Document type: Report
Title: D2.1- State of the Art
Work Package: WP2
Deliverable Number: D2.1
Editor(s): S. Gürses, K.U. Leuven
Dissemination level: Public
Preparation date: 30. September 2011
Version: 1.0

Legal Notice
All information included in this document is subject to change without notice. The Members of the IWT SBO SPION project make no warranty of any kind with regard to this document, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose. The Members of the IWT SBO SPION project shall not be held liable for errors contained herein or direct, indirect, special, incidental or consequential damages in connection with the furnishing, performance, or use of this material.
The IWT SBO SPION Project

<table>
<thead>
<tr>
<th>Nr.</th>
<th>Participant name</th>
<th>Country</th>
<th>Department</th>
<th>Participant role</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>K.U. Leuven</td>
<td>BE</td>
<td>COSIC/ESAT</td>
<td>Coordinator</td>
</tr>
<tr>
<td>2</td>
<td>K.U. Leuven</td>
<td>BE</td>
<td>DISTRINET</td>
<td>Partner</td>
</tr>
<tr>
<td>3</td>
<td>K.U. Leuven</td>
<td>BE</td>
<td>DTAI</td>
<td>Partner</td>
</tr>
<tr>
<td>4</td>
<td>K.U. Leuven</td>
<td>BE</td>
<td>ICRI</td>
<td>Partner</td>
</tr>
<tr>
<td>5</td>
<td>Vrije Universiteit Brussel</td>
<td>BE</td>
<td>SMIT</td>
<td>Partner</td>
</tr>
<tr>
<td>6</td>
<td>University of Ghent</td>
<td>BE</td>
<td>OWK</td>
<td>Partner</td>
</tr>
<tr>
<td>7</td>
<td>Carnegie Mellon Univer-</td>
<td>USA</td>
<td>Heinz</td>
<td>Partner</td>
</tr>
</tbody>
</table>

Contributors

<table>
<thead>
<tr>
<th>Name</th>
<th>Organisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Alessandro Acquisti</td>
<td>CMU</td>
</tr>
<tr>
<td>2 Brendan Van Alsenoy</td>
<td>K.U. Leuven</td>
</tr>
<tr>
<td>3 Ero Balsa</td>
<td>K.U. Leuven</td>
</tr>
<tr>
<td>4 Bettina Berendt</td>
<td>K.U. Leuven</td>
</tr>
<tr>
<td>5 Dave Clarke</td>
<td>K.U. Leuven</td>
</tr>
<tr>
<td>6 Claudia Diaz</td>
<td>K.U. Leuven</td>
</tr>
<tr>
<td>7 Bo Gao</td>
<td>K.U. Leuven</td>
</tr>
<tr>
<td>8 Seda Gürses</td>
<td>K.U. Leuven</td>
</tr>
<tr>
<td>9 Aleksandra Kuczerawy</td>
<td>K. U. Leuven</td>
</tr>
<tr>
<td>10 Jo Pierson</td>
<td>VUB</td>
</tr>
<tr>
<td>11 Frank Piessens</td>
<td>K.U. Leuven</td>
</tr>
<tr>
<td>12 Rula Sayaf</td>
<td>K.U. Leuven</td>
</tr>
<tr>
<td>13 Tammy Schellens</td>
<td>UGhent</td>
</tr>
<tr>
<td>14 Fred Stutzman</td>
<td>CMU</td>
</tr>
<tr>
<td>15 Ellen Vanderhoven</td>
<td>UGhent</td>
</tr>
<tr>
<td>16 Ralf De Wolf</td>
<td>VUB</td>
</tr>
</tbody>
</table>
Table of Contents

1 EXECUTIVE SUMMARY ................................................................. 9

2 INTRODUCTION................................................................................ 10
2.1 SCOPE AND OBJECTIVES 10

3 SOCIAL ASPECTS (SMIT) ............................................................... 13
3.1 STATE OF AFFAIRS AND EVOLUTION OF SNS 13
3.2 A SOCIAL PERSPECTIVE ON SNS 14
   3.2.1 Why do individuals use SNS? 14
   3.2.2 How does SNS effect the representation of self? 16
   3.2.3 What does this mean regarding privacy? 16
   3.2.4 The challenge 17
3.3 A COMMUNITY PERSPECTIVE ON SNS 17
   3.3.1 The meaning of community in present society 17
   3.3.2 The relationship between offline and online communities 18
   3.3.3 The meaning of community for PETs 20
   3.3.4 Tackling the challenge 21
3.4 SOCIAL REQUIREMENTS AND EVALUATION FRAMEWORK 22
3.5 CONCLUSION 23

4 BEHAVIORAL ASPECTS (CMU) ..................................................... 24
4.1 DISCLOSURES AND PRIVACY IN SOCIAL NETWORK SITES 24
   4.1.1 The Evolution of Social Network Sites 24
   4.1.2 Disclosures 24
   4.1.3 Privacy 25
4.2 PRIVACY, BEHAVIORAL ECONOMICS, AND SOFT PATERNALISTIC INTERVENTIONS 28
   4.2.1 From Hurdles in Privacy Decision Making to Soft Paternalism 28
   4.2.2 Privacy (and Security) Nudges in the Literature 30
   4.2.3 Privacy Nudges in the Industry 31
   4.2.4 Challenges 32

Version 1.0
# 5 Educational Aspects (OWK)

5.1 Evaluation Framework .................................................. 33

5.2 Main Existing Educational Packages Evaluated .................. 35

5.2.1 Primary School ....................................................... 35

5.2.2 Secondary School .................................................... 37

5.3 Gaps and Future Challenges ........................................... 40

# 6 Legal Aspects (ICRI)

6.1 Introduction ........................................................................ 42

6.2 Data Protection ............................................................... 42

6.2.1 Scope ........................................................................... 42

6.2.2 Data Protection Principles .......................................... 44

6.3 E-Privacy ........................................................................... 45

6.3.1 Scope ........................................................................... 45

6.3.2 Location Based Services ............................................ 46

6.3.3 Cookies ......................................................................... 46

6.3.4 Spam ............................................................................. 47

6.4 E-Commerce ....................................................................... 47

6.4.1 Scope ........................................................................... 47

6.4.2 Substantive Provisions .............................................. 48

6.4.3 Liability of Intermediaries .......................................... 48

6.4.4 (Voluntary) Codes of Conduct .................................. 50

6.5 Conclusion .......................................................................... 50

# 7 Access Control and Accountability Aspects (DISTRINET)

7.1 Access Control Models .................................................... 51

7.1.1 Preliminary Notions .................................................... 51

7.1.2 Classical ACMs .......................................................... 51

7.1.3 Access Control Models for Social Networks .................. 52

7.2 Access Control and Information Flow ............................... 56

7.2.1 Script access control and isolation .............................. 56

7.2.2 Fine-grained access control and information flow control 58
7.3 Accountability and Audit ........................................ 59
   7.3.1 Accountability Applications .................................. 59
   7.3.2 Audit Mechanisms ........................................... 60
7.4 Challenges .................................................................. 60

8 Confidentiality Aspects (COSIC) ..................................... 62
   8.1 Introduction ................................................................ 62
   8.2 Graph Anonymization .............................................. 62
      8.2.1 De-Anonymization Attacks .................................. 64
   8.3 Content Protection .................................................. 64
   8.4 Traffic Data Protection ............................................. 66
   8.5 Darknets .................................................................. 66
   8.6 Metrics .................................................................... 67
   8.7 Future Directions .................................................... 68

9 Feedback and Awareness Aspects (DTAI) ......................... 69
   9.1 Introduction ............................................................ 69
      9.1.1 The general idea of privacy feedback and awareness .... 69
      9.1.2 PFA tools in the context of SPION – complementary solution approaches 69
   9.2 What is a Privacy Feedback and Awareness Tool? ........... 70
   9.3 Existing PFA Tools ................................................... 71
      9.3.1 Conceptual developments in the scientific literature .... 71
      9.3.2 Available Tools ................................................ 71
      9.3.3 Reflections on the classification and the range of PFA tools ........ 80
      9.3.4 A note about the evaluation of PFA tools ................. 81
   9.4 Conclusion ............................................................. 81

10 Conclusion and Outlook ............................................... 83
   10.1 Interdisciplinary Mapping ......................................... 83
      10.1.1 Awareness and Educational Aspects ....................... 83
      10.1.2 Privacy Settings, Discrepancies and Control Aspects .... 85
      10.1.3 Accountability and Compliance Aspects ................... 86
      10.1.4 Social and Community Aspects ............................ 88
   10.2 Future Steps .......................................................... 89
Keyword List

State of the Art, Research Scope, Online Social Network Sites (SNS), Privacy, Security
Table 1: default

<table>
<thead>
<tr>
<th>ACM</th>
<th>Access Control Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>DoW</td>
<td>Description of Work</td>
</tr>
<tr>
<td>HCI</td>
<td>Human Computer Interaction</td>
</tr>
<tr>
<td>OSN</td>
<td>Online Social Network Sites</td>
</tr>
<tr>
<td>PFA</td>
<td>Privacy Feedback and Awareness Tools</td>
</tr>
<tr>
<td>RBAC</td>
<td>Role Based Access Control</td>
</tr>
<tr>
<td>ReBAC</td>
<td>Relationship Based Access Control</td>
</tr>
<tr>
<td>SPION</td>
<td>Security and Privacy for Online Social Networks (Project Title)</td>
</tr>
<tr>
<td>SNS</td>
<td>Online Social Network Sites</td>
</tr>
</tbody>
</table>
1 Executive Summary

The objective of this deliverable is to provide an overview of existing literature and case descriptions of social and community uses of online Social Network Services (SNS); summary of available educational solutions and empirical evidence of the efficiency and efficacy and the satisfaction they generate; analysis of legal frameworks applicable to SNS, and a review of confidentiality, access control and information flow, as well as feedback and awareness solutions. The Deliverable also includes an analysis of how the gaps and challenges in the different disciplines represented in the project are interrelated, mapping out research gaps and potentials for future interdisciplinary research on privacy and security in online Social Network Services.

Readers Guide: In Chapter 2 we introduce this deliverable’s objectives and scope. Chapters 3 through 5 are organized to present the state of the art and research challenges to privacy and security in social networks from the different social science perspectives represented by the partners of SPION. This is then followed by a survey of the legal frameworks relevant to security and privacy in online social networks in Chapter 6. Next, in Chapters 7 through 9 the technical partners of the project provide an overview of the technical proposals that address the different aspects of the identified challenges while identifying gaps with respect to the unique technical hurdles of privacy and security in online social networks. Finally, in Chapter 10 we conclude with an analysis of how the gaps in the different disciplines can be addressed in an interdisciplinary fashion in collaboration with the different project partners and provide an outlook into future steps of the project.
2 Introduction

2.1 Scope and Objectives

The document in your hands is the first deliverable of the SPION project, an interdisciplinary project on privacy and security in online Social Network Services (SNS) with seven partners representing various disciplines. The objective of this deliverable is to provide an overview of relevant state of the art in the respective fields upon which we will base the research and valorization activities planned in the next few years. However, calling this document a “state of the art” does not do justice to either the intricate process through which it has been written nor to the results presented here.

Producing a state of the art document with partners from varying disciplines starts with the questions: what approaches to analyzing the “state of the art” exist in that discipline and what are the objectives of such a review. Already when we authored the proposal of the SPION project, we identified substantial differences in what counts as a review of the state of the art in the different disciplines represented in the project. This diversity in practices has an effect on the approaches that are utilized by each partner, the output of such an analysis and eventually on the collaboration that can be based on the results of such an analysis. In the SPION Description of Work (DoW) it is stated that the objectives of Deliverable 2.1 Report on Research Activities (State of the art) are as follows:

- Overview of existing literature and case descriptions of social and community uses of SNS.
- Summary of available educational solutions and empirical evidence of the efficacy, efficiency and the satisfaction they generate.
- Analysis of legal frameworks applicable to SNS.
- Review of confidentiality, identity management, trust.

As expected in the proposal, the partners had different ways of surveying their state of the art and identifying their gaps and future research challenges. Specifically, SMIT and CMU focus on providing an overview of social and behavioral economics literature to explore conceptions of privacy and related challenges and to subsequently design their studies. OWK’s analysis focuses on an expansive review of existing educational packages about online privacy in Flanders. ICRI provides an evaluation of the relevant European legislation that would apply to SNS and the types of issues that arise when these are applied. The technical partners also prove their differences: while COSIC and DISTRINET mainly rely on previous proposals in research papers for their system models, DTAI enhance their review by including a survey and categorization of feedback and awareness tools recently introduced into social networks. By doing so, DTAI makes up for the fact that few research results are available in this cutting edge line of inquiry. Hence, the object through which the inquiry is done changed for each partner, from literature to educational packages, from legislation to existing software solutions.

A further layer of differences manifests itself in the definitions and approaches to privacy and security as topics of research. SMIT distinguishes between social privacy and instrumental privacy. The former is defined, according to Raynes-Goldie [2010] as ‘the control of information flow about how and when their personal information is shared with other people.’ The latter refers to data accessed by governments and corporations, e.g. using data mining and related statistical analysis methods [danah m. boyd and Hargittai, 2010]. In the behavioral economics research, the focus is not on giving a definition of privacy, but to help users in making the “right” decisions with respect to privacy, where privacy decisions are framed as being about balancing revelation and protection in ways that maximize a user’s welfare and minimize her future regrets. Both SMIT and CMU show recent results that indicate that privacy is not only a matter of individual decision making, but part of a social practice [Kaufman et al., 2008]. In OWK, the concern is with educational packages designed to
raise awareness against privacy risks that young people are exposed to, that could infringe upon their privacy, implicitly defining privacy risks as those activities that can pose a threat to young people.

The technical partners also utilize varying conceptions of privacy in their research, which run parallel to the different privacy research paradigms within computer science [Gürses and Berendt, 2010]. The research on traffic analysis protection, content protection, and graph anonymization by COSIC are typical examples of solutions in the privacy as confidentiality paradigm. The solutions in this paradigm are inspired by the definition of privacy as “the right to be let alone” [Warren and Brandeis, 1890]. Such solutions aim to create an individual autonomous sphere from which disclosure of information is by default prevented, or information is minimally disclosed in a way that cannot be linked back to the individual. The elaborate models of access control and information flow as studied by DISTRINET are representative of privacy as control, which includes solutions that rely on the definition of privacy as “the right of the individual to decide what information about himself should be communicated to others and under what circumstances” [Westin, 1970]. The main objective of these solutions is to provide individuals with autonomy, control and oversight over data collection, processing and disclosure. Finally, DTAI’s work on feedback and awareness tools are representative of the privacy as practice paradigm, which refers to solutions based on the definition of privacy as “the freedom from unreasonable constraints on the construction of one’s own identity” [Agre, 1999]. Assuming that privacy decisions are not made individually or a-priori, the goal of these solutions is to increase users’ and communities’ privacy awareness through feedback on information practices.

As we have sketched above, there is great variety of perspectives and approaches, significant contrasts in terminology and language, and different understandings of privacy (and security) represented by the partners in this project. These differences give this project its strength but also pose to it some challenges. Documents like the state of the art play an important role in bringing forth the strengths of interdisciplinary work, while also addressing the challenges of such a practice. The objective of this document is hence not “only” to provide the readers with an overview of the state of the art in our disciplines and fields. It is also a document that provides us, the partners, with an opportunity to relate our discussions into a common thread, to initiate a process in which we map and converge our vocabulary, and finally, to fine tune our research goals based on the state of affairs in the various disciplines. The final results of this convergence and conversation is summarized in Chapter 10, where we discuss points of intersection across the research fields represented in this project and gaps identified in the state of the art.

Next, we shortly describe the process through which this document was produced, and how we introduced various steps to address the challenges of interdisciplinary work and to utilize these perspectives to improve our research.

**Production Process:** We started early on with the organization of this deliverable. In the first 6 months of the project, each partner made a presentation to the others about the research topics they are working on. These presentations, and the following discussions, allowed the partners to get a better understanding of the different research areas, and to bring in their perspectives on the problems addressed by the presenting researchers.

Next, in June, 2010, each partner was asked to bring an outline of their state of the art to a project meeting and to engage in a discussion on the organization of the deliverable. At the end of this meeting, we set milestones for the different steps of the deliverable. In the first round, each partner prepared a first draft of their chapter in the Deliverable. This was followed by a project team meeting, where the authors received feedback from the editor of the Deliverable with respect to what seemed to be missing, alignment of the research with the project objectives and with the work of other partners. At the end of this round, the editor had an overview of the incoming packages and how they fit into the deliverable. This first impression was discussed and basic divergences were consolidated.

After this first review, the partners revised their text and submitted it for a second round. In this round, partners reviewed each others’ chapters. The assignments for reviewers were made such that for each chapter a scholar with a social science background and a scholar with a technical
background was a reviewer. Following this review, the use of terminology as well as differences in assumptions and approaches were discussed and consolidated over the mailing list.

Finally, the partners were asked to identify and discuss how the research gaps and their research interests interact with each other. Specifically, each partner was asked to discuss and document:

1. How topics that the researcher sees as gaps and challenges in her discipline are also re-occurring in the fields of the other partners.
2. How the gaps and challenges identified in the field of the researcher may be addressed through some of the research ideas that are in the other chapters.
3. Which possibilities the researcher sees for collaborations on common questions from the different perspectives
4. How the researcher thinks her research results may be interesting for the questions that the other partners are seeing as challenges.
5. Any other points the researcher thinks are relevant for the interaction given the state of the art in the different disciplines.

The results of these discussions and analyses were then consolidated and circulated for feedback and refinement.

Once the different editorial cycles were completed, we integrated the document into one whole. This required both editorial work, as well as formatting work. This intricate and cyclical process, as well as the content of this document, constitute this Deliverable.
3 Social Aspects (SMIT)

3.1 State of Affairs and Evolution of SNS

The identity layer seems to have nested itself permanently on the Internet. By acting towards others and self people create various types of identity. ‘An identity refers to a person’s location relative to others in the situation, the community, or the society as a whole’, according to Hewitt [2007]. Next to emotions and cognitions, identity can be seen as a third fraction of an individual that constitutes the self. Positioning the self used to take place in an entirely offline environment. Nowadays the processes of identity on Web 2.0 applications have become part of everyday normal life. Thus the Internet in the realm of mass self-communication [Castells, 2009] does not solely revolve around information gathering and communication anymore. SNS make it possible to present the own identity online, taking communication processes to a whole new level. In this section we want to present some remarkable statistics and trends regarding SNS, to demonstrate the strong interconnection between SNS and society. Over the following paragraphs the three most relevant topics that concern the people regarding SNS will be discussed: (1) the overall use; (2) the economic value and (3) privacy issues.

Facebook is currently the largest SNS [Hampton et al., 2011]. According to socialbakers.com, one of the biggest Facebook statistics portals in the world, the total amount of users is closing in to 700 million [“The Ultimate Answer”, 2011]. Hence, more than 1 out of 10 people in the world have a Facebook-account. When looking at numbers for Belgium we see that approximately 4.2 million of the Belgians use this SNS, with a growth of 9.90% last 6 months [“Facebook Statistics”, n.d.]. The digimeter, a research initiative that maps the use and ownership of media and ICT technologies in Flanders, states that almost two thirds of SNS users have an account on Facebook. Moreover, half of the Flemish people online have an account on a SNS [Schuurman and Schuurman, 2010]. Not only Facebook, but also Twitter has acquired a large market share nowadays. It is difficult to determine the precise amount of users on Twitter, but ‘tweets per day’ (TPD) give an indication of the usage of this medium. The average TPD in March 2010 was 50 million [“Twitter Statistics”, n.d.]. The average TPD in February 2011 was 140 million. The 280% increase of tweets gives a rather strong indication on what can be defined as the hype of today. It is uncertain whether this trend will sustain itself over time.

It seems not so long ago that other SNS, like Friendster and Myspace, were extolled as the revelation of 21th century. In May 2011 however, Friendster repositioned itself from a social network site to a social gaming site. Likewise the number of MySpace users has declined immensely [“Statistics Summary”, n.d.]. In Belgium a similar phenomenon has occurred for Netlog [Deckmyn, 2010]. Like Friendster, Massive Media (The company behind Netlog), is broadening its scope to gaming and dating [“netlog richt zijn pijlen”, 2011]. It seems that these old hands simply cannot compete with Facebook and Twitter today. Maybe the future will bring the same destiny for Facebook and Twitter, maybe not. Google+, the new social network site of Google, is a new competitor on the SNS market [Austin, 2011]. With 25 million users and growing, Google+ is the fastest website to reach that audience size [Barr, 2011]. Still, other sources state that its weekly visits are declining [Whitney, 2011]. One thing these evolutions teach us is that the identity and social relationship layers on the Internet are here to stay and continue to gain ground.

‘Social advertisements’, ‘advergaming’, ‘behavioral targeting’ and ‘profile targeting’ all illustrate on how social relationships are adapted to the principles of the market. In other words, the identity layer and economy go hand-in-hand together. [Thrift, 2005] refers to this as ‘knowing capitalism’. Once you are on a SNS, these platforms want the user to share information and stick to it [Chapman, 2010]. At the beginning of this year, Goldman Sachs Group investment valued Facebook at $50 billion [Van Tuil, 2011]. According to PrivCo, Facebook raised $1.5 billion in January 2011. This can be considered as one of the largest amounts of venture capital ever raised by a private company [Privco, 2011]. Now the question has shifted from ‘whether SNS will enter the exchange market’ to
Facebook 56.1%
Netlog 18.5%
LinkedIn 12.8%
Twitter 8.2%
MySpace 6.2%
Hyves 3.0%

Table 3.1: Amount of SNS accounts in Flanders. Data collected in the period between August 2010 and November 2010. Total sample: n 1403. Source: www.digimeter.be

‘when this will happen’. Nonetheless, SNS are based on capitalistic principles of commodification and market exchange and this does not seem to change. From an economic point of view we could state that people are not customers, but primarily products on SNS. Another fact that seems to stay.

To gain insight into the economical SNS landscape, we briefly discuss on how personal information is commodified for LinkedIn. Just as for other commercial SNS LinkedIn makes use of cookies and log files to collect data: the URL of the site you came from, your IP-address, type of browser, etc [Heyman and Pierson, 2011]. Likewise, information is gathered through web beacons. Heyman and Pierson [2011] state that ‘the PII gathering itself is a very clear manifestation of immaterial labour 2.0. Firstly, the PII is solicited in a way that is beneficial and interesting for the user. Secondly, users can use this pool of data for professional purposes in various ways that provide both stickiness and revenue. Lastly, this data is used for behavioural advertising, which is a third means of valorisation.’ The Silicon Alley Insider identifies three revenue streams: premium subscriptions, marketing solutions and hiring solutions [Yarow and Angelova, 2011]. It is beyond the scope of this state of the art to talk about all three revenue streams. Hence, we will focus on the first one: premium subscriptions. When subscribing on LinkedIn people may choose from two options: premium and basic. The former subscription gives you more access to all sorts of information: view profiles of over 100 million people; add notes and contacts into any profile; etc. The latter has fewer options. Through this process LinkedIn creates scarcity and deforms personal information into a commodity on the market.

The fact that identity is ubiquitously presented on the web combined with the fact that social media are commodified in an unsurpassed way raises privacy questions. Famous quotes from Eric Schmidt, CEO of Google and Mark Zuckerberg, CEO from Facebook should make the reader aware on how an economic point of view has changed privacy concerns. According to Schmidt ‘if you have something that you don’t want anyone to know, maybe you shouldn’t be doing it in the first place’ (December, 2009). Zuckerberg’s infamous quote states that ‘the privacy is largely false, but for most students, the privacy is good enough’ (October, 2006). This indicates that people’s privacy conceptions have changed through the advent of SNS. The example of Beacon or the introduction of the Newsfeed on Facebook illustrates this. It’s impossible the give a short overview of all disputes, let alone to become aware of all of them from a user perspective. But we can state that there is a lot of upheaval and confusion on what the problems are and how they should be handled [Picone, 2011]. We can conclude that the concept of privacy has changed, and is still changing. The architecture of typical SNS has created a platform that is ideal for marketing, but not per se for a human as a social being. A social and community perspective on SNS can further elaborate this.

3.2 A Social Perspective on SNS

3.2.1 Why do individuals use SNS?

Whenever an individual is born, it begins a journey of discovering the social world and culture it is embedded in. Gradually this environment becomes part of the individual’s acting, as if it were a
natural fact [Hewitt, 2007]. Hence, this self-evident process is not questioned or reflected upon, at least not all of the time. If this were the case, societal life itself would become literally unbearable. As society isn’t a static given but constant in motion, it has a profound influence on the identity formation of the individual. In current digital society such an important motion has occurred, namely a convergence between mass communication and personal communication. Castells [2009] defines this as mass self-communication. ‘On the one hand mass communication because social computing tools can potentially reach a global Internet audience. On the other hand self-communication because the message production is self-generated, the potential receiver(s) definition is self-directed and the message or content retrieval is self-selected’ [Pierson, 2010]. Hence, new platforms for acting are created. SNS can be seen as an important fraction of mass self-communication, which are prominent in the everyday life sphere of many individuals.

Danah M. Boyd and Ellison [2007] define SNS as ‘a web-based service that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system. The nature and nomenclature of these connections may vary from site to site.’ This fraction of mass self-communication has brought along many positive features. Zhao et al. [2008] state that the way people profile themselves in an offline environment differs considerably from an online environment, which can exert positive influences on the self-image and esteem. So they conceive identity construction as different in a nonymous online world (e.g. Facebook) and a nonymous offline world. A nonymous environment refers to an online context in which the offline identity is displayed. He uses the term nonymous to contrast with an anonymous online environment, in which people do not have to display their offline identity, e.g. online dating site match.com. Consequently, there does not have to be a connection to the offline world in the latter. In a nonymous online environment people have to display their offline identity, but have control over how it is displayed. It looks like people take advantage of this opportunity to ’stretch the truth a bit, for creating their hoped for possible selves’, as Zhao et al. [2008] would say. Hence, creativity is stimulated and social norms can be contested. In other words, the context is free to be filled. According to Danah M. Boyd [2011] networked publics, as SNS, can serve multiple purposes. ‘They can play a civic function, serving to gather people in a democracy. But they can also play a social role, enabling people to make sense of the world around them and understand their relationships to society.’ Livingstone [2005] describes them as a common understanding of the world, a shared identity; a claim to inclusiveness, a consensus regarding the collective interest.’ Especially teenagers seem to be positively disposed towards SNS. According to Danah M. Boyd [2008]: ‘teens’ struggle to access these spaces is driven by their desire for agency and status in a world defined by adults.’ To put in other words, a new life sphere is found without parental control. A place that isn’t subjected to rules and standards but their own. This agency aspect is also confirmed by Papacharissi and Mendelson [2011]: ‘the individual combines the affordances of both older and newer media to construct a social sphere that lends autonomy and fluidity to the way in which sociality is managed. A model of networked sociality emerges on online spaces, the architectural affordances of which inform human activity, by suggesting possibilities for interaction.’

Besides the opportunities SNS brings along, it seems that considering living ‘in’ media, rather than ‘with’ media is more appropriate in defining present use of social media [Deuze, 2011]. Especially because the use of SNS has grown immensely and ‘moved from niche phenomenon to mass adoption’ [Gross and Acquisti, 2005]. Different authors have pointed out to the importance of perceiving the online world as not being separated from the offline world [Papacharissi, 2005, Danah M. Boyd and Ellison, 2007, Zhao et al., 2008, Pierson, 2010, Deuze, 2011, Papacharissi and Mendelson, 2011, Parks, 2011]. The fact that it is very difficult to be anonymous on SNS confirms this process. Hence SNS are entangled in society and not floating around in a vacuum. To sum up, these positive features of SNS combined with its mass adoption indicate that SNS have occupied a prominent place in current society.
3.2.2 How does SNS effect the representation of self?

In his magnus opus Goffman tries to make an analogy between everyday life and dramaturgy. On stage an actor tries to play a credible role for his audience [Goffman, 1990]. In everyday life an individuals acting resembles that of a performance on stage. The only difference is that everyone is both actor and audience at the same time. This means that we are constantly managing our impression towards others. In the front stage of real life we can separate a setting or context of acting and the individuals presented within. What is important in this setting is that everybody has a clear view on the definition of the situation. Only when this condition is statisfied can adequate behaviours be made possible. With adequate behaviour, we mean behaviour that takes into account all different aspects that (can) influence behaviour in a certain context. According to Hewitt [2007]: ‘A definition of the situation is an organization of perception in which people assemble objects, meanings, and others, and act toward them in a coherent, organized way. A definition of the situation, in other words, organizes meanings in such a way that people can act individually and jointly’. A clear definition of the situation is exactly what is absent on SNS. There are many aspects an individual has to take into account, if it wants to perform adequate behaviours. In an offline world more or less clear barriers between contexts exist. Most of the time we know who is present in a situation, what conduct we ought to expect from others, what role we should perform, and where the situation is located. When mass self-communication enters the picture, this more or less clear context disappears. Throughout the rest of this part of the SOTA the different problems that impede the making of a clear definition of the situation will be discussed.

3.2.3 What does this mean regarding privacy?

When the definition of the situation isn’t clear, performances on SNS become difficult in relation to privacy on mainly two levels: social privacy and instrumental privacy. The former can be defined, according to Raynes-Goldie [2010] as ‘the control of information flow about how and when their personal information is shared with other people.’ The latter as their data being accessed by governments and corporations, e.g. data mining [danah m. boyd and Hargittai, 2010].

The overall problem with instrumental privacy in online environments is that people do not know what happens with their personally identifiable information (PII): Who gathers it? Why do they gather it? Can I do something about it? Individuals get lost in this maze and lack every ability to act in a meaningful way [Solove, 2001].

Although instrumental privacy is important, social privacy problems seem to be more important to the user. Context collision [Raynes-Goldie, 2010] or context collapse [danah m. boyd, 2008] represents a problem for social privacy. It refers to the blurring of contexts in an online environment, whereas in an offline environment more or less strict barriers can be distinguished. Combined elements of mass media and personal communication makes acquiring a proper self-presentation to multiple audiences hard for people. Fortunately they seem to be balancing personal and public information, avoiding certain topics and maintaining authenticity [danah m. boyd, 2008]. Another problem, related to context collision, is the phenomenon of forced disclosure [Rosen, 2001]. Forced disclosure refers to the ongoing process of clarifying private information through private information. Because a lot of self-representative information on social network sites is not put into context, the only way to clarify this seems by putting more information on these sites. This concept can be better understood when contrasting it to disclosure of public information. Rosen [2001] gives the example of a football player, whom everyone remembers for making a single boneheaded play. If this player wants to improves his image, he can do so by playing better the next time, instead of saying during a press conference that he has some problems with his wife. Hence, misjudgement on the basis of public information is more easily countered by behaving in accordance with the public role, according to Rosen [2001]. When private information is disclosed, the only way of clarifying this is by giving more private information, certainly in a situation with numerous context collisions. For example, when a person breaks up his relationship with someone and changes his status from ‘in a relationship’ to ‘single’ only a couple of people will know exactly what happened. The majority of people will not.
The only way to clarify this is by putting more private information on this platform. Thereby, being seen out of context reduces a person from an acting subject to an object. As Merton would say, ‘the pressure to live up to the details of all (and often conflicting) social norms would become literally unbearable’ (citation from Merton in [Rosen, 2001]). By constantly being watched – past, present and future actions – the self becomes less spontaneous and autonomous.

3.2.4 The challenge

This chapter makes clear that there are as many positive as negative features connected to SNS. In search of these features we did find a common denominator, which is a lack of a clear definition of the situation. This can be liberating and empowering for identity processes, but it also makes it hard to take into account privacy issues, regardless of whether one places greater value on social or instrumental privacy. We need to find a way to diminish the negative features, without jeopardising the positive ones. A community perspective can help in trying to achieve this challenge.

3.3 A Community Perspective on SNS

As the reader will notice the concept of a community is often contested in its nature. This section will focus on the value of offline communities in filling the empty/unclear context, described above. We will start by elaborating upon different perspectives on communities. Next, the relationship between online and offline environment will be theorised and demonstrated. The most important authors and their conclusions in community research will be reviewed. Finally, the relation between community and privacy will be elaborated.

3.3.1 The meaning of community in present society

Tönnies [1887] can be considered as one of the first theorists in defining community [Tönnies, 1887]. Communities, which he refers to as Gemeinschaft, can be described as a private and intimate place that stands for the basic needs of individuals, such as warmth, shelter, nurture, etc. Here the Wesen-wille can be seen as a central feature, where the individual serves the group. Society, which he refers to as Gesellschaft, is seen as a more rational and purposeful. In contrast to community, the individual itself is the most important feature. To put it otherwise, the Kurwille stands central. Tönnies can be categorised as a scholar with a strong desire for a premodern society. This he yearns, because an evolution to a society in which Gesellschaft prevails is predicted. Other classics, like de Tocqville and Durkheim foresaw a similar evolution [Kivisto, 2003].

In contemporary literature two main visions regarding community can be distinguished. One, which states that community is a diluted concept and ought to be forgotten in describing current society. Another, that tries to embed the concept of community in such a way that it still has its meaning. Postmodernists like Bauman, for example, aren’t satisfied with the concept of community. Bauman [2008] questions the existence and meaning of it. He sees identity as a surrogate for community, reducing the latter as an extension of the former. ‘Predestination’ was replaced with ‘life project’, fate with vocation and a ‘human nature’ into which one was born replaced with ‘identity’ which one needs to saw up and make fit. Turner and Cohen can be placed within the perspective of holding on to the concept of community. Turner sees community as an opposition to structure, an expression of the social nature of society [Delanty, 2003]. The moments in between, liminality as he calls it, is the expression of such a community. These moments refer to events of life not subjected to instrumental rationality, and create a powerful bonding between members of society. To put it differently, one obtains a feeling of belonging and relating to others when not being subjected to rules, laws, norms, etc. Then, in interaction with others, the community within the self reveals itself. Cohen has a pure hermeneutic approach on community. ‘Cohen defines community in terms of particular
kinds of awareness of reality; and as such community is a symbolization of boundaries by which the community differentiates itself from others’ [Delanty, 2003].

Following the more positive perspective on communities, the concept of community is often divided in two categories: a community of interest, and a community of place. Community of place refers to a geographical fixed community. A community of interest is based upon a common interest between members. It may be that both communities overlap each other. This teaches us that a community does not need to be anchored in a particular location, but can also exist in the virtual.

### 3.3.2 The relationship between offline and online communities

Rheingold [2000] can be seen as the first author using the concept of virtual communities. His thoughts should be captured in the spirit of time, that is when the Internet began to bloom in the consciousness of the public [Parks, 2011]. Rheingold saw the Internet as an alternative reality, with capacities to transform society [Delanty, 2003]. With his concept of virtual communities he refers to a non-existing offline community, solely rooted in cyberspace. ‘What is significant about his view on virtual communities is that virtual communities are ‘communities on the Net’. They do not exist in everyday life’, according to Delanty [2003]. Even further, the downfall of communities can be compensated by a virtual one.

Castells has a more interactionist view on the virtual and reality. New communities like virtual ones are built out of networks of social actors, be it individuals, families or social groups [Delanty, 2003]. In a global network society, like our own, spatial communities are replaced by a spaceless space of what we call the Internet. Castells [2000] states that ‘localities become disembodied from their cultural, historical, geographical meaning, and reintegrated into functional networks, or into image collages, inducing a space of flows that substitutes for the space of places.’ It is not the global network society itself that changed social relations, rather the individualism inherent in society. Communities can be defined as personalized communities embodied in networks and centred on the individual. Where Rheingolds view on virtual communities can be described as thick, Castells would definitely speak of thin communities. With ‘thin’ we refer to a virtual reality that is an addition to the offline reality, whereas ‘thick’ can be seen as an equivalent of the offline reality. This comparison brings forward another important concept relating to communities: social capital. Putnam and Bourdieu are probably the most prominent authors on this topic. Putnam defines social capital on a community level as ‘features of social organisation such as networks, norms and social trust that facilitate coordination and cooperation for mutual benefit’ [Baum and Ziersch, 2003]. Bourdieu holds a more individual definition on social capital, ‘the aggregate of the actual or potential resources which are linked to possession of a durable network of more or less institutionalised relationships of mutual acquaintance and recognition’ [Baum and Ziersch, 2003]. Within social capital we can make a difference between ‘bonding social capital’ and ‘bridging social capital’. The former refers to horizontal bonding between individuals of the same community. The latter to vertical bonding, which cuts through communities. In line with the difference between Rheingold’s and Castells’ view on communities, one can question what is more preferable. Common sense would tell us that bonding social capital is more important, but Granovetters paper on the strength of weak ties highlights the other side. Granovetter [1983] states that weak ties are more important in some situations, as for looking a job. Whether or not virtual communities can be labelled as thick or thin, both seem to be important for different reasons. In his revised copy on virtual communities Rheingold states the following ‘A social network with a mixture of strong ties, familial ties, lifelong friend ties, marital ties, business partner ties, is important for people to obtain the fundamentals of identity, affection, emotional and material support. But without a network of more superficial relationships, life would be harder and less fun in many ways. Weaker ties multiply people’s social capital, useful knowledge, ability to get things done’, according to Rheingold [2000].

Calhoun also views that mediated relationships are becoming more important, although we should not exaggerate these forms Delanty [2003]. Offline communities are supplemented by virtual togetherness, rather than substituted. Calhoun has a rather negative view on the capacity of virtual communities...
communities to enhance democratization. Changing patterns of mediated communication thus combine with the increasing compartmentalization of community to produce a deterioration in public discourse. We are aware of others (a notable accomplishment of mass media, as classically of cities), but we are not in discourse with them’, so [Calhoun, 1998] observed. Hence, Calhoun states that virtual communities contribute to a separate categorisation of individuals. This view closely follows the principle of the filter bubble, described by Pariser [2011]. The filter bubble describes how the Internet is tailored to the personal identity of the individual, leaving no room for other perspectives.

Although different views on the meaning and capacity of online communities exist, studies have revealed a clear connection between offline and online environment. In studying Facebook Lampe et al. [2006] found that it is used primarily for maintaining previous, offline relationships. When conducting research on MySpace, danah m. boyd and Ellison [2007] found that teenagers are motivated to go on SNS because their offline friends are there too. Friendster too is deeply connected to the participant’s offline social life, according to [danah m. boyd, 2006]. Miller and Slater [2000] state that almost every time online and offline spheres are almost always interconnected, thereby creating one reality. Parks [2011] when studying MySpace, stated that offline and online communities are linked in ways that we are only beginning to understand. Moreover, ‘... it may be more accurate to say that virtual communities are often simply the online extension of geographically situated offline communities.’ According to Pew research center only a small fraction of Facebook friends, are people we have never met offline. 89% of the friends we have on Facebook, we’ve met more than once offline [Hampton et al., 2011].

The focus of most research lies on the individual as a user when it comes to investigating online behaviour on SNS, not the community an sich. By this we refer to an individual embedded in a particular context. However research projects show just how much an online environment can learn from an offline community. Three projects will be discussed, to demonstrate how offline communities can contribute to the online experience: PICOS¹, Ketnet Kick² and Citizen media³.

• In the PICOS project the main goal was to achieve privacy-enhanced identity and trust features within communitysupporting services [Tschersich et al., 2011]. More specific, a mobile application with location sharing was developed for the angler’s community. Within this application it was possible to create an online identity. Moreover, partial online identities could be developed, which was strongly appreciated by the anglers. This means that the user 1) can create multiple identities; 2) each with their own profile; 3) which can be used in different contexts. The concept of partial identities follows closely the needs of the offline angler’s community to differentiate between contexts. Other concepts they put forward like, context blurring, access control and privacy advisor similarly have roots in the offline world.

• Van Lier and Pierson [2007] studied the interpretation of online communities by children. In particular they wanted to ascertain what constitutes a sense of community for them. They used ‘Ketnet Kick’, a game developed by the Flemish public broadcasting company VRT and gamesdeveloper Larian to study their research question. The overall conclusion was that children have created a sense of community online. But even more so, children brought the online community more or less ‘alive’ offline. Van Lier and Pierson [2007] state that if the community

---

¹Privacy and Identity Management for Community Services (PICOS) is an international research project, focused on mobile communities. Funded by the Seventh Framework Programme by the European Commission, they studied, developed, built and evaluated an open, privacy-respecting, trust-enabling identity management platform that supports the provision of community services by mobile communication service providers. For more information, please see the following site: http://www.picos-project.eu/

²The KETNET KICK project examined the determinants of offline and online communities. Moreover they tried to identify what children find important in a community, with a clear focus on where and how children experience the sense of community. The project was funded by the Vlaamse Radio- en Televisieomroep (VRT).

³The CITIZEN MEDIA project studied, developed and validated audio-visual systems to enable multiple non-professional users to co-create networked applications and experiences based on their own user-generated content. This project was funded by the Sixth Framework Programme by the European Commission. For more information, please see the following site: http://www.ist-citizenmedia.org/
wants to be real in the minds of children, it needs to be linked to their own life and experiences; offline interactions can be very helpful.’

- In the ‘citizen media project’ audio-visual systems from a non-professional user perspective were examined, developed and validated [Lievens et al., 2008]. A longitudinal research among members from two offline communities has led to the identification of various social requirements that steered the design and development of an online community application that could complement the offline community. Controlling the application, a display of their social heritage and identity as well as a mirror image of their offline community patterns needed to be established in this application. In other words, the respondents wanted to preserve their offline community.

3.3.3 The meaning of a community perspective in creating privacy enhancing tools for SNS

Several authors perceive the online world not as not being separate from the offline world. SNS are entangled in society and not floating around in a vacuum. Unfortunately the architecture of SNS does not allow sensing the community in the same way an offline world does. This influences the performance of adequate behaviours, not the least regarding privacy concerns. Research projects such as PICOS, Ketnet Kick, and CITIZEN MEDIA show the importance of offline community in developing online applications. Future research and technologies will have to take this fact into account and study how SNS should be established from a community point of view. Attempting to operationalize concepts of ‘liminality’ or ‘sense of belonging’ into technologies that preserve the privacy of the individual is not an easy task to do, but very much necessary. To demonstrate the importance of offline communities in creating privacy-enhancing tools some examples are given from the PhD research by Ralf De Wolf, based on preliminary results of interviews with adolescents between the age of 14 and 18.4

1. In an offline world, people can make a clear distinction between situations and communities. In an online environment, the individual is often confronted with a crossing-over between friends, family, acquaintances and offline communities. On Facebook, the majority of people that we have interviewed had classified their friends into one big group. Some of them, the minority, made a difference between family and friends, in what was visible to them. Common sense would teach us that dividing the one big group on Facebook into smaller ones, similar to offline structures of a community, would make it more easier for people to grasp their environment and act accordingly. Unfortunately, this appears to require too much effort. But even more importantly most of them liked the one big group for different reasons: letting all the people know what they did with whom, observing others, profiling their identities, etc. However, it should also be noted that all respondents did not post everything they had in mind, but figured out for themselves a strategy what could and couldn’t be posted. Hence, we could state that the adolescents preferred a sort of agency to the structure they are adapted to. However, using this strategy makes it difficult to allow adolescents attain feel the ‘community feeling’ they have offline. These results indicate that making structures for adolescents, or circles as you will, that can differentiate groups of people will not be used by this audience. At least not by the majority of people.

4Within this research the concept of context clarification is studied, as its interrelation with privacy. With the concept of context clarification we refer to the necessity of a clear context in conducting adequate behaviours. Hence, when the context in which people interact is clarified individuals will grasp their own behaviour more. To study this, online behaviours between two online platforms were compared: Facebook and Smartschool. The former needs no introduction, the latter can be considered as an online learning environment for students and teachers. A qualitative ethnographic study was performed, using half structured in-depth interviews combined with a thinking aloud method. The preliminary results indicate that there is no simple trade-off between identity and privacy.
2. According to the literature, and our own results confirm this, adolescents care more about social privacy than instrumental privacy. Often they do not know who gathers information or what is done with it. They simply do not question the internal structure they are using. Hence, there does not (yet) appear to be any demand for creating a technology that encrypts information exchanged between individuals on Facebook. Adolescents often go with the flow, but the ‘flow’ needs to be triggered. Warm experts\(^5\), such as people they trust within their own community, can maybe perform this task.

3. In examining online behaviour on Facebook a comparison was made with another online platform that is used within schools for educational reasons, namely Smartschool. It is expected that on Smartschool, within the clear context of school, scholars would act more responsible, regarding their privacy. Regarding social privacy this was certainly the case. For example, on forums they would pay more attention to what and how they posted it, even more than on Facebook. But regarding instrumental privacy, this was not the case. A plausible explanation is that the only thing they posted on Smartschool was schoolrelated and had nothing to do with their personal identity. Still, the majority of people wanted more control over their privacy settings, so they could decide themselves what the school can see and not. It is clear to us that the generalized others\(^6\) of the school had a profound influence on the behaviour of adolescents. Unfortunately, the limited agency granted to scholars regarding privacy settings combined with a lack of identity information on this platform, renders it uncertain whether the generalized others has such a significant influence on adolescents in relation to instrumental privacy. Future research will have to measure this in a different context, outside the school.

### 3.3.4 Tackling the challenge

In this chapter we have presented the concept of community in tackling the challenge described in previous chapter. ‘Community’ is often identified with a romantic premodern yearning. As a result it is often regarded as a diluted concept. We hope this chapter clarified the significance of community studies in achieving more adequate behaviours on SNS. Our research on communities will leave this premodern yearning and focus on communities of interest, bonding social capital, strong ties, thin communities and communities that can exist in the mind of the individual and that does not have to be tied down to a geographical location.

In ‘alone together’ Turkle \[2011\] described that the simulation of real life processes, e.g. SNS, will never suffice in constructing the social. In the ‘filter bubble’ Pariser \[2011\] described how an Internet tailored to the self leads to a closeness and narrow-mindedness. Integrating a sense of community and social identity on the Internet can tackle these individualistic problems. Moreover it can also tackle privacy problems, such as context collision [Raynes-Goldie, 2010], forced disclosure [Rosen, 2001], invisible audiences [danah m. boyd, 2008], unconsciousness of what is happening with our data [Solove, 2001], etc. By integrating processes of community on SNS the definition of the situation can become brighter for the individual. In doing so, more adequate behaviours should follow.

The preliminary results above show the importance of digging through the social environment of the users before technologies are implemented. The following section on social requirements will translate precisely how future technologies will be studied from a community perspective.

\(^5\) If we decompose the concept of warm experts, we can distinguish two concepts: ‘warm’ and ‘expert’. The former refers to people who are close to you and can be trusted, the latter refers to people who are aware of certain technologies and how they operate. Combine these two terms together and you get a warm expert.

\(^6\) The generalised others is best thought of as the imagined perspective of an imagined other, whether this other is the whole society, the community to which the individual belongs, or some smaller category of grouping of people, according to Hewitt \[2007\].
3.4 Social Requirements and Evaluation Framework

Our central question is: what can we learn from users in online and offline communities with regard to privacy issues within SNS. The goal is two-fold. Deriving (1) what communities an sich can do and (2) which social requirements are necessary in establishing privacy-enhancing tools. The latter objective can be interpreted in a two ways: a) communities steering the designing process of the technologies and b) technologies being implemented in a specific community. With this we want to emphasise the bilateral relation in the design of technology. A domestication framework can help in clarifying these goals.

A domestication framework looks at the use and integration of technologies in the everyday life of users. Under this approach, technologies are studied in the context in which they are used. A central concept within the domestication framework is consumption. Van den Broeck [2011] states that ‘...our domestic lives are more and more defined by our consumption of objects and meanings.’ Everyone does not consume technologies in the same manner. Instead, there are different meanings attached by the user to technology. By studying how the latter is consumed we get a grip on the social requirements of technology. This can be defined as ‘the requirements that are extracted from the social background and everyday life of people, with an emphasis on groups or communities and the social practices within’ (Vermeir et al, 2008). An important assumption within this framework is that the role of the users cannot be underestimated. It is the user and the context in which the technologies are used that define technology.

The process of how these social requirements will be measured from a community perspective is still undefined. Roughly speaking, following phases can be distinguished: (1) conceptual phase (2) experimental phase (3) evaluation phase. Before elaborating these three phases we want to point out that the line of examination of the usage of SNS and the interlinkage with privacy issues will be studied around three dimension: the individual – community dimension, the non-technological – technological dimension and the offline – online dimension. These dimensions will be apparent, more or less, through all phases.

Within the first phase a representation of privacy enhancing technologies will be given to the users. Within this phase, broader attention will be paid to the meaning users give to SNS, what behaviour is conducted, what risks are perceived, etc. In the second phase the technology that does not yet exist will be studied through a proxy technology assessment (PTA). The latter can be defined as ‘existing technologies that resemble as much as possible the functionalities under development’ [Pierson et al., 2006]. For the last phase, an evaluation framework will be set up that evaluates the different privacy-enhancing tools in a real life environment from both an individual and community perspective. Through this approach the development process will be steered and identity gaps looked at. With identity gaps we refer to features of privacy enhancing tools that do not take into account the identity layer on SNS. Privacy without depriving the individuals needs that is. Hence, the central question will be: to what extent are the social requirements for creating these tools are taken into account? The evaluation framework can, for example, take the following parameters for evaluation into account when implementing technologies in a certain community.

- Is the offline structure of the community taken into account?
- Does the technology take into account the usage content and the purposes of the community?
- Does the technology rely solely on trust in the technology itself or also on trust in each other?
- Does the technology add value to the members of the community?
- Does it take into account the risks they perceive as threatening?
3.5 Conclusion

Context collision [Raynes-Goldie, 2010] or context collapse [danah m. boyd, 2008], forced disclosure [Rosen, 2001], filter bubble [Pariser, 2011], invisible audiences [danah m. boyd, 2011], living alone together [Turkle, 2011], a blurring between public and private [danah m. boyd, 2008], unconsciousness of what is happening with our data [Solove, 2001] all refer to major gaps in the architecture of SNS. These makes it hard for users to interact, represent themselves and create communities and on top of that bear in mind their social and instrumental privacy. The architecture of SNS clearly is in need of a renovation. In achieving this a domestication framework centred around communities will be set up, by which we hope a clear definition of the situation for the user can be made, without touching upon the positive features overall represented on SNS. Jaron Lanier quotes the current situation of conduct on SNS as follows ‘If you overrely on social media, you become the image of you’ (Singer, 2011). We think this quote aptly summarizes well the opposite of what we want to achieve, namely a social identity and a sense of community on SNS that create a transparent online environment. The identity layer on SNS seems to only represent the individual needs. Hence, the image of the community needs to be restored. On the topic of privacy danah m. boyd [2011] stated recently that the solution to this puzzle will not be to restrict data collection or to enhance individual control over specific items of data, but to think long and hard about what happens as the data flows across networks and as the data is networked together. This requires moving beyond the individual and focusing on the collective.’ A collective perspective is exactly what this section and the related SPION research part is trying to promote. The social and community hatch of the SPION project shall try to embed a course of action beyond the individual and with an emphasis on the interconnection between offline and online world in achieving privacy. Regarding this, we want to end this section with a statement and nuance of Sherry Turkle on living with new technologies. New technologies, as SNS, in the online world can be considered as a simulation of processes common to mankind in an offline world. But according to Turkle [2011] ‘Simulation not only demands immersion but also creates a self that prefers simulation. Simulation offers relationships simpler than real life can provide. We become accustomed to the reductions and betrayals that prepare us for life with the robotic.’ With this we want to emphasize the importance of the offline world where the human is surrounded with faces and voices. When the architecture of SNS is improved and the context is filled, identity and privacy gaps will probably diminish, but the improvement of these technologies will never suffice on its own. Humans simply need others to designate themselves and others as a human being, which no technology can replace completely.
4 Behavioral Aspects (CMU)

4.1 Disclosures and Privacy in Social Network Sites

4.1.1 The Evolution of Social Network Sites

In recent years, the landscape of social network sites has been characterized by growth and evolution. According to 2010 statistics from the Pew Internet and American Life Project, virtually all US teenagers use social network sites. Once the purview of the young, social network sites are now used by almost half of all online adults, an approximate five-fold increase since 2005 [Lenhart et al., 2010]. At the same time, social network site platforms have evolved; Facebook and LinkedIn have experienced tremendous growth, while early-adopter sites such as Myspace and Friendster have contracted. The rise of micro-blogging services such as Twitter and Tumblr have threatened social network sites, forcing rapid development of new products and evolving norms of disclosure and privacy.

4.1.2 Disclosures

A large body of research on social network sites has explored the disclosure behavior of social network site users. Because social network sites thrive on peer-produced content, disclosure is often concomitant with social network site use. The definition of social network sites, posed by boyd and Ellison, illustrate the importance of disclosure. They state that social network sites allow the creation of “public or semi-public profiles within a bounded system,” they foster the articulation of lists of personal connections within the system, and they allow the transversal of these connection lists within the system [danah m. boyd and Ellison, 2007]. That is, a social network site is characterized by what you say about yourself, who you choose to publicly associate with, and increasingly, what your connections say about you.

The earliest studies of social network site use provided empirical evidence of the remarkable disclosure practices within the site. Work by Acquisti and Gross [2006] found that students in the Carnegie Mellon University Facebook network extensively shared sensitive information such as political views and sexual orientation in Facebook, and that information shared in Facebook was generally self-reported as valid. Other studies conducted at the time in different university networks, including Stutzman [2006] and Lampe et al. [2006], further evidenced the high degree of personal disclosure within social network sites. Large scale studies such as [Thelwall, 2008] and [Caverlee and Webb, 2008] provided evidence of similar disclosure phenomena in Myspace, once the leading social network site. These findings were corroborated by a national probability study conducted by [Lenhart and Madden, 2007].

Researchers have considered a range of motivations for disclosure in social network sites. Drawing on the work of Goffman [1959], Donath and danah m. boyd [2004] and danah m. boyd and Heer [2006] describe the use of a social network site as a performance of identity. Users of social network sites are therefore challenged to strategically present themselves, through the constructed profile, to their increasingly diverse network of social ties. As noted by Lampe et al. [2006], the motivations for use, and disclosure within, a social network site are a function of offline outcomes such as relational formation and deepening. Works by Bumgarner [2007] and Joinson [2008a] illustrate the social

---

motive of social network site use, that use of social network sites was driven by participant desire to connect and learn about one another. Without significant personal sharing in these sites, these motives of use would not be addressed.

For the past six years, social network site use has rapidly evolved as new participants joined the various platforms. Between 2005 and 2011, social network sites experienced remarkable growth in active users, as well as shifts in the popularity of key sites. In particular, we have observed a consolidation in the socially-oriented social network sites, with Facebook achieving market dominance. However, Facebook’s transition from a university-focused social network site for students to a global social network site has not been seamless. Changes to the network structure of Facebook, and unpopular moves by site governance, have raised privacy concerns for site users.

4.1.3 Privacy

A significant portion of previous research on social networking sites has focused on users’ privacy risks and associated concerns and practices.

Gross and Acquisti first highlighted risks such as stalking, identity theft, price discrimination, or blackmailing [Gross and Acquisti, 2005]. In addition, boyd and Ellison identified privacy risks such as damaged reputations, unwanted contacts, surveillance-like structures due to backtracking functions, harassment, and use of personal data by third-parties [danah m. boyd and Ellison, 2007]. However, findings from early empirical studies of student Facebook users in the United States also found evidence of behavior inconsistent with stated privacy concerns, excessive sharing of personal data, and rare changes to default privacy settings. For instance, in their above-mentioned 2005 study of Facebook users at Carnegie Mellon University, Gross and Acquisti found that the majority of users shared a large amount of personal data, and only a very small percentage of users changed their default privacy settings [Gross and Acquisti, 2005]. In a follow-up 2006 study, Gross and Acquisti found that even users who claimed to be concerned about privacy tended to reveal a great deal of their personal information — a discrepancy between stated privacy attitude and actual behavior [Acquisti and Gross, 2006]. Ellison et al. found that only 13% of Facebook profiles in the Michigan State University network were restricted to “friends only” in 2007 [Ellison et al., 2007], and this was confirmed by a longitudinal study of the same population from 2006 to 2008 [Lampe et al., 2008]. These studies only examined American college students. Hence, their results might not generalize to other populations. For instance, Joinson conducted a study with primarily non-student Facebook users in the U.K. in 2008 and found that the majority of the respondents (57.5%) reported having changed the default privacy settings [Joinson, 2008b]. More recent studies also seem to suggest that users are becoming more privacy concerned and more likely to change their privacy settings [Madden and Smith, 2009, danah m. boyd and Hargittai, 2010]. For instance, according to a 2009 report by the Pew Internet & American Life Project, 71% of SNS users between the ages of 18 and 29 reported changing their privacy settings [Madden and Smith, 2009].

When Facebook debuted in 2004, the site was segmented by university, so that university network membership created a meaningful privacy boundary. Although university networks could be quite large (e.g. Acquisti and Gross [2006], Stutzman [2006]), this boundary generally separated students from family, employers, and municipal law enforcement. Starting in 2006, Facebook gradually liberalized its policies for site membership, and began changing (and eventually highly discounting) the value of “networks” within the service. Facebook was rewarded for these moves, with adoption climbing through 500 million as of 2010[^1]. Facebook’s growth in popularity proved both an opportunity and challenge for users of the site. On the one hand, users can now articulate a greater portion of their “social graph” in the service, and take advantage of the benefits reaped from the establishment and maintenance of large weak-tie networks [Ellison et al., 2007]. On the other hand, the presence of a larger portion of users raises privacy challenges, such as the presence of multiple contextual networks in the site. Individuals may face challenges as they attempt to share information in the

[^1]: http://www.facebook.com/blog.php?post=409753352130 "500 Million Stories"
presence of coworkers, family, and distant friends in a single social network site (e.g. DiMicco and Millen [2007], Farnham and Churchill [2011], Skeels and Grudin [2009]). As a result, the known-audience and the expected audience in social network sites do not always overlap [danah m. boyd and Heer, 2006, Ellison et al., 2007, Lampe et al., 2008, Stutzman and Kramer-Duffield, 2010]. This can be intended to mean that within a system with hundreds of articulated connections, disclosures are intended for a subset of the audience. In most cases, one does not expect their disclosure to range beyond a certain subset of alters. The implication of this finding is often in collision with discourses that argue that disclosure in a socio-technical system is intended to be public.

The challenge of managing disclosure across multiple social contexts in a social network site has been referred to as “context collapse” [Manwick and Boyd, 2011]. On the one hand, Zeynep Tufekci has argued that because disclosure networks were so large, the concept of privacy was not only dead, it was a ‘zombie’\(^3\). At a functional level, Tufekci’s argument holds water. If an average of 350 people have the potential to see whatever one does online (within a socio-technical system), they have clearly “traded up” to less privacy. However, sociality is mediated; affordances such as the News Feed create an “illusory” privacy [Hoadley et al., 2010] that influences behavior. Finally, practice does not afford ongoing social surveillance of an entire network, but rather alters of particular situational interest. Indeed, the potential for large-scale surveillance exists, but does not occur in practice due to segmentation, non-participation and socio-technical affordance. On the other hand, in fact, Lampinen et al. [2009], Lampinen et al. [2011] have shown that individuals employ a range of strategies to manage multiple contexts in social network sites, including self-censorship and withdrawal of content, creating more inclusive group identities, and sharing different types of content in different spaces. In addition to these behavioral and mental strategies for context and privacy management, individuals also turn towards the application of privacy settings within the site. Numerous studies documented both increased use of privacy within Facebook by students [danah m. boyd and Hargittai, 2010, Lenhart et al., 2010, Lewis et al., 2008, Tufekci, 2008] and the contextual application of privacy settings in relation to perceived harms [Debatin et al., 2009a, Stutzman and Kramer-Duffield, 2010].

In addition to the shifting nature of networks in Facebook, changes to the interface and site policies have produced public backlash that has increased privacy awareness in social networks (e.g., [Grimmelmann, 2009, Hoadley et al., 2010, Hull et al., 2010]). These site-directed changes may affect disclosure by altering the level of trust individuals have in Facebook, which was often described as the more trusted social network site when compared with Myspace (Dwyer et al., 2007; Fogel & Nehmad, 2009). To combat the increases in privacy and decreased disclosure to a wide audience in the platform, Facebook has consistently changed the nature of sharing certain items in the platform, and the default sharing settings for new accounts\(^4\).

Demographics seem to affect privacy attitudes and behaviors of social network site users. Fogel and Nehmad found that in general men had less privacy concerns than their female counterparts, and thus tended to disclose more personal information such as telephone numbers and physical addresses on SNSs [Fogel and Nehmad, 2009]. Stutzman and Kramer-Duffield found that female users and users who have more Facebook friends are more likely to have friends-only profiles [Stutzman and Kramer-Duffield, 2010]. In a study of MySpace users, Gilbert et al. found that rural users have fewer friends and comments than urban users. Further, rural users, particularly women, have a higher level of privacy concern and use privacy settings more often than urban users [Gilbert et al., 2008]. boyd and Hargittai also found that individual characteristics such as Internet skill, frequency, and type of Facebook use are correlated with making modifications to privacy settings [danah m. boyd and Hargittai, 2010].

Users display more concern about sharing with their weak-tie friends than with outsiders or companies. Stutzman and Kramer-Duffield suggest that users adopt friends-only profiles mainly to deal with unintended disclosure to their weak ties rather than outsiders [Stutzman and Kramer-Duffield, 2010]. Raynes-Goldie found that users cared more about protecting information from members of various social circles, rather than protecting their information from companies [Raynes-Goldie, 2010].

\(^3\)http://technosociology.org/?p=35  
\(^4\)https://www.eff.org/deeplinks/2010/04/facebook-timeline “Facebook’s Eroding Privacy Policy: A Timeline”
Besmer and Lipford found most photo privacy concerns were about identity and impression management within the user's social circle. These photo-privacy concerns revolve around revelation of incriminating evidence (e.g., underage drinking), unflattering photos, and unwanted associates (e.g., ex-significant others) [Besmer and Lipford, 2009].

Social influence seems to play a role in privacy attitudes. In a 2008 study of Facebook users at Harvard University, Kaufman et al. found that students are more likely to have private profiles if their friends and roommates have them [Kaufman et al., 2008]. In a 2009 study of privacy settings on Flickr, Nov and Wattal found that sense of trust and sharing norms of a community positively affect community members’ privacy concerns and information sharing behavior [Nov and Wattal, 2009].

4.1.3.1 Coping Strategies to Reduce Privacy Risks

A number of strategies to counter or defuse privacy risks have been identified in the literature. Lampinen et al. found users “dividing the platform into separate spaces, using suitable channels of communication, and performing self-censorship” [Lampinen et al., 2009]. Similarly, in Lampe et al.’s study, some users reported active management of their profile, e.g., restricting who can see it and removing “sensitive” content. Their interview respondents reported incidents of minor embarrassment but did not report any strong negative consequences [Lampe et al., 2008]. Tufekci analyzed college students’ information disclosure behaviors on social networks and found that “students manage unwanted audience concerns by adjusting profile visibility and using nicknames but not by restricting the information within the profile” [Tufekci, 2008]. In a year-long ethnographic study of Facebook users in their 20s, Raynes-Goldie found various strategies including using aliases, deleting wall posts, untagging photos, and creating multiple accounts to circumvent Facebook’s default privacy settings [Raynes-Goldie, 2010].

4.1.3.2 Design Approaches to Privacy in Group Context

As noted in the literature review, one of the primary challenges of privacy in social network sites is the management of group context. Privacy risks emerge as individuals are forced to manage their disclosures between spheres of life, such as the professional and personal spheres. Recently, designers have attempted to address the problem of group context management in social network sites through the inclusion of technical features that enable the grouping of contacts. In 2007, Facebook launched a “Friends List” feature that allows individuals to aggregate friends according to individually-defined criteria, and then selectively disclose to these lists. Facebook’s system has proven highly complicated, which has limited the product’s success in providing privacy at the group level. Recently, Google introduced the Google+ social network, which aims to “bring the nuance and richness of real-life sharing to software.” Google+ has, by default, defined spheres of life where individuals can place their connections, and share accordingly. Google+ has received much credit for making the group management process more simple and intuitive, but privacy concerns, particularly related to the Google+ “real name” policy, lead influential critics to challenge the privacy gains of Google+.

\[^5\text{http://www.facebook.com/blog.php?post=7831767130}\]
4.2 Privacy, Behavioral Economics, and Soft Paternalistic Interventions

4.2.1 From Hurdles in Privacy Decision Making to Soft Paternalism

Privacy decision making has become more complex, as information systems have vastly expanded our ability to permanently share with others information about ourselves. Making the “right” privacy decision online — that is, balancing revelation and protection in ways that maximize a user’s welfare and minimize her future regrets — is difficult. The hurdles stem from a combination of factors: opposing or contradictory needs (such as the need for publicity intertwined with the need for privacy), inconsistent preferences and frames of judgment, and incomplete or asymmetric information about consequences, risks, or solutions associated with the provision of personal information. These hurdles are exacerbated in the context of Web 2.0 and social network sites in particular: consider, for instance, the power of statistical re-identification techniques, data mining, and statistical surveillance based on SNS data to make inferences about individuals that are as much sensitive as they are invisible, and surprising, to the individuals themselves [Acquisti and Gross, 2009].

In recent years, research in privacy (and security) usability has made significant progress towards overcoming some of these hurdles, building systems that allow users to efficiently manage their privacy and security preferences. Research in this area has investigated the role of notices and salient information [Tsai et al., 2007, McDonald and Cranor, 2009, Kelley et al., 2009, Egelman et al., 2009a, McDonald and Cranor, 2008], control [Tsai et al., 2009], and expressiveness [Sadeh et al., 2009] in empowering users in these domains.

However, while empowering users may be a desirable goal, neither information, nor control, nor expressiveness, per se, are guaranteed to allow users to achieve their stated preferences and maximize their welfare in privacy sensitive scenarios. Even useful and salient information can be ignored (for instance, nutrition labels indicating calories associated with the consumption of fast food do not lead individuals to change their eating habits [Downs et al., 2009]). Control, even when afforded, can be too hard to use (the Facebook online social networking system offers extensive privacy controls, but these controls require users to navigate through many layers of menus and access an array of confusing options), or in fact lead to the paradox where too many options lead to poorer decisions [Schwartz, 2005]. Furthermore, providing users more controls can do little to help them protect their privacy if the default settings of a system (for instance, Facebook) expose a large amount of users’ data. Finally, giving users the ability to expressively determine their ex-ante privacy preferences does not guarantee that they will not make decisions that they will later stand to regret. In other words, strategies such as relying on “responsibilizing” users, or providing users with more information, or making systems more transparent to users, offer no guarantee that privacy risks from social networks will be deflected and privacy concerns soothed.

The reason is that additional hurdles to privacy decision making stem from issues of bounded cognitive power (that limits our ability to consider or reflect on the consequences of privacy-relevant actions) and from various cognitive and behavioral biases (that is, systematic deviations from the theoretically rational decision process) [Camerer and Lowenstein, 2004] that research in behavioral economics has uncovered. Studies have documented systematic inconsistencies in people’s preferences in domains ranging from consumer choice [Simonson and Tversky, 1992] to the valuation of environmental amenities [Kahneman et al., 1993], showing that preferences are often labile and influenced by contextual factors [Slovic, 1995], leading often to sub-optimal choices that individuals...

---

stand to regret. For example, preferences depend on how choice alternatives are framed and how preferences are elicited [Tversky et al., 1990]. Research has further identified a range of mechanisms through which contextual factors influence decision making, including altering the salience of information [Tversky et al., 1988], the types of comparisons evoked [K. Hsee et al., 1999], and the types of memories that are brought to mind [Weber and Johnson, 2006]. Furthermore, problems of will power and lack of foresight plague individual decision making [Loewenstein and Haisley, 2007].

Paternalistic policies try to solve decision making hurdles by mandating decisions for individuals. Such policies are often heavy-handed and generate externalities [Loewenstein and Haisley, 2007]. “Soft paternalism,” on the other hand, avoids coercion. Soft paternalistic solutions try to affect individual behavior through technology or policy design: their goal is to design systems, rules, or mechanisms that counter known cognitive and behavioral biases and steer users in a direction believed to be more desirable (based on the user’s own prior judgement, or on external empirical validation), without impinging on her autonomy or freedom. The idea behind soft paternalism (or “nudging,” [Thaler and Sunstein, 2008]) is that lessons learnt about the psychological processes underlying behavior can be used to actually aid that behavior, by designing technologies and policies that enhance choice, without restricting it [Loewenstein and Haisley, 2007], with the goal of increasing individual and societal welfare. To do so, policy makers, technology designers, or users themselves may try to counter the fallacies and biases that affect privacy decision making, or even try and exploit those fallacies, in a way that does not force a solution on the user but nudges her, offering the user the option of more informed and more desirable choices.

In the way we use the term, nudges are intended as tools of soft-paternalism that can be used to 1) influence as well as, ideally, 2) help, assist, and therefore ameliorate privacy decision making. Since privacy decision making can be afflicted by different hurdles (such as asymmetric information, bounded rationality, and cognitive or behavioral biases [Acquisti, 2009]), nudges may come in many form: altering the incentives scheme associated with a certain behavior or technology; changing the information available to the user; modifying the interface of a system; re-design a public policy or a technological infrastructure; and so forth. Hence, soft paternalistic concepts could inform technology or policy design in order to help users protect their privacy online.

Such approach may still sound rather forceful, and may raise a legitimate question: who decides what is the “right” thing to do for the user when it comes to her privacy? After all, the success of online social networks demonstrates that many individuals, by and large, enjoy disclosing a lot to others - friends and strangers alike.

In reality, the study of soft paternalistic approaches to privacy decision making is much less heavy-handed than it may initially appear. In addition, such investigation does not necessarily imply or assume that less revelation is always the optimal or more desirable choice. In fact, soft paternalistic tools to improve decision making can be analyzed without having to make normative statements about the desirability of a given disclosure or non-disclosure, or the optimality of a given privacy-sensitive behavior.

First, the scientific investigation of what mechanisms are successful in affecting disclosure or privacy behavior is a different question from how those mechanisms should be used, and towards which outcome. Second, a distinction must be drawn between anticipating and countering the impact of cognitive and behavioral biases in privacy decision making, and influencing users towards specific privacy behaviors. In the former case, the goal is merely to identify a bias (that influences the user, sometimes without her knowledge, or awareness, or even against her own desires) and then counter or neutralize that bias, thus leaving the user free to make her decision about disclosing or protecting personal data without being subject to the influence of that bias. In the latter (and certainly more forceful) case, users’ behavior may indeed be influenced towards certain outcomes. In turn, such outcomes may be chosen according to different principles. For instance, (1) the outcomes may be chosen to be consistent with users’ stated preferences; or, (2) the outcomes may be chosen among the set that users are less likely to later regret, based on the analysis of other users’ behavior; or, (3) the outcomes may be estimated (based on sound evidence) to be welfare enhancing for the user.
4.2.2 Privacy (and Security) Nudges in the Literature

Currently, only a small number of recent studies and industry applications focus on mechanisms that may be categorized as nudges. We present a brief overview of them in this and the following sections, including – in addition to privacy studies – a few cases that relate to information security.

Different types of soft paternalistic interventions (or, in the language of this section, “nudges”) aimed at influencing privacy decision making are possible.

Some nudges simply consist of informing the user about privacy risks or solutions. This type of nudges is related to privacy research on informed consent. For instance, recent experimental research has shown that users are interested in protecting their privacy and may even pay for it, if appropriate tools and salient, simple, and compact privacy information are offered. Specifically, Tsai et al. [2010] studied the impact of making information about privacy practices on web sites more accessible to buyers. The results showed that some online customers are more likely to shop online from websites that exhibit more protective privacy policies. Additionally, those customers are willing to pay a premium for privacy. Furthermore, privacy indicators displayed at the moment an individual is shopping online may have an impact on users decisions. In particular, they increase the willingness to pay for privacy; however, if the indicator is provided only after the shopper has already chosen the website to buy at, it will not make the user to change decisions already made. This result concludes that timing is essential when trying to help people to protect their privacy [Egelman et al., 2009b]. Similarly, Gross and Acquisti [2005] found that priming Facebook users merely with questions about their online disclosure behavior and the visibility of their Facebook profiles was sufficient to trigger changes in their disclosure behavior.

Some nudges focus on making systems simpler to use. This type of nudges fall into the realm of research on privacy usability. For instance, application interface design can help users notice when changes in context generate changes in information flows and then help them to maintain their privacy [Hull et al., 2009].

However, other nudges may aim at countering specific cognitive and behavioral biases - such as neutralizing the effects of immediate gratification biases in privacy decision making [Acquisti, 2004] by altering the individual’s perception of the sequence of costs and benefits associated with revealing sensitive information. For instance, in the context of Web 2.0 location sharing applications, providing feedback to users whose location has been requested by others has been shown to have both positive and negative implications [Jedrzejczyk et al., 2010]. It can prevent excessive and unnecessary requests and hence protect people’s privacy. However, unless appropriate notifications are used, feedback receivers could also be annoyed. In addition, notifications may inhibit users from requesting others locations and hence affect system usage.

An area of research closely related to privacy decision making is the investigation of how users make security decisions.

Egelman et al. [2010] have shown that users are more tolerant of delays caused by security tools (such as antivirus software) when they are informed about the purpose and the risks covered by such tools. Villamarin-Salomón and Brustoloni [2010] have shown that rewards and punishments are more effective than traditional security dialogs in incentivizing “good” security behavior: with traditional security dialogs, users are not motivated to engage in secure behaviors because they do not face any positive (reward) or negative (punishment) consequence.

Context is also important for security decisions. Tools helping users to make better security decisions should know about the context before offering security choices. Further, a natural way to understand the context is by receiving feedback from users. Context-sensitive guidance using polymorphic and audited dialogs have successfully been used to encourage users to pay attention to computer warnings and take appropriate actions [Brustoloni and Villamarin-Salomón, 2007].

One of the main problems with securing the Internet is the lack of incentives Internet users have in implementing security measures due to negative externalities imposed by others and the low probability of actually suffering a catastrophic damage. Grossklags et al. [2010] analyzed two types of nudges: pay for outcome and pay for effort. According to the authors, different nudges are suitable
for different network scenarios.

4.2.3 Privacy Nudges in the Industry

Example of industry products or solutions that influence decision making in regards to privacy take various forms. Some of these solutions may be interpreted as soft paternalistic forms of privacy protection: they try to nudge towards privacy. They include privacy/security usability solutions, simplifications of privacy settings, or tests and delays before one can post information. More frequent, however, are the examples of products and solutions that nudge individuals to reveal more personal information. These include privacy defaults that are open, lack of usability in privacy settings interfaces, poorly designed warnings, and other rewards for sharing data or encouraging friends to share data.

4.2.3.1 Connections in social applications

Some applications provide information about who can see your data, who has seen your data, or how many people can see your data. For instance, Flickr.com - a video and image sharing website - provides information on each user-owned picture stating who can see it, followed by a link to edit the privacy settings for that individual picture. The information is both granular and accurate about who can access the data. This may be a nudge towards privacy, as users may decide to share certain photos with friends or family, and share other photos with everyone.

If the user has connected their email address, social networking sites often show the number of connections a user has. These connections can be called followers, friends, or ties. In some cases, connections can have access to all the user's information that is on the application. Twitter and Google Buzz are examples of sites that prominently show the number of connections. In the case of LinkedIn - a job searching tool - the user may prefer to add additional connections, even with people they don’t know well, in order to grow their job-searching network. These applications may nudge users towards increasing their connections and revealing more information. Indeed, several online social networks such as Facebook.com and LinkedIn.com periodically encourage users to add new connections by searching the user's email accounts for email contacts.

More people may have access to information than indicated by the number of connections. Connections may be able to share information with others, or the information may be publicly available. This is the case for Twitter, where re-tweets allow connections to pass on information without the individual's control. Facebook commonly shares an individual's information with friends of friends. Therefore, the number indicating the number of connections may mislead the user about the privacy of their data, decreasing the likelihood that the user will take a information-protective stance.

4.2.3.2 Privacy Settings

The privacy settings allowed in an application impact the user's ability to control how their information is shared. Both the default settings – a well known issue in the literature – and the usability of changing the settings create nudges towards, and away from, privacy [Mackay, 1991, Gross and Acquisti, 2005, Lai and Hui, 2006].

Some websites make privacy options very simple. For example, Pandora.com - an online music station - explicitly gives users two options regarding their profile page: make private or keep public. These options allow a user to choose without understanding complex details or settings. Conversely, the lack of granularity may encourage users to make everything public.

Several websites or software tools provide simple ratings of privacy settings. PrivacyCheck™ and ProfileWatch® give Facebook settings a scaled score of privacy, allowing a user to determine

---

how private their information is based on a single number. Other services provide a user-friendly layer on the Facebook privacy settings, allowing the user to change the settings. For example, Privacy Defender, \url{http://privacydefender.net}, provides a sliding color scale that allows the user to set their Facebook options as more or less private. These software services actively encourage stricter privacy settings.

4.2.3.3 Reduction of Information Disclosure

If an individual expects she may be likely to post information she may later regret, software exists to discourage her from doing so. Sophisticated users may choose to employ software tools to prevent excess disclosure. For example, the Social Media Sobriety Test\(^9\) and Mail Goggles on Gmail\(^10\) both allow the user to set certain hours of the week when they may typically embarrass themselves, such as weekend evenings after trips to the bar. During these hours, social network sites or Gmail may be blocked until the user can complete a dexterity or cognitive test. The user has the option to bypass the test.

Alternatively, a user may set up a warning system if a message is likely to be poorly interpreted. ToneCheck\(^11\) scans emails written in Outlook to discover whether the tone is off-putting, and will ask the user to confirm before sending it.

Other tools may discourage users from posting information by reminding the user who can see it. NetNanny a tool that parents can use to protect their children online. It will show a message every time a child posts on a social network. This message reminds the child that her parents will see the post as well.\(^12\)

4.2.4 Challenges

The literature on soft paternalism applied to privacy decision making is in its infancy, and therefore extremely scarce. The challenge we face, therefore, is to significantly expand our understanding of what forms of nudges can be effective in influencing disclosure and privacy behavior — as well as when, how, and under what conditions they should in fact be used.

By studying and understanding users’ behaviors and associated biases in Web 2.0 applications, we hope to suggest and test nudges that will help users make decisions that improve their satisfaction and well being. We are moving towards that goal by (first) understanding users’ needs, preferences, biases, and limitations about privacy, and (second) by using that information to evaluate the efficacy of techniques that counter, or exploit, biases to improve decision making. Ultimately, the objective is to study, design, and test systems that anticipate, and sometimes even exploit, those cognitive and behavioral biases that hamper users’ privacy and security decision making, and use these systems to “nudge” users towards certain behaviors that the users themselves have claimed to prefer, or which sound empirical evidence has demonstrated to be preferable. We hope that this type of work may inform the work of privacy and security technologists, providing insights and methods that go beyond better interfaces to revisit the strategies and assumptions underlying those systems.

---

\(^9\)See \url{http://socialmediasobrietytest.com}.

\(^10\)See \url{http://googlelabs.com}.

\(^11\)See \url{http://tonecheck.com}.

\(^12\)See \url{http://netnanny.com}.
5 Educational Aspects (OWK)

Children and adolescents are one of the main user groups of social network sites (SNS). Indeed, according to Facebook statistics (checkfacebook.com), in April 2011 38% of their Belgian users were under 24 years old. Moreover, recent studies show that in Flanders 87% of the Flemish teenagers have a profile on a SNS [Paulussen et al., 2010]. This young group of SNS-users is therefore an important target group for the SPION-project. How can we protect them from the problems resulting from privacy- and security issues that for now are inherent in using SNS? Education has been put forward as a solution by many authors working on the topic (eg. [Marwick et al., 2010, Patchin and Hinduja, 2010]. The scope of the educational part of the SPION-project is therefore to develop and evaluate educational packages that inform youngsters about the risks of the use of SNS. These packages may contain informative texts, games, stories, posters, comics, ... that can be implemented in in-class activities in secondary schools. Next to being informative, these packages also aim to change possible risky attitudes and behavior. A lot of educational packages have been developed already in Europe (for an extensive overview, see [Insafe, 2011]). As a start, we give an overview of some of these packages: which packages already exist in Flanders, what is their scope and to what extent are they effective? Afterwards, we conclude what this evaluation teaches us for what is relevant for the further process of our project.

5.1 Evaluation Framework

With regard to the scope of the educational packages, we try to find out what risks they tackle, using the overview of internet risks made by [DeMoor et al., 2008] (see Figure 5.1). This overview is made based on literature study, interviews with youngsters, parents and teachers [Walrave et al., 2010]. We chose this overview because of its exhaustiveness.

Figure 5.1: Risks youngsters might face using the internet [DeMoor et al., 2008].

Although this overview is made for internet use in general, it can be fully applied on the use of SNS as well. Indeed, youngsters might face some provocative content (eg. typical are hate-messages) or wrong information on SNS. Since it is a social network, there are also contact risks. SNS are for
example one of the most used ways for cyberbullying, next to instant messaging [Livingstone et al., 2011]. Moreover it's a way for sexual solicitation, by sending sexual messages [Livingstone et al., 2011]. There is also an offline-risk associated to the use of SNS, since 9% of European youngsters aged 9-16 has had offline contact with someone they first met online, in 62% of the cases on a SNS [Livingstone et al., 2011]. The privacy risks are clear as well, since SNS require members to provide some personal information to connect and communicate with others, and this while 26% of youngsters put their profile public and 28% put it only partially private using the privacy settings of their SNS [Livingstone et al., 2011]. This may also lead to commercial misuse of these personal data, since information can be shared with third companies via applications, and user behavior can be followed in order to provide target advertising and social advertising [Debatin et al., 2009b]. In the evaluation of the existing educational packages, we will try to find out which of these risks are being tackled. With regard to the effectiveness of these existing Flemish packages about the safe use of SNS, an extensive search in different academic databases showed us that there are no reports of empirical evaluation. Communication with authors of different existing packages led to the same conclusion. We did the same for educational packages about safe SNS use that were not in Flemish, but again we did not find anything. Yet, the impact of education in security problems is hotly debated, so we can not just take their effectiveness for granted. While some say education is pointless [Evers, 2006, Görling, 2006], positive effects have been found in some domains of internet security (eg. anti-phishing education, [Kumaraguru et al., 2010]).

Since no empirical evaluation could be found from the Flemish packages about safe SNS use, we will give a theoretical evaluation of these packages for now, using an evaluation framework for prevention packages in general described by [Nation et al., 2003]. These authors used a review-of-reviews approach to identify general principles of effective prevention campaigns that transcend specific content areas, this is characteristics of campaigns that show to be beneficial in helping youth to avoid numerous problems. Based on 35 reviews of studies of prevention campaigns in four different topics (substance abuse, risky sexual behavior, school failure and juvenile delinquency and violence) they defined nine principles, ordered in three categories:

- Principles related to program characteristics:
  - Comprehensive: multiple interventions, in different settings (combined parent, peer and school interventions support positive outcomes)
  - Varied teaching methods: active, skills-based component, interactive instruction, hands-on experience. Not relying too much on knowledge, information, or group discussions.
  - Sufficient Dosage: intense intervention, measured as the quantity and quality of the contact hours. Follow-up or booster session necessary to assure long term effects.
  - Theory driven: Based on etiological theories (focus on cause of the problem) and intervention theories (focus on the best method to lower risks).
  - Positive relationships: parent-child, peer-influences, significant others (peers, teachers, community members).

- Principles related to matching the program with the target population
  - Appropriately timed: Trying to have a maximal impact: early enough (before onset unwanted behaviour), but not too early (so that positive effects are not washed out before onset). Focus on changeable precursors of unwanted behavior. Tailored to the intellectual, cognitive and social development of the participants.
  - Socioculturally relevant: Taking into account the community norms, cultural beliefs and practices. The program must also address the needs of the target group.

- Principles related to implementation and evaluation of prevention programs:
  - Outcome evaluation: testing the effectivity of the program
-- Well-trained staff: Staff must be sensitive and competent. They must get sufficient training, support and supervision.

Most of these principles were already (independently) described by [Finkelhor and Luna, 1998], relying on different studies with even broader topics (also traffic safety, dental health, mental health, driver education, suicide). This confirms that these are consistent characteristics of effective prevention programs, that go beyond content. We might therefore conclude that these characteristics need to be part of any prevention program that tries to educate youngsters about privacy and security issues in SNS. Moreover Finkelhor and Luna [1998] also stated some factors that are typical for non-working preventions: the use of fear tactics, efforts to change attitudes alone and generalized approaches such as affective education. Thus we do not want to see this in any prevention programs on SNS.

As extra characteristics, we also discuss the lay-out and the usefulness (amount of detail, extra work for teachers...), since teachers report that these characteristics are important for effective dissemination Cannaerts [2011]. For this, we asked 8 experts in education to rate the different packages with respect to the attractiveness and usability of the package in a few items (e.g. I find this an attractive package) on a 7-point likert scale. In what follows we will describe some existing Flemish educational packages, have a closer look at their scope (using the summary of [DeMoor et al., 2008]) and evaluate them according to the characteristics summed above [Nation et al., 2003]. We need to mention the fact that this list of educational packages is not exhaustive. We focused on the packages made to use in a school setting (e.g. no websites, parent guides...). Moreover, not all packages were available to the authors at the time this overview was written. A summary of the evaluation can be found in Table 5.1 and Table 5.2.

5.2 Main existing educational packages evaluated

![Cover pages of the evaluated educational material.](image)

**Figure 5.2:** Cover pages of the evaluated educational material.

5.2.1 Primary School

1. **Play and Learn: Being Online**
   Developers: European Schoolnet, as part of the Insafe project (with for Belgium the partners ChildFocus and Clicksafe).
   Target population: 4-8 years old, exists in 13 different languages, including Dutch
   Content: small workbook, containing different exercises for children, regarding different aspects of being online (not only SNS) next to a guide for teachers and parents.
   Scope: making privacy and modern technology discussable between parents/teachers and
children. Tackles wrong information (adapted pictures), cyberbullying, privacy (difference between public/private, password, what to put on your profile). Provocative content, sexual solicitation, offline risks and commercial risks aren’t tackled.

Lay-out and usefulness: really attractive package, with nice, age-appropriate pictures (Mean rate attractiveness: 6.5). Ready-to-use, but without a detailed description of how to use the workbook (Mean rate usefulness: 5.3).

Theoretical evaluation: Most of the criteria of [Nation et al., 2003] are met (see Table 5.2), but a lot may depend on the implementation. There is e.g. the possibility to give this booklet to a child without further explanation, this way children probably will not understand the deeper level and the effectiveness will probably be close to zero. Nevertheless, when optimally used, the activities can be comprehensive and sufficiently dosed (there are a lot of exercises, but we should take in account the fact that they are not all about the same topic), appropriately timed (there are even ‘star’ indicators referring to difficulty, explaining what age is appropriate for what exercise), using interactive and skills-based teaching methods and emphasizing positive relationships between parents/teachers and children. The developers have also tried to make the booklets socioculturally relevant, by translating it into 13 languages. It has to be mentioned however that the content is exactly the same, independently of the cultural norms of the country (but they are all western), and that translations are not always carefully made (e.g. words from the Netherlands used in the Flemish booklet). Moreover there is no reference to any theoretical base and the outcome evaluation is scarce: there is a questionnaire on the website, with general questions to the teachers/parents about how much they and their children liked the package. A report of this evaluation could however not be found. Communication with the authors revealed that the exercises themselves have been evaluated with regard to difficulty, before publishing, but not with regard to their effectiveness of changing attitudes and behavior. As a last point it can also be stated that the available guide for parents and teachers might be a little too limited to speak of a well-trained staff, but then again this may depend on the implementation.

2. Kids in Cyberland

Developers: Sensoa vzw and Clicksafe (Childfocus), with support of the Flemish government and European Union.

Content: didactical marker, guidance for teachers, instructional materials (games, postulations, crosswords, online quiz, examples,...), background information and wordlist. Different teaching methods for different aspects of internet safety in three main themes: who am I on the internet (SNS), Me and the other on internet (mostly chat), surfing on the internet. There is a clear focus on sexuality in different parts of the package, since Sensoa vzw focuses on developing materials about relationships and sexuality.

Target population: Flemish youngsters in primary school aged: 10-12

Scope: to start up conversations about experiences and risks on the internet between children and adults. Tackles provocative content (porn,...), wrong information (finding reliable info), cyberbullying, sexual solicitation, privacy and offline contact risks. Only commercial risks are not included.

Lay-out and usefulness: Mostly text-based, but the instructional materials contain images of 2 kids, coming back in different exercises (Mean rate attractiveness: 4.3). The package is completely worked out, with a lot of details and explanations for teachers: it looks easy to implement (mean rate usefulness: 5.5).

Theoretical evaluation: A summary of the characteristics of Nation et al. [2003] can be found in Table 5.2. A lot of aspects are tackled and the teacher can choose between a lot of different teaching methods. There are also enough materials to have multiple, sufficiently dosed, interventions. Moreover, teachers are encouraged to give homework and let the children talk with their parents, so the package leaves the possibility to have interventions in multiple settings. It does not only encourage positive relations with the parents, but also with peers, in letting them work together, learning about respect for each other... Only with regard to the theoretical
base, there is some uncertainty. In the background information, there is only one reference to scientific literature. Because of this, it can be assumed that some content is based on theory. Yet, there is no reason to assume they have taken into account any theory about prevention. Most of the package items are appropriately timed. Questions can be asked however about the timing to learn about profile pages, since children under 13 are not allowed to have a profile page on most SNS. Yet, the package does contain images of how a profile looks like, and makes it easy to understand. A follow-up session might be good because of this young target population, but there is one available (Connected, see further). The chosen examples in the package are also socioculturally relevant: eg. Netlog (Belgian SNS), references to Kinder-en jongerentelefoon, Childfocus,.. A lot of effort has been done to match to the populations characteristics.

Concerning the outcome evaluation, no report could be found. Communication with the developers pointed out that there was no evaluation done due to a lack of funding. Finally, concerning the training of the staff, it can be expected that the teacher, with the help of the teacher guidelines, will be totally informed. These guidelines are clear, detailed and useful. At last we should mention that this package focuses on a positive message: show the opportunities, the fun things of the internet before telling something about the risks (no fear induction!) and without exaggerating these risks.

5.2.2 Secondary School

1. Op zoek naar mezelf in de 21ste eeuw (Finding myself in the 21st century)
   Developers: Flemish government, campus De Nayer and KHLeuven.
   Content: posters, didactical marker and learning materials (on paper and on computer). Contains three different packages: ‘Faceflap’, in which they make an offline paper profile, ‘Ooo mijn foto’ in which they adapt pictures of peers, and ‘Ik ben met veel’, using a digital tool. This way, the three packages each tackle a different aspect of safety on SNS.
   Target population: Flemish youngsters in secondary school aged: Faceflap: 11-16, Ooo mijn foto: 11-14, Ik ben met veel: 15-18 (ASO-TSO)
   Scope:

   - ‘Faceflap’: experiencing the fact that information about yourself on the net has a real impact on what others think about you.
   - ‘Ooo mijn foto’: understanding that photos are personal, that you should not adapt someone else’s photos without permission and what impact the wrong dissemination of pictures might have.
   - ‘Ik ben met veel’: How you can form an identity, what influences your identity, what is the difference between online identity and different physical identities and how can the absence of context bring problems with online identity.

   - Tackles wrong information (interpretation of (adapted) pictures and messages), cyber-bullying (adapted pictures), privacy (what to put on your profile, who to add as friends, pictures online, online identities, absence of contexts,...). Provocative content, sexual solicitation, offline risks and commercial risks are not tackled explicitly.

   Lay-out and usefulness: It seems like a lot of effort has been put in lay-out (pictures, posters, leaflets,..). Everything is combined in an attractive folder (mean rate attractiveness: 5.2). The didactical marker contains enough detail to make it useful for teachers (mean rate usefulness: 5.1).

   Theoretical evaluation: A summary of the characteristics of Nation et al. [2003] can be found in Table 5.2. These packages offer really attractive materials, with the choice between three
packages with different instruction methods, for different age groups. This gives the possibility
to provide a prevention with sufficient dosage (also over time) and varied teaching methods,
at the appropriate time. While it is possible to have multiple interventions over time, the pack-
ages are only oriented towards teachers in schools, and no extensions to other settings can be
found, so the comprehensiveness of the package could be extended. There are some refer-
ences to research in the learning materials and in the report about the project, but there is no
clear view on the theoretical base of the package. We might assume though, because of these
references, that the developers did have a view on an etiological theory of the problems they
assessed. Yet, there is no reason to assume that they used any intervention theories. With
regard to the sociocultural relevance, although the package is developed especially for Flemish
youngsters, there are a lot of parts in the 'we zijn met veel'-package that are written or spoken
(videos) in English. Finally, with regard to the implementation and evaluation characteristics,
the didactical fiche is really helpful for teachers. The developers have also put a lot of effort
in disseminating and setting up workshops to train teachers. The report about the project also
reveals the fact that they changed the materials several times in response to feedback from
teachers Cannaerts [2011]. There is a basic evaluation form online, as part of the 'we zijn met
veel'-package, for students to fill in after their session, mostly about their feelings regarding
technology. They also asked about their interests for the project, and how they liked it. 40
students completed the questionnaire, but an independent sample t-test pointed out that the
mean on most items was not different from the neutral 3 on a 5-point likert scale. It is hence dif-
cult to draw conclusions. An evaluation of the effectiveness of the project (change in attitudes,
behavior, knowledge) has not been done.

2. Think before you post
Developers: Childfocus, Clicksafe, as part of Safer Internet Day 2010
Content: didactical marker, instructional materials (workbook with examples, exercises, pic-
tures, schemes, ...), a little bit of background information. The package contains three main
themes: 1) anonymity, personal, private, public; 2) online identities and 3) it can go wrong
(risks), which are all tackling different aspects of safety online (broader than SNS).
Target population: not clarified: “children and adolescents”
Scope:

(a) raising awareness about information flow and the public character of internet, the risks of
sharing personal information and the advantages and disadvantages of staying anonymous

(b) Learn how to build an online identity, know that this can differ from a physical identity,
learn what to share with whom (incl. pictures)

(c) Learn what kind of behavior is risky, what behavior can count as cyberbullying and how to
cope with problems. Tackles all problems: provocative content, wrong info, cyberbullying,
sexual solicitation, privacy risks, offline risks, commercial risks.

Lay-out and usefulness: pictograms have been used, but few pictures are used in the in-
structional materials (mean rating attractiveness: 3.3). The didactical marker is detailed, but
because the information for teachers isn’t really elaborated, implementation might cause some
difficulties (mean rating usefulness: 4.6).

Theoretical evaluation: A summary of the characteristics of Nation et al. [2003] can be found
in Table 5.2. This package offers some ready to use lessons. While some of the parts are the
same as in “Kids in Cyberland” (also developed by childfocus), and the look is quite the same,
the evaluation of the whole package with regard to the characteristics of Nation et al. [2003]
isn’t as good. Children aren’t stimulated to work at home with their parents, which makes the
package less comprehensive and less focused on positive relationships. Still there are enough
materials to have a sufficient dosage of lessons with varied teaching methods. It seems based
on an etiological theory because of the resources in the background information-section. Yet,
there is still no reason to assume that there were any intervention theories taken into account.

Version 1.0
Moreover, there is no indication of an appropriate age to use this package. The socio-cultural relevance is ok, but not splendid, since there are no references to Flemish organizations, social networks,.. Concerning the training of the teachers, the information for teachers isn’t really elaborated, and the teacher might remain ignorant. An outcome evaluation has been asked, but communication with the developers pointed out that there was no evaluation done due to a lack of funding.

3. **Connected**

**Developers:** Sensoa vzw and Clicksafe with support of the Flemish government and European Union

**Content:** didactical marker, guidance for teachers, instructional materials (games, postulations, crosswords, online quiz, examples,...), background information and wordlist. This package seems to be an extension of Kids in Cyberland, focusing on different aspects of safe internet use. There is again a clear focus on sexuality in different parts of the package.

**Target population:** Flemish youngsters in secondary school (age not specified)

**Scope:** Educating youngsters to be critical internet users, with respect and attention for privacy. Teach them to know their boundaries, to show them and to respect them. Tackles provocative content (porn,..), wrong information (finding reliable info), cyberbullying, sexual solicitation, privacy and offline contact risks. There is no attention however for commercial risks.

**Layout and usefulness:** Only text, no images (mean rate attractiveness: 2.7). A lot of themes but no clear structure. Detailed, but not attractive to implement (mean rate usefulness: 4.2).

**Theoretical evaluation:** A summary of the characteristics of Nation et al. [2003] can be found in Table 5.2. Although this package is made by the same organizations as Kids in Cyberland and seems to be an extension of it, it doesn’t have the same precision of development and thoroughness of materials and explanation. Concerning the comprehensiveness of the package, we miss the encouragement of working in different settings, as in the other package. The rest of the program characteristics are just as good, but there is a lack of precision with respect to the age of the target group (secondary: 12-18 years old is a big range) and of references to Flemish organizations (there are some, but this was much more worked out in Kids in Cyberland). The guidelines for teachers are also not worked out as good as in the other package (even for some identical exercises!). Of course, the package might have a lot of potential, with Kids in Cyberland as a good example. Communication with the developers pointed out that there was no evaluation done due to a lack of funding.

<table>
<thead>
<tr>
<th></th>
<th>Play and Learn being online</th>
<th>Kids in Cyberland</th>
<th>Op zoek naar mezelf in de 21ste eeuw</th>
<th>Think before you post</th>
<th>Connected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Only SNS</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>How many risks? (/7)</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Avoids fear induction</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Lay-out</td>
<td>6.5</td>
<td>4.3</td>
<td>5.2</td>
<td>3.3</td>
<td>2.7</td>
</tr>
<tr>
<td>Usefulness/detail</td>
<td>5.3</td>
<td>5.5</td>
<td>5.1</td>
<td>4.6</td>
<td>4.2</td>
</tr>
</tbody>
</table>

**Table 5.1:** Summary of some extra characteristics of the described prevention programs

Packages downloadable from:

**Version 1.0**
### Program Characteristics

<table>
<thead>
<tr>
<th></th>
<th>Play and Learn being online</th>
<th>Kids in Cyberland</th>
<th>Op zoek naar mezelf in de 21ste eeuw</th>
<th>Think before you post</th>
<th>Connected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comprehensive</td>
<td>+</td>
<td>+</td>
<td>+/-</td>
<td>+/-</td>
<td>+/-</td>
</tr>
<tr>
<td>Varied teaching methods</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+/</td>
</tr>
<tr>
<td>Sufficient dosage</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+/</td>
</tr>
<tr>
<td>Theory Driven</td>
<td>?</td>
<td>+/-</td>
<td>+/-</td>
<td>+/-</td>
<td>+/-</td>
</tr>
<tr>
<td>Positive Relationships</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+/</td>
</tr>
</tbody>
</table>

### Matching Target Population

<table>
<thead>
<tr>
<th></th>
<th>Appropriately Timed</th>
<th>Socioculturally Relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appropriately Timed</td>
<td>+</td>
<td>+/</td>
</tr>
<tr>
<td>Socioculturally Relevant</td>
<td>+/-</td>
<td>+/-</td>
</tr>
</tbody>
</table>

### Implementation and Evaluation

<table>
<thead>
<tr>
<th></th>
<th>Outcome Evaluation</th>
<th>Well Trained Staff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome Evaluation</td>
<td>+/-</td>
<td>+/-</td>
</tr>
<tr>
<td>Well Trained Staff</td>
<td>+/</td>
<td>+/</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>6.5</th>
<th>7.5</th>
<th>6</th>
<th>4</th>
<th>5.5</th>
</tr>
</thead>
</table>

**Table 5.2:** Presence of the characteristics of good prevention programs, as described in [Nation et al., 2003](#)

- Op zoek naar mezelf in de 21ste eeuw. [www.opzoeknaarmezelf.be](http://www.opzoeknaarmezelf.be)

### 5.3 Gaps and Future Challenges

In Flanders, some nice and ready to use educational packages about safety on internet and SNS do already exist. However, in our evaluation we can see that there still seem to be some gaps. First of all, most packages are about safety on the internet in general, not only on SNS. Basically this doesn’t seem to be a problem, but this way a lot of risks typical for SNS are not tackled (eg. the risks discussed regarding the overview of [DeMoor et al., 2008](#)). Mostly, with regard to SNS, only privacy and wrong info (adapted pictures) are tackled, while the other risks are tackled in other contexts. This might cause a lack of awareness of these risks, while using SNS. Moreover, some typical aspects of SNS are often overlooked because of the focus on pictures and contact information, e.g. social advertising, impact of hate-messages or joining hate-groups, selling of personal data to third companies, identity-shaping content like pictures or messages about alcohol abuse, negative attitudes towards school or superiors, ... We can also see that commercial risks are only tackled in one of the packages, namely Think before you post. A more comprehensive approach, concerning the different risks, is designated.

While most packages have a lot of exercises, allowing different teaching methods, only few refer to real technical skills like changing privacy settings on SNS. Furthermore, when there is a reference to such a skill, there is no related explanation for the teachers. We believe an extension in this direction might be necessary, but might be challenging due to the daily change of SNS designs.

Version 1.0
Most packages already meet some criteria of Nation et al. [2003]. There is however room for amelioration, especially with regard to the outcome evaluation and the theoretical base of the program. Moreover there seems to be more need for dissemination: in a questionnaire we conducted with 739 teenagers (13-19 years old), 99.5 % of them reported never to have heard of any educational package concerning safer use of SNS. Of the 36 teachers we questioned, only 2 had ever heard of an existing package, of which one was an ICT-teacher.

In the further process of our project, we will try to contact the developers of the packages discussed, and ask them for their permission and willingness to work further on these existing packages. We will try to combine and puzzle, to create a comprehensive package with regard to the different risks in SNS. We will try to extend this with more technical exercises, eg. changing privacy settings. Finally, we will empirically evaluate the created package, by implementing it in classes. By using a pretest-posttest, quasi-experimental design, we will try to determine the effectiveness of the package, with regard to changing knowledge, attitudes and behavior and we’ll determine the characteristics that seem to be important to make the package effective (starting from the characteristics as described by Nation et al. [2003], but specifying this for our specific context). Furthermore, we will put efforts into disseminating the materials.
6 Legal Aspects (ICRI)

6.1 Introduction

A correct demarcation of the legal and regulatory framework applicable to SNS is essential in order to adequately assess the legal issues arising in relation to online social networks sites (SNS). Technological developments, the constant introduction of new types of services in SNS, and the lack of borders on the Internet are only few of the reasons that intensify the difficulties in defining the legal landscape with regard to online social networks. The main objectives of the legal research track within the SPION project are to:

- document the current legal framework applicable to online social networks;
- identify legal uncertainties, gaps and shortcomings of the existing framework;
- explore new legal mechanisms which enhance the accountability of SNS service providers and mitigate responsibilization of individuals;
- develop compliance assessment criteria; and
- provide guidance for privacy-friendly policies and default settings.

The legal research track in SPION shall focus on the following three areas of regulation: Data Protection, E-Privacy and E-Commerce. The first two areas were selected because they contain the main rights and obligations of users, SNS providers and third parties in relation to the processing of personal data. The third area, E-Commerce, was selected because it contains important liability exemptions which may be held applicable to the providers of SNS or related applications.

6.2 Data Protection

The main legal instrument on data protection in the EU is Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data. This Directive formulates the basic principles and rights of data subjects in relation to the processing of their personal data. It was implemented into Belgian law by means of the Law of 11 December 1998, which modified the existing Data Protection Act of 8 December 1992.

6.2.1 Scope

Directive 95/46/EC applies 'to the processing of personal data wholly or partly by automatic means, and to the processing otherwise than by automatic means of personal data which form part of a filing system or are intended to form part of a filing system' (article 3(1)). The scope of the Directive covers all automated processing of personal data, save for the areas excluded by article 3(2) of the Directive, namely:

2 Currently the Directive is under review. The European Commission plans to propose new legislation in the course of 2011.
3 B.S., 3 February 1999.
4 B.S., 18 March 1993.
• processing of personal data in the course of an activity which falls outside the scope of Community law, and to processing operations concerning public security, defence, State security and in areas of criminal law;

• processing of personal data by a natural person in the course of a purely personal or household activity.

Personal data is defined by article 2(a) of the Directive as any information relating to an identified or identifiable natural person (‘data subject’). An ‘identifiable person’ is a person ‘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’\(^\text{5}\). In order to determine whether a person is identifiable, account should be taken of ‘all the means likely reasonably to be used either by the data controller or by any other person to identify the said person’ (recital (26)). Finally, ‘processing of personal data’ is defined as any operation or set of operations which is performed upon personal data, whether or not by automatic means, such as collection, recording, organization, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, blocking, erasure or destruction.

A substantial amount of data processing that takes place in the context of SNS qualifies as personal data processing within the meaning of Directive 95/46/EC. Profile information, pictures, wall posts, friend lists, etc, shall all be considered as personal data provided they relate to an identified or identifiable individual. While the identifiability of the individuals concerned may be more straightforward in some cases than in others, one should always take into account that even where individuals do not appear to be directly identifiable (e.g., when an alias is used), they will often be indirectly identifiable through one or more other elements, such as the personal attributes listed in their profile (e.g., age, residence, etc.), their list of friends, traffic data (e.g., IP-addresses) or cookie data.

Once it has been established that a certain processing operation falls within the scope of the Directive, one must determine which entity (or entities) is (are) responsible for ensuring compliance. The Directive has incorporated two core concepts which serve to allocate responsibilities among actors involved in personal data processing, namely that of the ‘controller’ and the ‘processor’\(^\text{6}\). A controller is defined as the entity who alone, or jointly with others, determines the ‘purposes and means’ of the processing (article 2(d)). A ‘processor’ is defined as an entity who processes personal data on behalf of the data controller (article 2 (e)). Controllers carry the primary responsibility for ensuring compliance with the substantive provisions of the Directive. Processors on the other hand, shall as a rule only be indirectly accountable for the compliance obligations under Directive 95/46/EC\(^\text{7}\).

Technological and societal developments since the enactment of the Directive have made it increasingly difficult to apply the distinction between data controllers and data processors in practice\(^\text{8}\). More and more, entities define their respective roles and responsibilities in ways which do not always


\(^{6}\)The controller and processor concepts are also of central importance for a number of other legal questions, such as which national law(s) shall apply to a particular processing operation. See Article 29 Data Protection Working Party, ‘Opinion 1/2010 on the concepts of “controller and “processor”’, WP169, 16 February 2010, available at http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2010/wp169_en.pdf.


allow for an easy distinction between controllers and processors\(^9\). This is also the case with online social networks. SNS services and related applications typically involve a multitude of different actors (e.g., the SNS service provider, users, third-party application providers, service providers which have ‘like buttons’ on their website). The level and nature of involvement of each of these actors may vary considerably. As a result, the nature of their relationship towards one another under data protection law can take on many different forms: a controller-processor relationship, a controller-to-controller relationship, a relationship of joint control, etc. One of the main objectives of the legal research track within SPION will be to analyse the respective qualifications, rights and obligations of the different types of actors involved in online social networks. In doing so, we hope to not only clarify the existing framework, but also to identify areas in which the current definition of actors, roles and responsibilities could be improved\(^10\).

### 6.2.2 Data Protection Principles

Directive 95/46/EC embodies a number of data protection principles, which are considered to be the bedrock of the European Data Protection law. These principles are a set of rules, which are considered to be the necessary requirements with which any processing of personal data that falls within the scope of this Directive must comply.

- The first principle states that personal data must be processed fairly and lawfully (article 6(1)(a)). This means that personal data may only be collected and further processed by fair and lawful means. This provision also reiterates that data controllers must stay in line with other legal obligations, even outside of the Directive, regardless of whether these obligations are general, specific, statutory or contractual.

- A second basic data protection principle is the principle of legitimacy. This principle entails that the processing has to be based on one of the legitimate grounds listed in article 7 and/or 8 of the Directive\(^11\) (e.g., unambiguous consent by the data subject, processing necessary for the performance of a contract to which the data subject is a party).

- Pursuant to the finality principle (article 6(1)(b)), the collected personal data may only be processed if the purpose for which they were collected can be justified and may not be further processed in a manner incompatible with that purpose.

- The principle of proportionality requires that the processing of personal data should be limited to data that are adequate, relevant and not excessive (article 6(1)(c)). This means that data controllers are obliged to store only a minimum of data necessary to run their services. Additionally, personal data shall not be kept for longer than necessary for the purposes for which it was collected. After achieving the purpose for which the data was gathered, it should be rendered anonymous or destroyed (article 6(1)(e)).

- The data accuracy principle (article 6(1)(d)) provides that all personal data should be accurate and, where necessary, kept up to date. Data controllers are obliged to take every reasonable step to ensure that data which are inaccurate or incomplete are either erased or rectified.

- The principles of confidentiality and security (article 16 and 17) oblige the data controller to implement appropriate technical and organizational measures to ensure the confidentiality and security of processing, and to protect the data against any form of unlawful processing. Such

---

\(^9\)C. Kuner, o.c., 72.

\(^10\)See in particular SPION deliverables D6.2 (‘Rights and obligations of actors in social networking sites’) and D6.3 (‘Liability and accountability of actors in social networking sites’).

\(^11\)Whether one of the grounds of either art. 7 or art. 8 may be used depends on the nature of the personal data. Art. 7 relates to ‘normal’ personal data, whereas art. 8 outlines the possible legitimacy grounds for ‘special categories of data’ (i.e. data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and the processing of data concerning health or sex life).
measures must be appropriate to the risks represented by the processing and the nature of the data to be protected.

- Articles 10 et seq. set forth the transparency obligations of data controllers and list the rights of the data subjects. Underlying these provisions is the idea that the data subject should in principle be notified of the processing of her personal data (notice), or have immediate means of recourse towards the controller in case she feels her data are being processed improperly (right to rectification, erasure or blocking). Transparency of the processing is ensured also by the obligation to notify the respective national data protection authority before any data processing operation is carried out (article 18).

6.3 E-Privacy

The next relevant legal instrument is Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector, which is often referred to as the 'E-Privacy Directive'. This Directive was implemented into Belgian law by way of the Law of 13 June 2005 regarding Electronic Communications.

The E-Privacy Directive compliments the general rules of data protection provided by Directive 95/46/EC with the specific rules for the processing of personal data in the context of electronic communications. In addition, the E-Privacy Directive contains a set of generally applicable rules, e.g. with regard to spam or cookies.

6.3.1 Scope

The E-Privacy Directive applies to the processing of personal data in connection with the provision of publicly available electronic communications services in public communications networks in the EU, including public communications networks supporting data collection and identification devices (article 3). Many of its provisions therefore apply in first instances to the providers of such services. However, several of the provisions of the E-Privacy Directive have been drafted in such a way that their scope is not limited to the provisioning of publicly available electronic communications services as such, but may also apply to other types of services or actors.

Over the following subsection we will briefly elaborate upon three provisions which may be of particular relevance in the context of online social networks.

---

12 The transparency obligations of SNS providers shall be analyzed in the context of SPION Deliverable D6.1 ('Legal requirements for privacy-friendly model privacy policies').


14 B.S. 20 June 2005. This law in fact implements several different European directives. Most of the provisions of the E-Privacy Directive are implemented in the articles 108 et seq.

6.3.2 Location Based Services

Location-based services are services whose main functionality depends on the processing of location information. Location data are by the E-Privacy Directive defined as ‘any data processed in an electronic communications network or by an electronic communications service, indicating the geographic position of the terminal equipment of a user of a publicly available electronic communications service’ (article 2(c)). In digital mobile networks, location data are often processed to enable the transmission of communications. However, such data may also be useful in providing additional functionalities (e.g., location-based direct marketing). They may also be used to yield interesting applications related to SNS (e.g., a friend proximity service, sharing of location tracked through GPS).

The E-Privacy Directive regulates those location-based services which rely on the processing of location data other than traffic data offered via a public communications network or in a publicly available electronic communications service (article 9(1)). Specifically, such data may in principle only be processed when they are made anonymous, or with the consent of the users or subscribers, to the extent and for the duration necessary for the provision of a value added service. The service provider must inform the users or subscribers, prior to obtaining their consent, of the type of location data other than traffic data which will be processed, of the purposes and duration of the processing and whether the data will be transmitted to a third party for the purpose of providing the value added service. Users or subscribers shall be given the possibility to withdraw their consent for the processing of location data other than traffic data at any time.

6.3.3 Cookies

Online social networks, like many websites, often place ‘cookies’ on the devices of individuals visiting their webpages. Such cookies can serve a variety of purposes: personalization, tracking, session management, etc. The use of cookies is regulated by the E-Privacy Directive, specifically by article 5(3). The recent amendments to the EU Telecoms package modified the wording of this provision, which is often referred to as the “cookie rule”. The revised version of this provision entails that users and subscribers must in principle provide prior consent before the placement of (or subsequent access to) cookies on their computer. The individual concerned must in principle also have received clear and comprehensive information about the processing involved, in accordance with Directive 95/46/EC (cf. supra; section 6.2.2). The previous version of this provision was slightly less stringent, stating that the use of cookies was only allowed when the user was informed about it, in a clear and comprehensive way, in accordance with the 95/46/EC Directive, and was offered the right to object to such processing by the data controller. According to the Preamble of the amendment Directive 2009/136/EC, the user’s consent to processing may be expressed by using the appropriate settings of a browser or other application. See recital (66) of Directive 2009/136/EC. This is however a highly disputed issue, seeing as many users are not aware of their browser settings, making it questionable whether they have in fact expressed valid consent. The issue of informed consent and default settings shall be revisited as part of the legal research track in SPION, in particular in

---

16 See E. Kosta, o.c., 264.
17 Traffic data’ is defined by the E-Privacy Directive as ‘any data processed for the purpose of the conveyance of a communication on an electronic communications network or for the billing thereof (article 2(b)).
19 Article 5(3) provides two (narrow) exceptions to this rule, namely where the storage or access is (a) carried out for the sole purpose of carrying out the transmission of a communication over an electronic communications network, or (b) is strictly necessary in order for the provider of an information society service explicitly requested by the subscriber or user to provide the service.
20 See also E. Kosta, o.c., p. 254 et seq.
the context of deliverables D6.1 (‘Legal requirements for privacy-friendly model privacy policies’) and D6.4 (‘Guidelines for privacy-friendly default settings’).

6.3.4 Spam

The E-Privacy Directive also introduced a provision on unsolicited communication (article 13). The aim of this provision is to protect subscribers against any form of spam, which may impose a burden and/or cost on the recipient. Article 13 (in both the old and the amended versions of the E-Privacy Directive) states that the use of automatic calling machines, fax or electronic mail for the purposes of direct marketing may only be allowed with respect to subscribers or users who have given their prior consent. However, if a customer has provided his e-mail address when purchasing a product or a service, this address can be used by the same provider of the product or the service for direct marketing of its own similar products or services. This exception only applies under the condition that customers are clearly and distinctly given the opportunity to object to this practice, free of charge and in an easy manner.

6.4 E-Commerce

The third legal instrument which merits further elaboration is Directive 2000/31/EC, which is commonly referred to as the ‘E-Commerce Directive’. The aim of this Directive was to establish a clear and general framework for certain legal aspects of electronic commerce within the internal market. It was implemented into Belgium law by way of the Law of 11 March 2003 concerning certain legal aspects of information society services.

6.4.1 Scope

The E-Commerce Directive applies to the provisioning of ‘information society services’, which are defined as ‘any service normally provided for remuneration, at a distance, by electronic means and at the individual request of a recipient of services’ (article 2(a)). This definition covers a wide range of online services. Examples of information society services may include: on-line contracting, services providing transmission of information via communication networks, access to a communication network, hosting of information, search tools, etc. Online social networks and related applications shall in principle also qualify as information society services, as they are services which are:

- ‘normally provided for remuneration’: information society services are not restricted to services which are remunerated by their recipients. The element of remuneration refers to the existence of an economic activity, which means that they could be remunerated for example by advertising. Most SNS and related applications, even if they are provided free of charge to the end-user, satisfy this criterion;
• ‘provided at a distance’: the service provider and end-user are not simultaneously physically present in the same place while the service is being provided;
• ‘by electronic means’: the services provided by SNS and related applications are sent and received with the use of electronic equipment; and
• ‘at the individual request of the recipient’: the services provided by SNS and related applications primarily consist in the transmission of data on demand.

It is important to note that a number of services and legal issues are excluded from the scope of the E-Commerce Directive, such as: questions covered by the Data Protection Directive (article 1 (5)(b)); and questions relating to agreements or practices governed by cartel law (article 1 (5)(c)).

6.4.2 Substantive Provisions

The E-Commerce Directive regulates a variety of legal issues with regard to provision of the information society services, such as:

• Application of the Member States national provisions to the providers of information society services established on its territory. The rule is expressed by two principles, namely the rule of origin principle and the principle of freedom of services (article 3).
• Transparency of the information society services, as expressed through a number of information obligations (article 5, article 6, article 11)
• Conclusion of contracts by electronic means, which should not be deprived of legal effectiveness and validity on account of their having been made by electronic means (article 9).

Most relevant to our research however are the liability exemptions for intermediaries, as well as the provisions creating a framework for the development of voluntary codes of conduct. Each of these provisions will be elaborated briefly in the following subsections.

6.4.3 Liability of Intermediaries

Section 4 of the E-Commerce Directive regulates the liability of intermediary service providers. The Directive provides that intermediary service providers shall not be liable for actions that qualify as ‘mere conduit’ (article 12), ‘caching’ (article 13) or ‘hosting’ (article 14). In order to benefit from these exemptions, the providers of such services must comply with the conditions foreseen by each article. The scope of these exemptions has a horizontal nature, which means that they cover various types of illegal content (infringements on copyright law, defamation law, protection of minors, unfair commercial practices, etc.) and different kinds of liability (criminal, civil, direct, indirect).

The exemption for the ‘mere conduit’ service providers (article 12) is aimed at two types of information society services: services which consist of the transmission in a communication network of information provided by a recipient of the service (‘transmission services’); and services which consist of the provision of access to a communication network (‘access services’). The ‘mere conduit’ exemption applies only if the service provider: (a) did not initiate the transfer of data; (b) does not select the recipient of the data; and (c) does not select or modify the transmitted data.

According to recital (42), the exemptions provided by the Directive apply only to cases ‘where the activity of the information society service provider is limited to the technical process of operating and giving access

---

29The service provider would then not be able to claim that his role is limited to that of an intermediary.
30This requirement does not cover manipulations of a technical nature which take place in the course of the transmission as they do not alter the integrity of the information contained in the transmission (recital (43)).
to a communication network’. The liability exemption for mere conduit also extends to the automatic, intermediate and transient storage of the information transmitted in so far as this takes place for the sole purpose of carrying out the transmission in the communication network, and provided that the information is not stored for any period longer than is reasonably necessary for the transmission (article 12(2)).

Caching of information is an ‘automatic, intermediate and temporary storage of that information, performed for the sole purpose of making more efficient the information’s onward transmission to other recipients of the service upon their request’. Liability of the service provider is excluded only if he is in no way involved with the information transmitted (recital 43). Moreover, there are five additional conditions (article 13(1): (a) the provider cannot modify the information; (b) the provider has to comply with conditions on access to the information; (c) the provider must update the information regularly; (d) he may not interfere with the lawful use of technology that is used to measure the use of information; (e) and he must remove the cached information immediately upon obtaining knowledge that the initial source of the information is removed, access to it has been disabled, or that a court administrative authority have ordered such removal or disablement.

The most relevant liability exemption in the context of online social networks is the hosting exemption. Like caching, hosting also involves storage of information, but is said to occur when the storage is not merely ‘incidental’. Walden, ‘Directive 2000/31/EC D Directive on electronic commerce D Hosting’, in Bullesbach A., Poullet Y., Prins C. (eds.), Concise European IT law, Kluwer Law International Alphen aan den Rijn, 2005, p. 253. and is provided for a prolonged period of time. It may certainly be the primary object of the service. If an information society service consists of the storage of information provided by a recipient of the service, the provider shall not be liable on the condition that: (a) the provider does not have actual knowledge of illegal activity or information and, as regards claims for damages, is not aware of facts or circumstances from which the illegal activity or information is apparent; or (b) the provider, upon obtaining such knowledge or awareness, acts expeditiously to remove or to disable access to the information.

It has been said that the hosting exemption was originally aimed at ISP’s providing space on their internet servers for third parties’ websites, or bulletin boards or chat room services provided by the ISP itself (where the ISP only provides technical means for the users’ communication without interfering with the content being communicated between the users). However, the exemptions provided by the E-Commerce Directive are defined in functional terms (i.e. in terms of the activity being performed), not in terms of the qualification of the actor. While the European legislator arguably only envisioned providers whose services consisted mainly, if not exclusively, in the performance of operations of a strictly technical nature, the scope of the exemption may also be applied to other entities (provided the conditions set forth by art. 14 are met). As a result, the exemption may in principle benefit any type of service provider who stores content at the request of the recipient; including so-called ‘web 2.0’ service providers, provided the remaining conditions for the exemption are met. However, there remains a considerable amount of legal uncertainty concerning the exact scope of this exemption, and it is still undecided whether the provider of an online social network might be able to benefit from this exemption, and to which extent. This issue will also be analyzed


32 This would occur for example when technologies to keep track of the number of users visiting a web page showed less hits because of the caching activity (A. Lodder, I.c., p. 88).


further as part of the legal research track in SPION, in particular in the context of deliverable D6.3 ('Liability and accountability of actors in social networking sites').

6.4.4 (Voluntary) Codes of Conduct

Article 16 of the E-Commerce Directive encourages drawing up codes of conduct at the Community level.\(^{35}\) This provision seeks to help remove barriers to the development of cross-border services within the EU. These codes should contribute to the proper implementation of the general obligations imposed by the E-commerce Directive, and would apply to particular trades or professional or consumer associations and organizations (article 16(1)). The nature of the codes is voluntary and the interested parties are free to decide whether they adhere to them.\(^{36}\) Such codes can be very useful to help ensure compliance with other rules that apply to particular professions, trades and organizations.

6.5 Conclusion

Currently there is no specific legislation targeted towards social networking services as such. The privacy and data protection issues that arise in this context must therefore be dealt with by means of the existing national and European privacy and data protection legislation. Many privacy issues arise in the context of SNS, and therefore it is important to understand and delineate the relevant legal and regulatory instruments that exists today. The preceding sections have briefly elaborated upon the key regulatory instruments in this field. As indicated throughout this chapter, there are still a number of important issues which merit further research and elucidation. Such issues include (but are not limited to) the definition of actors and roles under the respective regulatory instruments, the requirements for informed consent under data protection and e-privacy regulations and the use of default settings.

Another key question in the context of social networks is whether the provider of the social networking platform may be held responsible for information that is exchanged on the social networking site or whether all the responsibility lies with its users. This major point of controversy has important implications when it is translated into legal liability. Notwithstanding the fact that there is a general demand for free speech on social networking sites, there are some limitations posed by interests such as the protection of minors, the prohibition of hate speech etc. The monitoring of information in social networking sites is similarly a major and very complex issue.

As already indicated, one of the main objectives of the legal research track within SPION will be to analyse the respective qualifications, rights and obligations of the different types of actors involved in online social networks. In doing so, we hope to not only clarify the existing regulatory framework, but also to identify areas in which the current definition of actors, roles and responsibilities could be improved.

\(^{36}\) Recital (49) E-Commerce Directive.
7 Access Control and Accountability Aspects (DIS-TRINET)

7.1 Access Control Models

Access Control is one of the oldest security concerns, and has been studied intensively since the early seventies. Access control models (ACMs) are essential to maintain security and privacy by mitigating the risks of unauthorized access. Access control mechanisms make decisions on whether or not a request for an action should be honored or denied by referencing the policies on who is authorized or prohibited to perform which specific actions on which objects. The underlying mechanisms differ according to type, infrastructure and context of the system, type of sharing and relations among users and the nature of data objects; all of these factors contribute to different degrees of complexity. Many diverse access control models exist, some models extend and evolve from previous ones with a higher complexity. The increasing complexity doesn't emerge from deficiencies in older ACMs, rather it naturally accompanies the continuous changes in systems and privacy concerns and requirements.

7.1.1 Preliminary Notions

An access control model defines the regulations of controlled access within a system. A subject is a system user represented by personal identification data in a profile. A subject creates and posts objects encapsulating data resources. Subjects are referred to as owners in some models to differentiate them from other subjects sharing access privileges on the same object or merely trying to perform an action on an object. Authorization is the process of defining what authority an identity of a user accrues in the system. An access rule or a policy is a high-level rule regulating access permissions. In order for a user to perform an action on an object the user is required to comply with a policy enforced in the system.

In social networks (SNS) users connect to each other creating a social graph, where relationships are denoted by connections or edges between nodes representing users. A path is a sequence of edges connecting two nodes. The distance between two nodes is the number of edges in the shortest path linking the two nodes.

In a social context, the notion of trust is a key attribute of SNs users relationships. Trust measures the degree of one user to rely on the actions and opinions of another user [Carminati et al., 2006]. Trust can be associated with users [Golbeck, 2005] or with users with respect to specific relationships involving them [Carminati et al., 2006]. Reputation is the average of trust ratings for a given user provided by other SN users [Golbeck, 2006, Levien, 2009]. Next we will review classical ACMs from which some SN ACMs emerged and evolved.

7.1.2 Classical ACMs

The most basic form of an ACM is Access Control Lists (ACLs) [Stiegler, 1979]. Each resource has a list mapping users to actions allowed to be performed on this resource. A major limitation is that every user is considered uniquely without any generalizations or groupings of users. Another seminal model on which most recent approaches are based is Lampson’s access control model [Lampson, 1974]. In this model, a reference monitor guards the access to objects, according to the current protection state represented by an access control matrix.

Access control models are classified into three main models in terms of administration [Chinaei and Barker, 2009]:

1. Mandatory ACM (MAC) [Bell and LaPadula, 1973]: is where the system is the central authority that constrains access according to objects and subjects classifications, and where individuals
cannot alter the constraints.

2. Discretionary ACM (DAC) [DC Latham, 1983]: sometimes referred to as identity-based access control (IBAC). In this model individual users can set access control policies themselves. DAC and MAC models are not mutually exclusive and can be jointly applied, as in the Chinese Wall model [Kessler, 1992].

3. Role-based ACM (RBAC) [Sandhu et al., 1996]: This model is an alternative to the previous two models. In RBAC users are assigned specific roles and are authorized to perform certain actions on certain resources based on those roles. RBAC covers ACL limitations mentioned earlier.

Classic policy models cannot be employed easily in collaborative applications. Some models are adapted with enhancements in terms of making policies more dynamic, context-based and supporting protection at varying levels of granularity [Tolone et al., 2005].

7.1.3 Access Control Models for Social Networks

Since the birth of SNs in 1992, the necessity to protect users’ objects from unauthorized access has increased. Several models have been proposed to address collaborative systems, usually focusing more on groupware-like applications. Notable models include Task-based access control [Thomas and Sandhu, 1994, 1998] and Team-based access control [Thomas, 1997]. As a generalization, context-based access control policies [Covington et al., 2001] can take a variety of context information, including time, location, task at hand, and so forth into account.

While such models are more appropriate for social networks, they are still lacking in several respects. Simple ACMs have the advantage of being straightforward, on the other hand, they are not flexible enough. To address these questions, some classical ACMs have been incorporated in SNs with specific aspects added to the underlying mechanisms. Other approaches emerged to fit different requirements of SNs.

In subsequent sections, we will review the state-of-the-art models in SNs context under the main mechanisms categories they belong to.

7.1.3.1 Role Based Access Control Models (RBAC)

In RBAC users are assigned roles for which permissions are associated appropriately. In the role-based models family, many models extend the essential role-based model RBAC96 [Sandhu et al., 1996].

One concern of RBAC is how to assign roles. In some cases a user with a specific role should not be assigned another role, otherwise privacy might be compromised. Sandhu et al. address the problem by proposing the separation of duty constraints [Chen and Sandhu, 1995] for which it is necessary to detect conflicts. Conflict resolution is proposed in the rule-based approach of Schaad [2001] by exploiting constraints and delegation mechanisms.

Tang et al. [2009] extend the decentralized administration role-based model ARBAC97 [Sandhu et al., 1999] to meet social networks requirements. The model looks at a social network as small interconnected communities. The access control module is comprised of a server to manage roles and a client to manage access permissions between users. When a user initiates an access request to an object owner, the server verifies the role relation between the two users. Then the owner checks with the client for available permissions of the accessor’s designated role.

The employment of RBAC for fine-grained access control is limited due to the difficulty of deciding what permissions should be associated with each static role in different contexts where roles are not distinctively separate and undergo continuous changes [bo Shen and Hong, 2006].
7.1.3.2 Attribute-based Access Control Model (ABAC)

ABACs were proposed to address the aforementioned limitations in RBAC. An ABAC incorporates attributes of subjects instead of roles. Access control policies are conditions on what attributes a user should have to gain access to an object. ABAC supports contexts and is flexible enough to integrate with other ACMs; for these reasons, it is widely incorporated into SNs.

Attribute-based encryption (ABE) is used in SNs to enable users to maintain access control policies without having to put their trust in the SNs providers. This model allows users to share keys with other users for encryption/decryption of data objects. The downside is that when a relation between two users is changed, a new key has to be generated and distributed again. Jahid et al. [2011] address this problem in the model EASiER, an Encryption-based Access Control model. The model exploits a proxy to resolve the key distribution problem upon revocation of a user, using the revocation scheme from CP-ABE [Naor and Pinkas, 2001]. The proxy is minimally trusted and cannot provide access to revoked users.

Squicciarini et al. [2007] employ an attribute-based authorizations to satisfy k-anonymity property [Sweeney, 2002]. A policy enforcer discloses access requirements and what credentials are needed from a credential submitter to access a resource. The credentials submitter provides a set of attributes or credentials that are indistinguishable from k other sets of attributes of counterparts. The model is based on a cryptographic protocol to ensure that credentials are not tracked and a submitter cannot be identified.

7.1.3.3 Trust-based Access Control Models

Rule-based models [Didriksen, 1997] apply to MAC models, where policies are rules of specific conditions that restrict access to a requested object. In SNs context, Trust-based access control models are rule-based models that incorporates the notion of trust to condition access to resources.

Carminati et al. propose the first rule-based model for social networks [Carminati et al., 2006]. Policies are composed of constraints on the type, depth and trust level of relationships. The model proposes a decentralized approach to support access control enforcement. Rules enforcement is performed at the client-side, where an accessor proves that it satisfies access conditions to the owner. Relationships authenticity is guaranteed by encoding the relationship information in certificates issued by a server and double signed by the two nodes involved in the relationship.


In some models, authorizations are classified as either positive and negative. The formal language EPAL [Ashley, 2003] allows to define positive and negative authorizations, however it is designed independently from any access control model. The same distinction of authorizations is expressed in Trust-involved Access Control (TAC) proposed by Wang and Sun [2010]. The model supports expressing complex privacy policies based on trust, relationship, access purposes and obligations. Access purposes are classified as intended and prohibited purposes and are structured in a hierarchical tree. Users are allowed to include user-defined restrictions in the access rules from which conflicts might arise.

7.1.3.4 Ontology-based Access Control Models

Semantic web technologies provide more dynamic and fine-grained access control policies using ontology representation [Carminati et al., 2006]. Ontology-based social network access control models have emerged to protect semantic-rich information. Most of these models exploit rule-based ACM.

Web Ontology Language (OWL) is adopted to describe user profiles, resources and relations in terms of ontologies. The Friend of a Friend (FOAF) ontology [Brickley and Miller, 2007] is a notion to
represent users and their relationships or friendships. This notion improves information access and dissemination in social networks.

Kruk [2004] quantify friendship relations in FOAF-Realm, an ontology-based model using the notion FOAF. The access control rules constrain friends’ access to resources based on the maximum friendship distance or group similarities distance in a SNs.

D-FOAF [Kruk et al., 2006] is a distributed identity management system for social networks based on the notion of FOAF. In this model relationships are associated with a trust level quantifying the level of friendship. This model, however and in contrast with other models, does not cover the composition of multiple policies associated with the same resource.

Similarly to the previous models, Villegas et al. exploit the notion of FOAF in a model that assumes all friendships are equal [Villegas et al., 2008]. In this model, the decision making approach learns from a pool of users’ control policies decisions and actions within the neighborhood of a user. Users are categorized within three categories according to their credibility and actions in the social network. As a result a user would fall into the trusted zone, marginally trusted zone or rejected zone. Moreover, the model facilitates a method of tracking for reposted objects and leakage.

Carminati et al. [2009a] propose a rule-based model that defines authorization, administration and filtering policies based on trust relationships. The model encodes user profiles, relationships, objects and actions in an ontology-based hierarchy. The model offers a high degree of flexibility in terms of defining hierarchies of user types, relations, objects and actions as well as adding new ontologies to the knowledge base. The hierarchical structure enables flexible rules propagation on concepts.

OSNAC, an Ontology-Based Access Control Model for Social Networking Systems [Masoumzadeh and Joshi, 2010], is more detailed than the previously discussed model [Carminati et al., 2009a]. Masoumzadeh et al. focus in this model on relationship protection. For this purpose, the concept annotation is incorporated in access rules, which is a type of a digital object representing a relationship between an object and a subject. Access rules define negative or positive authorization, which can be simple or advanced when composed by means of disjunctions or conjunctions yielding multiple authorization rules and delegation of authority from one user to another.

7.1.3.5 Relation-Based Access Control Model (ReBAC)

ReBAC is a term coined by [Gates, 2007] to refer to the paradigm of access control models based on interpersonal relationships. ReBAC protects the primitive relationships (e.g., friend) in policies composed of primitive relationships (e.g., friend-of-friend) [Fong, 2011]. A relational policy does not base authorization on users identities. Instead, it only consults the topological information in the social graph.

Anwar and Fong [2010] formalize Facebook-style access control model and a general family of Facebook-style social networks. A user u can define four types of policies to constrain actions of producing a search listing of u, traverse the relationships of u, communication with u and access to objects of u. An essential part of this relationship-based models is the local topological policies, which capture the particular relationship of an owner and an accessor. In a local policy the removal of an edge can alter an authorization decision. A global policy is a policy that is not local and the removal of an edge cannot alter its authorization decision. The model provides an alternative characterization of policy predicate in terms of monotonicity. A monotonic policy never denies access when adding an edge to the social graph, and never disables access deleting an edge. Thus, monotonic policies grant access to closely related users. In contrast, anti-monotonic policies grant access to distant users (e.g., a teenager disclosing photos for strangers but not to her parents).

The ReBAC model of Fong [2011] looks at the social network as a collection of assertions of users relationships that are context-dependent. To facilitate sharing of relationships across context, the contexts are captured in a tree-shaped hierarchy. Two relationships in two different branches cannot be activated at the same time, which is analogous to separation of duty in RBAC [Chen and Sandhu, 1995]. Another work of Fong et al. investigates the representational completeness of this model.
and shows that it is incomplete [Fong and Siahaan, 2011]. Monotonic policies [Fong et al., 2009] are introduced to make the model complete to be capable of expressing a number of Facebook-style policies and also applicable in decentralized networks [Fong and Siahaan, 2011].

Giunchiglia et al. [2008] propose a relation-based access model (ReiBAC) exploiting many features of other relation-based models. The key difference is that permissions of actions are expressed as relationships between classes of subjects and classes of objects. Access rules are the instantiations of these permissions. The model has many flexible features of defining ground and general rules on the n-arity relationships of subjects and objects, which guarantees that the rules will evolve with the changing dynamics of the system.

### 7.1.3.6 Shared Objects Protection

One feature of access control in SNs that is not addressed well in most of the models is support for multiple owners of relationship information. Different approaches and models have been investigated to address shared objects protection. Policies composition through mathematical operations is one of the approaches [Bonatti et al., 2002]. The relation-based model of Anwar and Fong [2010] defines policy combinators for representing complex policies from primitive ones of different users. Several competing approaches based on, for instance, metapolicies or policy combiner functions are surveyed in [De Capitani Di Vimercati et al., 2007]. Another approach focuses on turning policies into first-class objects that can be composed and manipulated at run-time. Polymer [Bauer et al., 2005] is the most representative system.

The Voting-Based model of Squicciarini et al. [2010] focuses on the problem of collective enforcement of privacy policies on shared data objects in social networks. By employing the Clarke-Tax voting protocol [Clarke, 1971], users are offered incentives and credits to promote truthfulness. The credits granted to an owner are proportional to the number of grants given by this user to other users co-owners.

In the same context, Wang and fung Leung [2004] propose a secure version of Clarke-Tax scheme using encryption techniques, implying a level of sophistication which may not be appropriate for most social network users.

### 7.1.3.7 Enforcement of Access Control

With the advent of applications that support partially trusted plugins, and multi-user applications that need to control access to application-managed resources (social network services are a prime example of such applications), application-specific access control policies have gained in importance. Application-specific policies can be modeled as security automata [Schneider, 2000]. A security automaton describes what traces of security-relevant events an application is allowed to generate, e.g., system calls, API method calls, or even individual machine instructions [Erlingsson and Schneider, 2000].

An interesting question is how to enforce such application-level policies. Several approaches have been proposed, based on run-time monitoring [Castrucci et al., 2008], program rewriting [Erlingsson and Irm, 2000], or the use of aspect orientation [Verhanneman et al., 2005]. Enforcement mechanisms are critical as it might be possible to enforce non-safety policies [Ligatti et al., 2005]. The fundamental limitations of each of the existing enforcement mechanisms have been investigated [Hamlen et al., 2006].

In social networks it is fundamental to securely enforce control policies. Carminati et al. [2009b] extend their earlier work [Carminati et al., 2006] by introducing semi-decentralized enforcement of access control at the owner’s node level. To gain access, an accessor generates a proof with a set of assertion for all access conditions of at least one access rule of the resource.
7.1.3.8 Evolving Access Control Model

Within the very broad requirements of privacy, data sharing and SNs structures, it is hard to employ one access control model to satisfy all requirements. Other privacy-preserving proposals emerge to guard access policies. Crescenzo and Lipton [2009] propose to implement an extra layer in a social network to guard privacy settings of users over time. The objective is to maximize the capability of sharing objects between users by automatically manipulating the visibility setting of an object. A data object is not considered to be sensitive on its own, rather the set of user objects may at some time become of a sensitive nature and hence comprises the privacy of the user. User privacy is maintained if at least one of the sensitive subset objects is set to be private. Evolving access control is a feature a user can trust to maintain her privacy settings automatically without extra actions or user involvement.

7.2 Access control and information flow for social network applications

An important contributor to the success of social networks is their support for third-party applications. Such applications provide social-enhanced features or functionality (e.g. quizzes) or social games. With such applications, the code provider is typically a third stakeholder, next to the social network provider and the end users. Since these applications are social-aware, they need access to private information of the user to reach their full potential. As a consequence, there is a strong need for security and privacy controls, and enforcing such controls is complex because of the many stakeholders involved.

In SPION, we will focus on this problem in the context of web-based online social networks (the majority of current social networks are web-based). In this context, third-party applications are typically developed in scripting languages like JavaScript, and access control or information flow control is addressed at the level of the scripting language. We discuss the state-of-the-art in providing access control and information flow control on scripts, and the challenges we will address in this project.

7.2.1 Script access control and isolation

Script isolation techniques use language based techniques to isolate a script (say one application) from other scripts (say other applications, or social network code itself). The general approach is restricting JavaScript to a subset, which adheres to the object-capability security model. This security model is based on the fact that separated objects have no capabilities and can only achieve capabilities on an object if they are handed a reference to that object. For example, if an object in the language has no reference to the Image object, it can not construct new images. By giving it a reference to the Image object, it obtains the capability to create images.

The three techniques presented here, i.e. ADsafe, Facebook JavaScript and Caja, follow this object-capability security model, thus achieving script separation, regardless of domain. Separation for DOM elements and built-in script objects is achieved using subset restrictions and run-time control over specific operations, such as DOM access. The isolated modules can interact using explicitly shared objects, which offer confidentiality and integrity. Mutual authentication can be implemented if desired.

7.2.1.1 ADsafe

The problem of script isolation first surfaced with the inclusion of JavaScript advertisements in web pages. The ADsafe subset [Crockford, 2009] is aimed at putting guest code, such as advertisements, in a web page, without suffering security consequences. This is achieved by restricting scripts to a
safe subset of JavaScript. Safe interaction with their environment, such as the DOM tree, is possible using a provided ADSAFe object.

ADsafe is not an active protection mechanism, but is enforced using a static code verification tool. This tool can determine whether a script adheres to the ADsafe subset or not, but will not actively rewrite code. Next to preventing access to the global object or well-known insecure language features, such as eval or with, ADsafe also prohibits the use of this, since it has subtle properties that can be used to obtain a reference to the global object.

In recent research on the security of JavaScript subsets, specific issues with ADsafe have been discovered [Maffeis and Taly, 2009]. These issues are minor design oversights, which do not break the fundamental model of the language. Continued formal verification is needed to prove that the ADsafe language fully adheres to the object-capability security model.

7.2.1.2 Facebook JavaScript

Facebook, the leading social networking site, supports safe incorporation of third-party applications through Facebook JavaScript (FBJS) [Facebook Developer Wiki, 2010], which is a secure JavaScript subset. FBJS is an active protection mechanism, which applies a rewriting process to normal JavaScript. This rewriting process includes rewriting variable and function names to a unique namespace, as well as defining Facebook-specific DOM objects, which do not implement insecure features. Remote communication is available through an Ajax object, which uses a server-side proxy to retrieve cross-domain content. More importantly, this retrieved content is rewritten to FBJS, to ensure continuous protection.

The major advantage of the approach taken by Facebook is the active protection mechanism, which allows the dynamic addition of content. This is particularly useful in mashup applications. The disadvantage however is that every request needs to go through the Facebook servers, which might not be feasible for each integrator.

Recent research on the security of JavaScript subsets has also identified issues with FBJS [Maffeis and Taly, 2009]. These issues do not have an impact on the fundamental model of the language, and can be further eliminated using strong formal models.

7.2.1.3 Caja

Caja [Miller et al., 2008], a safe JavaScript subset designed by Google, takes a similar approach to FBJS. It analyzes JavaScript to detect subset violations and it rewrites the code to create isolated modules, as well as to mediate DOM access. Caja is a fairly flexible subset, since it allows the use of this, albeit in a limited way. Caja does more than subsetting JavaScript, it also introduces a new feature: frozen objects. Frozen objects can not be changed, which makes them ideal for information sharing between components. Objects in the default global environment are automatically frozen.

An advantage of the way Caja is introduced is that it is aimed at supporting existing scripts, with some exceptions such as eval or with. This allows a gradual transition towards the Caja subset. Underneath, a second subset is defined, named Cajita. Cajita can be considered “Caja without this”, since this is considered a dangerous and unnecessary language feature. Cajita is meant to be the subset for writing new applications, while Caja is meant to be backwards compatible with current applications. Similar to FBJS, a server-side rewrite process ensures continuous protection of dynamic code.

Recent research on the security of JavaScript subsets has been able to prove that a subset based on Caja is capability safe [Maffeis et al., 2010]. This important result shows that a JavaScript subset can adhere to an object-capability security model, and can thus be used to achieve the proposed security requirements.

Caja has been used by several OpenSocial gadget integrators, such as Yahoo! Application Platform, Shindig, iGoogle, Code Wiki and Orkut.
7.2.2 Fine-grained access control and information flow control

The isolation techniques discussed above provide secure separation between components, but do not provide support for configurable policies that allow some form of sharing or flow of information. Security and privacy policies are essentially embedded in the code. Several research prototypes provide finer-grained control over security and privacy. In this section, we present three approaches which are aimed at providing fine-grained control over script behavior. The first approach focuses on enforcing a policy on JavaScript code, either with or without specific browser-side support. A second approach mediates access to specific objects, thus enabling the enforcement of a security policy. A third approach is aimed at enabling information flow control for JavaScript.

7.2.2.1 Policy Enforcement Techniques for JavaScript

ConScript enables the specification and enforcement of fine-grained security policies for JavaScript in the browser [Livshits and Meyerovich, 2009]. Such policies can be used to control the script behavior, such as disallowing calls to certain functions (e.g. eval), or preventing the script from accessing cookies. To ease the task of writing policies, ConScript supports automatic policy generation through static analysis of server-side code or run-time analysis of client-side code. Technically, ConScript supports the enforcement of security advice within the JavaScript engine. The advantage of this approach is its effectiveness, since all indirections and ambiguities, such as different paths to the same function, are eliminated inside the JavaScript engine.

Self-protecting JavaScript [Phung et al., 2009] provides similar security features, but does not require specific support within the browser. Policy enforcement is achieved by wrapping security-sensitive JavaScript operations before normal script execution. As a consequence of not depending on browser-support, this technique faces several challenges, such as covering all access paths to a specific function or preventing wrapped operations to be restored by the malicious script. Several of these issues have been addressed in a follow-up paper [Magazinius et al., 2010b], while others will be resolved in future research.

7.2.2.2 Mediating Access to Objects

Object views offer a fine-grained control over shared objects in a JavaScript environment [Meyerovich et al., 2010]. By creating and sharing a view of an object, instead of the full object, all calls to the object pass through the view, where a security policy can be enforced. An example application scenario is a document sharing policy, where the HTML document is a shared object. A view of this document can enforce the security policy, where a component can have read-only access to the entire DOM tree, and only gets write access to within its boundaries.

AdJail [Ter Louw et al., 2010] offers a technique to mediate access to advertisements, which are embedded as a DOM object. The technique was motivated by an incident on Facebook where advertisements on Facebook pages used private images (such as pictures of friends) as part of the advertisement. AdJail addresses this by executing ad scripts separately in a sandboxed environment, where they can cause no harm. In order to preserve the user experience and to enable ad-specific services, such as compatibility with ad network targeting algorithms or billing operations, a mediation technique selectively forwards specific operations, such as visualizing content and forwarding of user interface events, between the sandbox and hosting page.

7.2.2.3 Information Flow Control for JavaScript

A promising technique for addressing privacy in online social network applications is information flow control. A program is information flow secure if secret inputs to the program do not influence (flow into) public outputs. An information flow policy defines which inputs and outputs are considered secret or public. More generally, a policy has a partially ordered set (poset) of security or confidentiality
levels and labels input and output channels with such levels. The program is information flow secure if information only flows from inputs labeled $l_i$ to outputs labeled $l_o$ for $l_i \leq l_o$. In other words: information only flows upward, toward more confidential levels.

Information flow security has been studied intensely for several decades, and a wide variety of enforcement mechanisms has been proposed. Sabelfeld and Myers [2003] provide an extensive survey of static enforcement methods, and Guernic [2007] surveys dynamic methods. Several authors have already investigated the use of secure information flow techniques in the context of a browser, for instance to secure mashup composition [Magazinius et al., 2010a, Li et al., 2010], or to prevent private information to flow to advertisement providers [Ter Louw et al., 2010].

Devriese and Piessens [2010] introduced a novel secure multi-execution technique that proposes to execute a program multiple times, once for each security level, using special rules for I/O operations. The main advantage of this approach is that it is proved to be sound and precise, but it could have a substantial performance impact.

Very recently, Bohannon et al. [2009] proposed to replace the same-origin-policy in web browsers with information flow policies. They define the notion of reactive non-interference, an adaptation of the classic notion of information flow security to reactive systems, systems that perform asynchronous I/O such as web browsers. With such an integration in the browser, information flow control becomes potentially useful to web based online social network applications.

### 7.3 Accountability and Audit

Very often the more granular the privacy settings are the more difficult it is for users to configure. The users lack the feedback about what kind of threats exist in how data is being viewed; even when feedback information and awareness are available (see privacy feedback and awareness tools chapter) they are not well addressed in user defined policy settings in SNs [Spiekermann et al., 2001, Govani and Pashley, 2005, Cain et al., 2009]. This requires other mechanisms alongside access control to ensure security and privacy in general. Accountability and audit are key aspects to complement the correct usage and control of data as stated in the Bill of Rights, Article 7, “Data Control” [Gagnier, 2011].

Accountability is an important aspect in online life [Joan et al., 2011]. Accountability enables determining if an access policy has been violated and hold the violating entity responsible [Lampson, 2005]. It is a property of a system to verify actions performed on data and identify the responsible entity for misconduct [Castelluccia et al., 2011]. Audit is a strongly connected notion to accountability. In order to hold entities accountable for their actions a system has to continuously checks and assesses privacy compliance.

Weitzner et al. [2008] state that information accountability emerges by combining three functionalities in information systems: policy-aware transaction audit logs, policy language framework, and accountability reasoning tools to assist users using specific pieces of data in a purpose-based manner.

### 7.3.1 Accountability Applications

Some accountability applications are already available and distributed in the web and could be integrated in the accountability and audit tools, e.g., terrorist tracking applications [Lunt, September 29, 2003]. Data-Purpose Algebra could be employed for reasoning about heterogeneous policy rules between different applications [Hanson et al., 2007].

Various identity manipulations threat against satisfying identification and accountability in a system [Castelluccia et al., 2011]. Identification and accountability are implicitly addressed in the reputation-based ACM in social networks [Ali et al., 2007]. A user’s behavior is tracked over time and continuously contributes to a personal reputation value. In case of an undesired action, the user will be held
accountable automatically by the change of the reputation value.

In distributed decentralized systems, accountability is harder to implement [Druschel, 2008], especially when nodes are administered by third-parties. Achieving accountability requires each end-point to record all actions, such as message transmissions, and cryptographically link them to their source [Haebeleren and Kouznetsov, 2007]. All logs have to be secured for real time or future policy accountability assessment. These techniques, called pretty good packet authentication [Haebeleren and Rodrigues, 2008], are used for tracking failure, rather than privacy violations, but they have been applied to non-repudiation and denial of service attacks.

7.3.2 Audit Mechanisms

Simpson [2008] recommends the integration of audit functionality into SNs to analyze violations and anomalous actions of authorized users [Sandhu and Samarati, 1996]. In order to perform audit, relevant users’ decision information need to be kept during the running of the system until audit time, for example, by keeping secure logs [Schneier and Kelsey, 1999].

In the case of individuals managing a self-organized social network, they should be provided with solutions to determine the usage of their personal data. Rodriguez et al. propose digital rights management (DRM) schemes [Rodriguez et al., 2009], where users can flexibly determine constraints and conditions on the usage of data in the absence of centralized access control. Access rights can be expressed in a Rights Expression Language composed in a signed encrypted license. Accessed data is trackable by means of chain of licenses. One of the techniques to hold agents accountable is to run post factum analysis on the chain of licenses to detect violations.

Specific audit mechanisms have been proposed based on logic that could be integrated in SNs ACMs when both ACMs and DRMs fail to fully protect private data. Cederquist et al. [2005] propose an audit logic for expressing agent accountability with means of proof obligation to determine which policy an agent needs to justify actions. Vaughan et al. [2008] use formal authorization logics to specify access control policies. In this approach, proofs are used in the logic as evidence that the access-control decision has been made in accordance with the policy, as is the case in, for example, proof-carrying authorization [Appel and Felten, 1999]. Significantly, auditing using proofs as log entries can both reduce the trusted computing base and detect flaws in complex authorization policies. The programming languages Aura and Fable incorporate these ideas, which are checked using dependent types and inline reference monitoring [Jia et al., 2008, Swamy et al., 2008].

Sh@re is a rather different approach proposed by Gutierrez et al. [2009]. This approach is a negotiated audit and an attribute-based ACM [Yuan and Tong, 2005] that gives feedback to users on how data is being accessed. The system has three levels: no audit, anonymous audit and complete audit, to grant access for subjects of compatible audit level. If levels do not match the audit level is negotiated with the accessor.

7.4 Challenges

With data protection challenges arising everyday in social networks, different types of challenges need to be addressed by access control models. One of the major concerns is the gap between a SN users’ expectations and actual dissemination of their data objects. Moreover, users might not be able to achieve the privacy control level they want via composing various complex access control policies. This has to be addressed from two different angles. The first is enhancing access control models and their policies composition to be more flexible and easy to manipulate by different users. The second is the integration of privacy awareness and feedback tools to enable users to better assess and understand the privacy consequences of their access control policy decisions.

Access control models have specific approaches to representing data objects and users. RBAC and ABAC categorize users according to roles and attributes, which might not always be applicable in a social network context, especially considering dynamic groups and interrelationships among
users. ReBACs base access rules on dynamic relationships. In this approach it is a challenge to protect relationship information when needed during access control enforcement. One of the most prominent challenges of access control enforcement is when relationships and communities evolve through time while the policies do not, and consequently privacy is not preserved in the manner that the user initially desired. In addition, when some objects are shared among more than one user, there are many challenges in striking a good balance between all stakeholders’ privacy preferences.

Online social networks are increasingly deploying third-party applications that can access user’s private data. For instance, Facebook reports [Facebook, 2011] that Facebook users install 20 million(!) applications per day. While basic mechanisms for isolating applications are well understood (and can handle for instance advertisements in Javascript), these mechanisms fall short for social-enabled applications, that need more fine-grained control both over the access that these applications have to informations, as well as over how the information is used after access has been granted. In this project, we will develop novel security techniques for application access control and information flow control that can address these challenges.

Accountability raises the greater challenge of tracking users behavior, while other approaches and research focus on minimizing the tracking of users. Tracking users to hold them accountable for their actions poses transparency and anonymity challenges that conflict with access control models e.g. k-anonymity Attribute-Based Access Control [Squicciarini et al., 2007]. Relevant examples are cyber-bullying or whenever a crime is perpetrated. These cases help necessitate the need for accountability of users as well as accountability of service providers.
8 Confidentiality Aspects (COSIC)

8.1 Introduction

Since the appearance and popularization of online social networks (OSN), researchers from different fields have been trying to devise a means to keep confidential the contents and performed activities of the users of a social network. This comes as a natural response for privacy-aware users and organizations demanding better privacy guarantees in OSN [Hoffman, 2010]. In the SPION project, researchers from different communities are involved in providing users with a privacy-preserving way of benefiting from social networking services, e.g., teachers, legal scholars, psychologists or engineers. This means that the problem of privacy preservation in OSN are being tackled from various perspectives.

In the case of the technical community, in the last years researchers in areas like cryptography, graph theory and networking, among others, have been specially active dealing with the privacy issues arising from social networking. We shortly review the state-of-the-art on privacy preservation for OSN. We cover works performed in different technical areas including social graph anonymization, content data protection and traffic data protection.

We acknowledge that this is a very brief therefore limited review and refer the reader to other surveys in privacy-preserving social networking, e.g., [Beye et al., 2010, Pallis et al., 2011, Zhang et al., 2010, Zhou et al., 2008], for more details.

8.2 Graph Anonymization

OSN data has become a prized resource in multiple disciplines ranging from sociology, psychology, economy and law to computer science and engineering. OSN data help the research community to understand the users’ behavior, build models based on observed patterns, evaluate the propagation of information and even predict future events such as who is likely to become friends on a social network. However, at the same time the disclosure of these data represents a serious concern as it leads to the subsequent disclosure of the social network users’ sensitive personal information, which may include the users’ circle of friends, their membership to certain groups, the very fact that they belong to the social network and/or other attributes. In order to release OSNs’ data without invading the users’ privacy, different techniques for social graph anonymization has been proposed. The insight behind this approach is that it is not possible to undermine a user’s privacy as long as the user remains unidentifiable. Hence, the goal of anonymization is to prevent the re-identification of the users the social network in the released (anonymized) graph, i.e., the mapping between a node in the released dataset and an identifiable individual in the external dataset is concealed.

The simplest and most straightforward way of social graph anonymization is to substitute the users’ names with arbitrary unique pseudonyms. However, as noted by Backstrom et al. [2007], this rather simple way of anonymization, the so-called naïve anonymization [Hay et al., 2008], is not a guarantee of anonymity against an adversary that makes use of external information about the social network. Indeed, nodes in a social network can be re-identified based on contextual information such as their degree (the amount of friends they have in the social network) and the degree of their friends, i.e., the network structure itself.

Following these works, two main strategies beyond naïve anonymization have been proposed: k-anonymity and clustering or generalization techniques. The former are based on the notion proposed in [Samarati and Sweeney, 1998] for the anonymization of database records: a database is k-anonymous if every record is indistinguishable from k any other records. The latter is based on the simplification of the graph by means of “super-nodes”. A super-node represents a certain group of nodes, revealing not the topology of this group but only the number of nodes that were grouped.
and the edges between them, as well as the edges that connect it with other super-nodes. In this section we provide a review\(^1\) of these techniques, amongst others. All these proposals roughly present the following structure: First, they consider a certain adversarial model, which influences the effectiveness of their solution, i.e., considering weak (powerful) adversaries leads to frail (robust) anonymization techniques. Then, an anonymizing technique is proposed, based on some privacy notion, such as \(k\)-anonymity or randomization. Last, the technique is evaluated to assess the utility of the anonymized dataset and/or the degree of privacy provided by the proposed technique.

Da-Wei et al. propose in [Wang et al., 2006] a generalization of the privacy protection problem of tabulated data in a social network format. Their work aims to assess the privacy risk related to the release of social graph data. Singh and Zhan [Singh and Zhan, 2007] define two specific privacy breaches and a metric to measure the degree of privacy in anonymized social graphs. They define two local neighborhood breaches, node identity, i.e., a user is re-identified in the anonymized graph and edge inference, i.e., an adversary is able to determine whether or not two users are linked in the social graph. The metric, topological anonymity, based on the defined breaches and the \(k\)-anonymity notion is later reviewed in Sect. 8.6.

Hay et al. propose in [Hay et al., 2008, 2010] a new concept of anonymization of social graphs based on the notion of \(k\)-anonymity [Samarati and Sweeney, 1998]. As proposed in [Hay et al., 2008], an anonymized social graph satisfies \(k\)-anonymity if, for a given feature of a given node, (e.g., the degree of the node, the degree of a node's neighbors, etc.) there are in the graph at least \(k\) other nodes with such a feature. As a result of this, the best an adversary with external knowledge about the social graph can do is to identify a set of \(k\) candidate nodes in the published dataset who might correspond to an identifiable individual.

Zheleva and Getoor [2007] address the problem of link re-identification (i.e., inferring sensitive relationships from a social graph) through five different clustering techniques based on edge addition/removal operations.

Many other researchers formulate proposals based on the notion of \(k\)-anonymity. Liu and Terzi [Liu and Terzi, 2008] introduce the concept of a \(k\)-degree anonymous graph, i.e., a graph where for every node \(u\) there are at least other \(k - 1\) nodes with the same degree as \(u\). They also propose an anonymization strategy to build such graphs. Zhou and Pei [2008] go a step further considering an adversary that has knowledge about the neighbors of the target node. They propose a strategy to create a \(k\)-anonymous graph based on clustering and edge-addition techniques that takes into account the neighborhood of a node. Campan and Truta [2008] also propose a clustering method technique called edge generalization that tries to deliver a \(k\)-anonymity compliant graph, taking into account both the attributes and the neighborhood of a node.

With a different approach, Ying and Wu [Ying and Wu, 2008] propose two anonymization techniques that aim to minimize the disruption of the spectral properties of the original graph. Cormode et al. [Cormode et al., 2008] introduce a different and special technique called \((k, l)\)-groupings which can be used to anonymize bipartite graph data. However, no modification of the graph is done, so in the end this technique provides no more protection than naive anonymization against a structural attack. Besides, the applicability of this technique is restricted to bipartite graphs. This work is further extended by Bhagat et al [Bhagat et al., 2009].

Zhou and Pei [2011] go beyond \(k\)-anonymity and borrow the concept of \(l\)-diversity from the anonymity for relational databases literature [Machanavajjhala et al., 2007]. They propose a technique based on edge addition/deletion for the purpose of achieving a \(k\)-anonymous and \(l\)-diverse graph.

Recently, researchers have started studying the temporal evolution of social networks [Leskovec et al., 2008, Mislove et al., 2008, Viswanath et al., 2009], i.e., how the topology of the social network changes over time, and the weighted social graph [Kumpula et al., 2009, Toivonen et al., 2007], i.e., a social graph in which each edge is given a value that represents a notion such as trust, volume of information exchanged, etc. As a result current attempts on social graph anonymization techniques deal with the anonymization of, on the one hand, social networks at different stages of their evolution

\(^1\) Other reviews can be found in [Liu et al., 2008] by Liu et al., and [Zhou et al., 2008], by Zhou et al.
and, on the other hand, the edges’ weights.

Zou et al. [2009] present a comprehensive review of previous anonymization techniques, criticising that either most of them have been designed to either resist only one type of attack or either deliver an anonymized graph which suffers from major utility loss. They propose a new technique following the approach in [Hay et al., 2010] of automorphic equivalency: for every vertex in the released graph, there are always \( k - 1 \) other vertices with identical structural properties, no matter the amount of structural information the adversary has. Furthermore, they implement the mechanism (vertex ID generalization) to support dynamic network disclosure, i.e., to prevent an adversary from performing inferences from multiple releases of an evolving social network. Wu et al. [2010] also express their criticism regarding previous anonymization techniques for not being able to cope with resourceful adversaries. The anonymization technique they propose is conceptually similar to the one proposed by Zou et al.

Bhagat et al. [2010b,a] argue that releasing independently anonymized datasets of the same (evolving) social network might leak information if the adversary compares the different data releases. For this purpose, they extend with link prediction (i.e., a forecast of topological changes on the social network) mechanisms proposed in previous work [Bhagat et al., 2009] to address the threat of multiple data releases of a social network.

Liu et al. [2009] develop two strategies for the concealment of the weight of the edges in a social network while maintaining fundamental graph properties such as shortest path lengths. However, the adversary model they consider is unclear thus the privacy guarantees these mechanisms provide are uncertain.

Das et al. [2010b,a] present Anónimos, a linear programming based technique for the anonymization of weighted graphs. Their goals are to conceal the magnitude of each edge, and the indistinguishability when compared with other edges, including order relation of all the edges in the graph. Their technique is based on \( k \)-anonymity understood as follows: an edge of a node \( u \) with weight \( w \) is \( k \)-anonymous if there are other \( k - 1 \) edges in \( u \) with weight \( w \pm \mu \). At the same time, they claim to preserve fundamental properties of the weighted graph, such as the shortest path between nodes.

8.2.1 De-Anonymization Attacks

We began this section referring to the series of attacks performed by Backstrom et al. [2007] that demonstrated that naïve anonymization of social network data does not prevent the re-identification of the nodes. Narayanan and Shmatikov [2009] perform a de-anonymization attack over anonymized social network graphs based on publicly available auxiliary information, i.e., they re-identify the nodes of an anonymized social graph resorting to the publicly available social graph of another social network. They demonstrate that such attacks are feasible and very effective even when the degree of similarity between the two graphs is low, meaning that their attack succeeds even when the anonymization techniques heavily distort the original graph, rendering useless any conceived defenses known so far. Narayanan et al. [2011] further improve the attack by combining the use of auxiliary information with link prediction. Similarly, Wondracek et al. [2010] exploit group membership information that is publicly available on social networking sites. They demonstrate that knowing the groups to which a user belongs is often enough to re-identify a user in an anonymized social network with high certainty.

8.3 Content Protection

A promising line of research focuses on the protection of the users’ uploaded contents in OSNs. Initial solutions aimed to provide access control to the users’ resources independently from the specific OSN platforms used. The most recent solutions aim to provide robust and fine grained access control, preventing even the service provider from accessing the uploaded content data. The current section provides a review of these solutions.
Tootoonchian et al. proposed ImageLock [Tootoonchian et al., 2007], a Firefox extension that provides users with access control over their pictures in SNS. When a user uploads a picture to an SNS, ImageLock intercepts the picture and replaces it with a fake picture, uploading the latter to the SNS. The genuine picture is instead stored in a trusted server, i.e., a server that the user confides in, e.g., her own server. When authorized users request to see the real picture, ImageLock brings it from the trusted server to replace the fake picture. Later Tootoonchian et al. proposed Lockr [Tootoonchian et al., 2008, 2009]. Similarly to what ImageLock does with pictures, Lockr separates any content uploaded by the users from the OSN: Instead of uploading the content to the OSN, Lockr uploads it to a trusted server, together with an access control list (ACL) that specifies which social relationship a user must have with the content/ACL creator to be able to access the item. Therefore, in order to grant access to their contents, users issue social attestations, i.e., a public-key-based cryptographic credential which specifies the kind of relationship a user has with another. In short, as it is said in [Tootoonchian et al., 2008], "Lockr allows users to express access control policies based on social relationships". Yet, the main drawback of Lockr is that users still have to trust the third party server.

Guha et al. propose NOYB [Guha et al., 2008] (None Of Your Business), a tool to encrypt the users' information while preventing the adversary (i.e., the SNO) from detecting the users that do so. Instead of using traditional encryption or steganography, NOYB uses a special encryption scheme (substitution encryption). NOYB partitions the users' uploaded text into atoms. These atoms are then pseudorandomly substituted by other NOYB's users atoms. Each substitution is performed based on a dictionary, which is indexed by the encrypted genuine atoms. When a user wants to retrieve the genuine atom, she makes a reversed look up in the dictionary: the dictionary returns the index (the encrypted genuine atom) associated with the fake atom in the dictionary. Therefore, by decrypting that index (if the user has the key to decrypt it) the user obtains the genuine atom.

Lucas and Borisov propose flyByNight [Lucas and Borisov, 2009], an architecture for content protection based on the encryption of any data uploaded by the user to the service provider. However, the keys are managed by a platform-dependent application (e.g., in Facebook, a Facebook application) thus stored on the server side. This renders flyByNight insecure against active attacks performed by the service provider, i.e., the service provider can access the cryptographic keys thus decrypt all sensitive information stored by the flyByNight users.

Luo et al. propose FaceCloak [Luo et al., 2009], an application that substitutes the users' personal information in the SNS with fake data, storing the genuine information, encrypted, on a third party server. The fake data is in turn used as an index to the encrypted genuine information stored on the third party server. Therefore, through the use of different cryptographic keys users are able to selectively control who may access their data. Luo et al. suggest retrieving the fake information from dictionaries or sites like Wikipedia so that the fake data cannot be easily noticed by the service provider.

Baden et al. propose Persona [Baden et al., 2009], a cryptography-based mechanism to allow users to enforce access control in online social networks. Persona uses attribute-based encryption [Goyal et al., 2006] (ABE) and traditional cryptography to implement an advanced group encryption scheme [Kiayias et al., 2007]. ABE is based on attributes used to classify the contacts of a user in groups such as "relative", "co-worker" or "neighbor". ABE allows the generation of keys based on logical expressions of these attributes, e.g., a key for { "relative" AND "neighbor" } and a different key for { "relative" OR "neighbor" }. In short, users of Persona are able to define fine-grained access control lists without relying on any third party such as the SNO.

Conti et al. propose in [Conti et al., 2011] to use virtual private networks (VPN) for content protection in OSN. Users belonging to the VPN contribute to the privacy preservation of all members by storing the data locally on their computers. This decentralized approach is indeed the reason for a major drawback, i.e., the availability of content depends on whether or not the users who are storing the demanded content are online or not. On the other hand, users upload fake personal informa-

\[2\text{www.firefox.org}\]
\[3\text{www.wikipedia.org}\]
tion themselves to the OSN while sending the genuine information to members of the VPSN (virtual private social network) by out-of-band mechanisms, so that the SNO does not ban them from the site for posting encrypted or scrambled information. Similarly to other solutions [Baden et al., 2009, Beato et al., 2011], Conti et al. implement the replacement of fake information by means of a Firefox extension.

Beato et al. propose Scramble! [Beato et al., 2011], a Firefox extension that allows users to enforce access control over their data. Users of Scramble! can keep their data confidential (even from the service provider) by means of access control lists supported by hybrid-cryptography [Dent, 2005]. Scramble! also provides integrity to the users’ contents.

8.4 Traffic Data Protection

Most of the research community’s effort has been towards providing the users of OSN with content confidentiality protection. However, there are other privacy threats beyond content protection. Even when access to content is denied to unauthorized third parties through techniques such as encryption, the mere fact that two users are communicating may reveal sensitive private information. For instance, an OSN provider that observes frequent communication between two users is able to infer that those two users have a relationship of a certain importance or condition [Jernigan and Mistree, 2009, Lam et al., 2008, Mislove et al., 2010], even if the provider has no access to what they are actually saying.

Traffic analysis [Danezis, 2003] is therefore an important threat to be considered when trying to conceal the relationships between users. The importance of concealing the relationships between users has been studied in depth in other contexts [Korolova et al., 2008]. However, this line of research is still in its infancy: little work has been done to protect the social structure of the network undergoing a traffic analysis attack. Balsa proposed in [Balsa, 2010] a strategy based on dummy traffic [Berthold and Langos, 2002, Diaz and Preneel, 2004] to conceal the strength of the relationships between users or the very relationship itself. The strategy considers that users use a client-side plugin that sends encrypted messages to the users' friends along with dummy messages. Therefore, the adversary cannot distinguish both types of messages solely based on their appearance. The client arbitrarily forges relationships between the user and her friends and optionally between the user and other people that may be using the tool as well. As the forged relationships do not depend on the real relationships, if the amount of dummy traffic is sufficient, the existence or weight relationships is concealed. Balsa et al. [Balsa et al., 2012] further propose a technique based on information theory to determine the effectiveness of a dummy strategy as well as the amount of dummy traffic necessary to achieve a certain level of privacy protection.

8.5 Darknets

The concept of a darknet refers to private group of people communicating, i.e., a friend-to-friend network where people only connect to trusted friends preventing any observer external to the darknet from identifying the users participating on it. As put by Bethencourt et al. [Bethencourt et al., 2007], “each user only connects to her friends, trusting that her friends will not reveal her identity or existence in the network”. The concept was popularized by Biddle et al. [Biddle et al., 2002, 2003].

Darknets have derived fundamentally from anonymous communications systems [Berthold et al., 2000, Chaum, 1981, Dингledine et al., 2004, Reiter and Rubin, 1998] and peer-to-peer (P2P) systems [Schollmeier, 2001]. By definition, darknets are not necessarily an OSN. However, their networked structure makes them a good paradigm for the design of techniques and architectures for privacy-preserving OSN.

\footnote{It may be possible to distinguish real and dummy traffic due to other factors, e.g., timing.}
Napster\textsuperscript{5} was one of the first P2P systems for file sharing. Not designed with embedded privacy protection, Napster could nevertheless function as a model for designing distributed OSN (DOSN). Gnutella [Adar and Huberman, 2000, Ripeanu, 2001] and other technologies followed its lead, making P2P systems a very popular way of sharing files. Paradoxically, most of the attacks deployed against these systems exploit their lack of anonymity [Biddle et al., 2002, 2003]. Among some of the systems that do provide anonymity for content publishing are TAZ and The Rewebber [Goldberg and Wagner, 1998]. Both systems propose models for anonymous publishing on the Internet. Publius [Waldman et al., 2000] provides anonymous publishing, protects against censorship and enhances the integrity of the published content. The Eternity [Anderson, 1996] provides anonymous publishing and protection against denial-of-service attacks. Other systems, such as The Free Haven project [Dingledine et al., 2000], pursue a similar goal.

Clarke et al. proposed Freenet [Clarke et al., 2000] a P2P network that enables users to share (store and retrieve) content anonymously. Freenet uses collaborating nodes to replicate and forward data near anonymous requestors while deleting data that is no longer of interest. Popescu et al. propose Turtle [Popescu et al., 2004], a P2P architecture for privacy-preserving content sharing. Turtle presents a novel approach in the sense that instead of building trust relationships on top of P2P networks, it builds an overlay network on already existing trust relationships between the users, as they exist in OSNs.

These P2P privacy-preserving content sharing solutions have evolved into more specific solutions for social networking, namely, the so-called decentralized OSN. Yeung et al. [man Au Yeung et al., 2009] study how a decentralized architecture can help to solve privacy issues in OSN. Buchegger et al. [Buchegger and Datta, 2009] examine the challenges and opportunities of privacy preservation in OSN when shifting from a centralized to a decentralized model. Later on, Buchegger et al. propose PeerSon [Buchegger et al., 2009], a system that combines a P2P infrastructure with encryption as a means to achieve a privacy-preserving DOSN. Cutillo et al. propose Safebook [Cutillo et al., 2009a,b], an OSN based on a P2P architecture which leverages cryptography and trusted relationships to achieve privacy. Shakimov et al. [Shakimov et al., 2009] present three different schemes for DOSN which basically differ on where the user's data is stored. Aiello and Ruffo [Aiello and Ruffo, 2010] propose a framework based in distributed hash tables [Balakrishnan et al., 2003] (DHT) for the implementation of social networking services.

Other solutions that may be considered darknets are the so-called DC-nets: specific solutions for anonymous communications on fully connected networks, i.e., a network of $n$ participants each one connected to each other. DC-nets derive from Chaum's dining cryptographers problem [Chaum, 1988], where a group of $n$ participants can send messages to each other anonymously. These solutions include e.g., HerbivoreFS [Sirer et al., 2004], a file sharing system that provides anonymity.

On the other hand, Sorniotti and Molva [Sorniotti and Molva, 2010] worked on a framework to allow the creation of secret interest groups (SIG) in the OSN, i.e., using the existing infrastructure in the OSN, users belonging to a certain SIG are able to privately exchange information.

8.6 Metrics

Measuring the degree of privacy a user enjoys in an OSN is not a trivial task. Privacy in OSN has many different interpretations and implications, e.g., the meshed nature of a OSN implies that the degree of privacy of a user is dependent on the degree of privacy of other users.

Depending on the research field, different metrics have been proposed. Singh and Zhan [Singh and Zhan, 2007] propose topological anonymity to measure the degree of privacy provided by graph anonymization techniques. The topological anonymity metric is inspired in a $k$-anonymous notion and takes into account the degree and the clustering coefficient of the nodes of the network, evaluating the existence of groups of $\epsilon$ nodes with the same degree and similar clustering coefficient. This

\footnote{http://en.wikipedia.org/wiki/Napster}
metric is claimed to indicate how resistant the anonymized network is against node and edge privacy breaches (re-identifying a node and the existence of a relationship, respectively).

Campan and Truta [Campan and Truta, 2008] propose several metrics to measure the information loss incurred when using graph anonymization techniques, but it is not clear how these metrics relate to the privacy gain consequently achieved.

He and Chu [He and Chu, 2008] propose inference accuracy to evaluate the impact of certain social network features (such as the amount of publicly available information or the information the friends of a user disclose) on inferences made by an adversary, i.e., what is the effect of an observation made by the adversary on the ultimate conclusion she may draw as a result of using inference techniques. They define inference accuracy as the probability of success given by the inference. Particularly, they focus on Bayesian inference [Gelman et al., 2003].

Balsa et al. [Balsa et al., 2012] propose a metric based on information theory to evaluate the effectiveness of a dummy traffic strategy as a protection against traffic analysis. This metric is useful to model different adversaries, i.e., with different capabilities (depending on the amount of information available to the adversary) and to measure the effectiveness of the strategy when different amounts of dummy traffic are used.

8.7 Future Directions. Foreseeing New Challenges for Privacy in OSN

After the latest proposed attacks [Narayanan and Shmatikov, 2009, Narayan et al., 2011], it is unclear whether or not graph anonymization is actually feasible in the domain of any OSN. The research community still has to understand and acquire deeper knowledge about the relationship between the social graph and the privacy guarantees that a change in the topology can offer. Regarding content protection, current results indicate that we are in an exciting stage with increasingly advanced and complex solutions to protect users’ content, most of them based on cryptographic techniques. As for the traffic data protection research, it has not taken off yet, but the need to conceal sensitive relationships and the vulnerability of those under traffic analysis attacks indicates that this topic will become very important in the near future. On the other hand, there has not hitherto been any successful DOSN, although Diaspora® begins to look like the first of a future series of DOSN. There is a serious lack of metrics that allow us to evaluate the degree of privacy obtained by the different privacy preserving techniques currently devised. Therefore, it will be necessary to develop proper metrics that are useful to assess the level of privacy achieved by different techniques. Last but not least, studying how different techniques can be intertwined with each other is also a challenge for further research.

---

6https://joindiaspora.com/
9 Feedback and Awareness Aspects (DTAI)

9.1 Introduction

9.1.1 The general idea of privacy feedback and awareness

The idea of privacy feedback and awareness tools (PFA tools for short) can be described, with some simplification, as follows: The reason why many people disclose information on SNSs (Social Networking Services) or elsewhere on the Internet that they later regret is that they are unaware of the (semi-)public nature of this data disclosure. For example, they may think that uploading a photo corresponds to showing a print to a friend over coffee, when in fact their “privacy settings” allow every Web user to see this photo.

Making social network users aware of the public nature of disclosing their data – e.g. who can see it, how it can be distributed and used – would allow them to make more informed decisions about publication. Furthermore, they may choose to hide it (not upload it, upload it elsewhere, encrypt it, set restrictive privacy settings...) because they want to keep it private, or else choose to publish it because they regard it as a useful element of an identity they want to create in a (semi-)public space. In other words, increased awareness is expected to lead to more informed action choices and thereby to a more satisfactory managing of identities and their different public/private boundaries.

So how can people be made (more) aware? The basic idea is to give them feedback. The term “feedback” has today become nearly synonymous with “information having to do something with the recipient of the message”. Understood in its basic cybernetic sense, it is however more specific: first, it is information related to the output of the system itself (e.g. the user and the SNS platform's rules) – generally including feedback on the person's own prior choices – and second, it is information about a gap between the actual level and the reference level of a system parameter, which is used to alter the gap in some way [Ramaprasad, 1983]. As one example, a feedback in SNS can be a message alert from one of the user's friends, or the number of people who has viewed her profile, etc. In the former case, the system parameter can be a user's friendship. The actual level of the friendship between her friend who sent the message and her will probably be not as high as when she replies her friend's message. So she takes the action of replying in the hope of not losing/reducing the friendship with her friend. In the latter case, the system parameter can be her popularity on this SNS, if the number of people who viewed her profile page is not as high as she expects, she might be disappointed by her popularity and try to do something to change the current situation, e.g. chatting more with her friends or sharing more interesting photos or links. If the number of people visiting her profile is satisfying, she will probably continue to do what she is doing to maintain her popularity status.

Feedback and awareness tools are then information systems in which a computer software gives the feedback with the aim of raising awareness in a user or a user group, and privacy feedback and awareness tools focus on privacy. More details are given in Section 2.

9.1.2 PFA tools in the context of SPION – complementary solution approaches

SPION comprises a number of methods from different disciplines for analyzing and addressing/solving privacy problems. Within the latter, we can distinguish two basic approaches in the project. For simplicity, we will assume that the privacy problem concerns the dissemination of information that someone (for example the data subject) considers undesirable, such as the disclosure of one's profiling information in an SNS to non-friends. We also consider privacy practices that involve selective revealing of information, we will however allow for outcomes where this information does get transferred.

\[1\] For an example, see subsection 3.1
The first approach makes it possible to render the information dissemination technically impossible or prohibited, categorically or under well-defined circumstances. This approach is expressed in computational methods for security or access control, and in legal approaches (to the extent that they result in legislation etc.). The second approach attempts to affect the likelihood of the information dissemination via an educational route: the awareness of its consequences. This approach is expressed in educational materials prepared for a formal learning setting (secondary education). This is complemented by the approach of the DTAI subgroup of SPION, which targets informal learning settings (e.g., leisure-time computer usage), wider age groups (often including adults), and a “more open learning agenda”. A further contribution is the inclusion of machine processing and machine intelligence for generating the “material” in a user-and-situation-specific manner.

Knowledge about human behaviour and empirical and experimental evaluations of “solutions” will help us to validate these different “solution approaches”.

9.2 What is a privacy feedback and awareness tool?

By prepending a final enabler for measurement to the elements introduced in Section 1.1, we can summarize the essentially cybernetic understanding that defines a Feedback and Awareness (FA) tool (1.-4.) and a PFA tool (1.-5.), respectively:

1. Most activities that are related to online SNS and to privacy by definition take place on a computer. Computers log behaviour (e.g. posting something) as well as the data it generates (e.g. photos). Hence, privacy-relevant\(^2\) behaviour can be measured and recorded.

2. This information can be given as feedback. The feedback may also include information about the relation of the current data to a reference level\(^3\).

3. This feedback will make people aware of their actions and the consequences of their actions.

4. On the basis of this awareness, people can make more informed action choices.

5. The system outcome goal of this are privacy practices that are better than the current ones.\(^4\)

Two relevant caveats concern the strict cybernetic notion of feedback: the relation to the system-caused output, and the notion of a correct, normative reference value. A first relevant twist of the basic definition centered on an individual user is given by the history of the term “awareness” in computer science. The term has been used in the context of computer-supported cooperative work: Dourish and Bellotti ([Dourish and Bellotti, 1992], p. 107) defined it as “an understanding of the activities of others, which provides a context for your own activities”. Generalising from other people to a whole system of (social and technical) relations, Sohlenkamp ([Sohlenkamp, 1999], p. 41) defined awareness as “an understanding of the state of a system, including past activities, present status and future options”. This means that the activities in bulletpoint 1 of the above definition comprise not only the user’s own activities, but also their consequences which can be either directly relevant for the actor or indirectly. SNS, like computer-supported cooperative work, involve collaborative practices and collective decision-making. They may therefore be expected to likewise profit from awareness.

For PFA tools, these observations mean that “the system” can extend to anything from a specific SNS site, to the Internet or even an information-processing society as a whole: and that it is usually, but not necessarily always, a user-initiated action on which feedback is given. A second caveat is that FA tools tread a fine line between being informative, entertaining and normative/controlling. This can be seen when one looks at non-privacy-relevant feedback and awareness tools

---

\(^2\)By privacy-relevant, we refer to the data or actions that can be used to draw sensitive information that are not meant for disclosure.

\(^3\)For instance, the tool “Privacy Scanner” discussed in Section 9.3.2.2 computes a “privacy score” (scaling from 0% to 100%) of the user and of their friends. This suggests reference levels like “100%” or “the highest among my friends”. To what extent these are useful reference levels is another question.

\(^4\)By privacy practice, we mean any activities people perform that in some way impinge on their privacy.
such as www.manictime.com or www.rescuetime.com. Such FA tools focus –explicitly – on the system outcome of productivity (it differs from PFA in which the system outcome is for better privacy). Rescuetime.com claims that “on average [our tool] recovers 3 hours and 54 minutes worth of productive time per week per person”, by recording and playing back time spent with different applications, including presumably non-work-related applications like Facebook. This suggests that the reference value is to spend all one’s office time on work, or the distribution of time that other workers display, that longer use of Facebook is a deviation from this reference value, and that either the office workers themselves or their superiors can monitor and adjust work-time distribution.

However, one of the central tenets of SPION is that there is no generally applicable normative “reference value of privacy”. This general belief of course raises problems when one thinks about evaluating PFA tools. We will return to this issue in Section 5.

9.3 Existing PFA tools

9.3.1 Conceptual developments in the scientific literature

Privacy feedback and awareness tools are motivated by the increasing revelation and spread of highly personal (not only self)-profiling especially in today’s Social Media, and the observation that in spite of many people describing themselves as highly privacy-conscious when asked explicitly, online behavior differs markedly from these attitudes [Acquisti and Gross, 2006, Berendt et al., 2005]. Alongside this, the appeals to be more protective of one’s personal data scarcely have any effect; the idea of privacy feedback and awareness tools is to show users important consequences of potentially privacy-relevant activities they and others have performed.

Privacy awareness tools aim at fostering understanding and reflection: For example, [Lederer et al., 2004] suggest improving privacy sensitivity in systems through feedback that enhances users’ understanding of the privacy implications of their system use. This can be coupled with control mechanisms that allow users to conduct socially meaningful actions through them. These ideas have led to suggestions like the identityMirror [Liu et al., 2006], which learn and visualize a dynamic model of user’s identity and tastes. Similar ideas are embodied in the concept of privacy mirrors [Nguyen and Mynatt, 2002] or in the proposal for linkage control in identity management systems [Hansen, 2008].

9.3.2 Available Tools

On today’s Web, one can find a large number of tools (and sometimes scientific papers behind them) that realize, in different ways, the concepts that were described by the pioneers whose work was discussed in the previous section. In this subsection, we present a selection of such existing privacy-relevant FA tools.

The search strategy for compiling the tools to be included in the survey was to use a general search engine (Google) as well as a major social-networking site (Facebook) and two special-purpose websites: Information Aesthetics7 and Visual Complexity8. Information Aesthetics is a weblog that explores the symbiotic relationship between creative design and the field of information visualization. More specifically, it collects projects that represent data or information in original or intriguing ways. VisualComplexity.com intends to be a unified resource space for anyone interested in the visualization of complex networks. The website’s main goal is to leverage a critical understanding of different visualization methods, across a series of disciplines, as diverse as Biology, Social Networks or the World Wide Web.

---

7 This conceptualisation also points to the privacy issues that FA tools themselves may raise.
8 This subsection is based on [Berendt, 2011]

7 http://infosthetics.com/
8 http://www.visualcomplexity.com/vc/
On all sites, search was done via the keywords privacy, feedback, awareness, tool, social network, web, and visualization. The list of applications was filtered by a number of criteria to select a number of applications that are as “representative” as possible, yet can be described in the short available space. Criteria such as the environment (e.g. SNS, blogosphere, the Web in general, etc.), the data being analyzed (e.g. profile data, phone records, blog texts/comments, browser cookies, etc.) or the people being analyzed (e.g. a single user, several users, a specific web social media community, everyone on the Internet, etc.) are taken into account when selecting PFA tools, we try to make the selection diverse and representative, please see more in section 3.3. Aesthetics and usability of an application is also taken into account. With the same functionality, more “beautiful” and “usable” applications will be chosen. This search strategy necessarily has subjective components and cannot claim completeness of all existing applications (no search engine has full coverage of the Web), but it did produce a good overview.

9.3.2.1 What does it mean for feedback to refer to privacy?

Second, the investigation of tools showed that a vast range of tools may be said to incorporate aspects of what we defined as PFA tools: giving feedback on (and hopefully raising awareness of) privacy-relevant issues. A wide sense of the term could comprise any tool that shows any information about some data that could be privacy-relevant. This could even be a search engine (because it allows one to search for blogs in which people talk about personal matters), a news aggregator (because it may condense privacy-relevant information and thus lead to new awareness of it) or Facebook itself (because it makes one’s data behaviour explicit, which may lead to new awareness). A narrow sense of the term could comprise only tools that give feedback about one's so-called "privacy settings" in Facebook. We do believe that it is important to also look into the wider meaning of the term PFA tools in order to do justice to the notion of “privacy as practice”, which calls for an open investigation of practices of data creation, hiding and revealing. However, we also want to structure in order to not let the term become too vague.

We therefore decided to classify tool candidates by their purposes (privacy role), with some aiming at detecting and revealing privacy breaches, others at mirroring or summarizing privacy-relevant information for the user. We refer to the classification of [Gürses and Berendt, 2010] who identified the following breaches deemed relevant by current Facebook users: Breaches related to

1. Indeterminate Visibility: Indeterminate visibility denotes the problem of a user’s profile information being visible to others without the user’s explicit knowledge or approval.

2. Aggregation of Separated Digital Identities: Separation of digital identities denotes the construction of social identities by individuals that selectively reveal and conceal information in specific contexts and roles.

3. Misappropriation: the use of SNS data out of context or for previously undefined purposes.

Interestingly, several tools show that in order to think about problems such as visibility at all, one has to have or get an overview of “what one has done”. We call such tools

4. Mirroring: giving the user an overview of their data and/or actions.

In addition, we will also discuss a number of tools that

5. summarizes information and/or information flows of privacy-relevant information.

In addition, we categorized the selected applications based on two further criteria: time and space. The categorization based on time assumes a person has initiated or participated or is simply interested in a series of events by him/her-self on the web through time. We call this collection of

---

9 These purposes are not mutually exclusive.

10 A fourth breach type identified there relates to Contested Ownership: the explicit and implicit definitions of data ownership that may lead to privacy breaches. Tools dealing with this notion were not found here.
events associated with this user the activity stream [Hevia and Micciancio, 2008]. These events may include changes the user made to their profile page, the fact that the user added or ran a particular application on the social networking site, that they shared a news item, or that they communicated with one of their friends or commented in a forum, etc. Now that we have the axis of time, we are able to distinguish between the past and the future. We call a feedback and awareness tool that summarizes the history of a user using the web and presents the resulting statistics (regardless of what kind of form, plain texts or graphics) to the user of the Historical Viewer. We define another category called Ongoing Monitor: A feedback and awareness tool that informs the consequences of a user’s current actions or actions by others that are relevant to the user as well as predictions for the future (using e.g. what-if simulation). Note that the Ongoing Monitor can sometimes borrow some “history” to help the user better understand the current and possible future consequences. We can see the data for Historical Viewer to process are static since they are historical facts, while the data fed to Ongoing Monitor are dynamic, because of their ongoing nature.

The categorization based on space is meant to separate the user herself from the others on the web. There are data and events that are directly and/or immediately related to the user which are either initiated by or have direct effect on the user, no matter whether they are in the past or present. Let’s call this category micro-PFA (about me). There also other data and events that are not directly and/or immediately related to the user, the user simply has the interest to know what is there in the outside world, what is going on in the environment that the “virtual me” inhabits. Let’s call this category macro-PFA (about the world). Note that in order to know “about me”, we may borrow some “about the world” information\textsuperscript{11}. In the following, we order tools by their privacy role and characterize each in terms of its time and space class. These classes are not necessarily disjoint, but we have chosen the main characteristics.

9.3.2.2 Visibility

We will first investigate four tools that are Historical Viewers “about me”. They give feedback on the visibilities of one’s historical data or of one’s historical/current settings — which may have implications for future data.

• Privacy Mirror – demonstrating how much profile information 3rd party applications can access. Privacy Mirror itself is a 3rd party Facebook application. This application uses standard Facebook API to check the data items’ visibility in user’s profile. This application simply gives the user feedback on what is visible of her Facebook profile data to 3rd party applications. The application can be accessed via the following link: http://apps.facebook.com/privacy_mirror/

• Privacy Check – is very similar to Privacy Mirror, but besides telling what profile data is properly hidden and what is not, it also calculates a general score to measure the status of the user’s profile privacy settings. It can be found at: http://www.facebook.com/apps/application.php?id=106946932683898

• PrivacyScanner – scans user’s Facebook profile for offensive language and images like inappropriate posts by her drunk friends on the wall or her wild party photos. It returns an overall score (and nothing else). However, how this application defines “wildness” or “offensiveness” is not revealed in its description, nor is how it calculates the privacy scores on the user and his/her friends. What is interesting about this application is that it also scans the user’s friends, rates them as well and has the function to allow the user to inform his/her friends about their privacy status, which provides another level of awareness. The application can be accessed at http://www.facebook.com/apps/application.php?id=123663951038698&sk=info

• PViz – As stated in [Mazzia and Adar, 2011]\textsuperscript{12}, “PViz is an interface and system that cor-

\textsuperscript{11}Note that we can further split this categorization into two groups: “me” as event initiator or giver and “me” as a passive receiver. The former is “I do something to affect others or myself”; the latter is “others do something to affect me”.

\textsuperscript{12}To the best of our knowledge, this tool is currently not available online.
Figure 9.1: Coarse granularity view in PViz.

responds directly with the way users model groups and privacy policies applied to their networks. It allows the user to understand the visibility of her profile according to natural sub-groupings of friends, and at different levels of granularity. PViz is centered on a graphical display, which shows the user’s social network. Each node in the display represents a semantically meaningful sub-group of the user’s friends (a community) or an individual friend. Figure 1 shows a screenshot of PViz displaying Margaret’s social network. Inspecting the display shows that PViz has found five main communities of friends.

- **Collusion** – The idea is that when visiting website A, other websites B and C are possibly tracking the user’s browsing history, clicked items and at the same time and collecting the user's browser and computer info through A. This collected information is likely to be sufficient to uniquely identify the user. Collusion is an animated network combined with text, which reveals how data collection companies manage to leave cookies on the user’s computer during standard web browsing, and which also visualizes how these seemingly separate companies are actually secretly collaborating with each other. Figure 2 (each dot representing a website) shows that when visiting the imdb.com web site, several other websites are revealed to be trackers such as FullCircle and Google. Sites in red are confirmed trackers by privacychoice.org. Sites in gray are not, but this doesn’t necessarily mean they don’t collect data on the user. The graph is extended as the user continues browsing. This application reveals which websites are secretly tracking the user when she is browsing. However, the user is not able to know which website is tracking which specific information or turn the tracking off through this application. More detailed info can be found via: [http://www.toolness.com/wp/2011/07/collusion/](http://www.toolness.com/wp/2011/07/collusion/) and [http://donttrack.us/](http://donttrack.us/). An online demo: [http://collusion.toolness.org/](http://collusion.toolness.org/). The Firefox Add-on page: [https://secure.toolness.com/xpi/collusion.html](https://secure.toolness.com/xpi/collusion.html).

- **PrivacyWizard** [Fang and LeFevre, 2010, Fang et al., 2010] – A privacy wizard is a fine-grained tool that helps user configure which information is hidden from/ revealed to which friend semi-automatically, which means the user only has to configure a subset of the privacy settings manually, and the wizard will use its underlying machine learning model (a classifier) to automatically configure the rest. Two screenshots are shown in Figure 3. With an increasing number of friends that the user has manually checked, the machine learning model becomes more and more confident on recommending to the user good privacy configurations for the rest of her friends. 


- **deGeo** – A growing number of tools exploit the fact that publishing location data in any way may allow for sensitive inference. Thus, Please Rob Me or I Can Stalk You demonstrates how to

  However, this link doesn’t seem to work at the very moment, it directs you to the correct page but with a server-unavailable error.

  [urlhttp://pleaserobme.com/](http://pleaserobme.com/)

  [urlhttp://icanstalku.com/](http://icanstalku.com/)
Figure 9.2: Collusion reveals which websites are tracking the user when she is browsing.

Figure 9.3: Privacy Wizard learns through the configuration of privacy items for user’s friends, it becomes more confident when more friends are configured.
automatically identify empty homes (on the basis of information in Tweets) or people’s whereabouts (on the basis of uploaded photos’ meta data). There are also protection tools built on these ideas, for example – deGeo is a photo sharing privacy tool for iOS\textsuperscript{16} devices, of which the users can share photos without compromising locational privacy. deGeo works by removing the embedded geotags as well as associated EXIF\textsuperscript{17} metadata before sharing photos online or with friends since an iOS camera app stores the exact GPS location with each photo that a user takes. This application can be found via: http://itunes.apple.com/us/app/degeo-photo-sharing/id412472011?mt=8.

9.3.2.3 Mirroring

Based on current and/or user’s historical data and settings, the following Historical Viewers (Social Memories) and Ongoing Monitor (PRISM) give basic information “about me”: what have I done or what am I doing? coComment is a mixture in the dimensions of time and space, i.e. it runs on both current and historical data about both “me” and “the world”.

• Social Memories – A collection of Social Memory figures is automatically generated from user’s Facebook data, including the user’s profile data, album photos, status updates, friends’ profiles and messages, etc. Various infographics are used to illustrate data trends or patterns, e.g. most popular photo album, friends most tagged with user’s name, who the user photographed the most, distribution of friends, most active friends, most popular tags, friend gender distribution, status vs. responses, the events the user attended, weekly activity statistics, friend home towns, and so on. All the historical data in the user's account is used. This is an excellent application in the sense that it provides user a fairly comprehensive overview of one’s history on Facebook through clear, precise and intuitive infographics, in which users can learn their own past behaviors, their friends’ responses, etc. The application can be found via the link: http://www.facebook.com/socialmemories. Several screenshots show how this application looks in Figure 4:

• PRISM [Patil and Kobsa, 2010]– is short for PRIVacy-Sensitive Messaging system, a plugin for an open-source Instant Messaging system. It provides IM users with various visualizations that allow for greater visibility (to oneself) of one’s own actions in relation to one’s contacts (e.g., temporal patterns of login activity, periods of idleness). PRISM provides mechanisms for presenting oneself differently to various groups of contacts by selecting different impression-relevant settings\textsuperscript{18} for them.

• coComment – According to the website http://www.cocomment.com/, coComment is a service for managing, powering and researching conversations online. When using coComment, the user can keep track of her own comments as well as others’ of a specific article across large

\textsuperscript{16}http://en.wikipedia.org/wiki/IOS_(Apple)
\textsuperscript{17}Specification of the EXIF format: http://www.cipa.jp/english/hyoujunka/kikaku/pdf/DC-008-2010_E.pdf
\textsuperscript{18}such as “online”, “away” or “appear offline”, etc.
amount of and very diverse websites (from sharing sites like Youtube, Flickr to various kinds of blogs like BakingBites, Blogger, as long as there is public commenting section), share them with friends, and get notified when the user gets a response. The up-to-date notification of the comments of interest is done via coComment’s Firefox add-on: http://www.cocomment.com/tools/extension, however, this add-on is currently not compatible with Firefox 5.

9.3.2.4 Aggregation of separated identities

The following three tools take two very different approaches to the separation of identities. The first two focuses on the de-separation of virtual identities as a feature, a desired simplification of a multitude of virtual identities in various social networks. However, it could be used also as a feedback tool on the degree to which there is undesired de-separation. The second focuses on a user’s belief to be “hidden in a crowd” and thus able to separate their browsing identity from their “real-life, unique identity”. All tools are based on historical data or settings and thus are Historical Viewers, and all are micro-PFA.

- **MyConnections** – is a demo application built on Google’s Social Graph API, available from the Social Graph website. As stated on the website: “The Social Graph API now makes information about the public connections between people on the Web, expressed by XFN and FOAF markup and other publicly declared connections, easily available and useful for developers.” In the demo, only the URL(s) of the target website(s) is (are) needed as input, which can be e.g. her own or other people’s blog(s) or profile page(s). My Connections shows the list of web URLs that are connected to her that have been entered, including all contacts they are linked to and all contacts who have linked to them, using relationships like ‘friend’, ‘kin’ or ‘colleague’.

- **SiteConnectivity** – It is a demo application built on Google’s Social Graph API. Site Connectivity shows how different sites of a user are connected – since one user may have multiple profile pages or blogs, etc., which are probably interconnected via XFN or FOAF’s markups. The ‘Score’ column shows how well linked each site is to its peer sites - green for fully inter-linked, red for just one. A sample result is shown in Figure 6, with “kevinmarks.com” as the input URL. We see that Kevin’s Twitter and Flickr pages are fully connected whereas his sharing page on Google or the profile page on myOpenID is poorly connected.

- **Panopticlick** [Eckersley, 2010] – Panopticlick tests user’s browser to see how unique it is based on the information (such information includes browser type, browser plugin details, time zone, screen size, whether cookie is enabled) it will share with sites it visits. Such information is also called

---

19Further examples of this type of tools exist, e.g. ClaimID (http://claimid.com/) and AboutMe (https://about.me/).
21see My Connections
22http://gmpg.org/xfn/
23http://www.foaf-project.org/
“browser fingerprint. The user is given a uniqueness score (in bits) calculated based on the fingerprint, the higher the score is, the more identifiable the user becomes. Note that we put Panopticlick into the Ongoing Monitor category because it always measures user’s current browser fingerprint. However, user’s past browser fingerprint can be logged using this tool and become a fingerprint history of the user on the web. More information can be found on the website: http://panopticlick.eff.org/.

9.3.2.5 Information combination, information summary, information flow

The first three case studies and tools show the power of combining information that is available, and they help process available information by summarizing it. The last focuses on information flow. While the first has clear original privacy intent, this is not as clear for the others. Nonetheless, they present very relevant components of future privacy tools. For the time dimension, Tell-all telephone and Cascade are Historical Viewers, We Feel Fine and IMBuddy are Ongoing Monitors. For the space dimension, Tell-all telephone and IMBuddy are micro-PFA, We Feel Fine and Cascade are macro-PFA.

- **Tell-all telephone**—German politician Malte Spitz sued to have German telecoms giant Deutsche Telekom hand over six months of his phone data that he then made available to ZEIT ONLINE, who combined this geo-location data with information relating to his life as a politician, such as Twitter feeds, blog entries and websites, all of which is all freely available on the internet. This results in a privacy awareness tool (which can be found in http://www.zeit.de/datenschutz/malte-spitz-data-retention) specifically showing the mostly private activities of Malte Spitz from August 2009 to February 2010. From Figure 7 we see that surprisingly large amount of information can be revealed through the phone log files, such as the number of phone calls/SMS messages, duration of Internet connection, the route of the user with time stamps, even the location and the coverage of the tower. Although the purpose of such tool is to support Malte Spitz’s point that revealing phone logs is harmful to personal privacy, we can consider it as an augmented geo-location informer that collects user’s geo-location information on the web, and combines other public data of user, such as tweets, blogs to raise the user’s privacy awareness.

- **We Feel Fine** [Kamvar and Harris, 2011] – As stated on the website of We Feel Fine project: We Feel Fine has been harvesting human feelings from a large number of weblogs. At the core, We Feel Fine is a data collection engine that automatically scours the Internet every ten minutes, harvesting human feelings from a large number of blogs. Blog data comes from a variety of online sources, including LiveJournal, MSN Spaces, MySpace, Blogger, Flickr, Technorati, Feedster, Ice

---

24 In terms of contextual information that brings about the awareness of the surrounding virtual space of the user

25 Note that Tell-all telephone is strictly speaking not a PFA tool, rather, a project or demo, since it only renders one particular person’s geo-location data. However, we think the idea behind it is valid and useful for building a geo-location awareness privacy tool.

26 http://www.wefeelfine.org/
Figure 9.7: screenshots from We Feel Fine.

Rocket, and Google. The system searches the world’s newly posted blog entries for occurrences of the phrases “I feel” and “I am feeling”. When it finds such a phrase, it records the full sentence, up to the period, and identifies the “feeling” expressed in that sentence (e.g. sad, happy, depressed, etc.). Because blogs are structured in largely standard ways, the age, gender, and geographical location of the author can often be extracted and saved along with the sentence, as can the local weather conditions at the time the sentence was written. All of this information is saved. We Feel Fine only collects and displays data that was already posted publicly on the World Wide Web. We Feel Fine never associates individual human names with the feelings it displays, though it always provides a link to the blog from which any displayed sentence or picture is collected.

The idea of We Feel Fine is very interesting and innovative. It provides users with the up-to-the-very-moment feelings of the world. It is simple yet effective when it comes to searching for emotional status of a particular group of people using the “filter” interfaces. Figure 8 shows 6 screenshots (numbered from 1 to 6) of We Feel Fine. In Figure 8.1, each bubble represents a sentence or a picture that has recently been posted, and users can click on a bubble to reveal the “I feel” content. Figure 8.2 shows a user interface that helps user filter feelings based on feeling keywords, blogger’s gender, age, location, local weather and year. Figure 8.3 and Figure 8.4 show different forms of presentations of the top feeling key words within the last few hours. Figure 8.5 and Figure 8.6 show that feelings can be mapped according to gender and location. The application can be accessed via the link: http://www.wefeelfine.org/.

• **IMBuddy** [Hsieh et al., 2007] – is a contextual Instant Messenger that provides user with contextual information including interruptibility, location and current window focus of user’s friends on IMBuddy. E.g. a friend’s interruptibility is considered high if she is 60% available 10 minutes ago, location info can be “last seen off-campus 5 minutes ago” and window focus can be “last used firefox.exe 8 minutes ago”.

• **Cascade** – a 3D interactive visualization tool that reveals how information propagates through twitter messages, created by the NYTimes R&D department. More specifically, NYTimes articles
Figure 9.8: A cascade shows one key node (user) that helps the NYTimes article spread even more when many of his followers retweeted his tweet.

are tweeted or retweeted by twitter users (simply in the URL of the article or sometimes a URL with the user’s comments) on daily basis; Cascade aims to reveal the causal factors that determine how tweets about news stories are propagated in the most successful ways. For instance, the visualization can demonstrate how viral popularity is mostly influenced by the people who are followed by famous people, who in turn drive a lot of online discussions and retweet activities. An example is shown in Figure 5, the NYTimes article “America Goes Dark” is widely shared in twitter space via one important twitter user). Also on the top right corner of Figure 5 we can see the tweeting event graph with the horizontal axis as time and vertical axis as number of tweeting events about the article, a burst of tweeting happened at a specific time in history, which can be tracked by user. This tool helps user be more aware of the tweeting events related to NYTimes articles within the tweeter community. This is not a privacy feedback and awareness tool, but the idea of using cascade interactive visualization to inform user about the information flow on the web is very interesting and relevant to our research, since such kind of tools can be transformed to helps user identify the very information flow related to him/her self. For example, to see who has retweeted your tweet(s) or who is the key person that makes your shared photo album spread even wider through time. Cascade can be found at http://nytlabs.com/protect/projects/cascade.html.

9.3.3 Reflections on the classification and the range of PFA tools

Though we have created the classifications based on time and space, which are essentially continuous. There are no fine lines to separate the past and the presence, and “me” is living in the “world” while the “world” contains “me”. Sometimes the feedback is not at the very moment, but fairly up-to-date, like Cascade, sometimes it’s about one’s own actions and also about the others’, like coComment and Social Graph. Such classifications mean to emphasize that the factors “time” and “space” are important for PFA tools, however, there are many other factors that shape and differentiate these tools, for instance, the data that are analyzed by the tools are different. Privacy Mirror and Privacy Check process the profile data of the user. Social Memories process even more, it takes almost everything that a user has in her Facebook account. We Feel Fine harvests the data from the blogosphere on the Internet. Cascade searches through Twitter-sphere to find cascade structure on NYTimes related tweets. Collusion analyzes the browser cookies a user is using. Panopticlick analyzes the data that a user’s browser sends. coComment harvests and integrate the comments from various websites. Tell-all telephone visualizes Malte Spitz’s phone logs combined with his public online records, e.g. his tweets. Social Graph looks specifically for XFN or FOAF data of a given website to analyze its social connection status...

Another interesting observation is that at first sight, a user may think that others’ blogs, tweets and comments, etc. are not relevant to her individual privacy concern, as in We Feel Fine, Cascade and coComment. However, in order to protect one’s privacy, and ultimately, to positively construct one's
online identity, the user also needs to take into account the context. By context, we mean the virtual space that one inhabits and the information in it as well as how this information flows.

9.3.4 A note about the evaluation of PFA tools

Do PFA tools work? The answer to this comprises general principles and specific questions.

The general effectiveness of the feedback approach has been tested in a wide range of empirical and experimental demonstrations of the effects of feedback on human behaviour such as academic learning (where the awareness is constituted by metacognition of one’s learning activities) see e.g. [Berendt, 2006, Berendt and Brenstein, 2001, Gama, 2004], psychotherapy [Claiborn and Goodyear, 2005] and health-related areas [Lehrer, 1996]. Moreno et al. [2009] demonstrated that even mild feedback on public profile contents in online SNSs, delivered via email, can be effective in the sense of leading to a more “privacy-conscious self-portrayal”. However, this demonstration did not involve a tool. To the best of our knowledge, there is no corresponding evaluation of a PFA tool. To evaluate such a tool, one would need to investigate general usability criteria as well as the correctness and comprehensiveness of the statistics (etc.) shown. In addition, an approach based on machine learning should be evaluated based on machine-learning quality criteria such as accuracy and speed. (This was done for the Privacy Wizard.) An inspection of correctness and comprehensiveness of analyses done by a tool is only possible when the code is available, which is not always the case.

Privacy measures such as anonymity, unlinkability, undetectability, unobservability and pseudonymity have been evaluated formally [Pfitzmann and Hansen, 2009, 2008]. However, such evaluations do not directly address the effectiveness of feedback and awareness mechanisms. Acquisti and colleagues developed models to measure user's privacy gratification and rationality in privacy decision-making [Acquisti, 2004, Acquisti and Grossklags, 2005], which provides us a behavioral economic perspective for the evaluation of PFA tools.

The greatest challenge becomes obvious when one delves deeper into the question of usability: Which users should be supported by the PFA tools in which tasks? And how should the effectiveness of “solving” this task be measured? How should adaptivity be designed and evaluated? Assume that “the task” is always “to improve your privacy”. Then one could draw on a range of privacy metrics such as [Veeningen et al., 2010, Pfitzmann and Hansen, 2009, Hughes and Shmatikov, 2004, Hevia and Micciancio, 2008, Bonneau and Preibusch, 2009]. One challenge is that all these measures are disputed, which should not be surprising in an area that is still finding its definitions [Gürses and Berendt, 2010]. A second challenge is that privacy is by no means enforced if one considers informal learning settings where users define their own goals – for example, they may well use Google's Social Graph tools to aggregate separated virtual identities. In other words, not only the tool itself, but also usage contexts and user choices determine whether a given tool satisfies part 5 of the definition of a PFA tool in Section 2. Another challenge is the difficulty in measuring good interactivity and adaptivity of PFA tools. Weibelzahl and colleagues examined the existing and proposed new evaluation frameworks of interactive adaptive systems [Weibelzahl, 2001, Weibelzahl and Weber, 2002, Paramythis and Weibelzahl, 2005, Paramythis et al., 2010].

9.4 Conclusion

We have addressed the concerns on privacy loss in today’s social web, and we have presented the “feedback and awareness” approach. We discussed the basic concepts of awareness and awareness tools, and listed some existing privacy feedback and awareness tools. It was shown that Historical Viewer and Ongoing Monitor together can provide users with a comprehensive view of their privacy in time, so that they can trace back to see what has been done, have more control of their current actions and shape their future privacy with automatic predictions. Interestingly, most of the existing tools promote “the awareness about oneself” rather than “the awareness about others”, which is a clear sign that the notion of awareness has shifted from the CSCW context. However, the intrinsic
interrelatedness of privacy concerns especially in social networks [Gürses and Berendt, 2010] still tends to be largely ignored in today’s PFA tools.

In sum, today’s PFA tools give feedback and aim to raise awareness on a wide range of data and mostly their visibilities. Most of them restrict themselves to the display of settings, the (re)-presentation of data or descriptive statistics. However, as the PrivacyWizard illustrates, data mining can extend the scope of these applications and simulations for information inference by employing more sophisticated forms of induction and deduction for demonstrating the possible consequences of a user’s actions. Today, this approach remains under-researched. In [Gürses and Berendt, 2010], we have shown how the spread of visibility/accessibility of a user’s profile and relational data may be computed.

Although the studied awareness tools are great sources of reference for our future work, there is no (as far as we know) comprehensive privacy feedback and awareness tool available for the users. Existing tools tend to focus on one aspect of the problem, e.g. Privacy Wizard helps user tailor visible information user’s friends on Facebook, and We Feel Fine visualizes the emotion state of the bloggers on the web. The range of the privacy solutions is broad, but we want to ask ourselves what are the essential aspects to user’s privacy practice and protection, and what constitutes a comprehensive and effective feedback and awareness tool. No matter what the answers to the questions might be, applications that capture more of what is potentially harmful to user’s privacy than the state-of-the-art are needed. We have also argued that in the absence of unique usage scenarios and clear goals, evaluations of privacy enhancing technologies remain difficult. This core issue of privacy tools will be investigated further with the SPION partners.

Based on this summary of existing solutions, we will concentrate on conceptual, methodological and implementational improvements of privacy feedback and awareness tools along three dimensions. Taking into account the interrelatedness of privacy concerns and activities will be the first central concern. The second will be to leverage machine-learning/data-mining models for improved privacy feedback. Third, we will create interactive, adaptive user interfaces and meaningful, intuitive visualizations based on underlying data, examples of such applications are [Gao and Vanschoren, 2011, Gao and Berendt, 2011]. Fourth, together with other SPION partners, we will develop and apply new approaches to evaluate such tools.
10 Conclusion and Outlook

As we indicated in the Introduction, due to the interdisciplinary ambitions of the SPION project, this document goes beyond what is expected of a traditional state-of-the-art study. Namely, we used the process of producing this document as an opportunity for all partners to get to know each others’ objectives; to get inspired in defining future work; to refocus what is stated in the description of work; and, finally, to map the challenges and gaps on the topic of privacy and security in SNS together, e.g., to indicate where interdisciplinary work would be meaningful to address the complexity of problems identified by the different partners.

Chapters 3 through 9 gave a rather disciplinary account of the state-of-the-art. In this section, we provide an analysis of how the findings of the different partners interrelate. As we went through this exercise, certain topics emerged, under which it is possible to map the potential for interdisciplinary inquiry in SPION. Namely:

- Awareness and educational aspects.
- Privacy settings, discrepancies and control aspects.
- Accountability and compliance aspects.
- Social and community aspects.

In the following we summarize these different aspects and the points of interaction that the partners have identified. We expect this map to guide future research and valorization activities throughout the SPION project. We must nevertheless note, while we wish that we could address all the identified points, we are aware that given the time and resource limitations that only a subgroup of the points of interaction that are identified will be the focus of our activities.

10.1 Interdisciplinary Mapping

10.1.1 Awareness and Educational Aspects

One of the key themes that runs across the research interests of the SPION partners is the focus on raising user awareness through educational material, technical means, as well as soft paternalism (nudging). In literature about security in information systems we can distinguish three approaches to usable security: 1) using systems that ‘just work’ without requiring intervention of the user, 2) making security intuitive and easy to use, including warning systems, and 3) educating users [Cranor, 2008]. There are different challenges associated with each of these approaches.

One the one hand, some critics state that user education is a myth [Görling, 2006]. Actually with their critique, the researchers argue that user education can not be seen as the holy grail in which the user is given all responsibility for educating herself. Education in itself is insufficient and will only work in combination with other approaches. On the other hand, past experience and research in privacy and security shows that providing tools to users without educating them about how to use them may be more counterproductive than not providing them with tools at all. The partners can best address both of these challenges through iterative feedback sessions between the educational and technical partners.

In order to accomplish that, OWK can work together with the partners to reach a better understanding of user needs as well as technological affordances. An improved understanding of the issues at hand can be instrumental to developing educational materials that help raise awareness among youngsters about privacy issues on SNS. For example, the issues addressed by CMU, e.g., users' needs, preferences, biases, limitations in privacy decision making, can contribute to the development of effective educational materials. Further, the development of educational materials can
be informed by a better understanding of what behavioral mechanisms are more likely to be effective in assisting users’ behavior. At the same time, educational materials can offer a testbed to validate hypotheses about the effectiveness of nudging interventions.

Further, ICRI can provide OWK with the necessary input to make sure that these materials contain a basic outline of the rights and obligations of SNS users and providers. Youngsters can learn from the educational material not only about the legal requirements that bind SNS providers, but also about those requirements which require them to act responsibly when relating to others.

Next, input from technical partners with respect to tools that are available for the users to protect their privacy, as well as the practicalities when using these tools, can provide rich input for educational material. OWK can also provide feedback to the different technical partners with respect to notions that are difficult for users to understand or risks that are hard to assimilate. Such feedback can have a direct impact on the design of privacy enhancing technologies (PETs).

OWK’s work on user education on the social web is of even greater relevance to DTAI. DTAI’s objective is to develop solutions for secure and privacy-preserving systems that are interactive and educational. OWK’s study on empirical approach will be used to define evaluation principles from an educational perspective. These principles can be used to enforce the evaluation framework used by DTAI.

SMIT shares the common challenge with OWK and CMU of empowering the user, in this case youngsters. OWK’s objective to fill in the gaps of existing educational packages is a practical one that can benefit from the community oriented work done by SMIT. In developing a privacy manual and educational packages it is important to know how youngsters interpret risks regarding privacy on SNS. Through qualitative ethnographic studies SMIT hopes to grasp the meaning of these concepts for the users. In contrast, OWK starts from a quantitative perspective. A combination of the qualitative and quantitative methods can nicely complement each other throughout the project.

Comparable to OWK’s focus on raising awareness through educational material, DTAI focuses on raising user awareness on privacy issues and decision making through feedback tools. In this field, three main gaps are evident from the state of the art of Privacy Feedback and Awareness (PFA) tools. First, today’s PFA tools mostly focus on one aspect of privacy concerns, such as user’s browser fingerprint, user’s SNS profile visibility, user’s photos’ geo-location data, etc. The intrinsic inter-relatedness of different privacy concerns, and a comprehensive perspective on user’s privacy is either being ignored or underestimated. Second, Artificial Intelligence (AI, e.g. data mining, machine learning) has proven fruitful when used for PFA applications (Privacy Wizard), but integration of AI into today’s PFA tools is quite limited. Yet, the large quantity and complexity of user’s privacy data and actions on the web demand such integration. Third, in the absence of unique usage scenarios and clear goals, evaluations of PFA tools remain difficult.

There are a number of ways the challenges that DTAI is addressing can be tackled together with the SPION partners. SMIT’s work makes a distinction between society and community. The state-of-the-art demonstrates that, although debated and contested, the concept of community is important and necessary to simulate the social ecology on the web. The gaps in today’s architecture of SNS such as context collision, forced disclosure, etc., make it hard for users to represent themselves and create communities. PFA tools seem fit as a part of the solution to this problem. In order to develop appropriate tools that take into consideration the privacy concerns and activities of the users, the user community perspectives that will be studied by SMIT are essential. On the other hand, the evaluation framework developed by SMIT in order to analyze the different privacy enhancing solutions in real-life scenarios from both an individual and a community perspective can contribute to the evaluation gap that all technical partners are facing, including DTAI.

In addition, CMU will study users’ behavior and associated biases in Web 2.0 applications. Through this research, CMU hopes to identify and test nudges that will help users make online decisions that improve their satisfaction and well being. Although several streams of privacy, security, and HCI research are clearly related to research on nudging and soft paternalism (for instance: usable security and persuasive computing), few are the manuscripts that explicitly focus on soft paternalistic approaches to privacy decision making. The challenge we face, therefore, is to significantly expand
our poor understanding of what forms of nudges can be effective in influencing disclosure and privacy behavior — as well as when, how, and under what conditions they should in fact be used. This type of research interacts and overlaps with the research goals of DTAI, but also with the others.

The outcome of CMU’s work on nudging can also be very useful in developing user models and designing privacy tools. “Giving feedback and raising awareness about people’s privacy” is to some extent a precondition to “privacy nudging”. On the one hand, the target group of PFA models and applications is the social-web users, whose mindset and behavior have a direct effect on the development of PFA tools; on the other hand, the end results of people testing and using PFA tools can enrich CMU’s research context, while contributing to the evaluation of these tools. Moreover, analyzing the user behavior in the presence of PETs that are supposed to protect their privacy is a challenge and may bring interesting feedback to the design of PETs. Developing tools that can integrate both technological protection and behavioral nudges has great potential towards increasing the effectiveness of privacy preserving tools. Additionally, providing technology aware nudging may be an interesting area to increase the efficacy of both PETs and nudging.

10.1.2 Privacy Settings, Discrepancies and Control Aspects

Another theme that cuts across the different disciplines is how to address the discrepancy between SNS users’ expectations and actual dissemination of their data objects given the technical design of social networks. While this may seem like a technical point, i.e., a matter of access control, the partners recognize that the problem has many facets and can be tackled through the methods and theories of the different disciplines represented in the project.

As CMU also argues in their chapter, users might not be able to achieve the privacy control level they require via composing various complex access control policies. This problem has to be addressed from at least two different angles. The first is enhancing access control models and their policies composition to be more flexible and easy to manipulate by different users. Research on enhancing fine-grained access control models is conducted by DISTRINET. The second is the integration of privacy awareness and feedback tools to enable users to better assess and understand the privacy consequences of their access control policy decisions. This yields an interesting collaboration between DistriNet and DTAI to integrate tools from the two disciplines to enrich access control decision making with feedback and awareness of their activities and to grasp a better understanding of their consequences in online social networking life. PFA tools can be built on top of access control models for users to understand, analyze and assess the privacy consequences of their access control policy decisions. Another challenge that both DTAI and DistriNet can meet is the frequent conflict between accountability and anonymity. Collaboration on finding the trade-off to this conflict is an interesting topic of research.

COSIC outlines three major areas of research focus: graph anonymization, content protection and traffic data protection. These areas provide the foundational models for DTAI’s awareness solutions, especially when these require that users are more aware of the the possible inferences and they have to make decisions concerning their privacy. Furthermore, the gap in privacy metrics is addressed by COSIC, which is also an important part of the PFA evaluation. Ideas on this topic can be a point of collaboration between COSIC and DTAI.

Additionally, COSIC focuses on traffic data protection research and measuring the degree of privacy a user enjoys based on metrics. The targeted design assumes that the user will not notice the existence of the plugin that is developed. SMIT can use their methods to evaluate what will trigger the individual to use such PETs in the first place. Moreover, SMIT will explore what offline processes can stimulate the usage. One challenge COSIC is confronted with is that users are not aware of the capacities of the central instance holding the information they share with others on the platform they use, and of the implications of such data collections. With respect to such central instances, there is currently little room for active control other than disclosure avoidance. Maybe offline processes can stimulate the online behaviour of users to use PETs to communicate in a manner that keeps the communication confidential towards the central instance. This can be a challenge for both SMIT
and COSIC to explore. A better understanding of behavioral and cognitive factors that affect privacy decision making in SNS is also collinear with the goal of modeling user behavior accurately. The input from CMU on this topic is important for the design of PETs as proposed by COSIC.

Transparency and awareness tools can also be essential to provide the user with up-to-date information about her privacy online, what information she is sharing, with whom, where the information flows and so on. When privacy preserving tools like those proposed by COSIC are in place, this task is more important than ever. These tools usually perform certain activities in the background that non-savvy technology users may not be aware of. In this case, PFA tools may be used to provide users with a means to know to what extent they are being protected by the tool as well as the visible effects of that protection (e.g., fake messages posted for the sake of concealing real messages).

To contribute to an access control design that supports users in making informed decisions, input from CMU about how to manage data disclosures in different social contexts and what kind of nudges to employ may be of great value. Further, the research on access control decisions can expose users’ hurdles, and therefore identify areas where nudging intervention may be useful. Moreover, with the results from SMIT focusing on integrating notions of social identity and community into online life, it will be possible to fill the gaps identified by OWK with respect to SNS-specific educational packages that provide an overview of privacy settings and relevant feedback.

Access control models have specific approaches to representing data objects and users. RBAC and ABAC categorize users according to roles and attributes, which might not always be applicable in a social network context. This is especially true if we consider the dynamic groups and interrelationships among users. ReBACs access rules are based on dynamic relationships, however, this poses a challenge with respect to controlling the disclosure of relationship information during access control enforcement. The question of whether and how relationship based access control models can be enforced while controlling the disclosure of information about relationships draws another line of interaction with COSIC’s research on traffic analysis attacks and confidentiality of sensitive relationships. Further, whereas DistriNet approaches access control along the lines of traditional access control tools Stiegler [1979], COSIC performs access control along the lines of confidentiality based on cryptography. Therefore, the combination and mutual understanding of both approaches may bring great opportunities to both areas of research.

One of the most prominent challenges of access control enforcement refers to the fact that relationships and communities evolve throughout time while the policies do not. Consequently privacy is not preserved in the manner that the user initially desired. When some objects are shared among more than one user, there are many challenges in striking a good balance between all stakeholders’ privacy preferences. In this context, just like DTAI, DISTRINET is developing a tool that stimulates active behaviour of the individual. Therefore, it is necessary to explore what capabilities the users have and are willing to use. The idea of relationship based access control models is to take into account social processes in enhancing privacy, beyond the individual. Maybe, by grasping offline relationships the online ones can be better adjusted. This can be a common challenge for both SMIT and DISTRINET to study.

DISTRINET’s work on fine-grained access control models and default settings is also closely related to ICRI’s research during the first and last year of the project. ICRI will tackle default settings and access control from a legal perspective, focusing in particular on related issues of informed consent, finality, and confidentiality of processing. DistriNet will receive input from CMU to help ensure that users in fact make well-informed access control decisions. This is also important from a legal perspective, as these decisions may under certain conditions be seen as an expression of the user’s informed consent.

10.1.3 Accountability and Compliance Aspects

While privacy manifests itself as a social and technical topic, it also gets negotiated in policy making and formalized in legal code. This legal interpretation of the topic is influenced by social understandings of privacy as well as technical advances. Concurrently, compliance with legislation requires
rethinking of technologies, organizations and social interactions.

The main task of ICRI is to identify and tackle uncertainties, gaps and shortcomings of the existing (legal) framework of SNS, with a clear focus on enhancing accountability of SNS service providers and mitigating responsabilization of individuals. In order to fully complete this task, it is necessary to grasp the meaning of why individuals in a community setting use SNS and act as they do and how they perceive risks. In this way, the outcome of user research in offline and online communities as planned by SMIT can feed into regulatory discussions and lead to policy recommendations as they will be elaborated by ICRI.

Another manifestation of a crosscutting topic is accountability. Accountability is a joint concern of DISTRINET and ICRI, as both disciplines propose ways of addressing compliance and resulting needs of responsible entities. From a technical point of view, accountability approaches are concerned with ensuring the accountable use of a system through the tracking of user behavior (including the behavior of service providers), while other privacy approaches and research focuses on minimizing the tracking of users. Tracking users to hold them accountable can lead to transparency of activities towards a powerful few, leading to concerns with respect to anonymity requirements. These requirements, i.e., anonymity and accountability, may lead to conflicts with access control models e.g. k-anonymity Attribute-Based Access Control [Squicciarini et al., 2007]. DISTRINET plans to develop mechanisms which can help ensure that in case of a privacy breach, wrongdoers can be held accountable. This work is closely related to ICRI’s research concerning the obligations, liability and accountability of actors involved in SNS. While the direct overlap will be more limited than for access control and default settings, both partners will still be able to mutually inform each other’s work.

DTAI’s work on feedback and awareness tools will allow SNS users to readily ascertain how their data is being processed (with whom it is being shared, under which policies, etc). Such a tool has the potential to provide SNS users both with enhanced transparency as well as greater ease when seeking to exercise their rights as data subjects. This could be a win-win situation for both users and data controllers: users may easily obtain additional information, whereas data controllers may, at least initially, see the administrative burden for dealing with such requests reduced. In order to reach these objectives, DTAI and ICRI will work together closely to ensure that relevant information is included in the feedback tools. The transparency obligations of data controllers will be elaborated in the context of ICRI’s research.

DTAI may also benefit from ICRI’s expertise not only on the legal definitions of personal data, identifiable person, processing of personal data, etc. but also the data protection principles in general. With the input of ICRI, DTAI may identify the entities/actors and their responsibilities involved in a certain personal data processing operation. Based on this information, DTAI can better categorize the target user group and tailor functionalities accordingly. Detailed cases such as location-based services, cookies, spam, etc. will find their legal foundations in specific directives.

CMU aims to look at mechanisms which help ensure that users make well-informed access control decisions and adopt those settings which best match their actual preferences. This is important from a legal perspective, as these decisions may under certain conditions be seen as an expression of the user’s informed consent. The role of biases in decision making in return has implications for the debate on regulatory vs. self-regulatory solutions of privacy problems in online social networks, and therefore may inform research on the validity of informed consent and default settings in social networking sites.

Generally, the technical partners expect from ICRI assessments regarding the legal compliance of the devised technologies and the possibilities. The tools developed by COSIC seek to safeguard a number of core data protection principles; such as data minimization and confidentiality and security of processing. ICRI will be in close collaboration with them and provide them with up-to-date information regarding data protection and processing that may affect the design of effective PETs.
10.1.4 Social and Community Aspects

Discussions on privacy in social networks often depict the problem as one of a utility oriented individualistic decision process. Such framing of the problem may intensify the responsibilization of individuals, one of the main issues that we address with the project. It is also possible to argue that privacy concerns, rights and decisions are embedded in economic, socio-technical systems, and in social processes, and these constrain and enable individual and collective decision making. Given this perspective, one of the objectives of the project is to bring in the social processes into the analysis of privacy in online social networks by looking at how communities manage privacy offline and how these can be translated into the online experience of users of SNS. As brought forth by some of the partners, privacy problems can be distinguished as:

1. Social privacy problems: context collision, forced disclosure, blurring between private and public, the threatening of contextual integrity, invisible audiences, etc.

2. Instrumental privacy: data mining, behavioral advertising, collecting PII, etc.

SMIT is tackling these problems by putting forward community processes. A central assumption of SMIT is that not all individuals are equal and capable to be aware of (all) these privacy issues. Hence, SMIT focuses on the collective and social requirements of technologies. CMU and SMIT share a common challenge of making people more aware of their privacy settings through the actions of the users themselves (empowering). CMU proposes soft paternalism or nudging to make people aware of their privacy on SNS. Instead of relying on the “heavy-handed” principle, the focus is on the user perspective. Moreover CMU’s use of the concept of nudging not only focuses on awareness, but also on acting upon this. Even then, people are not always able to optimally interpret and act upon each (complex) privacy issue on SNS, especially those regarding instrumental privacy. A community perspective can help in achieving this, by focussing on social processes such as trust, and on warm experts.

Further, nudging mainly focuses on how existing technologies can be more transparent for the user. However, if we do not assume “existing technologies” as a given, agency through changing existing structures of technologies may also be included in the nudging strategy. Taken from this perspective, it is possible to think of nudging strategies that promote users to question the underlying structures, in addition to finding an optimal position for their well-being within existing structures. Moreover, nudging could be embedded in a social process whereas the users nudge each other. To understand community processes is one thing, to understand the individual within the group and his cognitive and behavioural biases is another. Whereas CMU focuses on the individual, SMIT focuses on the community. Together this promises to be a complementary and coherent framework for interdisciplinary research.

The technical partners focus on one of the three definitions of privacy: privacy as confidentiality (COSIC), privacy as practice (DTAI) and privacy as control (DISTRICT). One of the objectives in the project is to provide these technical partners with the social requirements that are necessary to develop these privacy-enhancing tools. The development of these technical solutions may benefit an iterative development process that diverges from the water-flow model, where requirements lead to technology which is then introduced to the world. SMIT and CMU will play an important role in this iterative process. Both partners can provide insights into behavioral aspects and compiling social requirements, but also evaluate these tools from a user perspective with respect to privacy, disclosure and identity. Specifically, SMIT could provide insight into how privacy is perceived from the community perspective, how different types of social media can alter the perception of privacy of a user, and how users’ actions impact the whole perception of privacy in the community, amongst many other studies. These insights are expected to generate a better understanding for designers of technologies.

DTAI targets solutions that are geared towards increasing the awareness of individual users. However, in the construction of an identity, it is necessary to include the reflection of our own behavior through the perspective of others. Through this process the self is socially constructed. SNS are in
that sense a simulation of real life processes, but many features (roles, structures, context,...) are not completely present on SNS. By constructing PFA tools that only take into account the relation between “technology and the individual”, a lot of other –offline– factors that can positively stimulate the individual are not considered. This can be a challenge for DTAI to address in collaboration SMIT.

10.2 Future Steps

The next step in the project is to elaborate privacy and security requirements, as well as social and educational requirements, and to analyze how these interact. The points of interaction analyzed above, as well as the extensive study of the state-of-the-art will be used to guide this process.
Bibliography


M. Anwar and P. W. L. Fong. An access control model for facebook-style social network systems. Technical Report 2010-959-08, Department of Computer Science, University of Calgary, Calgary, Alberta, Canada, Department of Computer Science, University of Calgary, Calgary, Alberta, Canada, 2010.


B. A. Bumgarner. You have been poked: Exploring the uses and gratifications of facebook among emerging adults. First Monday, 12(11), November 2007.


Version 1.0


S. Görling. The myth of user education. 2006.


Version 1.0


Insafe. Educational resources for teachers, 2011.


Version 1.0


