

Yaniv Altshuler · Yu
Armin B. Cremers ·
Alex Pentland *Editors*

Security Privacy in

Security and Privacy in Social Networks

Yaniv Altshuler • Yuval Elovici • Armin B. Cremers
Nadav Aharony • Alex Pentland
Editors

Security and Privacy in Social Networks

Editors

Yaniv Altshuler
MIT Media Lab
Cambridge, MA, USA

Yuval Elovici
Telekom Innovation Lab
Information systems engineering
Ben-Gurion University
Beer-Sheva, Israel

Armin B. Cremers
University of Bonn
Bonn, Germany

Nadav Aharoni
MIT Media Lab
Cambridge, MA, USA

Alex Pentland
MIT Media Lab
Cambridge, MA, USA

The book is based in part of works initially presented at the Workshop on Security and Privacy in Social Networks that was held in conjunction with the IEEE Social Computing conference 2012, Cambridge, MA.

ISBN 978-1-4614-4138-0

ISBN 978-1-4614-4139-7 (eBook)

DOI 10.1007/978-1-4614-4139-7

Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2012943943

© Springer Science+Business Media New York 2013

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

| | | |
|----------|--|------------|
| 1 | Introduction to Security and Privacy in Social Networks | 1 |
| | Yuval Elovici and Yaniv Altshuler | |
| 2 | Interdisciplinary Impact Analysis of Privacy in Social Networks | 7 |
| | Michael Netter, Sebastian Herbst, and Günther Pernul | |
| 3 | Recognizing Your Digital Friends | 27 |
| | Patrik Bichsel, Jan Camenisch, and Mario Verdicchio | |
| 4 | Encryption for Peer-to-Peer Social Networks | 47 |
| | Oleksandr Bodriagov and Sonja Buchegger | |
| 5 | Crowdsourcing and Ethics: The Employment of Crowdsourcing Workers for Tasks that Violate Privacy and Ethics | 67 |
| | Christopher G. Harris and Padmini Srinivasan | |
| 6 | The Effect of Social Status on Decision-Making and Prices in Financial Networks | 85 |
| | Yoel Krasny | |
| 7 | Stealing Reality: When Criminals Become Data Scientists (or Vice Versa) | 133 |
| | Yaniv Altshuler, Nadav Aharony, Yuval Elovici, Alex Pentland, and Manuel Cebrian | |
| 8 | Applications of k-Anonymity and ℓ-Diversity in Publishing Online Social Networks | 153 |
| | Na Li and Sajal K. Das | |
| 9 | Links Reconstruction Attack: Using Link Prediction Algorithms to Compromise Social Networks Privacy | 181 |
| | Michael Fire, Gilad Katz, Lior Rokach, and Yuval Elovici | |

| | | |
|-----------|---|------------|
| 10 | An Analysis of Anonymity in the Bitcoin System | 197 |
| | Fergal Reid and Martin Harrigan | |
| 11 | Privacy-Preserving Data Integration | |
| | Using Decoupled Data | 225 |
| | Hye-Chung Kum, Stanley Ahalt, and Darshana Pathak | |

Introduction to Security and Privacy in Social Networks

Yuval Elovici and Yaniv Altshuler

As the area of online social networking develops and many online services add social features to their offerings, the definition of online social networking services broadens. Online social networking services range from social interaction-centered sites such as *Facebook* or *MySpace*, to information dissemination-centric services such as *Twitter* or *Google Buzz*, to social interaction features added to existing sites and services such as *Flickr* or *Amazon*. Each of these services has different characteristics of social interaction and different vulnerabilities to attack.

The value of online social networking sites stems from the fact that people spend large amounts of time on these networks updating their personal profiles, browsing for social or professional interactions, or taking part in social-oriented online applications and events. People nowadays have become immersed in their preferred online social environments, creating an exciting entanglement between their real and virtual identities [1]. However, this immersion also holds great peril for users, their friends, and their employers, and may even endanger national security.

There is much information in the patterns of communication between users and their peers. These patterns are affected by many relationship and context factors and can be used in a reverse direction to infer the relationship and context. Later on, these relationships can be further used to deduce additional private information which was intended to remain undisclosed. A recent study carried out at MIT is said to be able to reveal the sexual orientation of Internet users based on social network contacts. In this example, the users whose privacy was compromised did not place

Y. Elovici (✉)

Telekom Innovation Lab, Information systems engineering, Ben-Gurion University,
P.O.B. 653, Beer-Sheva, Israel
e-mail: elovici@bgu.ac.il

Y. Altshuler

Human Dynamic Group, MIT Media Lab, 77 Mass. Ave, Cambridge, MA 02139, USA
e-mail: yanival@media.mit.edu

this information online, but rather disclosed their social interaction to users who apparently did disclose this information [2].

In other cases, this problem can become even worse due to the (false) assumption of users that information marked as “private” will remain private and will not be disclosed by the network. Indeed, although the operators of social networks rarely betray the confidence of their users, no security mechanism is perfect. Because these networks often use standard (and not necessarily updated) security methods, a determined attacker can sometimes gain access to such unauthorized information. The combination of sensitive private information managed by users who are not security-aware in an environment that is not hermetically sealed is a sure cause of frequent leaks of private information and identity thefts [3, 4].

This problem becomes even more threatening when viewed from the corporate (or even national) perspective. Users who possess sensitive commercial or security-related information are expected to be under strict control in their workplaces. However, while interacting virtually in social networks, these same people often tend to ignore precautions due to a false sense of intimacy and privacy, all the while being unaware of the damage their naive behavior may cause. Because it is hard (and sometimes illegal) to monitor the behavior of online social network users, these platforms constitute a significant threat to the safety and privacy of sensitive information. Hard to detect and almost impossible to prevent, leaks of business, military, or government data through social networks could become the security epidemic of the twenty-first century [5, 6].

This book aims to bring to the forefront innovative approaches for analyzing and enhancing the security and privacy dimensions of online social networks. To facilitate the transition of such methods from theory to practical mechanisms designed and deployed in existing online social networking services, we need to create a common language for use between researchers and practitioners in this new area, ranging from the theory of computational social sciences to conventional security and network engineering.

The rest of this book is divided into three parts covering three complementary themes and is structured as follows. The first part contains four studies that touch on fundamental aspects of security and privacy in social networks, raising and discussing topics such as the conceptual definition of identity in social networks and the interplay between ethics and crowdsourcing. The second part of the book is devoted to innovative mathematical models which link the social dimension of networks to existing privacy and network security issues. This section contains three studies which analyze different domains ranging from mobile networks to financial trading networks and demonstrating the essential differences between security issues in social and non-social environments. The third section focuses on specific case studies and presents an in-depth analysis of three unique examples of how “security-oriented research” is carried out in social domains and how it differs from similar efforts which do not take place in such environments.

Chapter “[Introduction to Security and Privacy in Social Networks](#)” introduces a multidimensional concept of privacy in social networks which delineates aspects of privacy along various legal, technical, and social dimensions. The privacy concept

thus developed is then visualized using tripartite diagrams which provide a quick orientation to this paradigm's strengths and weaknesses as demonstrated in social networks. The chapter then investigates how these properties evolve from the fact that information in the physical world decays over time, while in the online world, information is in principle permanently available. Although this chapter focuses on a qualitative analysis of this topic, a more quantitative metric that would clearly enhance comparability of privacy issues in different social networks and the tracking of improvements over time is envisioned for future development.

A key aspect of social networks is the digital identity (or identities) adopted by users to characterize and recognize themselves and others. At first glance, it may appear that users of social networks treat and use digital identities similarly to their "real-world" identities. However, the absence of physical contact enables people to create several identities, some of which may be anonymous. Furthermore, users of social networks search and acknowledge each other based mainly on attributes that they exchange through the infrastructure of the social network (which in turn can be further used to disguise one's true identity). Chapter "[Interdisciplinary Impact Analysis of Privacy in Social Networks](#)" sheds light on the fascinating topic of digital identities by presenting a basic conceptual framework that analyzes fundamental aspects of the use of identities in social networks and recommends possible methods to improve the use of such identities. The chapter begins by presenting basic concepts related to the differences between digital and real identities, followed by a discussion on the challenges of the digital facet. Next, solutions for security and privacy challenges relating to digital identities are presented. The chapter discusses the perception of the identity of an entity as a notion existing in the minds of other entities. This gives rise to the possibility of multiple identities for a single entity in different contexts, a phenomenon which is called "pseudonymity" and which is possibly or potentially available in the online world more readily than in the real world.

Chapter "[Recognizing Your Digital Friends](#)" presents an overview of the requirements for and comparisons of encryption schemes for social networking services based on a peer-to-peer (p2p) infrastructure (as opposed to centralized server architectures) and describes the challenges of p2p social networking architectures and their high-level requirements. The chapter then discusses the criteria by which p2p encryption systems should be evaluated and compared: efficiency, functionality, and privacy. Four examples of existing p2p social networking architectures are then reviewed (*PeerSoN*, *Safebook*, *Diaspora*, and *Persona*), which focus on encryption as a means of ensuring data confidentiality. This is followed by a comparative analysis of these architectures against the evaluation criteria presented earlier. In addition, this chapter contains a parallel discussion of the differences between broadcast encryption and predicate encryption techniques in the context of the p2p encryption challenge.

The first part of the book concludes with chapter "[Encryption for Peer-to-Peer Social Networks](#)," which thoroughly investigates various ethical issues with respect to the expanding field of crowdsourcing. This highly disruptive field involves the partitioning of a mission into many small pieces, each given to ad hoc employees

using an online platform. The rapid pace of this process enables fast completion of highly complex tasks at extremely low cost. Together with the anonymity of these platforms (which protects the identities of both the employers and the employee), this approach transforms crowdsourcing platforms into the equivalent of a super-computer network for a fraction of the cost. The number of potential applications is boundless, and several ethical questions arise. This chapter reviews recent developments in this area while examining some of these ethical challenges. In addition, chapter “[Encryption for Peer-to-Peer Social Networks](#)” studies the attitude of workers in crowdsourcing platforms (such as *MTurk*, *oDesk*, or *Elnance*) towards performing unethical tasks and asserts that, although many workers in several crowdsourcing platforms studied expressed reluctance to perform unethical tasks, in practice, many workers were willing to accept unethical tasks (especially if they were well paid). Simple but unethical tasks may include breaking into someone else’s email account and sending a fake email on behalf of that person, or faking a review of a commercial service. However, more elaborate large-scale uses may involve activities such as identification of demonstrators by police agencies or dictatorships. Interestingly, the results of an experiment detailed in the chapter hint that the anonymity provided by the crowdsourcing platform, the anticipated task consequence, and gender were not found to be influential. On the other hand, when the amount of monetary compensation offered increased, so did the willingness of workers to perform highly unethical tasks.

The second part of the book is introduced by chapter “[Crowdsourcing and Ethics](#)” and investigates how social networks influence the pricing of assets in the financial market. This influence is a result of the ongoing and unavoidable comparison of relative performance imposed on investors and traders because of the comprehensive integration of social networks into everyday life. Counterintuitively, this abundance of information may sometimes act to suppress of integrity in investment practice by pushing investors to adopt irrational investing strategies. For example, leading investors will in many cases be manipulated into buying risky assets knowingly at inflated prices. This chapter presents a mathematical model that studies these dynamics and suggests that the overpricing of risky assets that is often observed in the market is derived from these “social forces”.

Chapter “[The Effect of Social Status on Decision-Making and Prices in Financial Networks](#)” predicts the existence of new kinds of malicious attacks on communications and on mobile infrastructures that are targeted at extracting, not password or credit card information, but information about the relationships in a real-world social network and characteristic information about the individuals in the network. The chapter discusses the expected features of such attacks and explains the differences between these attacks and traditional types of attacks against data privacy. The chapter then presents a mathematical model of such attacks and predicts that they would be impossible (or very unlikely) to detect using most of the network monitoring tools used today. This problem is caused by the surprising fact that the best strategy for attackers seeking social information and habits is, counterintuitively, a very slow and nonaggressive strategy (in contrast to most of the known malware threats).

Many online social network (OSN) owners regularly publish data collected from their users' online activities to third parties such as sociologists or commercial companies. These third parties further mine the data and extract knowledge to serve their diverse purposes. In the process of publishing data to these third parties, network owners face a nontrivial challenge: how to preserve users' privacy while keeping the information useful to third parties. Failure to protect users' privacy may result in severely undermining the popularity of OSNs as well as restricting the amount of data that the OSN owners are willing to share with third parties. Chapter "[Stealing Reality: When Criminals Become Data Scientists \(or Vice Versa\)](#)" discusses this problem while focusing on the use of classical privacy preservation models originally developed to protect tabular data privacy, such as k -anonymity and l -diversity, to preserve users' privacy in the publication of OSN data. The history of these methods is reviewed, and their applicability is demonstrated.

The third part of the book examines specific case studies regarding the unique features of security and privacy in social networks. This section opens with a discussion of innovative methods for using machine-learning techniques to reconstruct the structure of unknown social networks. Using this method, publicly available information may be used to reveal concealed information, which severely compromises the users' privacy, anonymity, and trust in the network. Chapter "[Applications of \$k\$ -Anonymity and \$l\$ -Diversity in Publishing Online Social Networks](#)" presents the "link reconstruction attack," a method that is capable of inferring a user's connections to others with high accuracy. This attack may be used to detect connections that the user wanted to hide to preserve his privacy. We show that the concealment of one user's links is ineffective if it is not also done by others in the network and we present an analysis of the performance of various machine-learning algorithms for link predictions inside small communities.

In contrast to chapter "[Applications of \$k\$ -Anonymity and \$l\$ -Diversity in Publishing Online Social Networks](#)" which demonstrated an attack that can be executed on social networks to steal private information, chapter "[Links Reconstruction Attack](#)" analyzes this topic from a different angle by studying the *Bitcoin* peer-to-peer monetary exchange system. The degree of anonymity in the *Bitcoin* system, an electronic analog of cash in the online world, is investigated using data from transactions which are publicly available to ensure the integrity of the *Bitcoin* system. Using mainstream methods from network theory, this chapter demonstrates how this anonymous (at least in theory) payment system can be partially de-anonymized. This technique is then used to track the "flow" of large amounts of stolen monetary credits, thus demonstrating how the identity of the users responsible for this theft can be disclosed using this network-based analysis method.

As discussed in previous chapters of this book, integration between several data sources may lead to compromised data privacy through the use of certain network-based analysis methods. Chapter "[An Analysis of Anonymity in the Bitcoin System](#)" is devoted to exploring the record linkage problem and presents a scheme for the maintenance of data privacy when data and records from multiple databases are combined in a way which still allows record-linking information verification

services. The chapter begins by discussing two common modes of operation in this field, the de-identified and the fully trusted mode, and asserts that these approaches do not provide a definitive response to the needs of social data privacy. The chapter then reviews existing techniques and related work on record-linkage and privacy-preserving computations, pointing out the need for a new scheme for representing integrated data. The chapter contains a proposed model for a decoupled data architecture. The main technological concept studied in this chapter is the separation between identifying information and sensitive data, which needs to be protected. In this chapter, it is demonstrated how this decoupled data-access model can provide the same protection as de-identified data while at the same time being able to integrate data to support broad research in computational social sciences in a flexible manner. The study also tested the impact of different mechanisms for hindering inferences of identity when names are revealed for record-linkage purposes.

References

1. Onnela J-P, Reed-Tsochas F (2010) Spontaneous emergence of social influence in online systems. *Proc Nat Acad Sci* 107(4):18375–18380
2. Jernigan C, Mistree BFT (2009) Gaydar: Facebook friendships expose sexual orientation. *First Monday* 14(10)
3. Stana RM, Burton DR (2002) Identity theft: prevalence and cost appear to be growing. GAO-02-363. U.S. General Accounting Office, Washington, DC
4. Gross R, Acquisti A (2005) Information revelation and privacy in online social networks. In: *Proceedings of the 2005 ACM workshop on privacy in the electronic society*, Alexandria, pp 71–80
5. Brunner M, Hofinger H, Krauss C, Roblee C, Schoo P, Todt S (2010) Infiltrating critical infrastructures with next-generation attacks. *Fraunhofer Institute for Secure Information Technology (SIT)*, Munich
6. Krishnamurthy B, Wills CE (2009) On the leakage of personally identifiable information via online social networks. In: *Proceedings of the 2nd ACM workshop on online social networks*, New York, pp 7–12

Interdisciplinary Impact Analysis of Privacy in Social Networks

Michael Netter, Sebastian Herbst, and Günther Pernul

Abstract The rise of the social web has traditionally been accompanied by privacy concerns. Research on social web privacy has been conducted from various viewpoints including legal, social, and the computer sciences. In this chapter, we propose an interdisciplinary approach to capture the multidimensional concept of privacy. For this purpose, we developed a three-layered framework to systematically analyze the privacy impact of various research directions. In addition, we conducted an interdisciplinary literature analysis, highlighting areas for improvement as well dependencies between different research directions.

1 Introduction

Over the last decade, the evolution of the World Wide Web led to the significant growth of Online Social Networks (OSNs), which are receiving much attention in the research community. While social networks have always been an important part of daily life, the advent of Web 2.0 and its easy-to-use services increasingly shift social life to their online counterparts. OSNs provide an infrastructure for communication, information, and self-expression, as well as for building and maintaining relationships with other users.

The increase in relevance and the quantity of social web services has been accompanied by privacy concerns. On one hand, these worries have arisen due to the prevalent oligopolistic social web landscape with only a few service providers possessing large databases with millions of user profiles. On the other hand, privacy concerns focus on the challenges of presenting different facets of the self to different audiences, and to keep those views consistent. While this bears a

M. Netter (✉) • S. Herbst • G. Pernul

Department of Information Systems, University of Regensburg, Regensburg D-93040, Germany
e-mail: michael.netter@wiwi.uni-regensburg.de; sebastian.herbst@wiwi.uni-regensburg.de;
guenther.pernul@wiwi.uni-regensburg.de

resemblance to managing different appearances of the self in the real world, the inherent properties of mediated OSN communication (e.g., the permanency and searchability of personal information) places privacy at risk. Although privacy controls are in place to currently restrict access to personal data, users seem to be shortsighted with respect to future aspects of current behavior [1].

Both aforementioned areas of privacy have been studied extensively by researchers through various viewpoints such as law, the social sciences, and computer science. However, the ambiguous nature of privacy and the multiple definitions available impede a consistent view of the concept. Robert C. Post notes that privacy "... is a value so complex, so entangled in competing and contradictory dimensions, so engorged with various and distinct meanings, that I sometimes despair whether it can be usefully addressed at all." [2].

In this chapter, we stress the need to integrate insights from diverse areas of research on social web privacy. We contribute to this field by providing a framework with which to decompose social web privacy and systematically analyze the effects of different research directions. Subsequently, we applied the proposed framework to the body of research. Our results highlight areas for improvement as well as dependencies between different research directions, emphasizing the necessity to foster interdisciplinary research on social web privacy.

The remainder of this chapter is structured as follows. In Sect. 2, we give an overview of related work. In Sect. 3, we decompose social web privacy and transfer its components into a framework for analyzing the concept from different research directions. We apply our framework on the existing body of research, differentiating between privacy issues related to OSN users and OSN service providers in Sects. 4 and 5, respectively. Finally, in Sect. 6, we summarize our findings and highlight areas for future work.

2 Related Work

In this section, we present existing approaches that aim to integrate several research directions in order to create a holistic view of privacy. Approaches to particular aspects of privacy are discussed in our detailed impact analysis of the various privacy perspectives in Sects. 4 and 5.

Spiekermann and Cranor provide a framework with which to build privacy-friendly systems [3]. They distinguish between privacy-by-policy and privacy-by-architecture. The former is a legally-driven approach that focuses on notifying the user and obtaining consent prior to processing personal data. The latter is a technically-driven approach to minimize the collection of personal data without limiting functionality. However, their approach does not consider the social perspective of privacy and focuses on privacy in general, whereas our work examines social web privacy. The importance of social web privacy is acknowledged by the European Union, which is promoting several related research projects. For

example, PADGETS¹ uses an interdisciplinary approach to strengthen users' privacy while harnessing social network data for policy making. Similarly, the European research project PrimeLife² has developed a framework with which to analyze privacy issues related to other OSN users [4]. Project results show that privacy issues arise when legal or social norms are disregarded or technical safeguards are circumvented. Depending on the owner's initial categorization of personal data (private, semi-public, or public), the PrimeLife framework allows an estimation of potential privacy risks. Unlike our approach, this work does not take privacy threats stemming from OSN service providers into account, but solely focuses on user-related privacy issues. PRESCIENT,³ another EU-funded project, conducted an in-depth study of privacy conceptualizations [5]. It takes a legal, social, economic, and ethical perspective of privacy, highlighting similarities and interdependencies. This project's results provide useful insights to help understand the concept of privacy; however, the analyses do not follow a structured approach, as described in this chapter.

3 Proposed Three-Layered Framework

In this section, we give an overview of our proposed framework. The framework provides a general-purpose structure for social web privacy research domains. Subsequently, the concept of privacy is broken up into a set of characteristics that are used to conduct our impact analysis, as described in Sects. 4 and 5.

3.1 Overview

In their conceptualization of privacy in 1890 as “the right to be let alone,” Warren and Brandeis were one of the first to recognize the multidimensionality of the privacy concept [6]. Until then, privacy threats were primarily related to potential physical harm [7]. The rise of the information age led to a large number of privacy conceptualizations from a variety of directions such as the social sciences, law, architecture, urban design, health sciences, and computer and information sciences. In their work to structure the concept of privacy, Patil and Kobsa introduce three main perspectives from which to describe and analyze privacy [8]:

- *Legal*: This aspect focuses on laws and policies that protect the individual from corporations, governments, and other individuals. For example, the European

¹ <http://www.padgets.eu/>

² <http://www.primelife.eu/>

³ <http://www.prescient-project.eu/>

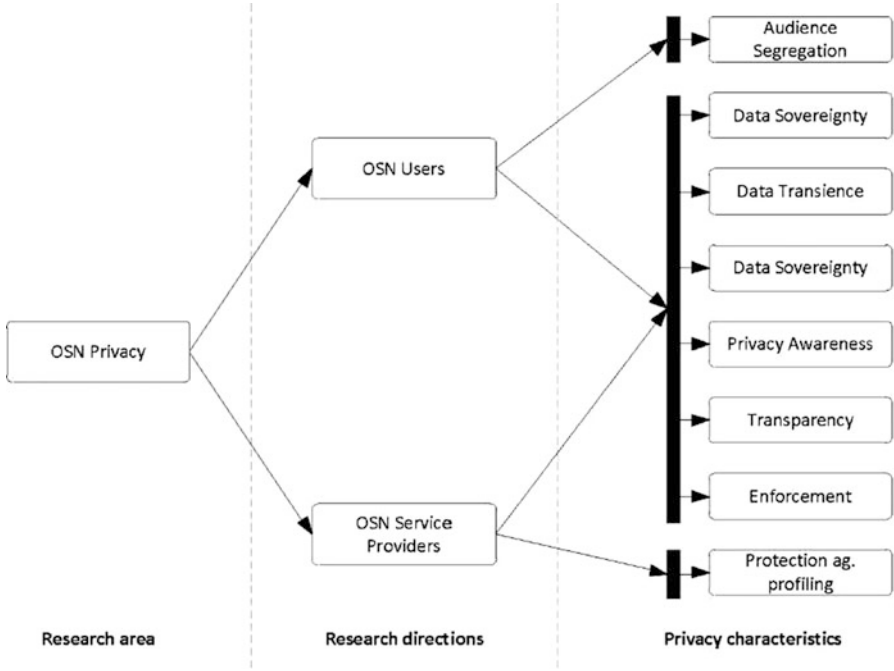


Fig. 1 Classification of OSN privacy research

Data Protection Framework promotes informational self-determination that emphasises an individual’s rights to control the collection and use of personal data [9].

- *Technical:* This aspect translates norms and regulations into technical specifications. The Platform for Privacy Preferences Project (P3P) is a popular example of enhancing the individual’s ability to control information disclosure by technical means [10].
- *Social:* This aspect concentrates on managing social relationships and the boundaries between private and public life. For instance, Nissenbaum describes privacy as contextual integrity, arguing that personal information is published within a well-defined social context [11]. Privacy is breached if personal information is available outside its intended context.

In this study, we adapt this three-layered view and extend it to cover privacy risks in online social networks. Typically, two distinct areas of research can be observed [12, 13] as depicted in Fig. 1:

- *OSN Service Providers:* Research in this direction includes the means to legally bind service providers to comply with current legislation, to increase end-user trust in service providers, and to provide technical safeguards; e.g., by cryptographic or steganographic means [14].

Table 1 Proposed three-layered framework for analyzing social web privacy

| | Privacy issues related to | |
|-----------|---|--|
| | OSN users | OSN service providers |
| Legal | International standards (Organisation for Economic Co-operation and Development (OECD) privacy principles, EU data protection framework), national laws | International standards (OECD privacy principles, EU data protection framework), national laws, privacy policies |
| Technical | Cryptography and steganography, privacy agents, fine-grained access control models, visualization of personal data | Cryptography and steganography, privacy agents |
| Social | Peer-group pressure, trust relationships, tie strength, privacy awareness | Privacy awareness, pressure of the media |

- *OSN Users*: This research aims to recreate the different social contexts of the real world; e.g., by supporting an individual to segment social streams for specific audiences, and by providing the means to have different digital identities [15].

The two aforementioned research directions are combined with the three perspectives on privacy (legal, technical, and social), resulting in our proposed framework. The framework is shown in Table 1, with the cells containing concepts that become relevant for their respective dimension. Note that the three dimensions are not mutually exclusive – they are interdependent. In Sect. 3.2, the two research directions (OSN service providers and OSN users) are further decomposed into a set of privacy characteristics.

3.2 Characteristics Used to Analyze Social Web Privacy

This section outlines fundamental characteristics of privacy derived from a literature review. These privacy characteristics are not exhaustive; rather, they aim to provide a solid foundation for analyzing the impact of the three perspectives on privacy. The characteristics are described in detail as follows.

3.2.1 Data Sovereignty

Data sovereignty describes the extent to which an individual is able to control the processing of his personal data [16]; i.e., his informational self-determination. Personal data in an OSN is typically available in a structured manner and can easily be copied, linked, aggregated, and transferred [4]. Consequently, it is difficult for an OSN user to control the flow of personal information, and thus privacy is placed at risk. The problem increases because the OSN typically lacks the spatial,

social, and temporal boundaries of the real world, which limits the flow of personal information by default [17].

3.2.2 Data Transience

Data transience relates to the loss of personal information over time, which can be considered a typical characteristic of real-world communication [4]. In contrast, the mediated communication of OSNs results in permanent storage of personal information. As Mayer-Schönberger noted, “Since the beginning of time, for us humans, forgetting has been the norm and remembering the exception. [...] Today, with the help of widespread technology, forgetting has become the exception, and remembering the default.” [18]. In addition, this permanency of personal information poses a great challenge to privacy, since we are no longer free to construct our future identities because contradictory information may be available online [19].

3.2.3 Protection Against Profiling

Protection against profiling subsumes an individual’s ability to prevent an adversary from collecting, aggregating, and linking personal data in order to create a digital dossier [20]. Such profiling threats are increased if secondary data such as location (e.g. from mobile phones) and connection logs are linked to existing OSN profiles [21]. The relevance of these threats is underlined by sophisticated attacks such as stealing-reality attacks [22]. The current landscape of social web service providers, with their targeted advertising-centered business models and large identity silos, adds to this threat.

3.2.4 Audience Segregation

Originally developed by Goffman [23], audience segregation states that each individual performs multiple and possibly conflicting roles in everyday life, and it needs to segregate the audiences for each role in a way that people from one audience cannot witness a role performance intended for another audience, thereby keeping a consistent self-image and maintaining privacy [24]. In current OSNs, contacts are typically classified as “friends,” making it difficult to selectively share personal information with a specific group of people. As a result, privacy is threatened because a large audience might have access to personal information.

3.2.5 Privacy Awareness

Privacy awareness encompasses the attention, perception, and cognition of the personal information others have received and how this information is or may be

processed [25]. An individual's awareness of privacy risk is a prerequisite for privacy-preserving behavior.

3.2.6 Transparency

With regard to OSN service providers, transparency describes the user's ability to be informed of processing and dissemination practices [26]. Taking a social point of view, transparency implies the ability of an individual to understand the flow of personal information within an OSN and to recognize contextual boundaries, which is important for contextual integrity [11].

3.2.7 Enforcement

Enforcement is an individual's means to bring his privacy preferences into force. With regard to OSN service providers and OSN users, it describes the extent to which an individual can control adherence to privacy settings and limitations [27].

3.2.8 Summary

Figure 1 provides a summary of the presented characteristics of privacy. Most properties apply to privacy issues related to social web users and service providers; audience segregation only applies to the former, and protection against profiling only applies to the latter.

3.3 Classification Scheme

The analysis of each privacy characteristic is based on a structured scheme. First, legal aspects are analyzed, highlighting their impact on privacy issues related to OSN users and OSN service providers. Second, the effects of existing technical approaches for enhancing social web privacy are discussed. Finally, the implications of social norms on strengthening privacy in a given scenario are examined.

Additionally, for each privacy characteristic, a visualization of the classification and the effect is provided. A tripartite diagram is used to represent the legal, technical and social dimensions. In this diagram, a colored circle represents the impact (dark blue indicates a major impact, mid-blue a medium impact, and light blue a minor impact).

4 Privacy Issues Related to Social Web Users

In this section, we describe an impact analysis of privacy issues related to OSN users. The results are summarized in Sect. 4.7.

4.1 Data Sovereignty

From a legal point of view, laws and policies applicable to governing the exchange and flow of personal information between people are typically not available. Thus, the legal dimension does not contribute to data sovereignty with regard to other OSN users (no impact).

In addition to the legal dimension of data sovereignty, several technical approaches have been proposed to support a context-sensitive disclosure of personal data in an attempt to strengthen data sovereignty. For example, access control models that enable the user to map their real world trust relationships to OSNs have been introduced [28]. Such technical approaches, in general, attempt to recreate real world social norms. Thus, they can be considered a useful means to strengthen data sovereignty, but their overall impact is minor due to their limited supportive character.

From a social point of view, data sovereignty is threatened if personal information is taken out of its intended context. Tagging people on pictures – a common feature of OSNs – is a typical example of losing control of personal data flow. Gross and Acquisti argue that social norms can strengthen data sovereignty if the fine-grained social relations of the real world can be transferred to OSNs, as these foster reliability and predictability in the behavior of other users [20]. However, adherence to social norms highly depends on the trust relationship between two users, which are commonly divided into weak ties and strong ties [29]. Strong ties typically reflect relations with well-known acquaintances, and an abuse of confidence is likely to have a negative impact on the associated real-world relationship [29]. In contrast, studies indicate that users tend to have increasingly weak ties in OSNs, lacking fine-grained social relations [30], [20]. Individuals are commonly viewed as “contacts” or are even called “friends.” Examining the impact on privacy issues related to other OSN users, unauthorized disclosure could primarily be regarded a social problem that relies on strong ties to be effective. As a consequence, the overall impact of the social aspect is medium, due to the aforementioned prevalent weak ties of current OSNs. Figure 2a illustrates our findings regarding data sovereignty.

4.2 Data Transience

Digitally mediated communication differs from real world communication; it adds persistence, searchability, replicability, and scalability by default [17]. However, other OSN users typically cannot be legally forced to delete voluntarily shared

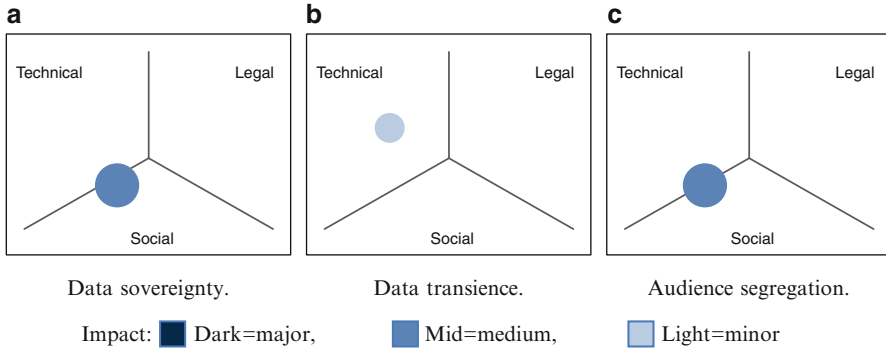


Fig. 2 OSN user privacy analysis (Part 1)

personal information after a given period of time. As a consequence, there is no legal impact on data transience regarding other users.

From a technical perspective, putting an expiry date on personal data is difficult because digital information that is eventually available can easily be copied. While approaches to technical data transience exist, successful attacks, as demonstrated in [31], substantiate their minor impact.

From a social point of view, the permanency of personal information in OSNs poses major challenges. According to Gross and Acquisti, OSN users are typically unaware of existing data storage periods [20]. Consequently, we deduce a lack of social norms regarding data persistence, and conclude that there is no impact stemming from social aspects. A summary of our results is shown in Fig. 2b.

4.3 Audience Segregation

Managing the presentation of the self to different audiences is a social challenge that is not governed by legal regulations (no impact). From a technical perspective, audience segregation is partially implemented in common OSNs (e.g., Facebook Groups⁴ and Google Circles⁵). In addition, audience segregation is starting to gain attention in the research community. The prototypical OSN Clique,⁶ developed within the PrimeLife project, for example, implements a fine-grained access control mechanism to present each audience with a different view on a user's identity [24]. Another approach presented in [32] automatically determines distinct audiences based on the user's relationships. In the current state, a medium impact of audience

⁴ <http://www.facebook.com>

⁵ <https://plus.google.com>

⁶ <http://clique.primelife.eu/>

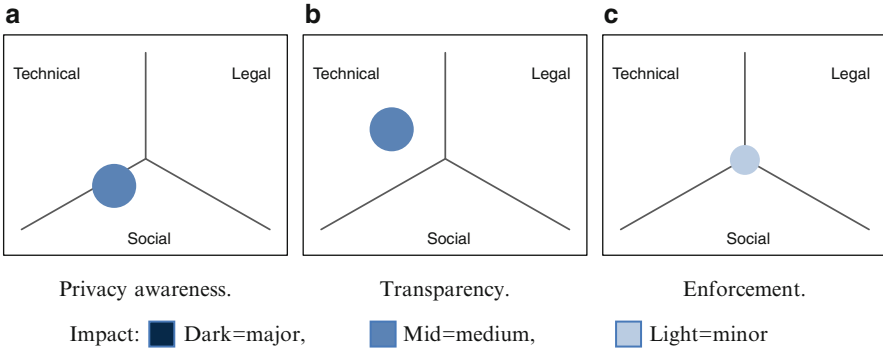


Fig. 3 OSN user privacy analysis (Part 2)

segregation on OSN user privacy can be deduced. However, increasing research activity indicates future growth of the importance of technical means.

From the social point of view, audience segregation is a useful concept that can be used to apply the theory of contextual integrity, as outlined in Sect. 3. Currently, however, audience segregation is not well supported in existing OSNs. Consequently, users resort to behavioral strategies such as choosing appropriate communication channels (e.g., private messages) and to mental strategies (e.g., self-censorship) [33]. Studies show that managing different audiences is a burden to many users, and is rarely applied [34]. Based on the results of the aforementioned studies, only a medium level of social impact of audience segregation on privacy can be inferred, as shown in Fig. 2c.

4.4 Privacy Awareness

Awareness is an important requirement of social web privacy that affects many of the characteristics presented in Sect. 3. However, from a regulatory point of view, OSN user awareness cannot be legally enforced (no impact).

Technical aspects such as usable user interfaces influence perceived privacy protection and the awareness of privacy risks [35]. However, similar to previous characteristics, technical aspects only have a supportive character with which to facilitate privacy awareness and draw attention to potential privacy violations (minor impact).

Privacy awareness is primarily a social concept with a gap existing between theoretical and practical privacy awareness [26]. Privacy awareness is backed by further studies indicating that OSN users frequently underestimate privacy risks and rarely use the available privacy settings [20, 36]. According to Acquisti, immediate gratification outweighs long-term privacy risk and leads to a myopic evaluation of privacy risks [37]. As illustrated in Fig. 3a, there is a medium level of social impact on privacy protection from other users due to the discrepancy between the theoretical and practical effects of privacy awareness.

4.5 *Transparency*

Although similar to privacy awareness, transparency aims to enhance a user's understanding of the propagation of personal data within an OSN to better protect the data from unauthorized access. From a legal perspective, an individual has few means with which to force other users to make their spreading of others' personal data transparent because, typically, no applicable regulations exist.

Taking a technical point of view, transparency-enhancing approaches focusing on logging and retrospective analysis of personal data disclosures have been proposed [38]. Additionally, it has been shown that weak ties and loose sharing preferences (e.g., friend-of-a-friend) may lead to a large personal network and non-transparent personal data spreading [20]. Technical approaches to visually improving personal network transparency have been proposed, underlining that transparency strongly depends on the OSN service provider and related application programming interfaces (APIs) [39]. Following this reasoning, we assigned a medium level of technical impact because many transparency mechanisms rely on APIs that are provided by OSN service providers.

Similar to the legal dimension, the spreading of personal information by other OSN users is typically not governed by social norms, leading to no social impact on transparency. The results of our analysis of data transparency are shown in Fig. 3b.












4.6 *Enforcement*

The enforcement of law is an inherent property of any legal system. In the context of social web privacy, an individual can seek an injunction if reputation-damaging information is published. However, legal remedies are not universally applicable to the social web. Following the European Court of Justice, legal protection requires personal information to be restricted to close friends and family members in order to be applicable [40]. In addition, legal remedies only allow the suing others after a privacy breach, thereby resulting in a minor overall impact of legal enforcement on privacy protection against other users.

A technical means of redress may have a positive impact on the enforcement of legal remedies. However, current OSNs differ widely in providing the technical means to address problems (e.g., cyber-bullying) [41]. Thus, technical means are considered to have only a supportive function with minor impact.

In investigating privacy enforcement from a social perspective, tie strength plays an important role. In some cases, a specific group of an individual's OSN (e.g., family members) may have established social norms that allow each member to employ peer-group pressure to enforce privacy interests [42]. Following the reasoning in [20] that relationships in OSNs often consist of weak ties, the effect of social norms on the enforcement of peer pressure can be considered minor. Figure 3c summarizes these findings.

Table 2 Summary of OSN user-related privacy impact analysis

| | Data sovereignty | Data transience | Audience segregation | Privacy awareness | Transparency | Enforcement |
|-----------|---|---|---|---|---|---|
| Legal | | | | | |  |
| Technical |  |  |  |  |  |  |
| Social |  | |  |  | |  |

Impact:  Dark = major,  Mid = medium,  Light = minor

4.7 Summary

Table 2 summarizes the results of our impact analysis using the proposed framework. This section has described how privacy protection from other social web users is predominately covered by social norms. This corresponds to the real world, where users mainly rely on selective sharing of personal data and highly differentiated relationships to ensure privacy. The mediated nature of OSNs (e.g., permanent storage and searchability of personal data) adds a new layer of complexity that influences privacy because the informational environment of OSNs is counterintuitive to the norms of personal data distribution in the real world. This often leads to a violation of contextual integrity [43]. Table 2 shows that technical approaches to privacy can be seen as a supportive means to translate social norms to the OSNs with potentially increasing importance in the future. On the contrary, legal measures play a minor role and are a last resort to retroactively punish privacy violations. These observations correspond to those of Strahilevitz, who suggested that the law does little to shape people’s actual expectations of privacy [44].

5 Privacy Issues Related to Service Providers

Following the analysis of privacy issues related to social web users, we considered the impact of service provider-related privacy issues in this section. These results were then summarized and integrated into our framework.

5.1 Data Sovereignty

To ensure data sovereignty, legal norms have been enacted to control the exploitation of personal data by OSN service providers [40]. For instance, according to the German Teleservices Act and the Federal Data Protection Act, service providers require a user’s explicit consent to use personal data for advertising purposes [40]. Furthermore, legal requirements for OSN service providers comprise the secure

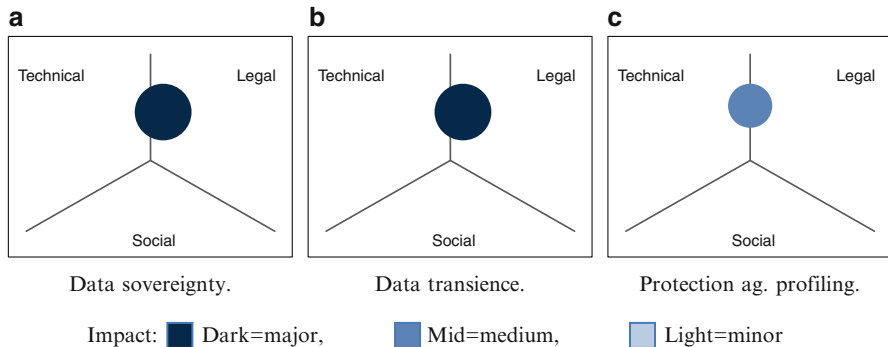


Fig. 4 OSN service provider privacy analysis (Part 1)

storage of personal data and exclusion of search indexes by default. Consequently, legal aspects have a high impact on strengthening an individual's data sovereignty.

From a technical point of view, several approaches to facilitate data sovereignty have been proposed (e.g. [14, 45]). These approaches rely on cryptographic and steganographic means to effectively protect an individual's personal data from service provider access. Although they can easily be integrated into current OSN, they commonly infringe the service provider's general terms and conditions because their business model typically relies on free access to personal data for advertising purposes [4]. Hence, despite the theoretical effectiveness of the aforementioned approaches, the practical difficulties lead to only a medium level of technical impact on data sovereignty.

Commonly, OSN users do not have any social relationship with OSN service providers. As a consequence, an individual cannot rely on social means to ensure service provider adherence to data sovereignty. Therefore, there is no impact from this dimension. Figure 4a shows that data sovereignty with regard to OSN service providers is mainly legally driven with a medium level of technical influence.

5.2 Data Transience

Similar to data sovereignty, data transience is fully covered by legal norms and regulations to be fulfilled by OSN service providers. Providers are required to entitle a user to delete all personal data stored in a OSN profile and to cancel his membership [40]. Similarly, the European Data Protection Framework requires personal data to be removed if the purpose for which the data was collected ceases to exist [9]. This places the user in a strong position and leads to a high legal impact on data transience.

Approaches described in [31] can be applied to technically enforce data transience with respect to OSN service providers. However, their general impact can be