiency.

Compliance with privacy legislation and 'accountability principle' should also address the means to avoid policy violations where policies are derived from compliance requirements. Compliance management (CM) also refers to standards, frameworks, and software used to ensure the monitoring and control of this policy enforcement. A good Compliance Management scheme should address many tasks derived from eventual inclusion of accountability principles in the European Directive revision. These can be roughly grouped around three main phases:

- Compliance engineering: this should address clear and unambiguous definition of privacy by design requirements and guidelines, as well as the translation of non-trivial regulatory compliance requirements (e.g. what is 'reasonable') expressed in natural language, into technical controls that can be deployed in operational infrastructure and can generate evidence which, at a later stage, enables privacy compliance risk assessment and eventual privacy auditing. A set of operational privacy policies should be used as an interface to an operational compliance infrastructure and internal control processes. There are different ways to assess the effectiveness of the measures taken by a controller (or in the case of outsourcing by a data processing entity), so these actions should be synchronised with the second group of actions around specific monitoring, assessment and auditing actions. In addition, the need to demonstrate (as mentioned in WP29 paper on accountability principles) will result in additional requirements on evidence representation, semantics, accessibility, aggregation etc.

- An operational infrastructure for evidence collection for privacy policy compliance: indicators of possible privacy policy violations that are tailored to measure levels of compliance are used in combination with software components for automated evidence collection and different types of internal controls. Parts of this infrastructure could include signalling, monitoring or evidence filtering components

- Assessment of privacy compliance and certification/auditing: in an ideal situation, users or data protection authorities should have the ability to continuously assess privacy compliance levels, not only for processes running on ICT systems at data controller premises, but also for those processes that run on external IT systems. Evidence aggregation, correlation, analysis, control refinement, risk re-assessment, etc., might be needed for this purpose

It should be stressed that compliance with accountability principles is also a risk mitigation approach and it might bring important benefits (e.g. service differentiation that increases reputation and trust). As WP29 **[18]** notices, one size does not fit all and for this reason it might be necessary to work on accountability profiling, where specific needs and a 'tier scheme' for different types of organisations can be developed. Data protection authorities need to play a role, especially in the medium and long term and in relation to certification or schemes for certified privacy assessment service providers.

2.1.3. Other issues

Personal data means any information related to the 'data subject'. Various data can be used to determine whether a person is "identifiable" and the concept of 'identifiability' is now heavily debated, including various discussions around whether qualification of IP addresses should be considered as personal data. The large amount of 'public' data freely available on the Internet is enabling faster 'de-anonymisation', so the cost of identification, often cited as the main factor to determine whether data can be considered personal, is continuously challenged. Besides the cost of identification, WP29 cites other factors that contribute to define 'identifiability', such as the advantage expected by the controller, interest at stake, risk of organisational dysfunctions etc. The nature of IP addresses, or other device related to data and its classification as personal data, would certainly have implications for the service provision chain and its stakeholders, especially if these data are classified in a different way across Member States. In this kind of situation, privacy options and settings might be mediated (e.g. privacy workflows that include an intermediate step), with each exception being carefully considered.

Recommendation #1. *Unambiguous definitions of personal data and identifiable information should be supported across Member States.*

Cross-country accountability, as well as the international application of the so-called data quality principle (requirements to ensure that the personal data processed are strictly required for such processing), is also raising doubts. It contains sub-principles that are often differently interpreted, such as unambiguous consent (e.g. vagueness in social networks), or reasonable expectations (e.g. derived use of data) of the data subject. One of the most important sub-principles and probably the most adequate for privacy by design is the data minimisation principle (see Section 3.1), which is expected to play an important role in the definition of which personal data appears to be necessary to use by whom. However, each country might have different considerations of what is "reasonable" in regard to a minimum set of data. Last, but not least, the conservation and data retention sub-principle that currently differs from country to country could present some technical difficulties when it comes to the monitoring of accountability across the whole service chain.

2.1.4. Conclusions

It is evident that the opportunities from the new technological developments, such as interactive, context-aware and location-based products and services, will also affect the future service chains as well as the objectives, roles and responsibilities of the business model stakeholders. Privacy sensitive information might contain, or might be linked to, a multitude of actors (e.g. patients, doctors, social workers, etc), devices (e.g. health-care terminals, PDAs, phones, computers, temperature sensors, etc), databases and applications. Moreover, the mere definition of what is privacy sensitive might differ according to business model stakeholder, or might depend on the country legislation and interpretation. In some circumstances even users can be data controllers (and not just data subjects). The shift in ICT usage and emerging business models is bringing more risks of security breaches of which data controllers are not aware. When a process is carried out on behalf of the controller, for example, the processor should provide guarantees with respect to technical security measures, as well as ensure compliance with these measures. When it comes to trust in the future service models, we should have in mind that trust is not necessarily aligned to the natural circles of trust that arise from "physical" service model chains. The stakeholders in the service chain do not necessarily share a common language, privacy policy or categorisation schema and they do not necessarily know how much trust to assign to an ad hoc partner in the same service chain.

The move towards services also increases the emphasis on relationships, negotiations, and agreements. This brings particular challenges for the area management and measurement of security and privacy. Rather than being predefined and fixed over long periods of time, as in conventional computing, the architecture of shared services and complex service compositions is defined in terms of the requirements specified for the end-to-end service functionality and the discovery, selection, choreography or adaptation processes applied across all participating services. These processes, hence the architecture, may also be governed by models, patterns, policies and contracts. Issues such as the 'accountability principle' may therefore limit ad hoc changes in service compositions, while patterns may constrain entire architectures that have been proved to work well in specific contexts.

Security and privacy policy monitoring, assessment, verification and audit are arguably the most important challenges of the Future Internet, since they allow users to trust that a service is performed as expected, no matter how many layers separate the party requesting the service from the service itself. Both the requesting service and the requested service may run in different virtual or shared environments, while end-to-end (E2E) may encompass different dimensions and architectural levels. For the party initially requesting the service, it is essential to be able to verify that the use of personal information throughout the entire value chain is only undertaken in accordance with the policy, as well as to detect when the process has been compromised. Monitoring of compliance and 'accountability principle' will therefore have an impact on architectures (distributed versus centralised), as well as on the overall cost.

2.1.5. Recommendations

Dealing With Dynamicity: a conceptual framework where one size does not fit all
Acceptance and assimilation of new business models for service provision is bringing change in attitudes, involving trust and privacy considerations. Conceptual frameworks such as contextual integrity[1] could be used for understanding privacy expectations and reasoning about norms in each place, time or other contextual dimension. This framework has also to consider dynamicity of risk and issues such as privacy risk appetite, perception of threats etc. Privacy Impact Assessment or similar schemes should not be static (executed once before the actual service invocation) but rather dynamic with a feedback loop (what happens when we change some privacy parameters?). The final result would be some sort of reconfigurable, context dependable privacy setting engine.

Recommendation #2. *Dynamicity of privacy, as well as subjectivity, should be investigated within conceptual frameworks for understanding privacy expectations and reasoning about norms in each place, time or other contextual dimension. This framework has to consider risk dimensions as well as schemes such as Privacy Impact Assessment and might also be linked to privacy decision support tools or privacy services (e.g. automated reconfiguration of privacy settings, enforcement of obligatory user consent, etc).*

Building for the future: anticipate impact
If issues such as "accountability principle", traceability or privacy auditing are finally adopted, we will need to rely on evidences, derived from ICT events and traces. This has to be anticipated in service design and service architectures.

Recommendation #3. *If issues such as "accountability principle", traceability or privacy auditing are finally adopted, we will need to rely on evidences, derived from ICT events and traces. This has to be anticipated in service design and service architectures.*

End-to-end is getting longer and more complex
Shared services, joint controllership, risk allocation etc. are all putting emphasis on relationships between stakeholders in the service chain. On the engineering side, indicators, metrics, semantics of privacy evidences (and corresponding event/traces) should be in place. On the operational infrastructure side, trust dynamicity, reputation management, contracts and policies are some of the elements that have to be in place to enable end-to-end accountability. Placement and access to the components that produce evidences relevant for privacy assurance are essential.

1 Privacy and Contextual Integrity: Framework and Applications, Barth, A.; Datta, A.; Mitchell, J.C.; Nissenbaum, H.; IEEE Security and Privacy, 2006, available at: http://www.nyu.edu/projects/nissenbaum/papers/ci.pdf

Recommendation #4. *Shared services, joint controllership, risk allocation etc. are all putting emphasis on relationships between stakeholders in the service chain. On the engineering side, indicators, metrics, semantics of privacy evidences (and corresponding event/traces) should be in place. On the operational infrastructure side, trust dynamicity, reputation management, contracts and policies are some of elements that have to be in place to enable end-to-end accountability. Architectural decisions about placement and access control to the components that generate events and traces, needed for evidences that support privacy assessment and assurance, should be consistent with organisational and service provision models.*

Assessment and assurance
Labelling has often been advocated as a possible solution. However, while ideas such as Quality Privacy Assurance (which could be an equivalent of QAA[2] scheme from STORK) sound attractive, there are many regulatory and organisational issues that should be solved first (who would be the cross-border authority?). Tools that support evidence-based assessment and assurance of privacy levels should be supported. One could also envisage special privacy enforcement services in the future, based on assessment results, but this is, however, a long term research topic.

Recommendation #5. *Privacy assessment, assurance, verification or enforcement should be evidence-based. These evidences might be derived from a number of sources, events and traces at different architectural layers.*

2.2. Behavioural tracking and profiling on the Internet

The main goal of this section is to present a state of the art of behavioural profiling on the Internet and to highlight some its potential privacy threats. This section is structured as follows: Subsection 2.2.1. introduces the concept of behavioural profiling. The following three subsections describe how profiling is performed by web sites, location-based services and social networks, respectively. Subsection 2.2.5. presents some of the existing tracking-prevention solutions. The last subsection concludes the report and proposes some recommendations.

2.2.1. Behavioural profiling

2.2.1.1. Definition

The concept of *behavioural profiling* (also known as 'targeting') consists of collecting and analysing several events, each attributable to a single originating entity, in order to gain information relating to the originating entity. It consists, in other words, of transforming data into knowledge [19]. Profiling involves collecting data (recording, storing and tracking) and searching it for identifying patterns (with the assistance of data mining algorithms). Profiling can be performed from:

- *Public data*: data publicly available on the Internet, such as on social networks or public forums.
- *Private data*: data privately collected by services through various forms, logs, cookies, web bugs, geolocalization services, or by other means.
- *Leaked data*: data that are private but leaked out via some inferring attacks, such as in [20, 21], or via data breaches.

2 QAA (Quality Authentication Assurance levels) in the STORK project [16].

An example of a behavioural targeting scenario is provided in [22]. A consumer shops online for an airline ticket to New York City. He or she searches for flights, but does not make any purchase and subsequently visits the web site of the local newspaper that displays ads offering tickets to New York. While no Personally Identifiable Information (PII) has been collected, his or her interest in airline tickets has been noted.

2.2.1.2. Motivations: why are we being profiled?

Profiles are very valuable for many companies in customising their services to suit their customers, in order to increase revenues. The clear intent of behavioural targeting is to track users over time and build up a profile of their interests, characteristics (such as gender, age and ethnicity) and shopping activities.

For example, advertising or publishing companies use behavioural targeting to display advertisements that closely reflect the user's interests. Online advertising systems are typically composed of three main entities: the advertiser, the publisher and the ad network [23]. The advertiser is the entity, for example a car manufacturer or a hotel, which wishes to advertise a product or service. The publisher is the entity, such as an online newspaper company, which owns one or several web sites and is willing to display advertisements and be paid for it. Finally, the ad network is the entity that collects advertisements from the advertisers and places them on publisher sites. If the user clicks on an advertisement, the ad network collects payment from the corresponding advertiser. There is, therefore, a strong incentive for the ad network to generate very accurate and complete profiles in order to maximise profit.

E-commerce sites also use behavioural tracking to recommend products that are likely to be of interest to users. For example, Amazon recommends products to online users based on the individual's past behaviour (personalised recommendation), on the past behaviour of similar users (social recommendation) and, of course, on the searched items (item recommendation) [24].

2.2.1.3. Profiling and privacy

It can be argued that the customisation resulting from profiling is also beneficial to the users, who only receive information relevant to their interest. However, it creates serious privacy concerns, since it allows some companies or institutions to gather and concentrate a huge amount of information about their customers, and about Internet users in general.

Privacy threats can be categorised in two dimensions:

1. *The personal information that is being (silently) collected or is leaked*. Users are often not aware of what data is collected about them. Profiling is often performed without their consent, and even knowledge. For example, most Internet users are unaware that third party aggregators are monitoring their browsing behaviour. [25].

2. *The destination to which the personal information is leaked*. Users are usually not informed about which of the data they provide to service providers, or publish on their favourite social network sites, are being sent/leaked to advertisers or others.

The danger is to move into a surveillance society or Internet, where all our online or physical activities will be recorded and correlated. Some companies offer various services that gather different types of information from users. The combination of all this information provides a powerful tool for the accurate profiling of users. For example, Google is one of the main third party aggregators and therefore tracks users across web sites [26]. In addition, it also runs the most popular search engine and, as such, stores web histories of users (i.e. their search requests), their map requests (i.e. their requests to the Google map service), their images and so on [21]. Web searches have often been shown to be sensitive [27]. It has actually been demonstrated that it is quite simple to derive the identity of a user from his or her web history [28]. Map requests also leak a lot of information, such as the user's home address or his or her favourite places. By combining these different types of information, Google is able to build very accurate profiles of the user. As argued in [19], "profiling shifts the balance of power between those that can afford profiling (mostly large organisations) and those that are being profiled (mostly individual citizens), because the profilers have a certain type of knowledge to which those profiled have no effective access."

3 Computational advertising is a new scientific sub-discipline whose main challenge is to find the best ad to present to a user engaged in a given context.

The advent of *ubiquitous advertising*, which can be seen as the application of *computational advertising*[3] to smart phones, will provide even more sources of profiling information [29]. With ubiquitous advertising, advertisements will not only be personalised to users' online profiles, but also to their *physical* profiles. Advertisements will be customised to users' locations, physical or intellectual activities, interactions and possibly moods. Since, as opposed to a regular computer, a mobile device is usually owned by a single person, more detailed and accurate profiles can be derived from his or her uses. It is also foreseen that sensors on phones will be able to infer users' food habits and preferences [29]. These new developments create serious privacy issues that need be studied more carefully [30].

The rest of this section considers three of the most popular Internet services, namely the web, location-based services and online social networks. It presents for each of them existing tracking and profiling mechanisms.

2.2.2. Web tracking and profiling

One of the main sources of information used for profiling comes from web tracking, i.e. tracking users across different visits or across different sites. Data collected include the sequence of visited sites and viewed pages, and the time spent on each page. This behavioural tracking is allowed and accepted because the collected data are supposedly anonymous, i.e. they do not contain any personally identifiable information (PII) such as name, address, phone number and so forth[4]. Web tracking is mainly performed by monitoring IP addresses, and using techniques such as cookies or the more powerful supercookies [31].

2.2.2.1. Cookies

A user who visits a web site composed of different objects imported from different servers generates multiple HTTP requests to numerous servers, controlled by different administrative domains. Very often a cookie is associated to each of these requests. This is a piece of text, stored by a user's web browser, which consists of one or more name-value pairs containing bits of information. It is set by a web site server and is often used to store user preferences, or as authentication tokens to maintain an authenticated session with a server. The cookie is sent back unchanged by the browser each time it accesses that web site and can therefore, be used by web sites to track users across visits.

Cookies should be sent only to the web sites that set them, or to servers in the same Internet domain. However, a web page may contain images, links, web bugs, HTML IFrame, Javascript, or other components stored on servers in other domains. Cookies that are set during retrieval of these components are called third party cookies[5], in contrast to first party cookies. Some sites, such as advertising companies, use third party cookies to track a user across multiple sites. In particular, an advertising company can track a user across all pages where it has placed advertising images or web bugs. Knowledge of the pages visited by a user allows the advertising company to target advertisements at the user's presumed preferences.

Note that first party cookies are sometimes useful; they allow users to visit a site without having to re-enter their configuration parameters. However, third party tracking raises serious privacy concerns, which are not hypothetical but real. The increasing presence and tracking of third party sites used for advertising and analytics has been demonstrated in a longitudinal study [26, 32]. This study showed that the penetration of the top 10 third parties grew from 40% in 2005 to 70% in 2008, and to over 70% in September 2009. Another study shows that not only are these third parties increasing their tracking of users, but that they can now link these traces with identifiers and personal information via online social networks [33]. In [22], a behavioural targeting study was performed on the levis.com site, the e-commerce site for the clothing line. The results showed that the web site contains a total of nine tracking tags that link to eight third party companies[6]. Javascripts are also used to collect users' information. Web sites often contain executable JavaScript files that are downloaded by visiting users. These files, in addition to their computations, sometimes update first party cookies and send information back to the servers. Javascripts have limited access to user data. However, they can access information stored in the browser, including cached objects and the history of visited links. Along with cookies and results of JavaScript execution, the tracking sites have all the regular information available in a typical HTTP request: sender's IP address, user-agent software information, current and previous URL (via referer header), email address (From header), language preference (Accept-Language header), etc.

4 We will see in a following section that this is not completely true.
5 Some sites included JavaScript code and third-party cookies from more than ten different tracking domains [oldcc2]
6 The largest third-party Ad-network companies include Advertising.com, Tacoda, DoubleClick and Omniture. Most of these networks are owned by Google, Yahoo, AOL or Microsoft. Since Ad-networks typically partner with many publishers, they can track users across several publishers and build their browsing profiles.

2.2.2.2. Supercookies and Evercookies

The use of tracking cookies is fairly ubiquitous and there are known techniques for avoiding them [34]. Therefore there exists a big impetus in the Internet tracking industry to discover and deploy more robust tracking mechanisms, often referred to as Supercookies [31]. One of the most prominent supercookies is the so-called 'Flash cookie', a type of cookie maintained by the Adobe Flash plug-in on behalf of Flash applications embedded in web pages [35]. Since these cookie files are stored outside of the browser's control, web browsers do not directly allow users to control them. In particular, users are not notified when such cookies are set, and these cookies never expire. Flash cookies can track users in all the ways that traditional HTTP cookies do, and they can be stored or retrieved whenever a user accesses a page containing a Flash application.

Flash cookies are extensively used by popular sites, often to circumvent users' HTTP cookie policies and privacy preferences. For example, it was found that some sites use HTTP and Flash cookies that contain redundant information [36]. Since Flash cookies do not expire, sites might automatically re-spawn HTTP cookies from Flash ones if they are deleted.

The persistence of Supercookies can be further improved, as illustrated by the recent evercookies [37]. This new type of cookie identifies a client even when standard cookies, Flash cookies, and others have been removed. This is accomplished by storing the cookie material in several types of storage mechanisms that are available on the local browser.

2.2.2.3. Browser fingerprinting

A recent study showed that browsers can be identified with a high degree of accuracy without cookies or other tracking technologies [38]. Every web browser provides enough unique information (user agent, fonts, screen resolution,...) to tell one from another. The study shows that a browser fingerprint is unique enough that it can, on average, identify a browser among a set of 286,777 other browsers. Browser fingerprinting is a powerful tool for tracking users and should be considered alongside IP addresses, cookies and supercookies as far as user trackability is concerned.

2.2.3. Location tracking and profiling

2.2.3.1. Location privacy

More and more systems and applications record users' locations and movements in public places. These systems provide very useful and appreciated services, and have come to be regarded as almost essential and inevitable. For example, RFID cards allow users to open doors or pay their underground ticket. GPS systems help users to navigate and find their way. Some services tell users where their friends are, or provide personalised services (such as indicating the closest restaurant or hotel). Some wireless parking meters send users a text message when their time is running out [39]. While the benefits provided by these systems are indisputable, they unfortunately pose a considerable threat to location privacy.

Location privacy is often defined as *the ability of an individual to move in public space with the expectation that their location will not be systematically and secretly recorded for later use* [39]. Location tracking is not a new problem, but new technologies (wireless networks, digital camera...) make it cheaper and easier to perform. It is this transformation to a world where location is collected pervasively, silently and cheaply that is worrisome [39].

2.2.3.2. Location based services

Already today, hundreds of millions of people permanently hold at least one mobile phone worldwide. It is predicted that smartphones will surpass PC sales within two years [40]. These mobile phones have increasing computational capacities and are equipped with multiple sensors, such as microphones, cameras, GPS, accelerometers, etc. As geolocated systems, they already enable individuals and communities to collect and share various kinds of data. Urban sensing is a new sensing paradigm leveraging users as part of a sensing infrastructure [41]. In the near future, several urban sensing applications are likely to appear, which will provide extra information about users [42]. Most average users are unaware of the extra information that is collected about them beyond the requested data, especially in the case of participatory sensing. For example, a picture taken by a user may reveal additional contextual information inferred from the background or the style of any associated text. A recent study showed that most people are unaware of the fact that the photos and videos taken with their smart phones or cameras contain geo-location information [43]. This information can be used to localise them while they are travelling, or even reveal their home address. This may be considered as a potential source of information leakage and may lead to a privacy breach if used for location tracking or in conjunction with data retrieved from OSN (Online Social Networks). The risk becomes higher as the border between OSN and LBS (Location-based Services) becomes fuzzier. For instance, OSN such as FourSquare[7] and Gowalla[8] are designed to encourage their users to share their geolocation data,.Information posted on social applications such as Twitter can be used to infer whether or not an individual is at home[9]. Other applications, such as Google Latitude[10], allow users to track the movements of their friends' cellphones and display their position on a map. In addition to social applications, there are other public sources of information that can be exploited by potential adversaries, such as the free geographic data provided by Google Maps[11], Yahoo!Maps[12] and Google Earth[13].

The W3C geolocation API, which is supported in the Firefox, Opera and Chrome browsers and in Internet Explorer via a plug-in, allows web sites to request geographical information for the client's device. With the approval of the user, the browser sends information like the client's IP address, MAC addresses of connected wireless access points and the cell ids of GSM/CDMA networks within range. With the help of a network location provider, such as Google Location Services, this information can be used to obtain an estimate of the client device's location. While the browser only sends this information to a web site with the user's explicit approval, few users realise the accuracy with which these services can often locate a device. For instance, Google Location Services rely on the MAC addresses of wireless access points detected during the Google Street View data collection to locate client devices to within the range of an 801.11 wireless base station (i.e., tens of meters).

Furthermore, a growing number of sites now provide public APIs to their geo-localised content. For example, Flickr, YouTube, and Twitter allow queries for results originating at a certain location. PicFog, for example, uses one of these APIs to provide real-time location-based search of images posted on Twitter. As shown in [43], these APIs can also be used to identify the current location of a user while he or she is away from home.

As another example of LBS, many projects are currently under study and a number of systems have already been deployed based on the 'Pay As You Drive' principle ('PAYD'). Such systems are of great interest to insurers, road operators ('Electronic Toll Pricing: ETP') as well as governments (see Section 3.5.2. for a brief discussion of current approaches). ETP intends to replace a flat road tax by a fee depending on the vehicle type, the actual vehicle use and taking into account parameters such as time, traffic and environmental conditions. The most intrusive solutions would involve the monitoring of all the whereabouts of all the drivers in the country.

7 http://foursquare.com/
8 http://gowalla.com/
9 http://pleaserobme.com/
10 http://www.google.com/latitude
11 http://maps.google.com/
12 http://maps.yahoo.com/
13 http://earth.google.com/

The emergence of *Reality Mining* raises even more privacy concerns [44]. As explained in [45], reality mining infers human relationship and behaviour from information collected by cell-phones. This information include data collected by cell-phone sensors, such as location or physical activity, and data recorded by phones themselves, such as the duration of the calls and the numbers dialled. Reality mining could help users identify things to do or new people to meet. It could also help to monitor health. For example, monitoring a phone's motion might reveal changes in gait, which could be an early indicator of ailments or depression. The idea of autonomous search is a first step toward reality mining. With *autonomous search*, the search engine will conduct searches for users without them having to manually type anything [40]. For example, a user could be walking down a street and receive personalised information about the places in the vicinity on his or her mobile phone, without having to click any buttons. While the promise of reality mining is great, the idea of collecting so much personal information naturally raises many questions about privacy.

2.2.4. Social network tracking and profiling

2.2.4.1. Online Social Networks

Online Social Networks (OSN) have gained an immense popularity in recent years. Social-based services such as Facebook, Twitter, MySpace[14] and Orkut[15], to name just a few, allow millions of individuals to share some of their personal information with a multitude of other entities, such as their friends, companies or even the public at large. The common characteristic of these OSN is that users can make contacts and easily share personal information on a large scale. More specifically, people can meet old as well as new friends (Facebook, MySpace), find new jobs (LinkedIn[16]), or receive and provide recommendations (Tribe[17]). In the near future, many more complex services are likely to appear, which will tap into the power of the social connection and personal information provided by OSN.

As the primary objective of most of these services is to make individuals or groups visible, people need to share personal information to ensure some form of identifiability. Hence, most OSN encourage users to publish personal information, which may enable anyone accessing this information to infer further private information, thus causing a privacy breach. On top of that, the majority of users are not only willing but also pleased to disclose their personal information to as many users as possible and some OSN make this information public by default. Moreover, compared to traditional off-line, real-life, social networks, OSN are usually larger and contain more ties. For instance, people easily classify thousands of users as 'friends', or as 'friends of friends', when they probably would not qualify some of these users as friends in their real life. These facts inherently entail the question of trust and privacy in OSN.

Generally, average users do not have a clear idea of who accesses their private information, or of what portion of it really needs to be accessed by applications. For instance, in Facebook, the terms of use of some applications clearly state that they can access any personal information provided by the user, even though it may not be required. Although most sites provide coarse-grained privacy controls, the majority of users do not use this feature because they find it too complex [46]. Moreover, these sites are permissive and allow anyone to access users' profile data, which means that, by default, it is accessible by any other user in the network. In addition, it is difficult for an average user to know and control users or groups of users who can access his information and to limit this access without losing the benefits of the various features of OSN.

Another problem stems from the fact that while a user's profile may be set to be inaccessible for other users, the friendship links and group affiliations often remain public. This public social information can leak further information about the private attributes of a profile. For instance, Ghetoor and Zheleva [47] have shown that the structure of the social network and group information leak indirectly a surprisingly large amount of personal information. Moreover, even if a user makes some parts of his or her profile private, the person's membership of a particular group remains publicly accessible from the group profile. Another study led by MIT students, called the Gaydar project, has shown that it is possible to predict with a fairly high accuracy the sexual preferences of an individual. This is possible even if his or her profile is private, just by looking at the number of gay friends it includes, compared with a person sampled randomly from the population [48].

14 http://www.myspace.com/
15 http://www.orkut.com/
16 http://www.linkedin.com/
17 http://www.tribe.net

Furthermore in [49], the authors show that, much like traditional web sites, third party aggregators track user activity pervasively in OSN. Third party domains are then not only able to track the web sites that a user visits, but also the OSN sites to which he or she connects. In a follow-up work [33], the same authors demonstrate that PII belonging to any user, such as name, gender, or OSN unique ID, is also being directly leaked to these third party servers via the OSN. This leakage happens via a combination of HTTP header information and cookies being sent to third party aggregators. This result implies that third parties are not just able to view the surfing habit of some users, but are also able to associate the habits with a specific habit and potentially gather much more personal information. This ability to link information across web sites and OSN raises important privacy concerns.

2.2.4.2. Mobile Online Social Networks

Mobile Online Social Networks (mOSN) have recently grown in popularity. Mobile devices provide ubiquitous access to the web and naturally to social networks. There are typically two classes of mobile OSN: (1) traditional OSN (such as Facebook, Twitter) that have created content and access mechanisms tailored to mobile devices, and (2) new mOSN, such as Foursquare and Loopts, created to deal with the new mobile context.

These new mOSN tend to customise their content to the location and the user's community (friends). For example, using the phone's self-location features, as well as information about the prior activities of the user's friends, some mOSN propose new places to explore or activities to try. Other mOSN allow a user to locate friends who are currently in his or her vicinity. The predominant concepts of new mOSN are *presence and location* [50]. *Presence* allows a user to know the current status of his or her friends. The indication of presence allows the expectation of a quick response. *Location* allows a user to locate his friends and obtain location-based services, such as the closest restaurants or hotels. A recent study showed that most mOSN leak some kind of private information to users within the same mOSN, to users within other OSN via the interconnect features and, more importantly, to third-party tracking sites [50]. In many cases, the data given out contained the user's precise location, his or her gender or name, and even subject's unique social networking identifier, which could allow third party sites to connect the records they keep of users' browsing behaviour with their profiles on the social networking sites.

The combination of location information, unique identifiers of devices, and traditional leakage of other personally identifiable information, now give third party aggregation sites the capacity to build a comprehensive and dynamic portrait of mobile online social network users.

2.2.5. Some existing solutions

Unfortunately, there is no easy way to use modern, cookie- and JavaScript-dependent websites and social networking sites and avoid tracking at the same time [51]. However, although not perfect [52], the private browsing modes of major browsers, that disable cookies, should be used when possible. Also, the popular Firefox NoScript extension should be considered. NoScript [53] is a Firefox add-on that allows executable content such as JavaScript to run only if it is being hosted on a trusted domain. Finally, anonymization networks, such as TOR [54], and network/web proxies that allow users to surf the Internet anonymously, mitigate some of the highlighted privacy issues (cf. Section 3.2).

As suggested in [35], privacy-invasive marketing practices need greater scrutiny. More research is needed to reveal how the other kinds of cookies described in [31] are also being used to track users. There is a lot of work to be done to bring these next-generation cookies even to the same level of visibility and control that users experience with regular HTTP cookies. Application and browser developers should do more to let users control how they are being tracked. However, this is not an easy task since, as shown previously, some of these tracking cookies, such as the Flash ones, are stored outside of the browser. The BetterPrivacy Firefox plugin tries to address this problem by finding Flash cookies on the hard drive and regularly deleting them.

2.2.6. Conclusions and recommendations

As illustrated in section 2.2., users are being constantly tracked and profiled when using the Internet. This profiling will increase with the development of ubiquitous advertising and personalised services.

In this context, it will be challenging to protect users' privacy. Some people argue that abstinence or withdrawal from the online word is the only method guaranteed to work [55], or that users should lower their privacy expectation. According to Eric Schmidt, CEO of Google, it is possible to identify a person from 14 of his or her photos and then search the web for more content about this user. Furthermore, he argues that, in the future, not only will we be able to identify a person, but also predict, from his or her messaging and location, where that person is going to go [56].

In order to protect privacy, users should be given the ability to control access and distribution of their personal data. Once data is used without the knowledge or consent of the user, privacy is clearly compromised. Solving these privacy issues will be beneficial not only to the users, but also to service providers. In fact, as argued in [30], users might react to this privacy fear by restricting the information they provide, or by giving false information. This would have the effect of limiting business and also affect the validity of customer databases and profiles.

Privacy issues in behavioural profiling are complex and cannot be treated exclusively by legal or technological means. There is a need for a true research approach that considers education, policy, legal and technological aspects.

Users should be informed about the potential danger of tracking and profiling. They should be educated about the privacy threats they are exposed to when they reveal their location, when they surf the Internet, or when they publish personal data on the Internet or social networks. They should be informed that any published information never disappears from the Internet and might eventually leak to unwanted destinations. Data is the pollution of the information age. It stays around forever [57, 58]. This data pollution creates many privacy concerns, since the lost content can be used to collect information about users without their consent.

Recommendation #6. *Users should be informed that any published information never disappears from the Internet and might eventually leak to unwanted destinations.*

Recommendation #7. *Privacy issues in behavioural profiling are complex and cannot be treated exclusively by legal or technological means. There is a new requirement for a true multidisciplinary research approach that considers education, policy, legal and technological aspects.*

In Europe, the data protection (95/46/EC) and the Privacy and Electronic Communications directives (2002/58/EC) provide some protection against marketing practices, such as profiling. The core principle of these directives is to guarantee that users are able to make an informed choice [30] and that they consent to being tracked using an 'opt-in' mechanism. However, as argued in [30], the critical aspect of permission-based tracking lies in the issue of *meaningful consent*. In particular, in order to make an informed choice, users must understand privacy policies, which are often very long and hard to decode. This issue is exacerbated on mobile phones because of the size of their screen. A simple, yet promising, approach is the *Do Not Track (DNT)* initiative [59]. DNT gives users a way to opt out of behavioural tracking universally. In its simplest form, DNT is implemented as a HTTP header. This contains a 'Do-Not-Track' flag that indicates to web sites the user's wish to opt out of tracking. This extension is simple to implement in the web browser and there is, in fact, already a Firefox add-on that implements such a header. However, this solution will only be effective if advertisers respect the user's preference not to be tracked. As discussed in [59], there are several possibilities to enforce it, ranging from self-regulation via the Network Advertising Initiative, to supervised self-regulation or direct regulation. Note that the American Federal Trade Commission (FTC) issued a report recently that suggests the implementation of the 'Do Not Track' mechanism [60].

The concept of *privacy certification*, which would label each site and service according to their profiling activities, should be further investigated.

Recommendation #8. *The concept of privacy certification, which would label each site and service according to their profiling activities, should be further investigated.*

Behavioural tracking is tolerated nowadays because it is supposedly only used to collect presumed non identifiable personally information, and therefore does not endanger user privacy. However, as argued in [22], anonymity does not necessarily equal privacy. By influencing his or her decision processes, behavioural tracking undermines user autonomy and therefore privacy. Furthermore, with all the data available on the Internet, any information that distinguishes one person from another can be used as reidentifying data. In other words, any 'anonymous' profile can potentially be de-anonymized, i.e. linked to a person. In the light of these new results, it is questionable whether the concepts of PII and 'anonymous' profiling still make sense, and should not be revisited [61]. These new de-anonymization results and emerging tracking mechanisms clearly need to be considered by privacy laws which, as argued in [62], might need to be updated. Privacy laws should clearly set the limits of behavioural tracking and profiling.

Tools that help users to make informed decisions about the publication of their data or their online activities should be developed. These tools should inform users whether the information to be published can potentially be combined with other data on the Internet to infer sensitive information, such as their identity [63]. ReclaimPrivacy is an example of such tools [64]. *ReclaimPrivacy* is an open tool for scanning Facebook privacy settings and warn users about settings that might be unexpectedly public. The CMU Privacy Nudge and PARC web-based inference projects are two other initiatives that go in the same direction [65, 66]. Their goals are to design software that essentially provides real time reminders to users about the privacy implications of what they are going to publish. They are based on computer science techniques, such as machine learning, natural language processing and text analysis, as well as disciplines like behavioural economics.

Recommendation #9. *Privacy and security should be considered early in the design process, particularly since many of the interactions in the backend infrastructure are essentially invisible to the end user, and also less frequently understood by those who are given the responsibility of protecting their interests.*

Users must also be able to choose what data is collected about them and must keep the right to access, modify and delete them. Users should be explicitly informed about how they are being tracked, how their data is being sent/leaked out of their social network sites, by advertisers or others, and the corresponding destination. For example, users should need to acknowledge usage of their location on a per-application basis, or even, for some applications, each time location information is used. Indeed, as argued in [43], a user might agree to have some of his or her photos geo-tagged. However, for other photos, for example family photos, he or she might oppose to it or might require the location resolution to be lowered.

Recommendation #10. *Users' identities (and other data, i.e. localisation) should be requested and used only when strictly necessary. Users should be aware of which data is being collected and how they are used.*

For example, data collection should be minimal and only performed when necessary. Services should potentially be distributed and open source to minimise data monitoring and collecting [67]. They should request and use users' identities only when strictly necessary. For example, most location-based services request users to provide their identity before offering their services. This is required for accounting and billing purposes. However, the only thing that service operators actually need is an anonymous proof that the user is a registered subscriber [39]. This can be achieved, without revealing the user's identity, by applying existing cryptographic primitives [68]. Finally, networks, and in particular the Internet of the future, should be designed to limit unnecessary data collection and give individuals control over their data [57, 58].

2.3. On monetising privacy

As can be seen in section 2.2., profiling and tracking are increasing and provide useful behavioural information for online marketing, advertising and targeted services. The profiling and tracking would not advance if such practices did not produce economic benefits. As users are willing to trade their personal information for direct benefits (for discounts, in the case of loyalty cards), or they do not have an alternative for certain services, new economic goods are available on the market: personal information.

Although ENISA recognises privacy as a basic human right, it cannot neglect the fact that in the online market personal data has an economic value; a clear understanding of the mechanisms/models used to monetise personal data is necessary as background information for any policy initiative in the area of privacy. The economics of privacy studies the trade-offs associated with the protection or revelation of personal information **[69, 70, 71, 72, 73, 74]**. Models in the economic literature analyse the welfare effects of information sharing, which may lead to increasing discrimination and behavioural targeting of individuals. Other works use assumptions regarding individual decisions, which are not always aligned with the stated attitude of the individuals regarding their personal information. For example, people might state that they are concerned with their privacy, but then share personal information for little benefit (the so-called 'privacy paradox'). The behavioural economics of privacy attempt to address such aspects **[72, 75]**.

Recent studies are confirming the apparent disjunction between privacy concerns and actual behaviour **[76]**, showing that trust is influencing people's behaviour and that people's attitudes towards privacy can be contextual, that is dependent on the specific situation. In some cases people are willing to ignore their privacy concerns in the context of a trusted requester, while a different attitude is shown in the case of an untrustworthy organisation or individual **[76]**.

In another study, published by the European Commission **[77]**, covering the economic benefits of privacy enhancing technologies (PETs), it has been noticed that in practice individuals are indifferent when it comes to actually purchasing PETs solutions, even if in surveys they show interest in paying a premium for privacy under certain conditions. Based on the same study, which focuses mostly on the data controllers' economic impact, it appears that economic benefit of PETs deployment needs to be assessed on a case-by-case basis; however as technologies mature and become better known the deployment rate could see an increase. The public sector has a role in helping data controllers to realise the benefits of PETs; this could be achieved if technologies are endorsed and there is support for development of PETs. The data controllers' decision to deploy PETs is determined by similar factors to those influencing the demands of individuals (fear of data loss, etc.) but also by the benefits given by the availability of personal data as an economic resource; using PETs might reduce these benefits.

A basic model for better understanding of the economic trade-offs in information sharing and the economic value of personal information has yet to be developed. Social networks and other services (e.g. online search engines) are providing opportunities for co-modification of personal information and for the use of consumer profiles for behavioural advertising.

There is a need for an economic model of privacy, which formalises the relative value of personal information (considering also the bargaining power of individuals, or of individuals being part of a group) and also the possibility of establishing a fee for certain online services. Data protection and information for the consumer about information sharing is of utmost importance in this area and we need to better understand how this is regulated in different European countries and what consequences these regulatory difference have for individuals.

Recommendation #11. *ENISA should support development of an economic model for monetising privacy.*

Recommendation #12. *Users should have access to online services with the right not to be profiled or tracked unless they have given their explicit consent. Moreover, they should be made aware of the scale and impact of the tracking and profiling.*

At the same time, support is required for policy recommendations that would allow both services available free-of-charge, and/or sustained by advertising and the provision of fee-based services.

ENISA should support these objectives in its future work.

2.4. Transparency-enhancing technologies

2.4.1. Background

Transparency of personal data processing is an important principle for the individual's privacy, as well as for a democratic society. A society, in which citizens could not know any longer who does, when, and in which situations, know what about them, would be contradictory to the right of informational self-determination. Hence, the privacy principle of transparency of personal data processing enforced by most western privacy laws, including the EU Data Protection Directive 95/46/EC [7], provides data subjects with extensive information and access rights. Pursuant to Art.10 EU Directive 95/46/EC, individuals about whom personal data are obtained have the right to information about at least the identity of the controller, data processing purposes and any further information necessary for guaranteeing fair data processing. If the data are not obtained from the data subject, the data subjects have the right to be notified about these details pursuant to Art.11. Further rights of the data subjects include the right of access to data (Art.12 a), the right to object to the processing of personal data (Art.14), and the right to correction, erasure or blocking of incorrect or illegally stored data (Art.12 (b)).

Transparency is also an important means for enhancing end user trust in applications. Social science researchers within the PRIME project[18] [78] stated that trust in an application can be enhanced if procedures are clear, transparent and reversible, so that users feel in control. This also corresponds to the findings of Trustguide [79], which provides guidelines on how cybertrust can be enhanced.

Transparency-enhancing technologies provide tools to the end users, or their proxies acting on behalf of the user's interests (such as data protection commissioners), for making personal data processing more transparent to them. A transparency-enhancing technology (or short: transparency technology or transparency tool) for privacy purposes can be defined as a technical tool that has one or more of the following characteristics (see also [80]):

1. It provides information about the intended collection, storage and/or data processing to the data subject, or a proxy acting on behalf of the data subject (e.g., a data protection commissioner), in order to enhance the data subject's privacy.

2. It provides the data subject with an overview of what personal data have been disclosed to which data controller under which policies.

3. It provides the data subject, or nominated proxy, with online access to his or her personal data, to information on how they have been processed and whether this was in line with privacy laws and/or negotiated policies, and/or to the logic of data processing in order to enhance the data subject's privacy.

4. It provides 'counter profiling' capabilities to the data subject, or proxy, helping him or her to 'guess' how the data match relevant group profiles, which may affect future opportunities or risks.

2.4.2. Example technologies

Transparency technologies with the first characteristic are tools and HCI (Human Computer Interaction) components that make the privacy policies of services sides more transparent, such as the P3P[19] privacy bird[20] or other P3P user agents. In section 2.5. on HCI (Human Computer Interaction) for policy display and informed consent, we will put forward HCI techniques for presenting the core information of privacy policies in a user-friendly and more transparent manner.

18 EU FP6 project PRIME (Privacy and Identity Management for Europe), www.prime-project.eu
19 P3P – Platform for Privacy Preferences Project, http://www.w3.org/P3P/
20 http://www.privacybird.org/

An important transparency tool fulfilling characteristics 2 and 3 is the data track that has been developed in the PRIME and PrimeLife[21] projects [81], [82]. The data track is a user side transparency tool, which includes both a history function and online access functions. The history function keeps for each transaction in which a user discloses personal data to a communication partner, a record for the user on which personal data are disclosed to whom (i.e. the identity of the controller), for which purposes, which credentials and/or pseudonyms have been used in this context, as well as the details of the negotiated or given privacy policy. These transaction records are stored at the user side in a secure manner (protected by the PRIME core). User-friendly search functionalities, which allow the user to easily get an overview about who has received what data about him or her, are included as well. The online access functions allow end users to exercise their rights to access their data at the remote services sides online and to correct or delete their data (as far as this is permitted by the services sides). In this way, they can compare what data have been disclosed by them to a services side with what data are still stored by the services side. This allows them to check whether data have been changed, processed or deleted (in accordance with data retention periods of the negotiated or given privacy policy). Online access is granted to a user if he or she can provide a unique transaction ID (currently implemented as a 16-byte random number), which is shared between the user (stored in his or her data track) and the services side for each transaction of personal data disclosure. In principle, this also allows anonymous or pseudonymous users to access their data.

Usability of the history and online access functions were main design goals for the data track. For the development of a usable data track, an iterative design based on four cycles of user interface prototyping, usability testing and refinements was followed (see also [81]).

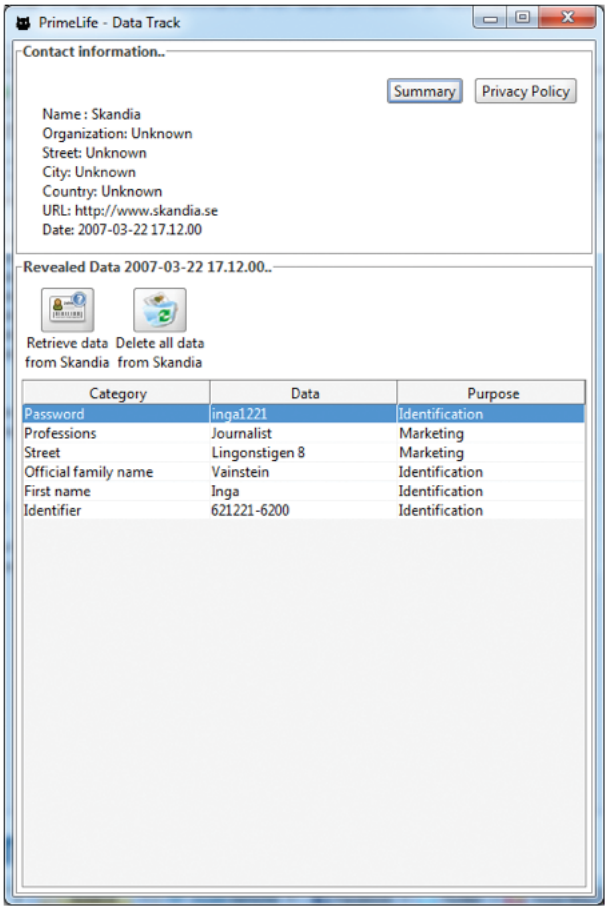

Figure 2.1.: Data Track window showing the details for one specific transaction of data disclosure.

21 EU FP7 project PrimeLife (Privacy and Identity Management for Life), www.primelife.eu

The window displayed in Figure 2.1., for instance, provides an example of the data track user interfaces that were the result of this iterative design process. It provides the user detailed information about what data were sent in that specific transaction, the data values, purposes for which the data were sent, as well as a link to the details of the negotiated privacy policy (see Figure 2.2.). After clicking the 'Retrieve data from <recipient>' button, the data values that are currently stored at the remote services side are listed as well (displayed in green when the values at the remote site match the data values that were previously disclosed and in red if there is a mismatch).

Further examples of transparency tools with the second characteristic (i.e. with history functionality) include 'iJournal' [83], 'iManager' [84] and Microsoft CardSpace [85]. 'iJournal', as part of Mozilla Privacy Enhancement Technologies (MozPETs), lets users define what data they want to keep track of and then analyses the released data for that information, thus keeping track on what information was released and when. 'iManager' was designed for use with PDAs and mobile phones and makes it possible for users to attach contexts (application, location and receiver) with identities, thus showing what data are handed out. However, to our understanding, there is no way of knowing what data have been released if the user dynamically changes his or her identity during a session or changes the settings of an identity. Thus, even though it keeps track of data that were released, it does not really have any history functionality. Microsoft CardSpace also has also some transaction tracking capabilities, although they do not include detailed information about negotiated policies.

Furthermore, there are some commercial and E-Government systems, which have services side functions that provide individuals with online access to their data and allow them to rectify and/or delete their data and thus fulfil (at least partly) the third characteristic. They are usually not only motivated by the goal to enhance the users' privacy, but are also seen as a means for improving data correctness and quality. Amazon's Recommendation Service, for instance, allows users to view which of their previous purchases were used to generate a recommendation and allows users to remove any purchases as input for their recommendation profile (see Figure 2.2.). It thus permits users to view and change parts of their profiles, even though the insight given into the processes and profile creation is rather limited (see also [80]). Also, Google Dashboard grants its users access to a summary of the data stored in a Google account, including account data and the users' search query history, which are however only a part of the personal data that Google processes. It does not provide any insight into how these data provided by the users (e.g. the users' search queries) have subsequently been processed by Google. Besides, access is provided only to authenticated (non-anonymous) users.

Figure 2.2.: Amazon's Recommendation Service, which allows users to view and change parts of their profile data

Further examples of transparency tools that allow users to view and control how their personal data have been processed and to check whether this is in compliance with a negotiated or given privacy policy are based on secure logging systems. Those that exist usually extend the Kelsey-Schneier log [86] and protect the integrity and confidentiality of the log data. Such a secure logging system and an automated privacy audit facility are key components of a privacy evidence approach proposed by [87]. This privacy evidence system allows a user to inspect all log entries that are recording actions of that user with a special view tool and allows him or her to send the log view created by that tool to the automated privacy audit component, which compares the log view with the privacy policy in order to construct privacy evidence. This evidence provides an indication to the user as to whether the privacy policy has been violated.

Unlinkability of log entries, which means that they should in particular not be stored in the sequence of their creation, is needed to prevent an adversary from correlating the log with other information sources, such as other external logs, which could allow him to identify data subjects to whom the entries refer (cf. Hedbom et al. 2010 [88]). Also, anonymous access is required to prevent an adversary from observing who views which log entries and by this concluding to whom the log entries refer. The secure and privacy-friendly Logging for eGovernment Services presented by Wouters et al., 2008 [89] addresses unlinkability of logs between logging systems in eGovernment, rather than the unlinkability of log entries within a log. Moreover, it does not address insider attacks, nor does it allow anonymous access to log entries.

Within the PrimeLife project, a secure logging system, which is part of the PRIME Core and based on multiple keyed hash chains, has been developed, which addresses these aspects of unlinkability of log entries and anonymous user access. In particular, it fulfils the following requirements (see [88], [82]):

- Only the data subject can decrypt log entries after they have been committed to the log
- A user can check the integrity of his log entries. A service provider can check the integrity of the whole log file
- It is not possible for an attacker to secretly modify log entries, which have been committed to the log before the attacker took over the system (forward integrity)
- It is practically impossible to link log entries, which refer to the same user
- For efficiency reasons, it should be possible for a data subject to read his log entries without the need to download and/or fully traverse the whole log database

To the best of our knowledge, there are no practical transparency enhancing tools yet fulfilling characteristic 4 of our definition above (see section 2.4.1.). However, the need and requirements of such tools, which can anticipate profiles that may be applied to a user and, based on this additional information, can perform a kind of counter profiling for the user, have been analysed within studies of the FIDIS project (see for instance [90]).

2.4.3. Conclusions

With the advance of modern communication technology, including sensor networks and ambient computing technology, transparency is increasingly at stake, and hence transparency-enhancing tools will gain more and more importance. In this section, we have argued that transparency is not only a basic legal privacy principle, but also a mean for enhancing trust. Characteristics of transparency-enhancing tools were listed, which can be seen as high-level requirements for implementing the privacy principle of transparency effectively.

Several research prototypes and commercial tools for enhancing transparency have been developed within recent years, from which a selection of the most relevant ones was presented. Even so, further research and development work is needed on transparency enhancing tools, which are designed to show how data have been processed, by whom and with what implications, in an easily understandable, user-friendly way.

Recommendation #13. *Further research and development work is needed on transparency enhancing tools, which are designed to show how data have been processed, by whom and with what implications, in an easily understandable, user-friendly way.*

Also, within the scope of cloud computing, more complex transparency enhancing tools need to be implemented to show how personal information is handled and spread among a number of different service providers. Increased transparency can also help to detect privacy breaches by service providers and this helps to make them accountable for their actions.

2.5. HCI (Human Computer Interaction) for policy display and informed consent

Privacy-enhancing identity management has the objective to (re-) establish users' control over their personal spheres. This implies that users can make informed choices about releases of personal data, the selection of credentials for proving personal properties, and about their privacy and trust policy settings. For enabling users to make well-informed decisions, user interfaces (UIs) are needed that inform them about the trustworthiness and the privacy policies of their communication partners, as well as the implications of personal data releases. These user interfaces should be informative, while at the same time not be perceived as intrusive. They also need to be intuitive, legally compliant and trustworthy.

In this section, we will provide HCI guidance on how end users can be informed about privacy policies in a user-friendly manner and how user interfaces can be designed to obtain really informed consent.

2.5.1. Multi-layered privacy policy display

Users need to be well informed about the implications when releasing personal data upon certain actions, such as login, registration, payments, etc. Art. 10 EU Directive 95/46/EC requires that data subjects in these situations are at least informed about what personal data are processed, by whom (i.e. the identity of the controller), and for what purposes. Today, many web sites simply provide a link to a privacy notice consisting of long legal phrases that are usually not comprehensible for most end users and very seldom read by them.

In order to make privacy policies and their core elements more understandable and transparent, the Art. 29 Working Party recommends that privacy policies are presented in a "multi-layered format under which each layer should offer individuals the information needed to understand their position and make decisions" [91].

The Art. 29 Working Party suggests that three layers of information be provided to individuals:

- The short notice (layer 1) must offer individuals the core information required under Article 10 of the Directive 95/46/EC, which includes at least what data are requested, the identity of the controller and the purpose of processing. In addition, a clear indication must be given as to how the individual can access additional information
- The condensed notice (layer 2) includes in addition all other relevant information required by Art. 10 of the directive, such as the recipients or categories of recipients, whether replies to questions are obligatory or voluntary and information about the individual's rights.
- The full notice (layer 3) includes in addition to layers 1 and 2 also 'national legal requirements and specificities'. Figures 2.3. and 2.4. provide examples for a short and a condensed privacy policy (see [91] – Appendices).

2.5.2. Policy icons for displaying policy components

The approach of multiple layers can be extended by adding standardised icons to the short notices at the top level for illustrating policy elements in an abbreviated and easily noticeable form. A first set of Creative Commons[22]-like policy icons was proposed by Rundle [92], but this did not match European privacy principles. For instance, her icon set included icons for indicating that a services side takes reasonable steps to keep a user's data secure and grants users the right to access their data. However, according to the EU Data Protection Directive 95/46/EC, services sides have a legal obligation to take reasonable measures to secure personal data (Art. 17) and to grant data subjects access to their data (Art. 10) – hence in Europe, these rights and obligations should anyhow be mandatory parts of privacy policies and thus do not need to be displayed prominently by icons. Another set of icons was later proposed by Mehldau [93], but this was not tested and was also partly based on metaphors (such as reordering of messages as a metaphor of anonymity), which we assume to be difficult to grasp for non-technical users. Further sets of icons for e-mail are currently being developed within the Privicons project[23] by researchers at Stanford University and PrimeLife. (see [94]). Also, the privacy icons project by Aza Raskin[24] is currently developing icons for making online policies more understandable. However, not many results have been published yet and the few icons that have been proposed so far are not all in line with EU privacy principles, but rather targeted at the US privacy legal system.

Within the scope of the PrimeLife EU project, an improved set of policy icons has been developed by the Independent Centre for Privacy Protection in Kiel, Germany, which can be used to illustrate core privacy policy statements, namely statements about what types of data are collected/processed, for what purposes, and what are the processing steps [94] [95]. A first intercultural comparison test of the PrimeLife policy icons was conducted at Karlstad University in the form of a paper mock up test with 17 Swedish and 17 Chinese students, which gave the first indications of which icons seem to be easily understandable and which require improvement. Icons which were easily understood by both Swedish and Chinese students were, for instance, the following ones displaying types of data (personal data, medical data, payment data), processing steps (storage, deletion), and the purpose, that is 'shipping'.

Figure 2.3.: Example of a Short Privacy Notice [91].

22 Creative Commons, website available at: http://creativecommons.org/
23 Privicons project, http://privicons.org/
24 http://www.drumbeat.org/project/privacy-icons

Figure 2.4.: Example of a Short Privacy Notice [91].

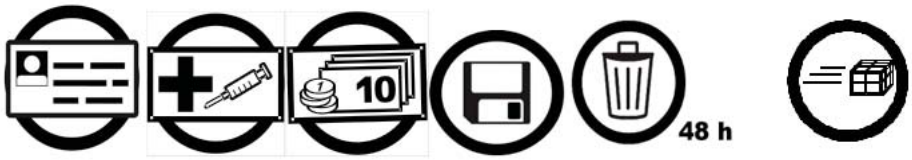

Figure 2.5.: Policy icons test conducted at Karlstad University. Easily understood icons

The following icons for the processing steps 'user tracking' and 'anonymization', are examples of icons which were not well understood by most test users:

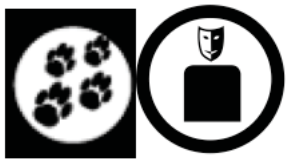

Figure 2.6.: Policy icons test conducted at Karlstad University. Not well understood icons

Our hypothesis, which however needs to be validated further, is that icons for technical concepts, which users are not very familiar with, are also more difficult to understand.

The tests also showed that participants with different cultural backgrounds had different understandings of some of the icons. For instance, while Swedish participants had no problems in understanding the 'post horn' as an icon for the function of 'shipping', this icon was not understood by their Chinese counterparts. A particular challenge is therefore to find icons that are well understood by different cultures.

An online survey was conducted at the Center for Usability Research (CURE) in Vienna with 70 participants, mainly from Europe (see also [95]). This test did not, however, evaluate the participants' understanding, but asked them to rate the icons and how easy to understand they perceived them to be, in comparison with alternative icons. In this online test, the icons depicted in Figure 2.5. received a rating as 'good' or 'very good' icons, which is in line with the findings of the icons test conducted at Karlstad University. An exception was, however, the floppy disk icon for storage, as many of the online test users considered a floppy disk to be an outdated medium and suggested using a pictogram of a CD ROM or USB device instead. The 'user tracking' icons depicted in Figure 2.6. received a low rating. The icon for anonymisation displayed in Figure 2.6. was, however, perceived as easy to understand. The results of both tests will be used to select a smaller set of icons, which will be elaborated further and/or will be candidates for standardisation.

2.5.3. Icons for displaying policy (mis-) match information

The icon of a bird in different colours and shapes and accompanied with different sounds has been used by the AT&T Privacy Bird[25] tool, which is a P3P user agent that allows the user to specify his or her privacy preferences regarding a web site's data handling policy. The privacy bird uses the traffic light metaphor for displaying information about the compliance of a site's policy with the user's preferences; if a site's policy meets the user's preferences, a small green bird icon in the browser's title bar emits a happy tweet after the page has been loaded. If the site violates the user's privacy preferences, the bird icon turns red and chirps a shrill warning when the page is first loaded. For sites that do not have a P3P policy, a yellow bird will appear. It is, however, questionable whether the traffic light is the right metaphor in this context, because having no privacy policy (symbolised by the yellow bird) can actually be regarded as worse than having a policy not matching the user's preferences (symbolised by the red bird).

Figure 2.7.: Privacy bird icons

Figure 2.8.: Icon for displaying a match and a mismatch between the user's privacy preferences (i.e. the user's 'settings') and the services side's privacy policy.

In the user interfaces of the PrimeLife policy engine, icons with matching or mis-matching puzzle pieces are currently used to illustrate whether or not the user's privacy preferences match a site's privacy policy (see Figure 2.8.).

25 http://www.privacybird.org/

2.5.4. Consent via click-through agreements[26]

"The data subject's consent is defined as any freely given specific and informed indication of his wishes by which the data subject signifies his agreement to personal data relating to him being processed" (EU Directive Data Protection Directive 95/46/EC). 'Informed' in this context means that the user fully understands what he is agreeing to and what implications this may have. A special challenge is the development of UI constructs for obtaining really informed and unambiguous user consent for the disclosure of personal data. Ordinary click-through dialog windows with long legal terms, which are often used, may cause users to click the 'I Accept' button too easily if the preference settings have filled in all the requested data for them. Putting up 'Are you really sure?' boxes does not resolve the problem as people may often click the 'I Accept' or 'OK' button even more automatically if they have to go through an extra dialogue box every time [96]. Also Dhamija et al. conclude that when confronted with dialog boxes, such as those for end-user license agreements, users tend to quickly skim the text and efficiently swat away the dialog boxes, without having read or understood what they had consented to [97].

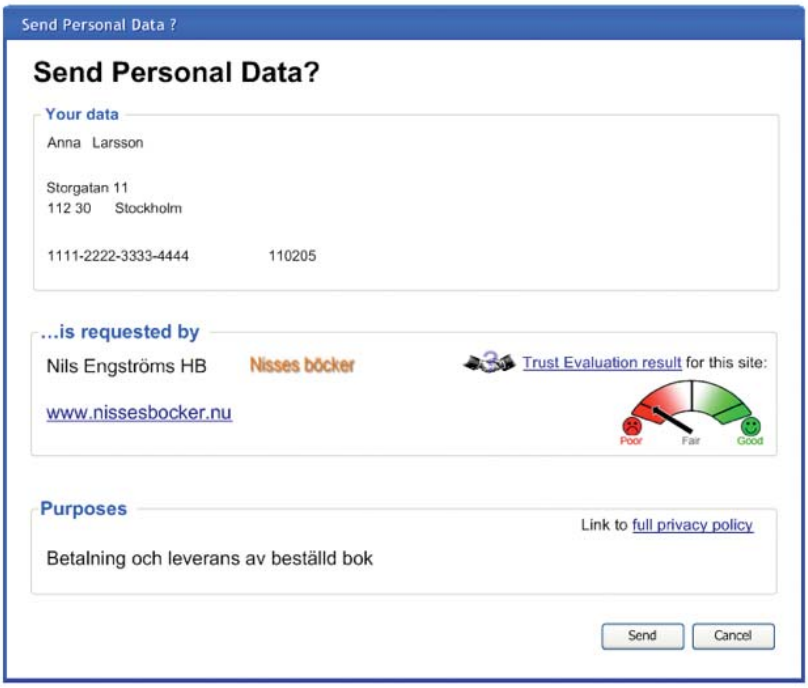

Figure 2.9.: Send Personal data? window as a JITCTA for obtaining informed consent (including trust evaluation result).

To address this problem, 'Just-In-Time-Click-Through Agreements' (JITCTAs), i.e. click-through agreements that instead of providing a large list of service terms confirm the user's understanding or consent on an 'as-needed basis', were developed within the EU FP 5 project PISA on 'Privacy Incorporated Software Agents' [98]. JITCTAs correspond to short privacy notices as defined by the Art. 29 Working Party (see above), which include information about what data are requested, the identity of the controller and the purpose of processing. The 'Send Personal data?' windows developed in the PRIME and PrimeLife EU projects correspond with their forms and content to a JITCTA.

26 In this study we do not cover all types of user in a position to give consent (children, adults with diminished responsibility, etc.).

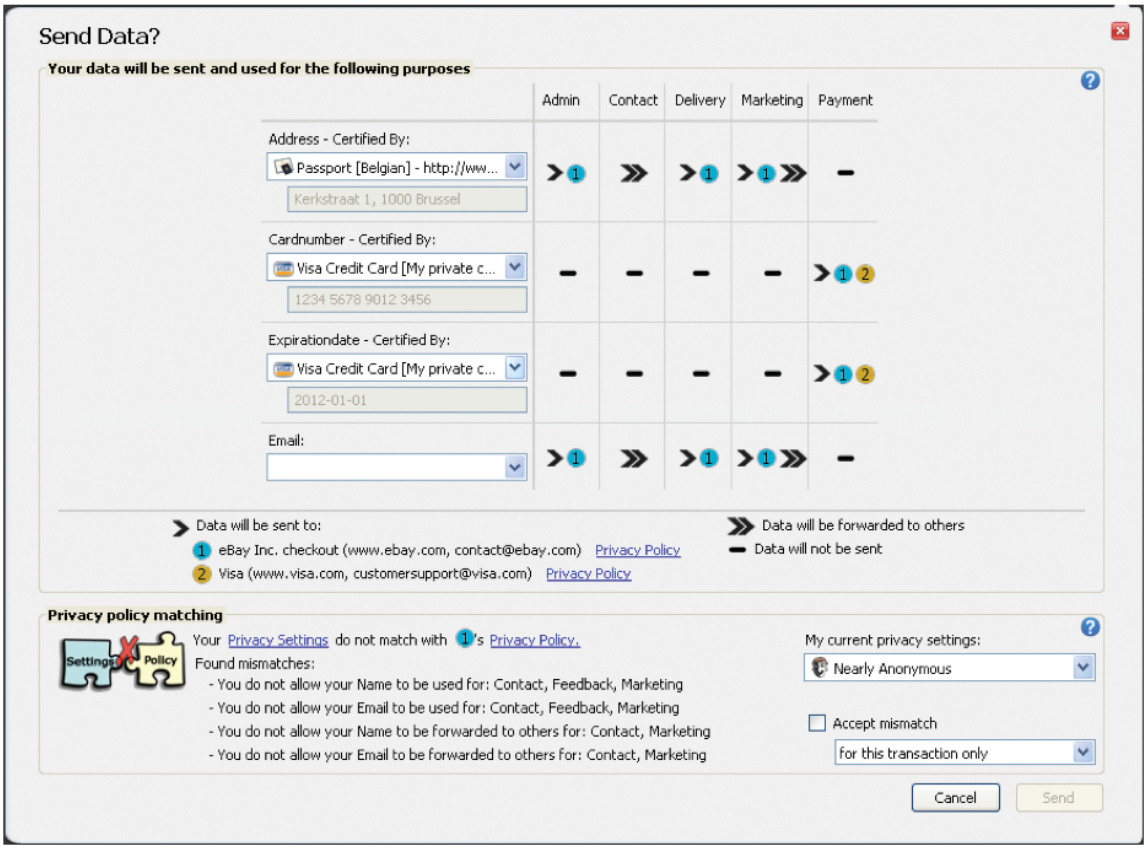

Figure 2.10.: PrimeLife "Send Data?" window (including information about policy (mis-) matches).

Figure 2.9. provides an example of a 'Send Personal Data?' dialogue window developed in the PrimeLife project, which also presents the result of a trust evaluation of the contacted services side. The user consents to disclose personal data to a services side by pressing the 'Send' button. Figure 2.10. provides another example of the 'Send Data?' dialogue, which presents information about what data is sent to whom for what purposes in a 2-dimensional grid, using a very similar presentation to that suggested by [99]. Besides, it also contains information about how far the displayed policy matches the user's privacy preferences.

2.5.5. Consent via menu choices

As mentioned above, click-through dialogs have the disadvantage that users tend to click the 'OK' or 'Send' button too easily, without having read the text. Presenting data items in cascading menus to select data or credentials, as shown in Figure 2.11. by the PRIME project mock up developed by IBM Research Zurich Lab, has the effect that the user must read the text carefully to make the menu choices (see also [100]). Hence, the user is forced to make more conscious selections. Such cascading context menus should also follow the Art. 29 WP recommendation for a multilayered structuring of privacy policies. However, this user interface design is not suitable if many data fields have to be filled in; rather, it is intended as a special feature for very simple data requests, where the user might have to select from among a few credentials asserting a specific data claim.

2.5.6. Consent via DaDAs

As a method for raising the consciousness about the nature of data disclosure, so-called 'Drag-and-Drop Agreements' (DaDAs) were elaborated in the PRIME project in the context of a town map-metaphor-based user interface paradigm [101], [100]. Symbols are used to represent personal data – and users can visibly drag-and-drop data symbols to icons representing the receivers. Here, the user has not only to pick a set of predefined data (corresponding to clicking 'I Accept', 'I Agree' or 'Send' in a pop-up window), but has also to choose the right personal data symbol(s) and drop them on the right receiver symbol. These explicit actions to some extent offer a guarantee for more conscious user consent.

Figure 2.11.: Prototype of a menu-based approach for selecting credentials developed within the PRIME project.

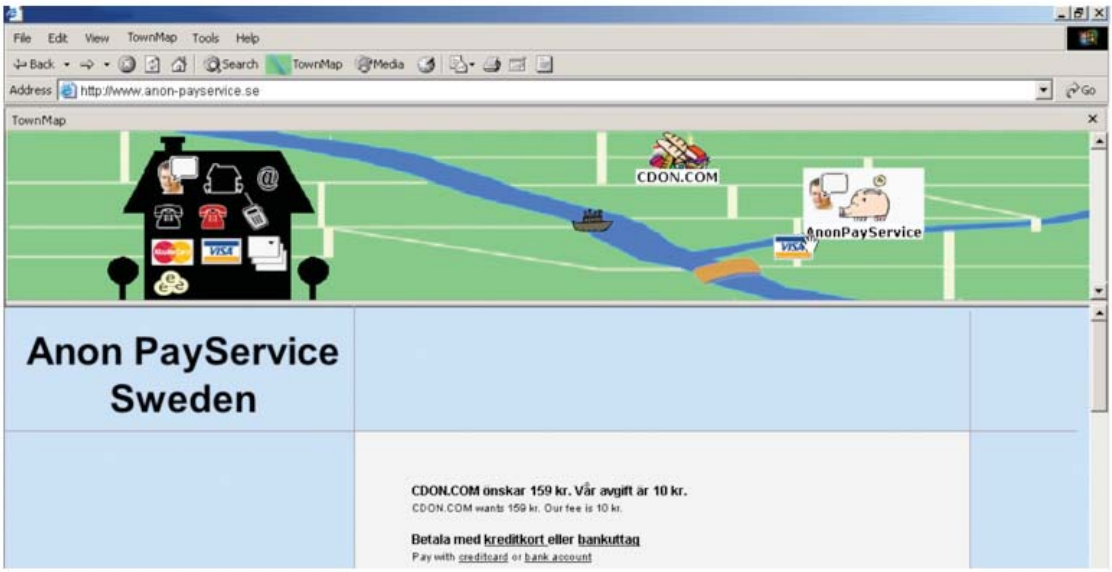

Figure 2.12.: DADA for obtaining informed consent for disclosing credit card data.

2.5.7. Conclusions

This section has presented HCI approaches for addressing the challenge of making privacy policies more transparent and of obtaining well informed user consent. Some of the approaches presented from the PrimeLife EU Project are still work in progress and are currently being further elaborated and tested. UIs for privacy policy display and for obtaining consent often appear in a context, where privacy is only a secondary interest and task for the user, while his or her main task is, for instance, to complete a shopping transaction. Hence, it is a challenge to raise the user's attention, so that he or she makes decisions that are well thought-through. Informed consent is, however, often a means for legitimising personal data processing. Without a user-friendly design of UIs for presenting policies and obtaining well-informed consent, users can easily be misled to consent to terms and conditions to which they would usually not agree.

Recommendation # 14. *Further interdisciplinary research and deployment is needed into the user-friendly design of user interfaces for presenting policies and obtaining well informed consent. For this, cultural differences have to be taken into consideration as well.*

3. Architecture side

"Architecture is politics", **Mitch Kapor**

This section sets out to describe architectural methods to protect privacy. One way to classify approaches to privacy is to distinguish between soft and hard privacy. Soft privacy is rooted in data protection legislation, e.g., the EU Data Protection. The individual user accepts that his data will be transferred to a third party. The basic premise is that after the transfer of the data, it is managed by the data controller. It is the responsibility of the data controller to ensure that data protection legislation is followed. This includes the fact that data can only be processed for a legitimate purpose, as well as taking in the transparency of data processing, as discussed in Section 2.4. In order to ensure that the data is properly protected, the data controller can use a broad range of measures, including security policies, access control, encryption, intrusion detection etc. A typical example is a 'sticky policy' (e.g. [102]) that is tightly coupled to the data. An important aspect of the soft privacy approach is accountability, which is discussed in Section 3.3. A typical setting where this approach is suitable is a health care system.I In order to treat a patient, medical staff require access to the patient's data. It is important to use a combination of technical, logical and organisational measures to prevent this data from leaking to outsiders, and in particular to insurance companies (except the data that is needed to refund the costs).

A limitation of the previous approach is that the user loses control over the data; he or she has to trust that the data controller behaves properly. Typically, he or she has very few tools available to verify that this is indeed the case. In the EU data protection regime, the data controller has to register the database with the data protection authority and indicate how the data is protected; however, most data protection authorities have very few resources to audit and enforce the correct implementation of policies. A different approach consists of building in the transfer of as few data as possible. This approach is known under various terms, which may not have exactly the same meaning: data minimisation, hard privacy and privacy by design. The main idea behind this approach is to reveal as little personal data as possible. The term `privacy by design', coined by Cavoukian [103], refers to a solution in which privacy is part of the architecture, rather than an add-on (cf. Section III.1). Other key principles of this approach are that it is proactive and preventative rather than reactive and remedial. Privacy should be offered by default and should include end-to-end lifecycle protection. In terms of architecture, one can distinguish between two approaches to privacy by design:

- Present your identity but hide what you are doing; this kind of approach can, for example, be used for road pricing (cf. Section 3.5.2.)
- Hide your identity but show what you are doing; this approach is typically used for privacy protection at the lower communication layers

Of course this is an oversimplification; in practice one may encounter hybrid solutions that mix both approaches. Moreover, in some contexts such as communication, it may be simply impossible to hide what one is doing, so there is no option but to hide the identity. Finally, one should always be aware that – unlike with data encryption – perfect hiding is not feasible; powerful and highly motivated opponents can always link and deanonymize actions by bringing together multiple sources of information and by using advanced data mining techniques. As a consequence, technological measures have to be complemented by protection originating in law and policies. However, this should not be used as an excuse to ignore the protection that can be provided by technology.

3.1. Architecture and privacy. Terminology and initiatives

The definition of personal data contained in Directive 95/46/EC [7] reads as follows: "*Personal data shall mean any information relating to an identified or identifiable natural person ("data subject"); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity*". Article 29 Working Party provided this opinion in 2007 on a common understanding of concept of personal data [104].

The terminology varies, based on legal jurisdiction. For instance, in the USA, the notions of personal information and *personally identifiable information (PII)* are used and defined by different regulatory authorities (U.S. Department of Homeland Security, U.S. Office of Management and Budget, states, etc.). Different definitions for personal data and personal information imply various ways to consider the design of privacy friendly systems and applications on the Internet.

In this section some initiatives addressing privacy are described.

The central part in any privacy considerations is the data subject. Identity management is an important topic in all online services, from eGovernment, eBanking or social networks. Some considerations are introduced in Section 3.2. regarding identity management and privacy.

When a user/consumer uses online services not only for information purposes, but also for sharing information, or perhaps for online shopping, he or she needs to know that means for accounting for the actions of the service provider exist for example, ordered services are delivered and repeated charges for the same goods are avoided). From the other perspective, the service provider needs to have means to hold the user accountable for his or her actions, and needs to be able to prove that the organisation respected the legal requirements regarding personal data processing. In Section 3.3., accountability – a key concept in ethics and governance – is presented from the technical perspective and in the context of privacy in the online environment.

Trust is an important topic in the online environment. Trust frameworks are presented in Section 3.4. with a clear focus on architectural design and resulting security and privacy properties.

Section 3.5. briefly presents some privacy preserving architectures. Section 3.6. raises the issues of privacy at the lower layers, a topic usually ignored in the design of privacy friendly architectures at the application layer.

Privacy by design and data minimisation principle

Privacy by design (PbD) [103] is often praised by lawyers as an essential step towards better privacy protection; in a world where privacy is jeopardised more and more by new information and communication technologies, the growing view is that part of the remedy should come from the technologies themselves. On the technological front, *privacy-enhancing technologies (PETs)* have been an active research topic in computer science during recent decades and a variety of techniques have been proposed (including anonymizers, identity management systems, privacy proxies, encryption mechanisms, filters, etc.). However, privacy by design is more than the use of PETs; it relies on the idea that privacy requirements should be taken into account in the early stages of the design of a system and can have a potential impact on its overall architecture. In other words, privacy by design represents a paradigm shift: prevention rather than cure.

Privacy by design is a concept originally developed by Ann Cavoukian [103], and has been recently promoted by the European Data Protection Supervisor (EDPS), who issued an opinion and submitted it to the Commission requesting further developments in regulatory framework, such as a 'Privacy by Design' approach that, according to EDPS, should be reflected in the EU data protection legal framework at different levels of laws and policy making [105]. Later in 2010, the European Commission published its 'Digital Agenda for Europe' [4] and throughout the document, the Commission underlines that trust in e-services necessarily includes confidence in the protection of privacy and personal data. The Commission also set as one of its key actions to "review the European data protection regulatory framework with a view to enhancing individuals' confidence and strengthening their rights by the end of 2010". It has also set out its intention to promote and progressively impose on goods and services providers the concept and notion of 'Privacy by Design'. As a reaction to these regulatory framework developments, industry group EOS (European Organisation for Security) is currently working on a White Paper that covers the main industry concerns related to this concept, as well as other privacy related legislation [106]. The need for the principle of PbD can be found, for example, in many different environments, from biometrics or body scanners to software or web services. The healthcare sector was mentioned in the EDPS paper as an example of centralised storage of patients' health related information, where the application of the PbD principle, such as the possibility of minimising data stored centrally, or limiting it to an index using encryption tools, anonymizing data once they are no longer required, etc, could lead to significant privacy improvements. Decentralised or semi-decentralised architectures limiting the disclosure of location data to a central point was mentioned as another example, in this case applied to emerging road charge ICT systems (where payments are based on km or road actually used). At a more technical level, this would translate into the requirement that all application programming interface (API) calls that request access to potentially sensitive information are designed to be asynchronous in order to enable the user agent to acquire the user's consent without prompts.

Data minimisation[27] principle, one of the most important sub-principles and probably the most adequate for privacy by design, is expected to play an important role in defining which personal data appears to be necessary to use by whom.

Other privacy by design mechanisms, besides obligatory minimisation, anonimization or forcing consent, could include enforcement such as blocking when 'prohibited' service calls are detected, an obligatory validation step when data is forwarded to be used for purposes other than the original use, obligatory separation (partial identity) etc. These requirements should also be 'embedded' in the orchestration and choreography mechanisms, which sometimes raise concerns regarding the threats against confidentiality, integrity, privacy, anonymity, and availability. Both web service requestors and providers may have privacy policy, requirements and properties that must be taken into account when composing web services. We refer to web service composition driven by privacy or privacy-aware composition. A web service provider, for example, may not want to accept requests issued by a specific IP address, or it may want to put some additional constraints on the composition that clash with the minimisation principle applied at service requester level.

Privacy impact assessment

Consulting practices like *Privacy Impact Assessments (PIAs)* are becoming more common and there are already similar practices for shared services (Federated PIA). The RFID industry in Europe, for example, has already started to work on a PIA framework **[107]** that would require RFID operator to conduct an assessment prior to deployment of an RFID application (although Article 29 Working Party **[108]** objected that their approach was not based on a proper risk assessment and that that approach was limited to a list of controls put in place by the RFID operators). In the case of software and service engineering, the whole lifecycle has to be re-designed in order to take these requirements from the beginning. A relatively recent appearance of secure software engineering should be enhanced with privacy aware software engineering.

When it comes to privacy, there are now some research efforts being made to automate aspects of configuration management. While obtaining information, for example, the configuration management system must uphold the user's privacy. Since notions of privacy may vary from user to user, the most common approach is to allow the user to specify what information can be accessible to whom. This, however, should be taken into account during requirements engineering. In secure software engineering **[109]**, privacy issues are often addressed along with other security requirements, rather than as a separate design criterion in the system development process **[110]**. However, there is a danger that privacy requirements are not fully captured, or that there is a conflict between the privacy-related and security-related requirements. The existing requirements engineering methodology shows that privacy and data protection issues are not properly addressed. There are few examples, such as the Non-Functional Requirements (NFR) framework **[111]** that treats both security and privacy requirements as non-functional (or quality) requirements and models them as soft goals. Most of these approaches do not support the notions and terms, such as secondary use, purpose, identifiability, etc. Another approach has been presented in **[112]**, where heuristics has been used for identifying the goals that systems must achieve and thereby derive privacy policies. Requirements engineering is only the first phase of software engineering lifecycle.

Some other approaches (e.g. Vanish project **[113]**) address some of the PbD challenges by ensuring that data stored in clouds will self-destruct after a period of time. Similar approach is presented in **[57]** as a method to reduce pollution on the Internet and improve privacy by giving back to users control over their data.

3.2. Identity management and privacy

During the past decade we have witnessed rapid developments in the area of identity management infrastructures and standards; however, the effective uptake of these standards has been limited.

From a cryptographic architecture point of view, one can distinguish between three approaches. The first two use symmetric and public key cryptology respectively:

27 www.privacybydesign.ca/content/.../pbd-implement-7found-principles.pdf

- In the simplest setting, only symmetric cryptography is being used; all entities share a secret key with a central party, the identity provider. If a user wants to authenticate to a relying party, he or she requests a temporary key from the identity provider. The identity provider sends a token to the user that contains this temporary key in a protected format and perhaps some user attributes. A second token with the same information is produced for the relying party. This approach is typically only suitable for closed systems. Attempts to expand such systems to multiple identity providers turn out to be very hard to manage. An example of such a system is Kerberos.

- In a more advanced setting, the identity provider has a public-private key pair; it is assumed that all relying parties know the public key of the identity provider. When a user wants to authenticate to a relying party, he or she first authenticates to the identity provider; this authentication process can use a password, a challenge response protocol using a shared secret key or a public-privacy key pair. If the authentication is successful, the identity provider produces a signed token that certifies the identity and other attributes of the user; it also contains key material that allows the user to authenticate him or herself to the relying party (e.g. the users receives a public-private key pair and the relying party receives the corresponding public key). This approach is more open; any relying party who has an authenticated copy of the public key of the identity provider can authenticate the users. This kind of approach is flexible in terms of the communication patterns. SAML is a standard that implements this approach.

These technologies have the property that, for every session with the relying party, the identity provider is contacted; the identity provider knows which services the user is requesting, when and how many times. This presents a serious privacy risk that seems inherent to the design choices. The problem can be mitigated in two ways:

- by caching the tokens, the user can hide how many times and when he or she accesses a specific service (but not that a particular service is being used)

- users can have multiple identity providers and multiple identities; in this way they can reduce the information available to an individual identity provider (and limit profiling)

In the most sophisticated approach to identity management, all entities need public-private key pairs. The user receives a digital signature during the registration phase, compiled by the identity provider from its identity and a number of attributes; in this context, such a signature is called a credential (however, in identity management, credentials are typically plain lists of attributes, so this term can be confusing). When the user wants to interact with a relying party, he or she will not show this signature. Instead, the user will prove a logical statement about the attributes in the credential (e.g. prove that the user is older than 18, holds a driver's licence and lives in a city with ZIP code 34xxx). The proof does not consist of handing over the signature; it can be mathematically shown that no other information is revealed than the statement itself; such proofs are called zero-knowledge proofs of knowledge (ZK). Two examples of these systems are Idemix (designed by IBM) [114] and U-prove (designed by Credentica and acquired by Microsoft) [115]. With this ZK approach, the relying party only acquires minimal information. Moreover, if the credentials are `multiple show', the identity provider learns very little information about the activities of the user. The integration of this approach into applications has been pioneered by the EU funded projects Prime and PrimeLife.

There are other approaches to identity management, such as OpenID, that use self-certified claims. OpenID started from the observation that some users can be identified by their blogs. OAuth is a mechanism to delegate access to services. While OpenID and OAuth are easy to use, both architectures have privacy issues similar to the first two types of identity management systems.

A broader perspective on identity management is offered in Section 3.4. on Trust frameworks. For an in-depth study of identity management and the management of multiple identities, the reader is referred to the ENISA Report on Management of multiple electronic identities [116].

Recommendation #15. *Support research on privacy-friendly identity architectures that minimise concentration of information and prevent unnecessary linking, while guaranteeing accountability.*

3.3. Accountability

Accountability is a key concept in ethics and governance, and fundamental to the role of individuals and organisations in society and in the online world. Here, we are concerned with a technical interpretation of accountability, as it pertains to the actions of a data processing system. (Note, however, that the actions of a data processing system can often be causally linked to a responsible individual or organisation that owns or operates the system.)

3.3.1. Definitions

Accountability in the technical sense of the term is a property of a data processing system. In its most general form, accountability offers three capabilities:

1. It allows users, operators and third parties to verify a posteriori if the system has performed a data processing task as expected (validation).

2. In case of a deviation from the expected behaviour (fault), it reveals which component is responsible (attribution).

3. It produces evidence that can be used to convince a third party that a fault has or has not occurred (evidence).

Not all accountability systems incorporate all three capabilities. For instance, some systems provide validation and attribution, but not evidence. Attribution alone is being considered, for instance, as a mechanism to establish the origin of Internet packets [117, 118].

Accountability is relevant in scenarios where an individual or organisation entrusts their data or computation to (or depends on) a processing system that is owned and operated by a third party (e.g., web services, cloud computing, federated systems like the Internet). In such a system, users and participants wish to verify that their computation was carried out and their data processed correctly, either in case of suspicion, or as a matter of course. Moreover, in the case of a fault, users wish to be able to determine the responsible party.

Accountability differs from fault-tolerance in that it does not attempt to mask faults. Accountability differs from fault detection in that it provides evidence and that it covers, in its most general form, arbitrary (including Byzantine) faults.

Accountability systems are related to reputation systems. While accountability is concerned with reliably and precisely accounting for past actions of a system, reputation systems maintain subjective evidence of online users' past behaviour, based on the evaluation of an individual's actions by other users.

Accountability and reputation systems depend on a sufficiently strong form of identity. In particular, an individual must not be able to act under multiple identities, or change his or her identity freely over time. Otherwise, the individual could evade accountability for a fault, or escape a bad reputation simply by switching identities (whitewashing) [119].

3.3.2. Types of accountability

Different levels of accountability can be usefully defined in the online world. Legal Accountability requires that identities can be mapped to an individual or an organisation. It implies that the individual or organisation can be held legally responsible for his or her actions.

For *pseudonym accountability*, it is sufficient to ensure that each individual or organisation uses only one, permanent identity. The identity can be pseudonymous, i.e. the mapping to the individual or organisation it represents can be hidden. Under pseudonym accountability, an identity can be black-listed or banned, restricting the individual's ability to use the identity. For instance an auction site would be able to ban a fraudulent buyer or seller and be sure that the culprit was unable to re-join the site with a different identity.

Note: If the mapping from identity to individual is additionally placed in escrow [120], then the individual can also be held legally accountable under the specific conditions that require the escrow agent to release the mapping.

Accountability can ascertain two properties of a data processing task:

- *Integrity*: the task produces the correct result (relative to its state and a sequence of inputs)
- *Confidentiality*: the task does not leak confidential data to unauthorised parties

To account for integrity, it suffices to verify that a task produces the correct result. To account for confidentiality, a system must monitor the flow of information comprehensively, which is unquestionably harder than verifying that the result is correct.

Accounting for integrity: a number of accountability mechanisms that cover integrity have been developed or proposed in recent years. Many provide accountability for specific applications and systems [121], [122], [123], [124]. Other systems provide accountability for general distributed systems [125], [126].

Accounting for confidentiality (a.k.a. information accountability): there is little work on accounting for confidentiality (as opposed to ensuring confidentiality). Unlike accounting for integrity, which is known to be unquestionably more efficient than fault-tolerance, it is not known whether accounting for information use can be accomplished more efficiently than controlling information use.

Researchers have contended that the use of information should be transparent, so that it is possible to determine whether a particular use is consistent with applicable policies, and be able to hold individuals and organisations accountable for any misuse [127]. Moreover, it has been argued that in an increasingly connected world, this principle of *information accountability* is a more realistic goal to achieve than controlling the use of information [127, 128], and governments are increasingly adopting policies consistent with this principle.

3.3.3. Threats against accountability

Because accountability depends on certain properties of identities, threats against accountability include certain threats against identity management systems, including:

- Sybil attacks: an individual uses multiple identities within a given domain, e.g. to bias the reputation of identities or the rating of objects (ballot stuffing), to circumvent per-identity resource limitations, to control all replicas of an object in a p2p storage system, or to break the anonymity in a p2p mix-net by aggregating information from different vantage points
- Whitewashing: an individual changes identity to escape accountability, e.g. to shed the bad reputation resulting from a fraudulent transaction at an auction site
- Attacks against authentication (e.g. impersonation): an attacker breaks the authentication system to be able to act under the cloak of another individual's identity
- Shared identity attack: several individuals share one identity within a domain, e.g., to avoid per-identity payments or contributions, or to accumulate the rewards for several individuals' actions

In addition, there are threats against the integrity of the collection, storage and transmission of accountability records:

- Falsifying record collection: an attacker omits, falsifies or fabricates accountability records, so that the record does not accurately reflect the actions a system has actually performed
- Tampering with records: an attacker tampers with accountability records during storage or transmission
- Destroying or suppressing the transmission of records: an attacker destroys accountability records, or prevents the records from being transmitted to an authorised auditor

3.3.4. Existing reputation systems

Reputation systems have been extensively researched and appear in many forms in the Internet. Examples include the buyer/seller reputation system in eBay, and the TrustedSource system, which provides reputation scores for IP addresses, DNS domains and Web URLs. Simple reputation systems merely count the number of positive, negative or neutral votes, while more sophisticated systems also consider the reputation of the voter, or the value of the transaction or object that is being evaluated by a vote.

Current reputation systems are vulnerable to a host of attacks. For instance, the eBay reputation system is currently of the simple type. In addition, the eBay account creation process makes it easy for users to create new accounts, opening the door for Sybil and white-washing attacks. For instance, in one common attack, a user creates an account, builds a good reputation based on a large number of penny-item transactions, then defrauds an unsuspecting buyer with a single, high-value transaction and finally abandons the account. The attack is further facilitated by sellers who offer inexpensive items and guarantee a positive evaluation.

3.3.5. Existing work on integrity accountability

Accountability for integrity in distributed systems is a relatively recent subject of research **[129, 130, 122, 121, 123, 125, 124, 126]**. Early work focused on making the case for accountability, later work contributed mechanisms for specific applications like p2p content distribution and cooperative storage, while the most recent work provides integrity accountability for general distributed systems and even for arbitrary binary software images that execute inside an *accountable virtual machine (AVM)*.

One element of a practical accountability mechanism is a tamper-evident log of a distributed execution **[131, 125]**; note that secure logging is also discussed in Section 2.4.2. Briefly, each participating node in a distributed system maintains a local log of all the messages it sends and receives, along with certain other events. The log entries are connected by a hash chain, i.e. each log entry includes a secure hash of the previous entry. Each message includes a cryptographic signature, covering the message contents and headers, and the hash of the most recent entry in the log (which describes the message being sent). In this way, whenever a node sends a message, it commits to the entire sequence of events recorded in the log up to and including the message transmission.

This log is tamper evident: given the local log of a node Alice and the set of messages that other nodes have received from Alice, an arbitrary node Bob can decide whether the log accurately reflects the sequence of actions of Alice, or otherwise prove that Alice has tampered with her log. The log is correct if (and only if) all messages that were sent by Alice appear in the log, are consistent with the entries recorded in the log, and there is a sequence of log entries in Alice's log that connects any pair of messages sent by Alice. Moreover, given the union of logs of all participating nodes, Bob can decide whether the combined log accurately reflects the distributed computation, or pin-point any node that has tampered with its log.

Given the combined, validated log of a distributed computation, an auditor can now verify that the log is consistent with a correct execution of the distributed system. For instance, the auditor can check the log for consistency with a set of invariants, or with a partial or full specification of expected behaviour. A simple and general form of fault-detection is state-machine replay, because it uses a reference implementation as the implicit specification of correct behaviour.

The *PeerReview system* **[125]**, for instance, records additional information in the tamper-evident log to ensure that each node's execution can be replayed deterministically, using a reference implementation of the node's expected software. The replay is able to detect any fault, attack and misbehaviour that involves a modification of the node software or its expected configuration. Moreover, one can show that any such fault in a distributed system can be detected, as long as each node's log is checked and replayed eventually by a trusted node.

Accountable Virtual Machines (AVM) is a generalisation that works for arbitrary, unmodified software images. Here, the software image of each node of a distributed system is run inside an accountable virtual machine. The virtual machine monitor records information to ensure deterministic replay of the virtual machine in a tamper-evident log.

3.3.6. Existing work on information accountability

There is relatively little work on technical information accountability, i.e. on technical means to ensure that information use is accounted for reliably. A general approach is to make sure information is exclusively stored at (or accessed via) agents who can be trusted to keep accurate records of information use. The record may additionally be tamper-resistant or tamper-evident. Unfortunately, in systems where information is shared and disseminated widely, it is hard to prevent attackers from creating unauthorised copies of information and using it in ways that escape accountability.

On the other hand, there is considerable work on controlling access, flow, and use of information, ranging from conventional access control mechanisms, operating systems that control information flow, to systems that control the information that can be inferred from multiple queries to a database, e.g. based on differential privacy (differential privacy attempts to maximise the accuracy of databases while minimising the chances of identifying the records) [132]. Such systems can be combined with tamper-evident or tamper-resistant logs to provide reliable information accountability [133]. Controlling information use appears to be the only known reliable and general method of accounting for information use.

3.3.7. Conclusions

Research on accountability for large-scale information systems is still in its infancy. There is some initial work on *integrity accountability*, but further progress is needed regarding the efficiency of accountability mechanisms (e.g. reducing the cost of maintaining tamper-evident logs, reducing the cost of automatic fault detection), and on managing the tension between accountability and privacy.

The tension between accountability and privacy must be managed carefully. In general, the system should expose information only to legitimate auditors, and only the information strictly necessary for the auditor to verify the properties of interest. By this standard, current accountability techniques are rather verbose. In particular, techniques based on state-machine replay may expose much more information about the execution than necessary – this is a price we pay for the power of verifying an execution without a formal specification of the relevant properties.

Further research should consider how the required information can be automatically inferred from a high-level statement of the properties that the system should account for, and limit exposure of information to that which is strictly required.

Another aspect of limiting the privacy cost of accountability is identity management. In many online systems, pseudonym accountability is sufficient, and there is no need to reveal the individual responsible for an online activity. In such systems, black-listing or eviction from a domain is a sufficient deterrent for undesirable behaviour.

There is very little work on reliable technical means of ensuring information accountability. Given the importance of this topic as part of the strategy to ensure appropriate use of private information, fundamental research on this topic is urgently needed.

Recommendation #16. *Research on information accountability technology should be promoted, aimed at the technical ability to hold information processors accountable for their storage, use and dissemination of third-party data.*

3.4. Trust frameworks

3.4.1. Introduction

In the design of communication systems there has always been the desire to offer users the ability to access services provided by one provider through an already established communication relationship without another provider. One user-perceived benefit is the single login experience that allows users to re-use their long-term credentials established with a single (or a small number of) identity providers.

Throughout this section we will look at various examples of such technical systems, some of which are widely deployed and others that are still at their early stages of deployment. Since each of them defines their own terminology we will introduce a high level description sufficient to illustrate the basic actors these distributed systems are facing. The main challenge these distributed systems have in common is to scale the trust relationships. As we will describe later in the text there is a relationship between the architectural design and the resulting security and privacy properties. As problems (both from a security and a privacy point of view) and usage needs change, there is a desire to evolve the chosen trust framework model as well.

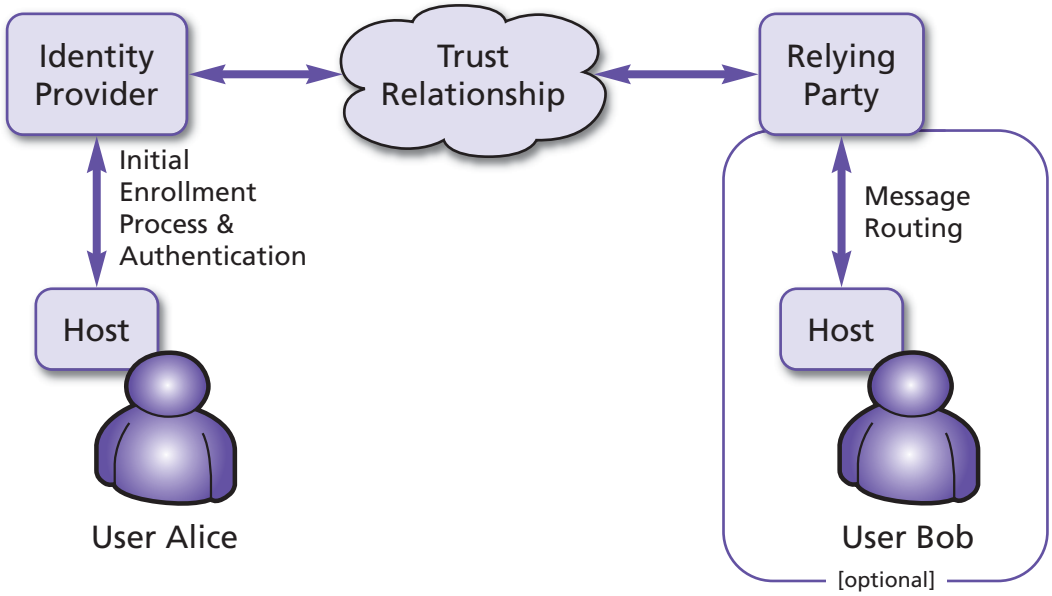

Figure 3.1.: Conceptualizing Trust Frameworks

Figure 3.1. shows the high-level architecture with terminology taken from [134] we use throughout this chapter to illustrate the different architectures. The relying party obtains identity information from the identity provider (also called asserting party) but wants to ensure that this information is genuine and has been verified for correctness and accuracy. For example, a statement about a person's creditworthiness is valuable when provided by well-known credit bureaus, such as Equifax or TransUnion. Self-asserted claims by the user are on the other hand less useful as a basis for decision making in financial transactions. The asserting party, on the other hand, may want to control the distribution of information to different relying parties, based on various criteria, including the user's privacy preferences.

The 'trust relationship' cloud in the middle between the identity provider and the relying party can be as loosely defined as "An association comprising any number of relying parties and asserting parties." The term used for this trust relationship construct varies as the technical and the operational characteristics differ. The terms ‘federation’, or ‘trust framework’ are commonly used terminology.

Note: for the purpose of this document, the term ‘trust framework’ is interchangeable with ‘federation’, with the former term being more frequently used in the web identity management community.

To further clarify the term ‘federation’, E. Maler advocated that “technology [that] is capable of distributing identity information and delegating identity-related tasks (at least including authentication) across domain boundaries. The organizational distances between the parties can be arbitrarily large” and that “federating identities doesn’t necessarily imply the existence of a circle of trust (a federation of business partners ...)” [135].

The communication interaction that happens in that backend infrastructure is not user visible, but there is a corresponding user/end host facing interaction that largely depends on the specific application environment being considered. We will illustrate a number of those exchanges in subsections. Figure 1 does not aim to restrict the interaction sequence between the different parties and there are indeed multiple choices about which we do not discuss the pros and cons in this chapter.

3.4.2. Email architecture

The email architecture is an interesting example because it is widely deployed and has evolved considerably over time. RFC 5598 [136] says, "Over its thirty-five-year history, Internet Mail has changed significantly in scale and complexity, as it has become a global infrastructure service". Figure 3.2. shows the architecture using the terminology from RFC 5598, with the author in the role responsible for creating the message, one or multiple mediators that provide message routing functionality, and the recipient as the consumer of the message. It is important to point out that the design of the email architecture allows the mediator to see the content of the message (unless it is encrypted). Over time, when the number of security problems with email delivery increased (e.g. the amount of email spam), enhancements were provided to offer identity assertions by the author as well as the mediator. An example of this work is the DomainKey Identified Mails (DKIM) signatures [137] developed by the IETF DKIM working group. The main idea is that a mediator, who authenticates the author of the mail message, attaches an assertion to the mail message delivered to the recipient. The recipient verifies the signature and can be assured that it was indeed provided by the indicated domain.

Filtering of mail messages at the recipient happens at different levels; filtering provided by a specific recipient (in the case of user provided policies) as well as by the administrator of the recipient's domain. In this last case, the judgment of what a legitimate author's domain is depends heavily on the identity of the domain with the associated repudiation. In addition to the usage of domain names, used as signed with DKIM IP addresses and IP address ranges are used.

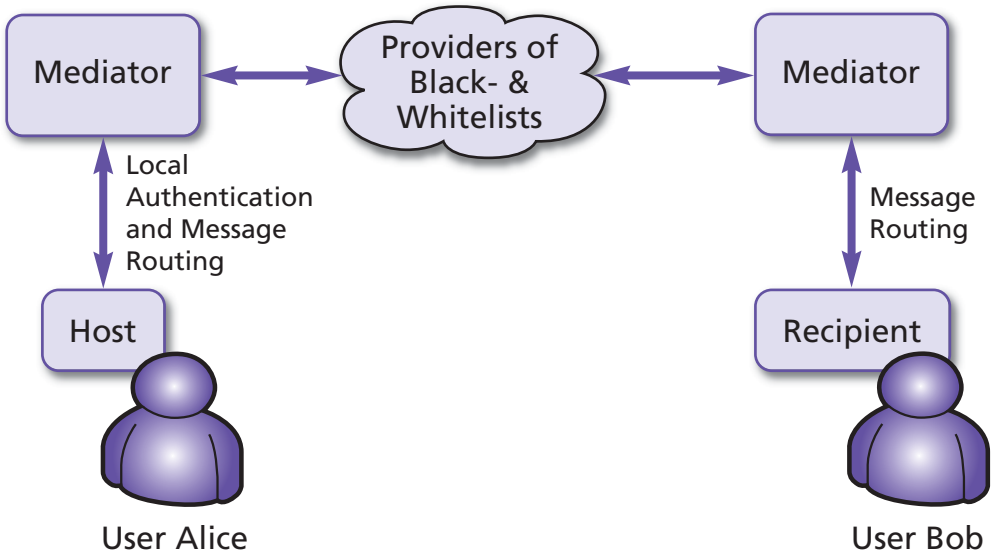

Figure 3.2.: Simplified email architecture

The federation concept sounds alien for the email architecture, since email was developed with a model of full meshed connectivity in mind, where mediators could transmit mails to any recipient by only determining the mail server in the recipient's domain.

This model has changed with the introduction of blacklists that were provided by independent organisations, such as SpamHaus [138]. Recently, whitelist functionality has been added as well.

Domain Assurance Council is another company that provides 'federation' services similar to SpamHaus, trying to advance the state of the art regarding email security. Initially, it was setup to provide a central association for organisations that vouch for:

- the identity of the owners of domain names
- the types of services that those owners use or provide associated with the domain names
- the owners' conformance to good principles for those services.

Their work led in the development of the "Vouch by Reference" specification now published as RFC 5518 [139]. The core idea of the specification is to use the DNS to store whitelists of domains. It relies on secure email authentication mechanisms, such as DKIM. The idea is that certain associations already have established ways of knowing who their 'members' are and these groups can essentially be re-used. For example, the Federal Deposit Insurance Corporation (FDIC) knows about financial institutions and, using this mechanism, they could store the list of banks in the DNS, for example, at fdic.gov.

The use of DNS was chosen to allow real-time verification. Large mail systems deal with over a billion (10^9) messages a day, so their lookup processes have to be fast. The DNS is fast enough, since various DNS based blacklist mechanisms are in use already.

Hence, a recipient would use a DNS lookup to determine whether the sender's domain name is contained in the whitelist. If a certain company behaves inappropriately then their name would be removed from the whitelist and email recipients would instantly know about the updated whitelist.

The Vouch by Reference deployment is still at an early stage, since it depends on the widespread deployment of DKIM.

The main challenge with email security is that the receiver of mail somehow has to determine whether the author or the author's domain is genuine and, moreover, that it sends mail the recipient wants. There are plenty of places that send spam under their own names.

For successful filtering, a strong identity mechanism is the pre-requisite, but someone has to assert that the respective mail server from which the mail comes fulfils certain security expectations. The intuitive approach would be to have assessors who go to companies that would like to become members and perform an audit. However, this is fairly expensive and most companies would not be willing to pay for such an audit.

Additionally, the expectation of the email infrastructure was that everyone can set up a mail server and participate in the exchange of mails.

Hence, various sophisticated procedures have been developed for testing whether a new member application should be granted and the company added to the white list. The exact algorithms are typically confidential, but usually consist of information publicly available on the Internet (such as domain names and IP addresses being used) and further information that can be easily provided by the applicant.

Many mail providers have collected log files over a very long period of time that can be utilised in decision-making, as historical information offers evidence of past behaviour of applicants. This information serves a purpose similar to a Criminal Records Bureau check.

The main advantage of such an algorithmic verification, which does not require human interaction, is in its lower cost. It is interesting to note that these blacklist/whitelist providers do not have to be involved in the email conversation itself. However, they need to obtain information about spam. The work in the IETF Messaging Abuse Reporting Format (MARF) working group [140] aims to standardise the Abuse Reporting Format (ARF) of a widely used mechanism for reporting email spam between mail providers.

Note that whitelisting is not a pre-requisite for using ARF. ARF reports are sent from lots of mail systems, including AOL, Hotmail, Time-Warner Roadrunner, and Comcast. With these reports, the sender is expecting to see less mail that causes user complaints from a well behaved mail provider. Places that have their own whitelists generally provide ARF reports to people on the whitelist, but one can easily have one without the other.

3.4.3. Voice over IP and Instant Messaging architectures

The Voice over IP and Instant Messaging architecture is an interesting example for observing the evolution of federations in a specific application domain. The initial goal in the design of SIP was to build on the success of email with a federation that relied on dynamic DNS-based discovery of the destination proxy that does not require business agreements, or other forms of contract, to be in place. At the time of SIP development, email spam problems did, however, already exist and therefore lots of concerns were raised in the area of unwanted traffic for real-time communication. It is hard to analyse whether these concerns were real or just placeholders for fears about changing business models, but the work on SIP interconnection/peering went in a different direction, namely that where peering providers create pair-wise business arrangements that impact SIP message routing and even the routing of voice traffic itself.

XConnect [141] is an example of such a company providing VoIP interconnection services. Historically, the GSMA [142] has provided interconnection services for the PSTN and their IPX [143] model follows the same paradigm. Unfortunately, many of the contractual details for these commercial peering providers are not publicly available.

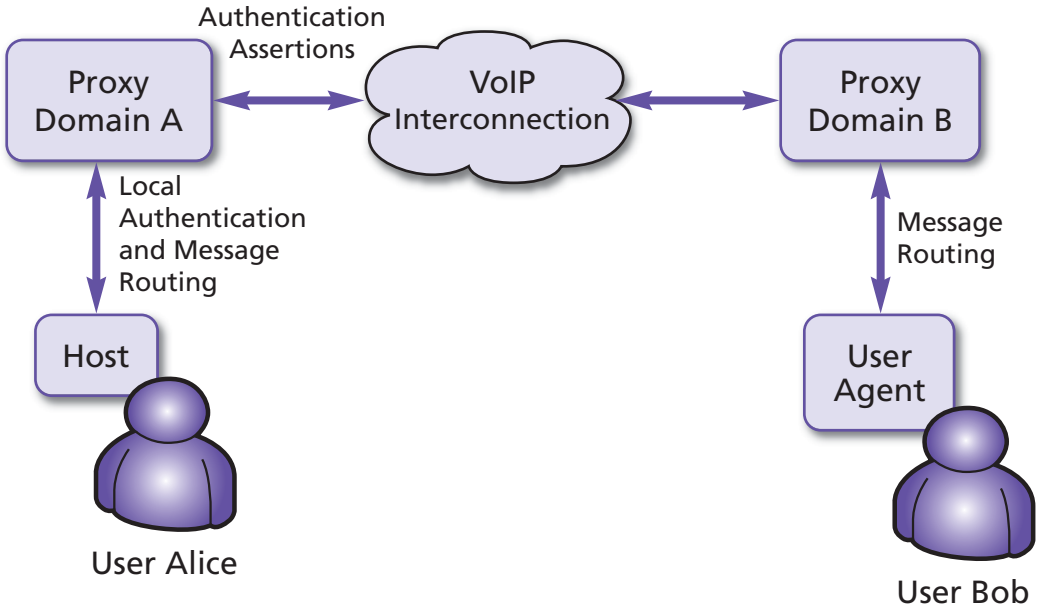

Figure 3.3.: VoIP architecture

From a standardisation point of view, a number of activities are ongoing in the area of VoIP peering, inside and outside the IETF. A number of working groups have been created in the IETF, including SPEERMINT [144], DRINKS [145], and MARTINI [146], trying to develop extensions to the peering space.

Lendl published an overview document [147] in the SPEERMINT working group that tried to offer some explanation as to why the initially envisioned federation goal did not lead to the desired outcome. It is also worth noting that one of the arguments being provided is in relation to unwanted traffic and security risks. A more generic study exploring the topic of interconnection is available with [148]. Clearly the VoIP interconnection architecture is heavily impacted by the PSTN interconnection history.

Figure 3.3. shows the architecture of VoIP with the interconnection providers playing the role of signalling message relays (and optionally voice traffic relays). The placement of these entities as intermediaries had impact on the deployment of the VoIP identity solution, using transitive trust for end-to-end security, as well as on the deployment of protocol extensions.

Recently Rosenberg and Jennings published a series of documents providing an alternative solution to voice over IP peering called VIPR [149]. VIPR is a revolutionary idea that claims to offers a migration path from the PSTN to IP-based calling and aims to solve the E.164 address ownership, a topic that was unresolved for a long time. VIPR relies on a technique for storing authorised mappings from phone numbers to domains in a distributed form, involving Distributed Hash Tables. As such, it does not rely on intermediaries offering signalling message (nor voice traffic) relay services and security functionality.

3.4.4. Web-based federations

The growth of HTTP-based application deployment, due to the increasing functionality of browser based implementations (including the support of JavaScript), has led to the desire to offer identity sharing infrastructures, as well as delegated authentication environments, for secure data access. With the deployment of OpenID, identity providers started to provide better consent-based interactions to let users influence data sharing from the identity provider towards relaying parties.

Figure 3.4. shows a high-level description of the architecture where a user interacts with a relying party (a regular website). Instead of providing credentials to that relying party, the user is redirected to the identity provider for log-in, where the consent for subsequent data sharing between the identity provider and the relying party is requested. Once the initial exchange is completed, the identity provider and the relying party interact based on the agreements established as part of the trust framework agreements.

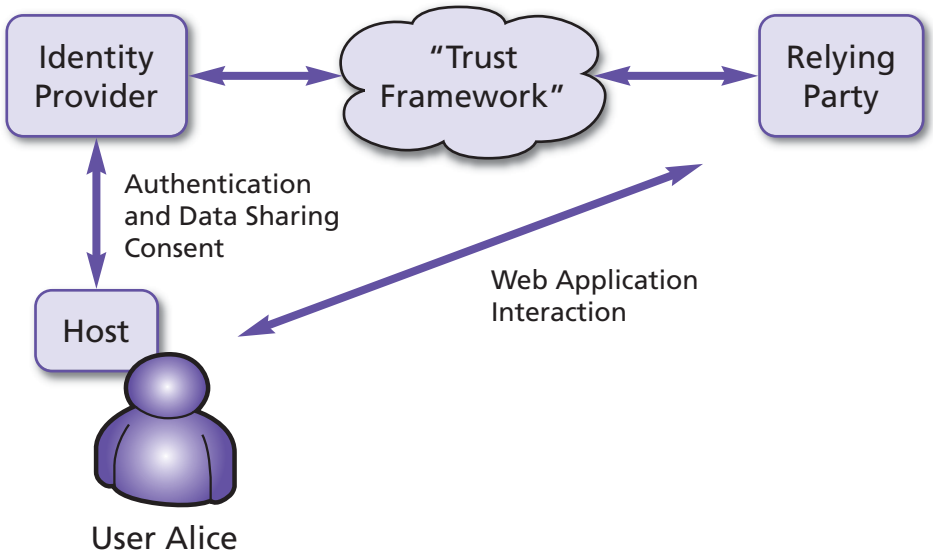

Figure 3.4.: WebSSO Architecture

inCommon

Another example of a federation for education and research is inCommon [150], which focuses on the United States. Like other federations, certain membership criteria have to be fulfilled (see [151]). Typical for such a non-commercial federation is the transparency. For example, the meta data of the inCommon federation is available online, see [151], and serves as an example of what technical configuration information is shared with the members.

WAYF

WAYF [152] is an education and university federation based in Denmark. Unlike other federations that enable direct protocol exchanges between the identity provider and the relying party, WAYF uses a model with an intermediary that provides protocol translation, as well as verification of exchanged information. The design principle is stated on the WAYF website [152]; WAYF ensures that the connected services only get the minimal set of information about you that is needed - and that you give your consent to have it passed on. Their intermediary only caches user identity and attribute-related data for a limited period of time (namely 8 hours) to get the single-sign effect and that information which passes by for any purposes other than forwarding it between the IdP and the RP.

In [153] Simonsen and Madsen explain the different design considerations for constructing a federation, focusing on technical and operational aspects. A video recording about the presentation is available as well [154].

Information about the contracts with IdPs and RPs are public and can be found at [155]. Additionally, information about the shared metadata [156], as well as the documents required for joining the WAYF federation [157], is available for download.

A legal analysis was done to determine the obligations for the WAYF federation when running such an intermediary and their conclusion was that they were operating as a data processor (rather than as a data controller).

An important aspect when joining the federation by an IdP/RPs is that they have to opt-out of services, if they do not want to provide them. The service of a RP being offered is analysed and the information being required for that service is then exchanged later on. There is no dynamic request of an attribute list by the RP during a protocol exchange with the IdP.

Interestingly, the user consent information is stored centrally in the WAYF intermediary rather than at the IdP itself. In their federation setup, IdP did not want to take on this task and they were happy that the user consent information was stored centrally rather than independently with each IdP. The user establishes the consent when he/she uses a WebSSO interaction at the RP for the first time. The consent is established for a specific purpose and information about the shared attributes and even their values is indicated. Currently, the values in these attributes are largely static and, if their values change (for whatever reason), then a new consent is solicited from the user.

EDUGAIN

eduGAIN [158] is an inter-federation, i.e. a federation of federations, focusing on educational institutions deploying web-based technologies. It prescribes policy that binds the participating federations together.

3.4.5. Authentication, Authorisation and Accounting (AAA) frameworks

eduroam (EDUcation ROAMing) [159] allows users from participating academic institutions secure Internet access at any Eduroam-enabled institution. The architecture that enables this is based on a number of technologies and agreements, which together provide a network access single sign-on experience. From a technology point of view EAP and RADIUS are utilised.

Figure 3.5. shows the architecture with a host accessing an access network, such as a WLAN, on a university campus. Typically, this initial exchange utilises special link layer frames, such as IEEE 802.1X, and terminates at the network access server (NAS), which takes on the role of the relying party. In a roaming context, the user is unknown to the local network and to the NAS and hence the exchange is relayed to the user's AAA server. The dotted arrow indicates the logical authentication exchange that is executed between the end host and the AAA server without active involvement of the NAS. At the end of the exchange, information is exchanged between the AAA server and the NAS, in both directions. The AAA server provides provisioning instructions to the NAS, together with authorisation information and keying material, whereas the NAS often provides accounting information back to the AAA server over the lifetime of the network attachment. The accounting aspect is naturally more important for a commercial network, where users are charged based on their usage rather than for a university setup.

Figure 3.5. refers to the arrangement between the two administrative domains deploying the AAA server and the NAS as ‘clearing house’, a term that is often used in commercial AAA federations. Typically, the clearing house establishes contractual agreements between the organisations involved, agrees on the protocol extensions used, distributes security information and clarifies how roaming, accounting and settlement are provided. The Wimax Roaming Exchange (WRX) is an example of a commercial roaming setup, about which more detailed information can be found here [160].

The EduRoam consortium [159] also provides information about its federation structure, technical requirements and conditions for membership. They can be found here [161].

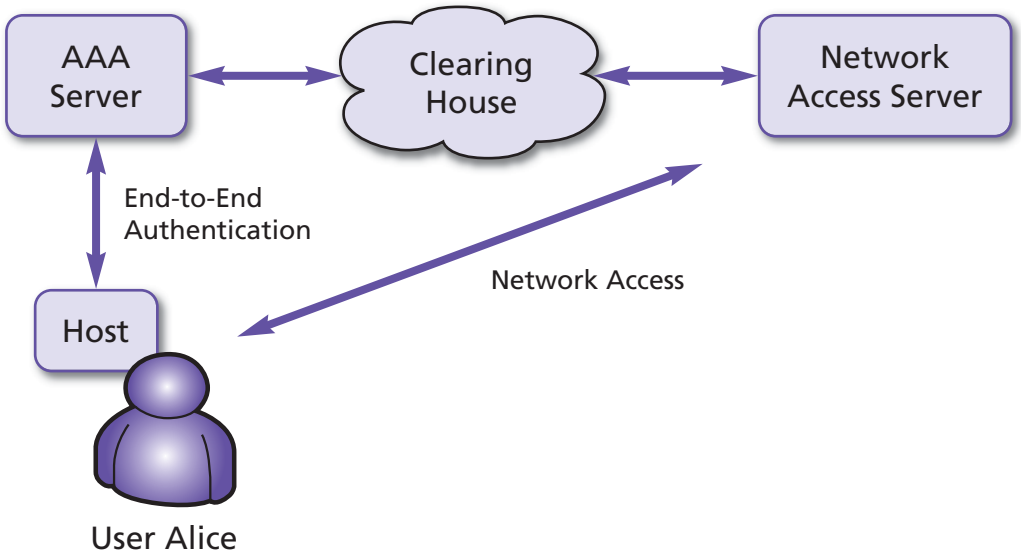

Figure 3.5.: AAA architecture

3.4.6. Payment card industry

In the debate about desirable properties of trust frameworks, the payment card industry is often mentioned as a successful example of a highly scalable and reliable architecture. Some even go as far as to suggest using it as a blueprint for building future identity management systems. Figure 3.6. shows a simplified description of its architecture, where merchants and issuers are interconnected with the help of payment networks that allow management of the number of business relationships between merchants and issuers. To reduce fraud and to provide more confidence in the system, the payment networks are mandated to offer a certain amount of security, as described in the Payment Card Industry (PCI) security standards [162]. Establishing industry-accepted security guidelines and ensuring that they are met is an important first step. However, another attractive property is the liability guarantee offered by the Federal Deposit Insurance Corporation (FDIC). In the case of credit card and banking fraud, liability for customers is limited to the first $50 of loss [163]. It is worthwhile to consider how such a concept could be applied to the Internet identity management world (or other environments), where an independent organisation offers an insurance to end users in case they become victims of security and privacy related incidents. In such a model, identity providers and relying parties have to deposit money with this independent organisation, thereby being likely to demand that a certain level of security and privacy requirements be in place (potentially verified via audits).

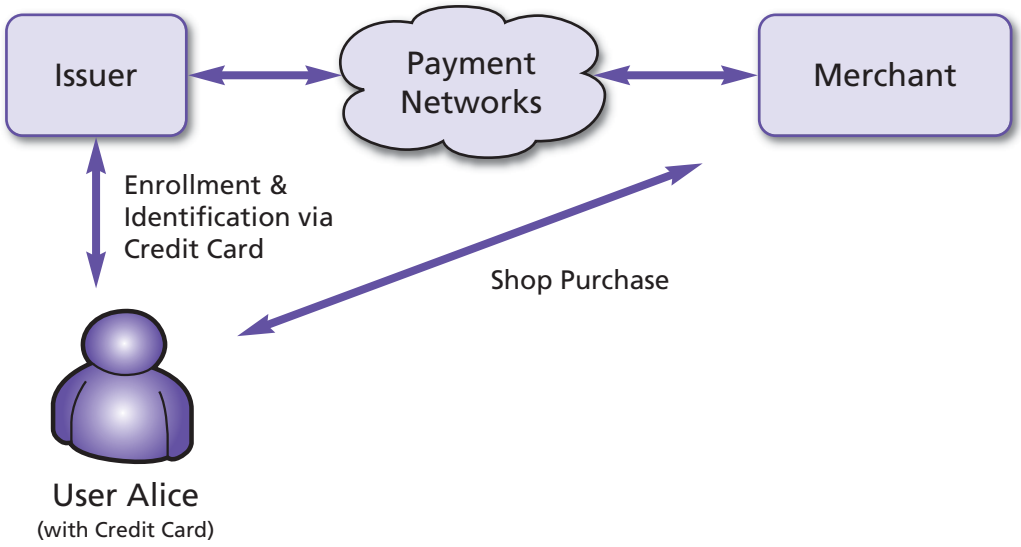

Figure 3.6.: Simplified payment card model

While it seems to be tempting to re-use an existing and well-established model, there are also concerns. The payment card model has indeed only been successful for payment-related transactions, rather than serving a role as a general-purpose data sharing architecture. The Internet identity management community would like to keep the barriers of entry for identity providers, as well as for relying parties, low, both from a technical as well as from a financial point of view. The payment card model requires a substantial investment from the participating parties.

Finally, transparency and openness are concepts promoted in the privacy and identity management world, but recent incidents in the payment industry give the impression that these established systems operate differently. For example, the Belgian-based Society for Worldwide Interbank Telecommunication (SWIFT), which is used for most international payment transactions, has provided information about payment transactions to the US Treasury Department for several years, without disclosing these practices to other stakeholders [164 SWIFT]. Upon these business practices being revealed to the press in June 2006, various data protection authorities questioned the classification of the payment intermediary as a data processor. In a later memo, the EC Article 29 Data Protection Working Party issued an opinion [165 01935/06/EN] concerning this case and concluded that SWIFT acted in the role of a data controller, which implied much greater responsibility for privacy protection.

3.4.7. Conclusions

Designing systems that scale to the size of the Internet is hard. Over the last 25 years, a number of distributed systems have been designed that accomplished such scale to an impressive degree. Section 3.4. takes a first step to highlight their similarities and their differences. Terminology differences prevent protocol designers from taking advantage of past efforts.

Recommendation #17. *Privacy and security should be considered early in the design process, particularly since many of the interactions in the backend infrastructure are essentially invisible to the end user, and also less frequently understood by those who are given the responsibility of protecting their interests.*

With the renewed interest in the design of the WebSSO solution, an entire industry has focused its attention on ‘trust frameworks’. Ever increasing privacy violations on the Internet require privacy and security to be considered early in the design process, particularly since many of the interactions in the backend infrastructure are essentially invisible to the end user, and also less frequently understood by those who are given the responsibility of protecting their interests.

3.4.8. Further investigations

The Centre of Democracy and Technology (CDT) published a long list of questions in its report on issues surrounding user-centric identity management [166]. Many of the questions raised directly concern the design and operation of trust frameworks; further research is needed for these two areas, namely the design and architecture of the frameworks and their operation.

Unfortunately, the research community cannot easily evaluate trust frameworks deployed today, since very little information about their internal structure and operational policies is known to the outside world. It is understandable that certain contractual agreements cannot be disclosed, but there is still a concern about the degree of transparency.

Recommendation #18. *Trust frameworks need further evaluation. Their internal structure and operation should be visible.*

We believe that the technical and research community need to be provided with more information, as already demanded by data protection authorities, for example in the recent publication of the Madrid resolution [167]. For the purpose of this report, the following privacy principles seem to be applicable:

- *Accountability principle:* The collection of personal information includes a duty of care for its protection. From this principle others can be derived, such as the principles regarding proportionally, data quality, consent, purpose limitations, etc.

- *Openness principle:* Making information about the policies and practices relating to the management of personally identifiable information available is an important component of accountability

- *Security principle:* This principle highlights the responsibility for securing systems handling personally identifiable information throughout its entire lifecycle

We believe that treating entities that are part of a trust framework as data processors is insufficient for building a solid foundation for future Internet identity management systems when it comes to the processing of personally identifiable data. Upcoming trust framework deployments will be generic rather than purpose built for dedicated applications (such as for payment, or network access only). This provides greater flexibility, but at the same time creates greater risk of privacy violations. Researching suitable models for improved customer confidence in the digital world is necessary and therefore entities that are part of the trust framework need to share the complete set of privacy obligations.

Recommendation #19. *Researching suitable models for improved customer confidence in the digital world is necessary and therefore entities that are part of the trust framework need to share the complete set of privacy obligations.*

3.5. Examples of privacy preserving architectures and trust frameworks

In this section we present architectural options for three applications: privacy-friendly advertising, location privacy, and smart metering.

3.5.1. Privacy-preserving advertising

Today, the basic principles of data privacy in advertising are: 1) minimise what needs to be collected in cloud servers, 2) protect data that has been collected, and 3) retain data only as long as necessary. In this approach it is tacitly assumed firstly that data cannot be used unless it is collected, and secondly that privacy necessarily entails compromising utility. There is a growing body of research suggesting that this trade-off between privacy and utility need not be suffered in many cases. The central premise of the research is that user data may be stored locally on the user's own computer, instead of in cloud servers. Applications that use this data run in the user's computer instead of in cloud servers.

Behavioural advertising is an application that looks promising under this client-centric privacy model. Three research projects have produced client-centric privacy designs for behavioural advertising; Adnostic (Stanford/NYU) [168], Nurikabe (Maryland/Microsoft), and Privad (MPI-SWS) [169]. Although these designs differ in many ways, they all share the following basic design principles. Firstly, the software that profiles the user runs on the user's computer, and stores the user profile locally. This profile never leaves the user's computer. Secondly, more advertisements than will actually be displayed to the user are transmitted to the user's computer. These ads may, for instance, span a variety of user demographics or interests. The ads shown to the user are locally selected from among these ads, and locally inserted into adboxes on web pages. As a result, it is not necessary for the ad network cloud servers to know the user profile for ad selection. Of course, the ad network does need to know which ads were viewed or clicked, and on which websites the ads appeared. This is necessary to charge the advertisers, and to pay the websites. Each design does this reporting in a privacy-preserving manner through the use of proxies and encryption. For instance, in Privad, these reports are encrypted with the public key of the ad network, and delivered to the ad network through proxies. The proxies do not see the content of the reports, and the ad network does not know which client delivered each report. These reports are unlinkable, thus preventing the ad network from compiling user profiles from the reports.

The above designs hide user data from the ad network. This is not enough. It is also important to hide user data from the advertisers themselves. This was demonstrated recently in a highly-publicised case where researchers showed that Facebook reveals users' sexual preferences to advertisers without the users' awareness. This happens even for users who keep their sexual preference settings private. The problem stems from the fact that Facebook allows advertisers to target ads based on sexual preference. Advertisers exploit this by targeting gay users for products and services that are not overtly related to being gay (i.e. a nursing school). Simply not allowing targeting on sexual preference or other sensitive attributes is not an ideal solution. A gay dating service, after all, should be able to target gays. The proxy and encryption techniques used to hide users from ad networks can also be used to hide them from advertisers, but only to an extent. Eventually a user may need to supply PII to the advertiser, for instance to purchase a product or service. A more fundamental approach is privacy-neutral advertising. The idea here is that the advertiser should learn no more about the user than is learned through the user's expressed interest in the advertiser's product or service. If, for instance, a user clicks on an ad for a gay dating service, then the advertiser may indeed deduce that the user is gay. If a user clicks on an ad for a nursing school, however, the advertiser should only be able to deduce that the user is interested in nursing. The advertising model here is that the advertiser, rather than target ads to demographics or keywords, simply provides the ad and the landing page to the ad network. From this, the ad network determines to which users the ad will be shown. Indeed, Google is experimenting with a similar approach--not for the purpose of privacy, however, but to save the advertiser the trouble of deciding which keywords to target.

3.5.2. Location privacy

Location-based services use information about the geographic location of an individual or device. As the cost to provide reliable and detailed location information decreases, and the potential benefits for both users and service providers are enormous, location based services are becoming increasingly important on the Internet. Location based services are discussed from a tracking viewpoint in Section 2.2.3.

Location can be deduced from multiple sources. Firstly, one can deduce location information based on the IP address (using GeoIP, http://www.geoiptool.com/); this will provide a rough indication of the location of a user. However, NAT (Network Address Translation) will make location information less precise and tunnelling techniques can be effective at obfuscating the information completely. Nevertheless, the network access provider knows the exact network attachment point of the end host; this allows for a more precise determination of the location. With additional help from the end host, the location information can be made more accurate. The technologies that can be used on the host include GPS (and in the future Galileo) and network based techniques. Note however that GPS signals can be spoofed by an adversary; the Galileo system intends to offer secure localisation services. Network based techniques can also be attacked; the conclusion of Gill et al. [170] is that the simple delay-based mechanisms are less susceptible to covert tampering than the more advanced and accurate topology-aware methods.

If one assumes that service providers need location data, the data protection approach consists of attaching sticky policies to this data; the GeoProvi architecture takes this approach [171]. A privacy by design (see Section 3.1) approach has two options: the first approach consists of trying to decouple identities from location data. For this purpose, Beresford and Stajano introduce in [172] the concept of mix zones, which are equivalent to mix servers for email. While the concept is elegant, it is not clear how much protection can be offered, given all the side channels that are present (e.g. surveillance cameras); similar to mixes, mix zones are vulnerable to advanced statistical attacks. The other privacy by design alternative consists of shifting the application to the device at the user's side. Rather than sending location data to a central server, the data is processed locally in a tamper resistant box that contains all the necessary data. Solutions have been developed and implemented for insurance pricing (Pay As You Drive or PAYD) and for road pricing. In this case, the black box contains a GPS receiver, maps, GSM communication, and cryptographic hardware to protect the data. For insurance pricing, the device outputs at the end of each month the cost of the insurance, digitally signed by the device; for road pricing, the device outputs the signed fee for the usage of the road. The design of the protocols is rather subtle as one needs to offer the user the possibility to audit the costs based on the detailed trajectories, yet this information should not be available to the service provider, to a police officer who stops the car for a routine check, or to the car maintenance workshop. The service provider on the other hand wants to verify that the black box has not been switched off, locked up in a Faraday cage, or fed wrong GPS data. It is rather surprising that a careful design in this case can result in a cost-effective application, without any central database containing sensitive location data.

3.5.3. Smart metering

There is currently substantial interest in the 'smart grid', a concept referring to the modernisation of the existing electrical grid. The goal is to introduce smart meters that have bidirectional communication with the utilities; moreover, these meters offer fine grained readouts (every 15 minutes or even more frequently) in combination with flexible pricing. One can also hope that these meters may reduce fraud and improve energy management. It has been pointed out that the architecture of several smart grid projects has serious privacy issues (see e.g.. Anderson and Fuloria [173] and McDaniel and McLaughlin [174]). As an example, the electricity consumption pattern reveals many details about the inhabitants of the house; it is obvious that this information would be highly valuable to potential burglars or could be used to profile users. A legal perspective on the privacy in the smart grid is offered by Cavoukian et al. [175] and Quinn [176]. This section does not attempt to cover all the solutions; it just offers some pointers to technological approaches that can improve the privacy of consumers, while respecting the requirements of the operator. Wagner et al. [177] propose a privacy-aware framework for the smart grid based on semantic web technologies. Garica and Jacobs [178] consider an aggregation technique in which the power consumption of all the households in a neighbourhood is combined; in order to protect the interests of all the parties, homomorphic encryption is used. Rial and Danezis propose in [179] a smart metering architecture that can sustain a broad range of pricing policies. Their design requires a tamper resistant meter with a secure digital signature; other cryptographic calculations can be performed outside the metrological unit. Their solution takes a `privacy by design' approach that leaks minimal information to third parties yet provides unforgeable bills based on complex dynamic tariff policies; their solution relies on zero-knowledge proofs.

Recommendation #20. *Support research and best practices in privacy friendly architectures in a broad range of applications. If appropriate, enforce some of the architectural approaches by legislation.*

3.6. Privacy at lower layers

There is a strong awareness of the need for a privacy-friendly architecture at the application layer; this requirement has been taken into account during the design of some of the recent identity management schemes. However, these efforts will only have very limited impact if unique identifiers are used at lower layers of the protocol stack, such as IP addresses, MAC addresses, IMEI numbers etc. This kind of information may be used to link all service interactions. Other lower layer information that may be revealed includes unique IC identifiers, and physical characteristics of the IC or of the antenna. (This problem has been discussed to some extent in Section 2.2. on behavioural tracking and profiling on the Internet; examples include cookies, browser fingerprinting and location tracking). The issues related to geolocation are even more complex; they are treated in Section 3.5.2.

The technologies to offer privacy protection at lower layers are typically much less understood than privacy at the application layer. As it is impossible to hide network data against a powerful opponent, the goal is to hide the identities of the communicating parties. A simple solution is the use of proxies. Mix networks form a more advanced solution; while there is an academic effort on the design and analysis of mix networks, and there is some limited deployment, there is a strong need for more fundamental research in this area to create a deeper understanding of the security properties under various adversarial models.

Technologies that will be discussed in this section include anonymous remailers (e.g. Mixminion) and onion routing (e.g., Tor); some comments will be offered on anonymous peer to peer networks. For more details on the first two types, the reader is referred to [180].

Very few of these systems have been designed with law enforcement access requirements (conditional anonymity); while it seems to be very difficult to design systems that offer privacy at lower layers, offering access to selected parties under well-controlled conditions is even more difficult and seems elusive for now.

3.6.1. Definitions and basic notions

Any communication has at least one sender and one recipient. In privacy at lower layers, one may want to hide either the sender to recipient, or the recipient to the sender or a combination of both. In other contexts, sender and receiver may want to be authenticated to each other, but they will want to hide their communication from network observers. In terms of protection, one can distinguish between several dimensions; we follow here the definitional work by Pfitzmann and Hansen [181], who have over the years carefully integrated many contributions from the community. For some definitions we follow IS 15408 [182].

Anonymity. The anonymity of a subject requires that there is an appropriate set of subjects with potentially the same attributes. Anonymity can then be defined as the state of being not identifiable within a set of subjects, the anonymity set. For every action (e.g., sending, receiving), the anonymity set consists of the entities who might have caused this action.

Unlinkability. This property considers a context in which users make multiple use of resources or services; a system offers unlinkability if third parties are not able to decide whether the same user was involved during any two actions.

Unobservability. This property considers a situation in which there are items of interest (e.g. messages flowing on a channel), but an observer cannot decide whether a particular action is taking place or not (e.g. if any sender from a particular set is sending a message). Similarly to the anonymity set, one defines the unobservability set as the set of subjects about which the observer has to decide.

Pseudonymity. This corresponds to replacing real IDs (e.g. firstname/lastname or national register number) by a pseudonym that can be restricted to specific contexts or specific time intervals. Pseudonyms offer a weaker protection, but they allow their owners to build a pseudonymous reputation over time. Moreover, it is quite straightforward to make them conditional or accountable; it suffices to store the mapping between real IDs and pseudonyms at a trusted third party.

For confidentiality and authenticity of data, there are now clear attack models and quantitative definitions of security; it is relatively straightforward to determine whether a given mathematical construction, or even its implementation in hardware or software, achieves a certain security goal under a specific attack model, provided that some mathematical assumptions hold. For privacy at lower layers, also known as traffic flow confidentiality, there is a much broader range of attack models and security goals; even creating a partial ordering in this space is hard.

Typically, the opponent knows an a priori distribution of the members of the anonymity set before any action; after a number of observations, the opponent can revise his distribution using Bayesian inference. One can measure the security by the mathematical properties of the distribution; the more natural choice is the entropy (as proposed in [183, 184]), the min-entropy (inversely proportional to the maximum probability in the distribution), the reduction in entropy, etc. A second dimension considers the level of access of an opponent. A network is typically structured as a set of end nodes (e.g. senders and receivers) and a set of intermediate nodes. A global opponent can observe all communication links; one can expect that the traffic on these links is protected using authenticated encryption, so the opponent may only observe whether or not messages are being sent (they could be dummy messages). A more powerful opponent can compromise a subset of the intermediate nodes; here one can distinguish between temporary or long-term corruption.

Even if an opponent can only monitor the end nodes, persistent communication between two end nodes will eventually be detected just by correlating the activity at these end nodes; the only way to preclude such an intersection attack [185] is to maintain a constant level of communication between any two parties using dummy messages; in most contexts the cost of such an approach is prohibitive.

In the following we review technologies for high latency communication (such as email) in Section 3.2.2. and for low latency communications in Section 3.2.3.

3.6.2. High Latency communication: remailers and mixes

The first anonymous remailer was a trusted mail relay, anon.penet.fi, which offered anonymous and pseudonymous email accounts in 1993. After a few years, it became obvious that such a server was not robust against either technical or legal attacks; in this last case, the owner of the server is forced to disclose the real identities of the users, based on a court procedure. While several more robust remailers have been developed, it is clear today that a resilient solution requires a refinement of the seminal ideas by Chaum, who introduced the concept of a mix in 1985 [186]. A mix is a special remailer that decouples input and output by collecting messages in batches – perhaps together with dummy messages – and by transforming them using cryptography. The construction of mixes has been refined over the years due to a succession of new ideas and sophisticated attacks, not unlike the development of modern block ciphers and stream ciphers.

The state of the art anonymous remailer today is Mixminion [187], which transports a fixed size message anonymously over a set of Mixminion remailers; the system supports sender anonymity, receiver anonymity via single-use reply blocks, and bi-directional anonymity by composing the two mechanisms. It is essential for the security of Mixminion that the intermediate remailers do not know too much. They are not aware of their position on the path of the message, nor of the total length of the path; moreover, they cannot distinguish between original messages or replies. The security of Mixminion requires that all clients must know the full network of remailers, which is achieved using a distributed directory service. In practice it remains difficult to decide in a worldwide network which nodes are honest; reputation-based systems (cf. Section 3.3.) can play a role here.

Mix networks are vulnerable to powerful attacks, such as the "blending attack" [188], in which an opponent mixes his own messages with the message under attack; unless special techniques are used, it becomes very easy to determine the source of these messages. Mixes can be made more robust by using advanced cryptographic techniques. Verifiability allows anyone to check the correct operation of the mix without leaking information on the messages; threshold cryptography can be used to distribute the trust and reduce the impact of the compromise of a mix. Dummy messages increase the bandwidth, or the delay, but improve robustness. Finally one can consider various network topologies, depending on the model of the adversary and the type of mixes. This is a very active area of research, in which one tries to optimise security and/or increase performance (as throughput versus delay).

3.6.3. Low latency communication: onion routing

In [189], Pfitzmann et al. designed a system to anonymize ISDN telephone conversations. This design could be considered practical, from an engineering point of view, since it met the requirements and constraints of the ISDN network. Later the design was generalised to provide a framework for real-time, low-latency, mixed communications in [190]. Finally, many of the design ideas from both ISDN and real time mixes were adapted for anonymous web browsing and called Web Mixes [191]. Part of the design has been implemented as a web anonymizing proxy, JAP. A recent analysis [192] of the cryptographic protocols employed in JAP has however uncovered vulnerabilities that could be exploited to break its anonymity properties, including lack of replay and integrity protection.

Onion routing is a more secure approach to offering anonymity in low-latency settings, such as web browsing. Rather than processing each part of a message independently, in onion routing circuits are established and used to facilitate bidirectional communication. Onion routing uses hybrid cryptography; public key technologies are deployed to establish a secure route and symmetric point-to-point encryption is then used on this route. In order to set up a route, a client downloads the list of internal nodes and selects three nodes on this list; the message is then encrypted with a key known to each of the intermediate nodes. During the transmission, each intermediate node 'peels off' one layer of encryption, resulting in the name 'onion routing'. Note that intermediate nodes in onion routes do not attempt to mask the relation between packets entering and leaving the node. The most advanced and widely deployed system that uses onion routing is Tor, released in 2004. Unlike the best mix networks, onion routing is vulnerable to passive attacks. The architecture of the Internet, with powerful hubs and bundling of links, makes it rather vulnerable; surprisingly, connections can even be de-anonymized by observing nodes outside the onion routing path.

Some interesting work was done on the performance and usability of Tor in 2007 [193, 194]. The main conclusion from this work was that the performance was rather slow and the usability left room for improvement. Since then, many improvements have been made to usability; one example is the Torbutton add-on for Firefox in combination with Privoxy. Moreover, currently between 100,000 and 300,000 people use Tor per day; some claim that the poor performance is a result of too many people wanting to use Tor and consuming all the bandwidth[28].

Several attempts have been made to create secure onion routing through peer-to-peer networks. The advantage of peer-to-peer networks is that they offer architectures with large, but transient, numbers of nodes; this is an attractive property for anonymous communications. One of the main challenges in these designs is to ensure that an adversary who controls only a small fraction of the peers cannot break the security and anonymity properties of a large fraction of the communications. Most of the recent research in this area has focused on structured networks, for example, see [195, 196].

These structured systems build anonymous routes by looking up nodes to route through in a distributed hash table (DHT). By exploiting the structure implicit in the DHT and by adding randomness plus redundancy to the lookups, they make it difficult to either determine or direct the node discovery and selection done by the route builder. However, these systems turn out to be less secure than anticipated. Mittal and Borisov demonstrated in [197] that Salsa [196] has significant leaks in the lookups associated with routing-node discovery and selection; these leaks allow compromise of route-selection security by an adversary composed of a much smaller fraction of the total network than had previously been thought. They show an interesting trade-off between mechanisms to prevent active attacks and passive observations of lookup. More redundancy helps to resist active attacks, but leaks more information.

While peer-to-peer architectures remain a promising approach for implementing anonymous communication infrastructures, due to their superior scalability, the design of such systems with robust anonymity properties remains a problem.

Recommendation #21. *Support research on advanced solutions for lower layer privacy. This also includes legal dimensions, scalability and deployment, measurement and monitoring.*

28 More details on the performance of Tor can be found at https://metrics.torproject.org/

4. Concluding remarks

To understand the current status of privacy on the Internet, it is essential to take its history into consideration. The Internet is very large and is utilised by many stakeholders in different ways. Sometimes the interests of these stakeholders in the Internet milieu are in conflict with each other, or are not transparent. Even worse, some stakeholders express positions on one topic that are in direct conflict with their position on a different topic.

> As an example, governments on one hand demand accountability, data minimisation, and better privacy protection for citizens, but on the other hand demand more access to data for their own purpose in the area of lawful intercept.

Understanding the current long list of privacy challenges requires an analysis of the incentives of various actors to violate basic privacy principles, such as the 'OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data' developed 30 years ago. While media has paid a lot of attention to privacy-related matters during the last few years, most of it is focused on a small sub-set of the Internet, namely the end user services, such as social networks. There are, however, many other applications being used on the Internet every day, such as enterprise applications, governmental applications, and many more that do not make use of the browser-based web, as experienced by many end users.

Many of the privacy challenges discussed in this document are focusing on the browser-based web and one can indeed argue that data minimisation is not the design goal for many of the companies operating in this space. Instead, they seek to ensure that their business opportunities are not restricted in any way and behavioural advertising is one of the core funding models for these services. Many end users do indeed enjoy the large range of free services, yet are surprised about the type of profiling that is occurring.

Developments in the Internet related to the semantic web, Web 2.0, Internet of Things, and Cloud Computing have one aspect in common; they envision that more data sharing will happen. Cross-site and cross-domain data sharing is best summarised with the term 'mash-up'.

Many innovations happen in the application domain where information previously kept in silos can now be combined, where information can be made available in real-time, and where more information is moved from the offline to the online world.

How can the Internet of the future offer more privacy preserving properties for end users when data sharing offers such great promises, when many enterprises build their business model on it, when governments enter the area of cyber-attacks? How can we ensure the speed of innovation we see on the Internet today without the negative side effects regarding security and privacy?

Keeping the balance between many conflicting goals will be challenging and will require a multi-stakeholder dialogue. Having the research community, enterprises, governments, regulators, data protection authorities, as well as the standards developing community, work together will help us to gain better insight into a topic with so many trade-offs.

Clearly, privacy challenges cannot only be solved by technological means. There is a need for a multi-disciplinary approach that considers economic factors, education, legal and technological aspects, further details of which are given below.

Education. Providing users with ways to understand complex technology, as is clearly the case with the Internet, is a demanding task. Nevertheless, users must be informed about the potential risks of services they use on the Internet. With new services being made available all the time, this education effort requires life-long learning.

Users should be educated about the privacy threats they are exposed to when they reveal their location, when they surf the Internet, or when they publish personal data on the Internet or social networks. They should be informed that any published information never disappears from the Internet and might eventually leak to unwanted destinations. Education is clearly an essential action to improve users' privacy on the Internet.

Legal. In Europe, the data protection (95/46/EC) and the Privacy and Electronic Communications directives (2002/58/EC) provide some protection against marketing practices, such as profiling. The core principle of these directives is to guarantee that users are able to make an informed choice [30] and that they consent to being tracked using an 'opt-in' mechanism. However, as argued in [30], the critical aspect of permission-based tracking lies in meaningful consent. In particular, in order to make an informed choice, users must understand privacy policies, which are often very long and written for a different audience (often for lawyers rather than for end users). This issue is exacerbated on mobile phones because of the size of their screen.

Policy. Behavioural tracking is tolerated today because it is supposedly only used to collect non identifier personal information, and therefore does not endanger user privacy. However, as argued in [22], anonymity does not necessarily equal privacy. By influencing his or her decision processes, behavioural tracking undermines user autonomy and therefore privacy. Furthermore, with all the data available on the Internet, any information that distinguishes one person from another can be used for reidentifying data. In other words, any 'anonymous' profile can potentially be de-anonymized, i.e. linked to a person. In the light of these new results, it is questionable whether the concepts of PII and 'anonymous' profiling still make sense, and should not be revisited [61]. These new de-anonymization results and emerging tracking mechanisms clearly need to be considered by privacy laws which, as argued in [62], might need to be updated. Privacy laws should clearly set the limits of behavioural tracking and profiling.

Technology. Research into new privacy-preserving tools is important, as the amount and sophistication of data collection about individuals increases. A recent example of such a research project is the work on differential privacy [132]. In this context it is important not only to focus on tools dealing with the end-user visible technologies, such as those related to the web, social networks and mobile phones. It is equally important to research tools for enhancing privacy in embedded systems and at the lower layers of the communication stack. Tools that help users making informed decisions about the publication of their data or their online activities need to be developed. Users must also be able to choose what data is collected about them. They must keep the right to access, modify and delete it. Users should be explicitly informed about how they are being tracked, how their data is being sent/leaked out of their social network sites, by advertisers or others, and the corresponding destination. For example, users should have to acknowledge usage of their location on a per-application basis, or even, for some applications, each time location information is used. Indeed, as argued in [43], a user might agree to have some of his or her photos geo-tagged. However, for other photos, for example family photos, he or she might be opposed to it, or might require the location resolution to be lowered.

As argued by Bruce Schneier in [58], "Privacy is not something that appears naturally online. It must be deliberately architected." The technical community nevertheless has to get a better understanding what it means to design privacy into protocols, architectures and systems. With the recent Internet Privacy workshop[29] a first step has been taken to investigate privacy by design. Many more efforts by the technical community will, however, be necessary to reach at least the level of understanding we see today with security, where a common language is available, threat models are understood, and desired security properties are known by designers.

29 "How can Technology help to improve Privacy on the Internet?" workshop organised by IAB (Internet Architecture Board) in December 2010, webpage available at: http://www.iab.org/about/workshops/privacy/

List of recommendations

	PAGE	TYPE OF RECOMMENDATION: EDUCATION / TRAINING	LEGAL / POLICY	TECHNOLOGY
1. Unambiguous definitions of personal data and identifiable information should be supported across Member States.	14		✓	
2. Dynamicity of privacy, as well as subjectivity, should be investigated within conceptual frameworks for understanding privacy expectations and reasoning about norms in each place, time or other contextual dimension. This framework has to consider risk dimensions as well as schemes such as Privacy Impact Assessment and might also be linked to privacy decision support tools or privacy services (e.g. automated reconfiguration of privacy settings, enforcement of obligatory user consent, etc).	15		✓	
3. If issues such as "accountability principle", traceability or privacy auditing are finally adopted, we will need to rely on evidences, derived from ICT events and traces. This has to be anticipated in service design and service architectures.	15		✓	✓
4. Shared services, joint controllership, risk allocation etc. are all putting emphasis on relationships between stakeholders in the service chain. On the engineering side, indicators, metrics, semantics of privacy evidences (and corresponding event/traces) should be in place. On the operational infrastructure side, trust dynamicity, reputation management, contracts and policies are some of elements that have to be in place to enable end-to-end accountability. Architectural decisions about placement and access control to the components that generate events and traces, needed for evidences that support privacy assessment and assurance, should be consistent with organisational and service provision models.	16		✓	✓
5. Privacy assessment, assurance, verification or enforcement should be evidence-based. These evidences might be derived from a number of sources, events and traces at different architectural layers.	16		✓	
6. Users should be informed that any published information never disappears from the Internet and might eventually leak to unwanted destinations.	23	✓	✓	
7. Privacy issues in behavioural profiling are complex and cannot be treated exclusively by legal or technological means. There is a new requirement for a true multidisciplinary research approach that considers education, policy, legal and technological aspects.	23	✓	✓	✓
8. The concept of privacy certification, which would label each site and service according to their profiling activities, should be further investigated.	24		✓	
9. Privacy and security should be considered early in the design process, particularly since many of the interactions in the backend infrastructure are essentially invisible to the end user, and also less frequently understood by those who are given the responsibility of protecting their interests.	24			✓

	PAGE	TYPE OF RECOMMENDATION		
		EDUCATION / TRAINING	LEGAL / POLICY	TECHNOLOGY
10. Users' identities (and other data, i.e. localisation) should be requested and used only when strictly necessary. Users should be aware of which data is being collected and how they are used.	**24**		✓	
11. ENISA should support development of an economic model for monetising privacy.	**25**		✓	✓
12. Users should have access to online services with the right not to be profiled or tracked unless they have given their explicit consent. Moreover, they should be made aware of the scale and impact of the tracking and profiling.	**25**	✓	✓	
13. Further research and development work is needed on transparency enhancing tools, which are designed to show how data have been processed, by whom and with what implications, in an easily understandable, user-friendly way.	**30**			✓
14. Further interdisciplinary research and deployment is needed into the user-friendly design of user interfaces for presenting policies and obtaining well informed consent. For this, cultural differences have to be taken into consideration as well.	**37**			✓
15. Support research on privacy-friendly identity architectures that minimise concentration of information and prevent unnecessary linking, while guaranteeing accountability.	**41**			✓
16. Research on information accountability technology should be promoted, aimed at the technical ability to hold information processors accountable for their storage, use and dissemination of third-party data.	**45**			✓
17. Privacy and security should be considered early in the design process, particularly since many of the interactions in the backend infrastructure are essentially invisible to the end user, and also less frequently understood by those who are given the responsibility of protecting their interests.	**53**			✓
18. Trust frameworks need further evaluation. Their internal structure and operation should be visible.	**53**			✓
19. Researching suitable models for improved customer confidence in the digital world is necessary and therefore entities that are part of the trust framework need to share the complete set of privacy obligations.	**54**			✓
20. Support research and best practices in privacy friendly architectures in a broad range of applications. If appropriate, enforce some of the architectural approaches by legislation.	**56**		✓	✓
21. Support research on advanced solutions for lower layer privacy. This also includes legal dimensions, scalability and deployment, measurement and monitoring.	**59**		✓	✓

5. References

[1] EUROSTAT, E-government usage by individuals, November, 2010, most recent data from 2009, available at: http://epp.eurostat.ec.europa.eu/portal/page/portal/product_details/dataset?p_product_code=TSDGO330

[2] EUROSTAT, E-government usage by enterprises, November, 2010, most recent data from 2009, available at: http://epp.eurostat.ec.europa.eu/portal/page/portal/product_details/dataset?p_product_code=TSIIR140

[3] EUROSTAT, Internet usage in 2009 - Households and Individuals, - Issue number 46/2009, December 2009, available at: http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-QA-09-046/EN/KS-QA-09-046-EN.PDF

[4] A Digital Agenda for Europe, COM(2010)245, May, 2010, available at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:52010DC0245%2801%29:EN:NOT

[5] Treaty of Lisbon, Consolidated versions of The Treaty on European Union and The Treaty on the Functioning of the European Union, Charter of Fundamental Rights of the European Union, 2010, available at: http://europa.eu/lisbon_treaty/full_text/index_en.htm

[6] Convention for the Protection of Human Rights and Fundamental Freedoms, The Council of Europe, European Court of Human Rights, with last amendments into force on 1 June 2010, available at: http://www.echr.coe.int/NR/rdonlyres/D5CC24A7-DC13-4318-B457-5C9014916D7A/0/ENG_CONV.pdf

[7] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, 1995, available at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML

[8] Regulation (EC) No 45/2001 of the European Parliament and of the Council of 18 December 2000 on the protection of individuals with regard to the processing of personal data by the Community institutions and bodies and on the free movement of such data, available at: http://ec.europa.eu/justice/policies/privacy/docs/application/286_en.pdf

[9] Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications), available at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32002L0058:EN:NOT

[10] Council Framework Decision 2008/977/JHA of 27 November 2008 on the protection of personal data processed in the framework of police and judicial cooperation in criminal matters, available at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:350:0060:01:EN:HTML

[11] Directive 2009/136/EC of the European Parliament and of the Council of 25 November 2009 amending Directive 2002/22/EC on universal service and users' rights relating to electronic communications networks and services, Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector and Regulation (EC) No 2006/2004 on cooperation between national authorities responsible for the enforcement of consumer protection laws, available at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32009L0136:EN:NOT

[12] ENISA, Work Programme 2010, available at: http://www.enisa.europa.eu/media/key-documents/enisa-work-programme-2010

[13] ENISA, Survey of accountability, trust, consent, tracking, security and privacy mechanisms in online environments, 2010, available at: http://www.enisa.europa.eu/act/it/library

[14] Article 29 Working Party, Opinion 1/2010 on the concepts of "controller" and "processor", 2010, available at: http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2010/wp169_en.pdf

[15] PICOS project webpage, Privacy and Identity Management for Community Services, available at: http://www.picos-project.eu/

[16] STORK project webpage, Secure idenTity acrOss boRders linKed, available at: https://www.eid-stork.eu/

[17] STORK project, D2.1 - Framework Mapping of Technical/Organisational Issues to a Quality Scheme, available at: https://www.eid-stork.eu/

[18] Article 29 Working Party, Opinion 3/2010 on the principle of accountability, 2010, available at: http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2010/wp173_en.pdf

[19] M. Hildebrandt, Profiling: from data to knowledge, In Datenschutz und Datensicherheit - DuD Volume 30, Number 9, pp.548-552, available at: http://www.springerlink.com/content/k313374344883l21/

[20] G. Wondracek, T. Holz, E. Kirda, and C. Kruegel, A Practical Attack to De-anonymize Social Network Users, in Proc. IEEE Symposium on Security and Privacy, 2010, pp.223-238.

[21] Claude Castelluccia, Emiliano De Cristofaro, Daniele Perito, Private Information Disclosure from Web Searches (or how to reconstruct users' search histories), in Proceedings of the 2010 Privacy Enhancing Technologies Symposium (PETS), 2010, LNCS 6205, pp. 38-55.

[22] Catherine Dwyer. Behavioral targeting: A case study of consumer tracking on levis.com, in Fifteen Americas Conference on Information Systems, 2009, available at: http://www.ftc.gov/os/comments/privacyroundtable/544506-00046.pdf

[23] V. Toubiana, A. Narayanan, D. Boneh, H. Nissenbaum, and S. Barocas, Adnostic: Privacy Preserving Targeted Advertising, ISOC Network and Distributed System Security Symposium (NDSS), 2010, available at: http://crypto.stanford.edu/adnostic/adnostic-ndss.pdf

[24] R. MacManus, A Guide to Recommender Systems, January 2009, available at: http://www.readwriteweb.com/archives/recommender_systems.php

[25] B. Krishnamurthy, C. Wills, Generating a privacy footprint on the internet, in IMC '06: Proceedings of the 6th ACM SIGCOMM conference on Internet measurement, 2006, available at: http://portal.acm.org/citation.cfm?id=1177080.1177088

[26] B. Krishnamurthy, C. Wills. Privacy diffusion on the web: a longitudinal perspective. In WWW '09: Proceedings of the 18th international conference on World wide web, ACM, 2009.

[27] G. Conti, E. Sobiesk, An honest man has nothing to fear: user perceptions on web-based information disclosure, In Proceedings of the 3rd symposium on Usable privacy and security SOUPS'07, 2007, pp. 112–121.

[28] M. Barbaro, T. Zeller, A face is exposed for AOL searcher no. 4417749, New York Times, August, 2006 available at: http://www.nytimes.com/2006/08/09/technology/09aol.html

[29] John Krumm, Ubiquitous advertising: The killer application for the 21st century,. IEEE Pervasive Computing, January 2010

[30] Evelyne Beatrix Cleff, Privacy issues in mobile advertising, International Review of Law, Computers & Technology, Volume 21, Issue 3 November 2007, pages 225 - 236

[31] Katherine McKinley, Cleaning up after cookies, Technical report, iSEC PARTNERS, December, 2008, 12 pages, available at: https://www.isecpartners.com/files/iSEC_Cleaning_Up_After_Cookies.pdf

[32] B. Krishnamurthy and C. Wills. Privacy diffusion on the web: a longitudinal perspective (updated graphs), September 2009, available at: http://www.ftc.gov/os/comments/privacyroundtable/544506-00009.pdf

[33] B. Krishnamurthy and C. Wills. On the leakage of personally identifiable information via online social networks. In WOSN '09: the second workshop on Online social networks, 2009.

[34] Pam Dixon, World Privacy Forum, Consumer tips: How to opt-out of cookies that track you, 2009, available at: http://www.worldprivacyforum.org/cookieoptout.html

[35] Seth Schoen, EFF, New Cookie Technologies: Harder to See and Remove, Widely Used to Track You, September, 2009, available at: http://www.eff.org/deeplinks/2009/09/new-cookie-technologies-harder-see-and-remove-wide

[36] Soltani Ashkan, Canty Shannon, Mayo Quentin, Thomas Lauren, and Jay Hoofnagle Chris, Flash cookies and privacy, Technical report, University of California, Berkeley, 2009, available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1446862

[37] Evercookie–never forget, available at: http://samy.pl/evercookie/

[38] Eckersley Peter, How unique is your web browser? in Proceedings of the 2010 Privacy Enhancing Technologies Symposium (PETS), 2010, LNCS 6205, pp1-18.

[39] A. Blumberg, P. Eckersley, On Locational Privacy, and How to Avoid Losing it Forever. 2009, available at: http://www.eff.org/wp/locational-privacy

[40] Clint Boulton, Google CEO Schmidt Pitches Autonomous Search, Flirts with AI, 2010, available at: http://www.eweek.com/c/a/Search-Engines/Google-CEO-Schmidt-Pitches-Autonomous-Search-Flirts-with-AI-259984/

[41] A. T. Campbell, S. B. Eisenman, N. D. Lane, E. Miluzzo, and R. A. Peterson. People-centric urban sensing (invited paper). In Second ACM/IEEE International Conference on Wireless Internet, 2006

[42] E. Miluzzo, N.E. Lane, K. Fodor, R. Peterson, H. Lu, M. Musolesi, S. B. Eis, X. Zheng, S. EisenMan, and A.T. Campbell. Sensing meets mobile social networks: The design, implementation and evaluation of the cenceme application. In 6th ACM Conference on Embedded Networked Sensor Systems (SenSys '08), Nov. 2008

[43] G. Friedland and R. Sommer. Cybercasing the joint: On the privacy implication of geo-tagging. In Usenix Workshop on Hot Topics in Security, 2010

[44] Reality mining, in Technology Review, MIT, March/April 2008., available at: http://www.technologyreview.com/read_article.aspx?id=20247&ch= specialsections&sc=emerging08&pg=1

[45] Kate Greene, Reality mining, in MIT Technology review, 2008, available at: http://www.technologyreview.com/read_article.aspx?id=20247&ch=%20specialsections&sc=emerging08&pg=1&a=f

[46] R. Gross, A. Acquisti, and H. Heinz. Information revelation and privacy in online social networks. In WPES, 2005

[47] E. Zheleva and L. Getoor. To join or not to join: The illusion of privacy in social networks with mixed public and private user profiles. In International World Wide Web Conference (WWW), 2009.

[48] C. Y. Johnson, Project Gaydar at MIT, an experiment identifies which students are gay, raising new questions about online privacy, 2009, available at: http://www.boston.com/bostonglobe/ideas/articles/2009/09/20/project_gaydar_an_mit_experiment_raises_new_questions_about_online_privacy/

[49] B. Krishnamurthy and C. Wills. Characterizing privacy in online social networks. In WOSN '08: Proceedings of the first workshop on Online social networks, 2008

[50] B. Krishnamurthy and C. Wills. Privacy leakage in mobile online social networks. In WOSN '10: Proceedings of the third workshop on Online social networks, 2010.

[51] Peter Eckersley, EFF, How Online Tracking Companies Know Most of What You Do Online, 2009, available at: https://www.eff.org/deeplinks/2009/09/online-trackers-and-social-networks

[52] G. Aggrawal, E. Bursztein, C. Jackson, and D. Boneh, An analysis of private browsing modes in modern browsers, In Proc. of 19th Usenix Security Symposium, 2010.

[53] Firefox NoScript Extension, 2010, available at: http://noscript.net/

[54] R. Dingledine, N. Mathewson, and P. Syverson. , in usenix security symposium, 2004

[55] Greg Conti. Googling Security: How Much Does Google Know About You? Addison-Wesley, 2009

[56] Google CEO Schmidt:"people aren't ready for the technology revolution",. http://www.readwriteweb.com/archives/google_ceo_schmidt_people_arent_ready_for_the_tech.php, August 2010

[57] C. Castelluccia and D. Kaafar. Ocn: Owner-centric networking. In Future Internet Security and Trust (FIST) workshop, July 2009

[58] Bruce Schneier. Architecture of privacy. IEEE Security and Privacy, 2009.

[59] Do Not Track Explained. http://33bits.org/2010/09/20/do-not-track-explained/, February 2010

[60] Protecting Consumer Privacy in an Era of Rapid Change, Preliminary FTC Staff Report, Federal Trade Commission, December 2010.

[61] A. Narayanan and V. Shmatikov. Myths and fallacies of personally identifiable information. Communications of ACM, 53(6):24–26, June 2010

[62] Paul Ohm. Broken promises of privacy: responding to the surprising failure of anonymization. UCLA Law Review, 2010 (to appear).

[63] M. Chew, D. Balfanz, and B. Laurie. (under)mining privacy in social networks. In Web 2.0 Security and Privacy workshop, 2008.

[64] Facebook Settings Scanner, http://www.reclaimprivacy.org/, 2010.

[65] Redrawing the Route to Online Privacy.
http://www.nytimes.com/2010/02/28/ technology/internet/28unbox.html, February 2010.

[66] J. Staddon, P. Golle, and B. Zimny. Web-based inference detection. In usenix security symposium, 2007.

[67] The Diaspora project. http://www.joindiaspora.com/, 2010.

[68] G. Zhong, I. Goldberg, and U. Hengartner. Louis, lester and pierre: Three protocols for location privacy. In Proceedings of the 2007 Privacy Enhancing Technologies Symposium (PETS), 2007

[69] Curtis R. Taylor, Consumer Privacy and the Market for Customer Information, RAND Journal of Economics, Vol. 35, No. 4, pp. 631-650.

[70] Benjamin Hermalin & Michael Katz, Privacy, property rights and efficiency: The economics of privacy as secrecy, Quantitative Marketing and Economics, Vol. 4, No. 3, pp. 209-239.

[71] Giacomo Calzolari and Alessandro Pavan, On the optimality of privacy in sequential contracting, Journal of Economic Theory, Vo. 130, No. 1, pp. 168-204.

[72] Patricia A. Norberg, Daniel R. Horne, and David A. Horne, The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors, Journal of Consumer Affairs, Vol. 41, No. 1, pp. 100-126.

[73] Jentzsch, Nicola, Financial Privacy – An International Comparison of Credit Reporting Systems (Springer-Verlag, Heidelberg), 2. revised edition.

[74] Alessandro Acquisti, From the Economics to the Behavioral Economics of Privacy: A Note, ICEB, LNCS 6005, pp. 23-26, 2010.

[75] John, Leslie K. , Acquisti, Alessandro and Loewenstein, George F., The Best of Strangers: Context Dependent Willingness to Divulge Personal Information, July, 2009, available at SSRN: ttp://ssrn.com/abstract=1430482

[76] Joinson, Adam N., Reips, Ulf-Dietrich, Buchanan, Tom Schofield, and Carina B. Paine, Privacy, Trust, and Self-Disclosure Online, Human-Computer Interaction, 25: 1, 2010, pp. 1-24, available at:
http://www.informaworld.com/smpp/section?content=a920445699&fulltext=713240928

[77] The European Commission - DG Justice, Freedom and Security, Study on the economic benefits of privacy enhancing technologies (PETs), July 2010, available at:
http://ec.europa.eu/justice/policies/privacy/studies/index_en.htm

[78] Ronald Leenes, Miriam Lips, R. Poels, Marcel Hoogwout. User aspects of Privacy and Identity Management in Online Environments: Towards a theoretical model of social factors. In PRIME Framework V1 (chapter 9), Editors: Fischer-Hübner, S., Andersson, Ch., Holleboom, T., PRIME project Deliverable D14.1.a, June 2005.

[79] H. Lacohée, S. Crane, A. Pippen. Trustguide: Final Report. October 2006.

[80] Hans Hedbom, 2009. A survey on transparency tools for privacy purposes. Proceedings of the 4th FIDIS/IFIP Summer School. Brno, September 2008, published by Springer, 2009.

[81] Erik Wästlund, Simone Fischer-Hübner. End User Transparency Tools: UI Prototypes. PrimeLife Deliverable D4.2.2, June 2010. www.primelife.eu

[82] Simone Fischer-Hübner,Marit Hansen, Hans Hedbom. Technik für mehr Transparenz. Zeitschrift für Datenschutz und Informationssicherheit, DIGMA, Vol. 10, No. 1, March 2001.

[83] L. Brückner, M. Voss. MozPETs – a Privacy Enhanced Web Browser. Proceedings of the Third Annual Conference on Privacy and Trust (PST05), 2005, Canada.

[84] U. Jendricke, M. Kreutzer, A. Zugenmaier. Mobile Identity Management. Workshop on Security in Ubiquitous Computing. UBICOMP 2002, Göteborg, Sweden.

[85] D. Chappell. Introducing Windows CardSpace, Windows Vista Technical Articles, 2006.

[86] Bruce Schneier, J. Kelsey. Cryptographic Support for Secure Logs on Untrusted Machines. The Seventh USENIX Security Symposium Proceedings, USENIX Press, 1998, 53-62

[87] S. Sackmann, J. Strüker, J. and R. Accorsi,. Personalization in privacy-aware highly dynamic systems. Communications of the ACM, 49(9), September 2006.

[88] Hans Hedbom, Tobias Pulls, Peter Hjärtquist, Andreas Lavén. Adding Secure Transparency Logging to the PRIME Core.Privacy and Identity Management for Life, 5th IFIP WG 9.2,9.6/11.7,11.4,11.6 / PrimeLife International Summer School, Nice, France, 2009, Revised Selected Papers, Springer 2010.

[89] Karel Wouters, Koens Simoens, D. Lathouwers, Bart Preneel. Secure and Privacy-Friendly Logging for eGovernment Services. 3rd International Conference on Availability, Reliability and Security (ARES 2008), IEEE, 2008, 1091-1096.

[90] Mireille Hildebrandt (Ed.), FIDIS Deliverable D 7.12: Biometric behavioral profiling and transparency enhancing tools, 2009, www.fidis.net.

[91] Article 29 Data Protection Working Party. Opinion on More Harmonised Information provisions. 11987/04/EN WP 100, November 25 2004, available online: http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2004/wp100_en.pdf. Appendices are available at: http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2004/wp100a_en.pdf

[92] Mary Rundle, "International Data Protection and Digital Identity Management Tools", presentation at IGF 2006, Privacy Workshop I, Athens, 2006, available online: http://identityproject.lse.ac.uk/mary.pdf

[93] Matthias Mehldau, Iconset for Data-Privacy Declarations v0.1, 2007, available online: http://netzpolitik.org/wp-upload/data-privacy-icons-v01.pdf

[94] Simone Fischer-Hübner, Harald Zwingelberg. UI Prototypes: Policy Administrtaion and Presentation – Version 2. PrimeLife Deliverable D4.3.2, June 2010. www.primelife.eu

[95] Leif Holtz, Katharina Nocun, Marit Hansen., Displaying privacy information with icons. Proceedings, PrimeLife/IFIP Summer School 2010, Helsingborg, 2-6 August 2010, to be published by Springer.

[96] J. Raskin. The Humane Interface – New Directions for Designing Interactive Systems. ACM Press, New York, 2000.

[97] R. Dhamija, L. Dusseault. The Seven Flaws of Identity Management: Usability and Security Challenges. IEEE Security and Privacy, vol. 6, no. 2, pp. 24-29, Mar/Apr, 2008.

[98] Andrew Patrick, S. Kenny. From Privacy Legislation to Interface Design: Implementing Information Privacy in Human-Computer Interaction. Privacy Enhancing Technologies Workshop (PET2003), Dresden/Germany, 2003.

[99] P. Kelly, J. Bresee, L. Cranor, R. Reeder, "A 'Nutrition Label' for Privacy", Symposium On Usable privacy (SOUPS) 2009, July 15-17, 2009, Mountain View, CA, USA.

[100] Simone Fischer-Hübner, John Sören Pettersson, Mike Bergmann, Marit Hansen, Siani Pearson, Marco Casassa Mont, HCI Designs for Privacy-enhancing Identity Management,"Digital Privacy: Theory, Technologies and Practices ", Book Editors: Alessandro Acquisti, Sabrina De Capitani di Vimercati, Stefanos Gritzalis and Costas Lambrinoudakis, Auerbach Publications (Taylor and Francis Group), 2008.

[101] John Sören. Pettersson, Simone Fischer-Hübner, Ninni Danielsson, Jenny Nilsson, Mike Bergmann, Thomas Kriegelstein, Sebastian Clauss, Henry Krasemann, "Making PRIME Usable", Proceedings of the Symposium of Usable Privacy and Security (SOUPS), 4-6. June 2005, Carnegie Mellon University, ACM Digital Library.

[102] M.C. Mont, S. Pearson, and P. Bramhall, Towards accountable management of identity and privacy: sticky policies and enforceable tracing services, 14th International Workshop on Database and Expert Systems Applications (DEXA'03), pages 377-382. IEEE Computer Society, 2003.

[103] Ann Cavoukian, Privacy by Design … Take the Challenge, 2009, available at http://www.privacybydesign.ca/publications/pbd-the-book/

[104] Article 29 Working Party, Opinion 4/2007 on the concept of personal data, 2007, available at: http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2007/wp136_en.pdf

[105] EDPS (The European Data Protection Supervisor), Opinion of the European Data Protection Supervisor on Promoting Trust in the Information Society by Fostering Data Protection and Privacy, 2010, available at: http://www.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/Opinions/2010/10-03-19_Trust_Information_Society_EN.pdf

[106] EOS (European Organisation for Security), White Paper, webpage: http://www.eos-eu.com/LinkClick.aspx?fileticket=PuxKMfvsHmk=&tabid=290

[107] Industry Proposal, Privacy and Data Protection Impact Assessment Framework for RFID Applications, 2010, available at: http://ec.europa.eu/information_society/policy/rfid/documents/d31031industrypia.pdf

[108] Article 29 Working Party, Opinion 5/2010 on the Industry Proposal for a Privacy and Data Protection Impact Assessment Framework for RFID Applications, 2010, available at: http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2010/wp175_en.pdf

[109] http://cordis.europa.eu/fetch?CALLER=PROJ_ICT&ACTION=D&CAT=PROJ&RCN=95924

[110] E. Kavakli, C. Kalloniatis, P. Loucopoulos, S. Gritzalis, Incorporating privacy requirements into the system design process: The pris conceptual framework, Internet Research 16 (2).

[111] L. K. Chung, B. A. Nixon, E. S. K. Yu, J. Mylopoulos, Non-Functional Requirements in Software Engineering, Kluwer Publishing, 2000.

[112] A. I. Ant_on, J. B. Earp, A requirements taxonomy for reducing Web site privacy vulnerabilities, REJ 9 (3) (2004) 169{185.

[113] Vanish, Self-Destructing Digital Data, project webpage: http://vanish.cs.washington.edu/

[114] Jan Camenisch, Els Van Herreweghen: Design and implementation of the idemix anonymous credential system. ACM Conference on Computer and Communications Security, pages 21-30, ACM 2002.

[115] Stefan A. Brands, Rethinking public key infrastructures and digital certificates. MIT Press. 2000.

[116] ENISA report on Management of Multiple Identities, 2010, available at: http://www.enisa.europa.eu/act/it/library

[117] David G. Andersen, Hari Balakrishnan, Nick Feamster, Teemu Koponen, Daekyeong Moon, Scott Shenker. Accountable Internet Protocol (AIP). SIGCOMM, 2008.

[118] David D. Clark, Susan Landau. Untangling Attribution. Workshop on Deterring CyberAttacks: Informing Strategies and Developing Options for U.S. Policy, 2010.

[119] JR Douceur. The Sybill attack. In Proceedings of the 1st International Workshop on Peer-to-Peer Systems (IPTPS '02),, Cambridge, MA, USA, March 2002.!

[120] Joe Kilian, Erez Petrank. Identity Escrow. Advances in Cryptology — CRYPTO '98, Springer Verlag, pp. 169—185.

[121] Katerina Argyraki, Petros Maniatis, Olga Irzak, Subramanian Ashish, Scott Shenker. Loss and Delay Accountability for the Internet. IEEE International Conference on Network Protocols (ICNP), October 2007.

[122] Aydan R. Yumerefendi, Jeffrey S. Chase. Strong Accountability for Network Storage. USENIX Conference of File and Storage Technologies (FAST), 2007.

[123] Nikolaos Michalakis, Robert Soulè, Robert Grimm. Ensuring Content Integrity for Untrusted Peer-to-Peer Content Distribution Networks. USENIX Symposium on Networked Systems Design and Implementation (NSDI), 2007.

[124] Andreas Haeberlen, Ioannis Avramopoulos, Jennifer Rexford, Peter Druschel. NetReview: Detecting when interdomain routing goes wrong. 6th Symposium on Networked Systems Design and Implementation (NSDI), 2009.

[125] Andreas Haeberlen, Petr Kuznetsov, Peter Druschel. PeerReview: Practical Accountability for Distributed Systems. ACM Symposium on Operating Systems Principles (SOSP), 2007.

[126] Andreas Haeberlen, Paarijaat Aditya, Rodrigo Rodrigues, Peter Druschel. Accountable Virtual Machines. USENIX Symposium on Operating Systems Design and Implementation (OSDI), 2010.

[127] Daniel J. Weitzner, Harold Abelson, Tim Berners-Lee, Joan Feigenbaum, James Hendler, Gerald Jay Sussman. Information Accountability. Communications of the ACM (CACM), 51(60, June 2008.

[128] L. Kagal, H. Abelson. Access control is an inadequate framework for privacy protection. W3C Privacy Workshop, 2010.

[129] Butler Lampson. Computer security in the real world. Annual Computer Security Applications Conference, 2000.

[130] Aydan R. Yumerefendi, Jeffrey S. Chase. Trust but verify: accountability for network services. ACM SIGOPS European Workshop, 2004.

[131] Petros Maniatis, Mary Baker. Secure History Preservation Through Timeline Entanglement. USENIX Security Symposium, 2002.

[132] C. Dwork. Differential privacy. In 33rd International Colloquium on Automata, Languages and Programming (ICALP), 2006

[133] Teresa Lunt, Paul Aoki, Dirk Balfanz, Glenn Durfee, Philippe Golle, Diana Smetters, Jessica Staddon, Jim Thornton, Tomas Uribe. Protecting the Privacy of Individuals in Terrorist Tracking Applications. Technical Report AFRL-IF-RS-TR-2005-131, US Air Force Research Laboratory, 2005.

[134] J. Howlett, S. Hartmann, H. Tschofenig, E. Lear, "Application Bridging for Federated Access Beyond Web (ABFAB) Architecture", draft-lear-abfab-arch-00 (work in progress), October 2010.

[135] E. Maler, "Account linking and the F-word", blog post available at: http://www.xmlgrrl.com/blog/2007/07/17/account-linking-and-the-f-word/

[136] D. Crocker, "Internet Mail Architecture", RFC 5598, July 2009.

[137] E. Allman, J. Callas, M. Delany, M. Libbey, J. Fenton, M. Thomas, "DomainKeys Identified Mail (DKIM) Signatures", RFC 4871, May 2007.

[138] "The SpamHaus Project", Nov. 2010, available at http://www.spamhaus.org

[139] P. Hoffman, J. Levine, A. Hathcock, "Vouch By Reference", RFC 5518, April 2009.

[140] Working Group MARF, IETF Messaging Abuse Reporting Format (MARF) available at: http://datatracker.ietf.org/wg/marf/charter/

[141] "XConnect", Nov. 2010, available at: http://www.xconnect.net/

[142] "GSM Association", Nov. 2010, available at: http://www.gsmworld.com/

[143] Wikipedia, "IP eXchange (IPX)", Nov. 2010, available at: http://en.wikipedia.org/wiki/Ip_exchange

[144] IETF, "Session PEERing for Multimedia INTerconnect (SPEERMINT) working group", Nov. 2010, available at: http://www.ietf.org/dyn/wg/charter/speermint-charter

[145] IETF, "Data for Reachability of Inter/tra-NetworK SIP (drinks) working group", Nov. 2010, available at: https://datatracker.ietf.org/wg/drinks/charter/

[146] IETF, Multiple AoR reachabiliTy InformatioN Indication (martini) working group", Nov. 2010, available at: https://datatracker.ietf.org/wg/martini/charter/

[147] O. Lendl, "VoIP Peering: Background and Assumptions", draft-lendl-speermint-background-02.txt (work in progress), Oct. 2008.

[148] S. Marcus, D. Elixman, "The Future of IP Interconnection: Technical, Economic, and Public Policy Aspects", Jan. 2008, available at: http://ec.europa.eu/information_society/policy/ecomm/doc/library/ext_studies/future_ip_intercon/ip_intercon_study_final.pdf

[149] J. Rosenberg, C. Jennings, M. Petit-Huguenin: "Verification Involving PSTN Reachability: Requirements and Architecture Overview", draft-rosenberg-dispatch-vipr-overview-01 (work in progress), Oct. 2010.

[150] "inCommon Federation", Nov. 2010, available at http://www.incommonfederation.org

[151] "inCommon Meta Data Repository", Nov. 2010, available at: http://wayf.incommonfederation.org/InCommon/InCommon-metadata.xml

[152] "Where are you from (WAYF) Federation", Nov. 2010, available at: http://wayf.dk

[153] D. Simonsen, J. Madsen, "Trusted third party based ID federation, lowering the bar for connecting and enhancing privacy", Aug. 2009, available at http://wayf.dk/wayfweb/articles_attchmt/2009_08_05%20TNC2009paper_WAYF_David_Simonsen_and_Jacob-Steen_Madsen.pdf

[154] D. Simonsen: "Recording of the WAYF presentation at the TERENA 2009 Networking Conference", 2009, http://tnc2009.terena.org/media/archivec1e3.html?stream=2B (presentation starts at the 33rd minute of the video).

[155] WAYF, "WAYF Contracts", Nov. 2010, available at: http://wayf.dk/wayfweb/contracts.html

[156] WAYF, "WAYF Meta Data", Nov. 2010, available at: http://wayf.dk/wayfweb/meta_data.html

[157] WAYF, "How to get my institution connected", Nov. 2010, available at: http://wayf.dk/wayfweb/how_to_get_my_institution_connected.html

[158] Geant, EduGain, web page available at: http://www.edugain.org/

[159] "Education Roaming (EduRoam) Federation", Nov. 2010, available at: http://www.eduroam.org

[160] Wimax Forum, "WiMAX FORUM® ROAMING GUIDELINES, Release 1.0 Version 2", June 2009, available at http://www.wimaxforum.org/sites/wimaxforum.org/files/technical_document/2009/08/WMF-T40-001-R010v02_Roaming-Guidelines.pdf

[161] EduRoam, Deliverable DS5.1.1: eduroam Service Definition and Implementation Plan", Jan. 2008, available at: http://www.eduroam.org/downloads/docs/GN2-07-327v2-DS5_1_1-_eduroam_Service_Definition.pdf

[162] "Payment Card Industry Data Security Standard", Nov. 2010, available at: http://en.wikipedia.org/wiki/Payment_Card_Industry_Data_Security_Standard

[163], Federal Deposit Insurance Corporation (FDIC), available at: http://www.fdic.gov/

[164] SWIFT website, http://www.swift.com/about_swift/press_room/press_releases/press_releases_archive/subpoenaed_swiftmessagedata_adequatelyprotected.page

[165],Swift Affair, http://ec.europa.eu/justice/policies/privacy/news/docs/PR_Swift_Affair_23_11_06_en.pdf

[166] CDT, "CDT Discusses Key Policies Issues Surrounding User-Centric Identity Management", Nov. 2009, available at: http://cdt.org/policy/cdt-discusses-key-policies-issues-surrounding-user-centric-identity-management

[167] International Data Protection Authorities, "International Standards on the Protection of Personal Data and Privacy: The Madrid Resolution", Nov. 2009, available at: http://www.gov.im/lib/docs/odps//madridresolutionnov09.pdf

[168] Vincent Toubiana, Arvind Narayanan, Dan Boneh, Helen Nissenbaum, Solon Barocas. Adnostic: Privacy Preserving Targeted Advertising. Annual Network and Distributed Systems Security Symposium (NDSS), 2010.

[169] Saikat Guha, Bin Cheng, Alexey Reznichenko, Paul Francis. Privad: Practical Privacy in Online Advertising. Max Planck Institute for Software Systems Technical report MPI-SWS-2010-001.

[170] Philippa Gill, Yashar Ganjali, and David Lie, Due, Dude, where's that IP? Circumventing measurement-based IP geolocation, Proceedings of the 19th USENIX Security Symposium, 16 pages, USENIX Association 2010.

[171] R. Barnes, M. Lepinski, A. Cooper, J. Morris, H. Schofenig, H. Schulzrinne, An Architecture for Location and Location Privacy in Internet Applications Internet-Draft , October 11, 2010.

[172] Alastair R. Beresford, Frank Stajano, Mix Zones: User privacy in location-aware services, Pervasive Computing and Communication Security (PerSec), pages 127-131, IEEE, Mar 2004.

[173] R. Anderson and S. Fuloria, On the security economics of electricity metering, in The Ninth Workshop on the Economics of Information Security, 2010.

[174] P. McDaniel and S. McLaughlin, Security and privacy challenges in the smart grid, IEEE Security and Privacy, vol. 7, pages 75–77, 2009

[175] A. Cavoukian, J. Polonetsky, and C. Wolf, "Smart privacy for the smart grid: embedding privacy into the design of electricity conservation," Identity in the information Society, 2009.

[176] E.L. Quinn, "Privacy and the new energy infrastructure," SSRN eLibrary, 2009.

[177] A. Wagner, S. Speiser, O. Raabe, and A. Harth, "Linked data for a privacy-aware smart grid," INFORMATIK 2010 Workshop – Informatik für die Energiesysteme der Zukunft, 2010.

[178].F.D. Garica and B. Jacobs, "Privacy-friendly energy-metering via homomorphic encryption," Radboud University Nijmegen, Technical Report, February 2010.

[179] Alfredo Rial and George Danezis, Privacy-preserving smart metering, http://research.microsoft.com/pubs/141726/main.pdf, 2010

[180] George Danezis, Claudia Diaz, and Paul Syverson, Systems for Anonymous Communication. In CRC Handbook of Financial Cryptography and Security, CRC Cryptography and Network Security Series, B. Rosenberg, and D. Stinson (Eds.), Chapman & Hall, 61 pages (in print), 2010

[181] Andreas Pfitzmann and Marit Kohntopp. Anonymity, unobservability, and pseudonymity – A proposal for terminology. In Hannes Federrath, editor, Designing Privacy Enhancing Technologies: International Workshop on Design Issues in Anonymity and Unobservability, pages 1-9. Springer- Verlag, LNCS 2009, July 2000.

[182] ISO 15408 Information technology – Security techniques – Evaluation criteria for IT security –Part 1: Introduction and general model, 2005.

[183] Andrei Serjantov and George Danezis. Towards an information theoretic metric for anonymity. In Roger Dingledine and Paul Syverson, editors, Privacy Enhancing Technologies: Second International Workshop, PET 2002, pages 41-53, San Francisco, CA, USA, April 2002. Springer-Verlag, LNCS 2482.

[184] Claudia Diaz, Stefaan Seys, Joris Claessens, and Bart Preneel. Towards measuring anonymity. In Roger Dingledine and Paul Syverson, editors, Privacy Enhancing Technologies: Second International Workshop, PET 2002, pages 54-68, San Francisco, CA, USA, April 2002. Springer-Verlag, LNCS 2482.

[185] Oliver Berthold, Andreas Pfitzmann, and Ronny Standtke. The disadvantages of free MIX routes and how to overcome them. In Hannes Federrath, editor, Designing Privacy Enhancing Technologies: International Workshop on Design Issues in Anonymity and Unobservability, pages 30-45. Springer-Verlag, LNCS 2009, July 2000.

[186] David Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. Communications of the ACM, 4(2):84-88, February 1981.

[187] George Danezis, Roger Dingledine, and Nick Mathewson. Mixminion: Design of a type III anonymous remailer protocol. In Proceedings, IEEE Symposium on Security and Privacy, pages 2-15, Berkeley, CA, May 2003. IEEE Computer Society.

[188] Andrei Serjantov, Roger Dingledine, and Paul Syverson. From a trickle to a flood: Active attacks on several mix types. In Fabien A.P. Petitcolas, editor, Information Hiding: 5th International Workshop, IH 2002, pages 36-52. Springer-Verlag, LNCS 2578, 2002.

[189] Andreas Pfitzmann, Birgit Pfitzmann, and Michael Waidner. ISDNMIXes: Untraceable communication with very small bandwidth overhead. In Wolfgang Eelsberg, Hans Werner Meuer, and Günter Müller, editors, Kommunikation in Verteilten Systemen, Grundlagen, Anwendungen, Betrieb, GI/ITG-Fachtagung, volume 267 of Informatik-Fachberichte, pages 451-463. Springer-Verlag, February 1991.

[190] Anja Jerichow, Jan Müller, Andreas Pfitzmann, Birgit Pfitzmann, and Michael Waidner. Real-time MIXes: A bandwidth-efficient anonymity protocol. IEEE Journal on Selected Areas in Communications, 16(4):495-509, May 1998.

[191] Oliver Berthold, Hannes Federrath, and Stefan Köpsell. Web MIXes: A system for anonymous and unobservable Internet access. In Hannes Federrath, editor, Designing Privacy Enhancing Technologies: International Workshop on Design Issues in Anonymity and Unobservability, pages 115-129. Springer-Verlag, LNCS 2009, July 2000.

[192] Benedikt Westermann, Rolf Wendolsky, Lexi Pimenidis, Dogan Kesdogan. Cryptographic Protocol Analysis of AN.ON. In Financial Cryptography (FC'10), pages 114-128, Springer-Verlag, LNCS 6052, 2010.

[193] Rolf Wendolsky, Dominik Herrmann, and Hannes Federrath, Performance Comparison of Low-Latency Anonymisation Services from a User Perspective. Privacy Enhancing Technologies, pages 233-253, Springer-Verlag, 2007.

[194] Jeremy Clark, Paul C. Van Oorschot, Carlisle Adams, An Examination of Tor Interfaces and Deployability, Proceedings of the 3rd Symposium on Usable Privacy and Security, SOUPS 2007, Lorrie Faith Cranor, Ed., page 41-51, ACM, 2007.

[195] Nikita Borisov. Anonymous Routing in Structured Peer-to-Peer Overlays. PhD thesis, University of California, Berkeley, 2005.

[196] Arjun Nambiar and Matthew Wright. Salsa: A structured approach to large-scale anonymity. In Rebecca N. Wright, Sabrina De Capitani di Vimercati, and Vitaly Shmatikov, editors, CCS'06: Proceedings of the 13th ACM Conference on Computer and Communications Security, pages 17-26. ACM Press, 2006.

[197] Prateek Mittal and Nikita Borisov. Information leaks in structured peer-to-peer anonymous communication systems. In Paul Syverson, Somesh Jha, and Xiaolan Zhang, editors, CCS'08: Proceedings of the 15th ACM Conference on Computer and Communications Security, pages 267-278. ACM Press, 2008.

PO Box 1309 71001 Heraklion Greece
Tel: +30 2810 391 280 Fax: +30 2810 391 410
Email: info@enisa.europa.eu
www.enisa.europa.eu