Privacy-preserving Infrastructure for Social Identity Management

Dissertation zur Erlangung des Grades eines
Doktors der Wirtschaftswissenschaften (Dr. rer. pol.)

eingereicht an der
Fakultät für Wirtschaftswissenschaften
der Universität Regensburg

vorgelegt von
Michael Netter

Berichterstatter:

Prof. Dr. Günther Pernul

Prof. Dr. Guido Schryen

Tag der Disputation:

25. Juli 2013

Berichte aus der Wirtschaftsinformatik

Michael Netter

Privacy-preserving Infrastructure for Social Identity Management

Shaker Verlag
Aachen 2013

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

Zugl.: Regensburg, Univ., Diss., 2013

Copyright Shaker Verlag 2013
All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publishers.

Printed in Germany.

ISBN 978-3-8440-2216-2
ISSN 1438-8081

Shaker Verlag GmbH • P.O. BOX 101818 • D-52018 Aachen
Phone: 0049/2407/9596-0 • Telefax: 0049/2407/9596-9
Internet: www.shaker.de • e-mail: info@shaker.de

Foreword

Participation in Social Network Sites (SNSs) has dramatically increased in recent years in both personal and professional life. Today, services such as Facebook, Twitter, and Google+ allow hundreds of millions of individuals to create and maintain online profiles and to share personal information with their contacts as well as with strangers. A recent study on privacy in SNSs shows that 94 % of the users regard protection of personal information as important. However, at the same time, nearly 60 % of the users state that they have lost control of what kind of personal information they have disclosed, how it is collected and used on the Internet, and what effect different SNS privacy settings have. Users, often willingly, share personal identifying information about themselves but do not have a clear idea of who can access their private information today and in the future. There are several factors contributing to this apparent lack of control. The most important might be: (a) The business model of most SNSs, which is built around collecting personal information, analyzing the user's behavior, aggregating and using the data for customized online marketing, and even selling the data to third parties such as advertisers and application developers. (b) Complex privacy settings that are difficult for users to understand and change make it virtually impossible for the average user to exercise actual control over his/her information even in a working environment he/she is familiar with. (c) Missing cooperation of SNS service providers, who often only after public pressure adopt privacy components into their service infrastructures.

In order to enhance privacy for SNSs users, in this book a vision and corresponding technology is developed. The vision consists of a conceptualization of privacy, including answers to questions such as what are the theoretical foundations of Social Identity Management (SIdM), what are strengths and weaknesses of current systems, what are the requirements for privacy-preserving SIdM and how can a corresponding process and its activities be designed and structured. The technology developed includes a detailed mapping of the requirements onto technical components, their integration into a privacy-preserving infrastructure for SIdM as well as its prototypical implementation.

In addition to its original contribution to new knowledge, this book also covers the state-of-the-art and recent innovative developments in privacy-enhancing technology for social networking. Whether you are a member of the staff of a SNS service provider, who is willing to address privacy challenges, a interested student or researcher, or even a general user of a SNS and interested and concerned about your privacy, this book is highly recommended for study because it will provide you with a comprehensive treatment of all the major challenges involved.

Regensburg, August 2013

Prof. Dr. Günther Pernul
Department of Information Systems
University of Regensburg, Germany

Acknowledgments

Working on a PhD is inherently a solitary endeavor. However, as with most creative processes, this could not have been done without the support of many people around me to whom I would like to express my deepest gratitude here.

First of all, I would like to thank my advisor Prof. Dr. Günther Pernul for his help on every step of the way, from providing me with this wonderful opportunity to pushing me to finally finish the dissertation. The freedom he gave me allowed me to work on a topic of my own choosing and explore my research interests. At the same time, he led me back on track when necessary and helped me to grow as a researcher. I also thank my second advisor Prof. Dr. Guido Schryen for his insightful feedback and constructive comments on this dissertation.

The Department of Information Systems has provided a stimulating environment in which to work and my thanks go to those who helped me along the way. In particular, I thank Dr. Ludwig Fuchs for sharing the office with me for more than two years, for tirelessly proofreading this work and commenting on various drafts, but above all, for his friendship. I am grateful for our countless discussions, his constructive criticism and especially for improving the presentation of my ideas. I have also been fortunate to work with Dr. Moritz Riesner. His remarkable ability to distill the essence of a problem and understand its nature while I was still stumbling in the dark has never ceased to amaze me and has changed many of our publications for the better. At the same time, his dry sense of humor made our collaboration even more enjoyable. I further wish to thank Michael Weber for being my office mate and a friend for the last two years. Thank you for patiently discussing much of the content of this dissertation and your proofreading. Besides, I am indebted to Dr. Christoph Fritsch, who has always been a source of sound advice and who helped me with several programming issues. I further thank Sabri Hassan for being both an enthusiastic supervisee and colleague, whose work has greatly influenced this dissertation. Moreover, I owe thanks to the students at the University of Regensburg for being a source of inspiration and for offering external viewpoints. I especially want to mention Christian Richthammer, Sebastian Herbst, Tobias Amann, Tobias Burger, Tobias Glas, Sebastian Gottwald, Paul Krugel, Johannes Sänger, Santiago Suppan, and Tao Yang.

Finally, I owe thanks my family. My parents, who unknowingly put me on the road to this PhD by raising my interest in computers in the late 80s. I am deeply grateful for their patience, encouragement and loving support especially during challenging times, allowing me to pursue my interests and trust that it will all work out OK. Moreover, I thank my sister Julia for the many hours she spent proofreading this work and for positively influencing many aspects of my writing.

Regensburg, August 2013 Michael Netter

Contents

1 Introduction 1

1.1 Motivation and Background . . . 1

1.2 Research Questions . . . 3

1.3 Research Methodology . . . 5

1.3.1 IS Research Paradigms . . . 5

1.3.2 Design Science . . . 6

1.3.3 Design Research Process . . . 8

1.4 Research Overview . . . 10

1.5 Chapter Structure . . . 13

1.6 List of Publications . . . 15

I Fundamentals 17

2 Social Web 19

2.1 Web 2.0 . . . 21

2.2 Social Software . . . 22

2.2.1 Characteristics and Existing Challenges . . . 23

2.2.2 The PADGETS Project . . . 24

2.2.2.1 Concept, Methodology, and Platform . . . 24

2.2.2.2 Security and Privacy as a Major Focus . . . 26

2.3 Social Network Sites . . . 27

2.3.1 Defining Social Network Sites . . . 28

2.3.2 Evolution of SNSs in Practice and Research . . . 29

2.3.3 SNS Characteristics . . . 31

2.3.3.1 SNS Architectures . . . 31

2.3.3.2 SNS Stakeholders . . . 33

2.4 SNSs in the Context of this Work . . . 35

3 Privacy 37

3.1 Privacy Paradigms . . . 38

3.1.1 Legal Perspective . . . 39

3.1.2 Technical Perspective 40
3.1.3 Social Perspective 40
3.2 Privacy Adversaries 41
3.2.1 SNS Service Providers 41
3.2.2 SNS Users 42
3.3 Privacy Objectives 42
3.3.1 Awareness 43
3.3.2 Control 44
3.4 Framework for SNS Privacy 44
3.4.1 Privacy Protection Strategies from the Viewpoint of SNS Users 44
3.4.1.1 Privacy Protection Against Other SNS Users 45
3.4.1.2 Privacy Protection Against SNS Service Providers . . 46
3.4.2 Framing Privacy for this Work 47

4 Identity Management 49

4.1 Digital Identity Management 50
4.1.1 Basic Building Blocks of Identities 50
4.1.2 Identity Models 52
4.1.3 IdM Functions and Technologies 53
4.1.4 IdM Types 55
4.2 Social Identity Management 56
4.2.1 Theoretical Foundations of SIdM 57
4.3 SIdM Concepts on SNSs 59
4.3.1 Identity Management on SNSs 59
4.3.2 Relationship Management on SNSs 62
4.3.3 Access Management on SNSs 65
4.3.4 Context on SNSs 68

II Requirement Elicitation and Related Work 71

5 Requirements for a Privacy-preserving SIdM Infrastructure 73

5.1 Requirement Elicitation 74
5.1.1 Identity Management Requirements 74
5.1.2 Relationship Management Requirements 75
5.1.3 Access Management Requirements 76
5.1.4 Infrastructural Requirements 77
5.2 Literature Review and Comparision 78
5.2.1 Methodology 78
5.2.1.1 Input 78
5.2.1.2 Processing 79
5.2.1.3 Output 80

5.2.2 Results of the Comparison 80
5.2.2.1 Limitations 80
5.2.2.2 Identity Management Shortcomings 81
5.2.2.3 Relationship Management Shortcomings 81
5.2.2.4 Access Management Shortcomings 82
5.2.2.5 Infrastructural Shortcomings 83

6 Impact of SNS-mediated Communication on SIdM 85
6.1 Visibility Dimensions and Information Sharing Domains 86
6.1.1 Default Information Spreading 87
6.1.2 Active Information Sharing 88
6.1.3 Past Information Availability 88
6.2 Conception and Design of the Study 89
6.2.1 Setup Phase 90
6.2.2 Default Information Spreading Questionnaire 91
6.2.3 Active Information Sharing Questionnaire 92
6.2.4 Past Information Availability Questionnaire 93
6.3 Results of the Study 94
6.3.1 Demographics and Data Statistics 94
6.3.2 Default Information Spreading 95
6.3.3 Active Information Sharing 97
6.3.4 Past Information Availability 99
6.3.5 Discussion of Results 100

III Proposed Privacy-preserving Social Identity Management Infrastructure 103

7 Privacy-preserving SIdM Fundamentals and Concepts 105
7.1 Information Systems Improvement 106
7.1.1 Theoretical Considerations 106
7.1.2 Principles 107
7.2 Pillars of a PPSIdM Infrastructure 108
7.2.1 Autonomous and User-controlled Environment 109
7.2.2 SIdM Tools 111
7.3 SIdM Process Model 112
7.3.1 Identity Management Process 114
7.3.2 Relationship Management Process 115
7.3.3 Access Management Process 116
7.3.4 Translation Process 116
7.3.5 SNS Integration Process 117

8 Privacy-preserving SIdM Infrastructure 119

8.1 Overview . . . 119

8.2 PPSIdM Provider Characteristics . . . 122

8.2.1 URL-based Identities . . . 122

8.2.2 Contact Authentication and Authorization . . . 123

8.2.3 SNS Integration Concept . . . 125

8.2.4 SNS Integration Workflow . . . 126

8.2.5 Summary . . . 126

8.3 Identity Facet Manager Component . . . 128

8.3.1 Design Principles . . . 128

8.3.2 Status Quo . . . 129

8.3.3 A Generic Model for Social Identity Representation . . . 130

8.3.4 Summary . . . 132

8.4 Assisted Audiences Component . . . 132

8.4.1 Data Analysis Fundamentals . . . 133

8.4.2 Import . . . 136

8.4.3 Clustering . . . 136

8.4.3.1 Relationship-based Assisted Audiences . . . 136

8.4.3.2 Access Policy-based Assisted Audiences . . . 139

8.4.3.3 Combining Both Approaches . . . 141

8.4.4 Manual Refinement . . . 142

8.4.5 Summary . . . 142

8.5 CA-RBAC Component . . . 142

8.5.1 Status Quo . . . 142

8.5.2 Dynamic and Context-aware Access Control . . . 143

8.5.3 Conceptualization of Context on SNSs . . . 145

8.5.4 Defining Contextual Constraints . . . 146

8.5.5 Conceptual Overview of CA-RBAC . . . 147

8.5.6 Example of Sharing of Personal Information using CA-RBAC . 149

8.5.7 Summary . . . 150

8.6 Access Policy Visualization Component . . . 150

8.6.1 Status Quo and Current Challenges . . . 151

8.6.2 Theoretical Considerations . . . 152

8.6.3 Information Visualization . . . 153

8.6.4 Access Control Policy Visualization . . . 154

8.6.4.1 Disclosure Browser . . . 155

8.6.4.2 Access Policy Grid . . . 156

8.6.5 Summary . . . 159

9 Prototype 161

9.1 Architectural Overview . . . 162

9.2 User Interface . . . 162

9.3 Components . . . 165

9.3.1 Identity Facet Manager . . . 165

9.3.2 Assisted Audiences . . . 167

9.3.3 CA-RBAC . . . 170

9.3.4 Access Policy Visualization . . . 171

9.3.5 Web Services . . . 174

9.3.6 Browser Plugin . . . 176

IV Evaluation and Conclusions 179

10 Evaluation 181

10.1 Scenario-based Evaluation . . . 182

10.1.1 Identity Facet and Attribute Management . . . 184

10.1.2 Relationship Management . . . 186

10.1.3 Access Management . . . 188

10.1.4 Translation Phase . . . 191

10.1.4.1 Intra-Identity Facet Review . . . 191

10.1.4.2 Inter-Identity Facet Review . . . 192

10.1.5 Existing SNS Landscape Integration . . . 194

10.1.6 CA-RBAC Evaluation from the Perspective of Contacts . . . 197

10.2 Requirements Framework Comparison . . . 198

11 Conclusions 205

Appendices 209

A AX Schema Attribute Types 211

B OAuth Application Scenario Responses 215

Bibliography 217

List of Figures

1.1 Complementary research cycle of behavioral science and design science (based on [HC10]) 6
1.2 IS research framework (adopted from [HMPR04]) 7
1.3 Conceptualization of the problem space 11
1.4 Chapter structure 14

2.1 Classification of Web 2.0-related concepts 20
2.2 Fundamental Web 2.0 characteristics (based on [RK07]) 21
2.3 Classification of selected types of Social Software (adapted from [Hip06]) 23
2.4 PADGETS methodology (based on [CGFL10]) 25
2.5 PADGETS platform (source: [AGK$^+$11]) 26
2.6 Basic concepts of social networks 27
2.7 Launch dates of selected SNSs 29
2.8 Publications on DBLP dealing with SNSs per year 30
2.9 SNS architectures (based on [Bar64]) 31
2.10 SNS stakeholders from a data flow and service perspective (based on [Kum10]) 33
2.11 Relation between SNS stakeholders, their goals, and core concepts . . 35

3.1 Privacy perspectives on SNSs 38
3.2 Privacy mirror implementation on Google+ 43

4.1 Relationship between identity-related concepts (based on [Tel10] and [JP05]) 50
4.2 3-legged OAuth scheme (based on [HL10]) 55
4.3 Foundations of SIdM 58
4.4 Concepts of SIdM on SNSs 59
4.5 Process of managing social identities (based on [Win05]) 60
4.6 Identity creation on SNSs 61
4.7 Transformation of overlapping audiences 62
4.8 SNS audience types 63
4.9 Types of permission assignments 66

4.10 Contexts on SNSs 68

5.1 Literature review methodology (adopted from [LE06]) 78

6.1 Visibility dimensions on SNSs on a conceptual level 86
6.2 Study design 89
6.3 Facebook authorization dialog 90
6.4 Conceptualization of questionnaires 91
6.5 Determination of discrepancies between perceived, actual, and preferred default visibility settings 92
6.6 Binary determination of discrepancies between perceived, actual, and preferred visibility settings for shared items 93
6.7 False positive and false negative rate for expected (SQ3) and desired visibility (SQ4) 98

7.1 Two pillars of a PPSIdM Infrastructure 109
7.2 Conceptual overview of the proposed PPSIdM Infrastructure 110
7.3 SIdM cycle to synchronize self-presentation in the physical world and on SNSs 112
7.4 Hierarchical structure of the PPSIdM processes 113
7.5 Identity management process 114
7.6 Relationship management process 115
7.7 Access management process 116
7.8 Translation process 117
7.9 SNS integration process 118

8.1 PPSIdM Infrastructure overview 120
8.2 URL-based identities 122
8.3 Adapted OAuth authentication and authorization workflow 124
8.4 SNS integration mechanisms 125
8.5 Overall SNS integration workflow 127
8.6 Generic identity representation model 131
8.7 Assisted Audiences process 136
8.8 Sample social graph 137
8.9 Interpretation of Assisted Audiences results 139
8.10 Sample access rights assignment 140
8.11 Temporal view on static assignments of access rights 144
8.12 Temporal view on dynamic, context-aware adaption of access rights . . 145
8.13 CA-RBAC model 147
8.14 Process of permission reduction in CA-RBAC 149
8.15 Transparency and usability on SNSs 153
8.16 Conceptualization of dedicated views of the Disclosure Browser 156

8.17 Access Policy Grid generation 158

9.1 Prototype architecture 163
9.2 PPSIdM Provider user interface structure 164
9.3 1st level navigation bar 164
9.4 Attribute and attribute type definition 166
9.5 Identity Facet Manager user interface 167
9.6 Assisted Audiences user interface 169
9.7 CA-RBAC user interface 170
9.8 Disclosure Browser user interface (Audience View) 172
9.9 Access Policy Grid construction 173
9.10 Access Policy Grid user interface 174
9.11 Firefox Browser Plugin implementation 176

10.1 Strategic design science research evaluation framework (adopted from [PHBV08]) 182
10.2 Social graph (Anna Clark) 183
10.3 Group trust value assignment 187
10.4 Assignment of audiences to identity facets 188
10.5 Definition of contextual constraints 189
10.6 Disclosure Browser (Contact View) 191
10.7 Sorted Access Policy Grid showing all contact-permission assignments 192
10.8 Suspicious permission examination and revision 193
10.9 Integration of URL-based references into existing SNSs 195
10.10SNS profile requests 196

Listings

9.1 XML-based attribute export (excerpt) 167

9.2 API friend list request on Facebook using FQL 168

9.3 Facebook friend list response in JSON format (excerpt) 168

9.4 Mutual friend request using FQL 169

9.5 Web service interface description (simplified WADL document) 175

B.1 OAuth Response (Chris Martin) 215

B.2 OAuth Response (Julie Matlock) 216

List of Tables

3.1 Framework for conceptualizing SNS privacy 45

4.1 Classification of identity models (based on [TM11] and [Ben05]) 53
4.2 Privacy impact of intended and actual audience 64

5.1 Overview of identity management requirements 75
5.2 Overview of relationship management requirements 75
5.3 Overview of access management requirements 76
5.4 Overview of infrastructural requirements 77
5.5 Comparison of existing SIdM approaches 82

6.1 Comparison of information sharing dimensions in the physical world and on SNSs 87
6.2 Participant demographics 94
6.3 Distribution of employed privacy settings for examined item categories (percentages) 95
6.4 Perceived default settings (SQ1) (percentages) 96
6.5 Preferred default settings (SQ2) (percentages) 97
6.6 Preferred action for previously shared items (SQ6) 100

7.1 IS improvement principles 107

8.1 Identity management-related limitations of existing SNSs and research approaches 129
8.2 Connection between attribute-related concepts 130
8.3 2 × 2 contingency table 134
8.4 Matrix representation of mutual relationships 137
8.5 Similarity matrix of contacts 138
8.6 Reduced contact cluster similarity matrix 138
8.7 Classification of visualization types 154

9.1 Prototype implementation status 165

10.1 Identity facets and attributes (Anna Clark) 185

10.2 Generated contact clusters 186
10.3 Heterogeneity between contact clusters 186
10.4 Contact-permission assignments (overview) 190
10.5 Effective authorizations at t_0 (Colleagues) 197
10.6 Effective authorizations at t_1 (Colleagues) 198
10.7 Requirements-based evaluation 202

A.1 AX Schema attribute types 211

List of Abbreviations

AJAX	Asynchronous JavaScript and **XML**	20
API	Application Programming Interface	90
AX	Attribute Exchange	131
CMC	Computer-mediated Communication	7
DOM	Document Object Model	164
CA-RBAC	Context-aware Role-based Access Control	121
DBLP	Digital Bibliography and Library Project	30
FoF	Friends-of-Friends	67
GUI	Graphical User Interface	162
HTTP	Hypertext Transfer Protocol	123
HCI	Human–computer Interaction	7
IdM	Identity Management	4
IDaaS	Identity as a Service	174
IdMS	Identity Management System	51
IdMI	Identity Management Infrastructure	52
IdP	Identity Provider	53
IS	Information Systems	5
FQL	Facebook Query Language	168
JSON	JavaScript Object Notation	168
OOP	Object-oriented Programming	130
P2P	Peer-to-Peer	30
MVC	Model-View-Controller	162
GWT	Google Web Toolkit	162
PADGETS	Policy Gadgets Mashing underlying Group Knowledge in Web 2.0 Media	24
P3P	Platform for Privacy Preferences	40
PPSIdM	Privacy-preserving Social Identity Management	3
PET	Privacy-Enhancing Technologies	34
SNS	Social Network Site	1
RBAC	Role-based Access Control	54
RPA	Reflective Policy Assessment	151
REST	Representational State Transfer	174
SSO	Single Sign-On	52
SIdM	Social Identity Management	2
UPA	User-Permission Assignment	65
URL	Uniform Resource Locator	110
WADL	Web Application Description Language	175
WWW	World Wide Web	7
XUL	**XML** User Interface Language	177
XML	eXtensible Markup Language	132