Privacy Papers for Policy Makers

2011

The publication of "Privacy Papers for Policy Makers" was supported by AT&T and Microsoft.

September 7, 2011

We are delighted to provide you with the second annual compilation entitled "Privacy Papers for Policy Makers," showcasing leading analytical thinking about current and emerging privacy issues.

The works featured and digested in the enclosed were selected by members of the Advisory Board of the Future of Privacy Forum (scholars, privacy advocates and Chief Privacy Officers) based on criteria emphasizing clarity, practicality and overall utility. There is a great deal of important and useful scholarship on privacy today, and choosing the works to be included in "Privacy Papers" was a difficult task but we think our Advisory Board has chosen well, and has put together a diverse and thought-provoking collection. Two of the papers were selected by the chairpersons of the annual Privacy Law Scholars Conference as recipients of the IAPP award for the best papers presented at the 2011 Conference.

We hope this relevant and timely scholarship helps inform and stimulate thinking among policy makers and policy influentials in the US and around the world, with whom we are sharing "Privacy Papers."

One of the goals of the Future of Privacy Forum is to convene thought leaders and to share new ways to think about privacy. We want to thank AT&T and Microsoft for their special support of the "Privacy Papers" project which furthers that goal.

Sincerely yours,

Christopher Wolf
Founder and Co-chair

Jules Polonetsky
Director and Co-chair

Future of Privacy Forum Advisory Board

Alessandro Acquisti
Associate Professor of Information Technology and Public Policy at the Heinz College, Carnegie Mellon University

Jim Adler
Chief Privacy Officer & General Manager, Data Systems, Intelius

Ellen Agress
Senior Vice President and Deputy General Counsel, News Corporation

Annie I. Antón
Professor of Computer Science, College of Engineering, North Carolina State University

Kenneth A. Bamberger
Professor of Law, Berkeley School of Law

Elise Berkower
Associate General Counsel, Privacy, The Nielsen Company

Joan (Jodie) Z. Bernstein
Counsel, Kelley Drye & Warren, LLP and former director of the Bureau of Consumer Protection at the Federal Trade Commission

Michael Blum
General Counsel, Quantcast

Bruce Boyden
Assistant Professor of Law, Marquette University Law School

Allen Brandt
Corporate Counsel, Data Privacy & Protection, Graduate Management Admission Council (GMAC)

Jim Brock
CEO, PrivacyChoice

Kathryn C. Brown
Senior Vice President, Public Policy Development and Corporate Responsibility, Verizon

James M. Byrne
Chief Privacy Officer, Lockheed Martin Corporation

Ryan Calo
Director, Consumer Privacy Project, Center for Internet & Society at the Stanford Law School

Dr. Ann Cavoukian
Ontario Privacy Commissioner

Brian Chase
General Counsel, Foursquare Labs, Inc.

Danielle Citron
Professor of Law, University of Maryland Law School

Maureen Cooney
Senior Counsel and Deputy Chief Privacy Officer, Sprint Nextel

Lorrie Faith Cranor
Associate Professor of Computer Science and Engineering, Carnegie Mellon University

Mary Culnan
Slade Professor of Management and Information Technology, Bentley University

Simon Davies
Director, Privacy International

Michelle De Mooy
Senior Associate, National Priorities, Consumer Action

Elizabeth Denham
Information and Privacy Commissioner for British Columbia

Michelle Dennedy
Former Chief Governance Officer, Cloud Computing, Sun Microsystems

Benjamin Edelman
Assistant Professor, Harvard Business School

Keith Enright
Senior Corporate Counsel, Privacy, Google

Leigh Feldman
SVP, Senior Privacy Executive Global Compliance Risk – Enterprise Privacy, Bank of America

Michael Fertik
Founder and Chief Executive Officer, Reputation.com

Eric Friedberg
Co-President, Stroz Friedberg

Scott Goss
Senior Privacy Counsel, Qualcomm

Sean Hanley
Director of Compliance, Zynga Game Network, Inc.

Leslie Harris
Center for Democracy and Technology (CDT)

David Hoffman
Director of Security Policy and Global Privacy Officer, Intel

Marcia Hoffman
Staff Attorney, Electronic Frontier Foundation

Andy Holleman
Chief Privacy Officer and Associate General Counsel, Qwest

Chris Hoofnagle
Director, Berkeley Center for Law & Technology's information privacy programs and senior fellow to the Samuelson Law, Technology & Public Policy Clinic

Jeff Jarvis
Associate Professor; Director of the Interactive Program, Director of the Tow-Knight Center for Entrepreneurial Journalism at the City University of New York

Pamela Jones Harbour
Former Federal Trade Commissioner; Partner,Fulbright & Jaworski LLP

Sandra Hughes
Global Privacy Executive, Procter & Gamble

David Kahan
General Counsel, Jumptap

Nuala O'Connor Kelly
Senior Counsel, Information Governance & Privacy, GE

Ian Kerr
Canada Research Chair in Ethics, Law & Technology, University of Ottawa, Faculty of Law

Brian Knapp
Chief Privacy Officer and Vice President, Corporate Affairs, Loopt

Jerry Kovach
Senior Vice President, External Affairs, Neustar

Fernando Laguarda
Vice President, External Affairs and Policy Counselor, Time Warner Cable

Barbara Lawler
Chief Privacy Officer, Intuit

Gerard Lewis
Senior Counsel and Chief Privacy Officer, Comcast

Chris Lin
Executive Vice President, General Counsel and Chief Privacy Officer, comScore, Inc.

Brendon Lynch
Chief Privacy Officer, Microsoft

Fran Maier
President and Executive Chair, TRUSTe

Jennifer Mardosz
Senior Vice President and Chief Privacy Officer, MySpace

William McGeveran
Associate Professor, University of Minnesota Law School

Terry McQuay
President, Nymity

Rena Mears
Partner, Deloitte & Touche LLP, Global & U.S. Leader Privacy and Data Protection

Scott Meyer
CEO, Evidon

Doug Miller
Executive Director, Consumer Advocacy & Privacy, AOL

Scott Nelson
Executive Vice President and Chief Operating Officer, TruEffect

Oren Netzer
Founder & CEO, DoubleVerify Inc.

Paul Ohm
Associate Professor of Law and Telecommunications, University of Colorado Law School

Adam Palmer
Cyber Security Officer, Symantec

Harriet Pearson
Chief Privacy Officer & VP Regulatory Policy, IBM

Robert Quinn
Chief Privacy Officer and Senior Vice President for Federal Regulatory, AT&T

MeMe Rasmussen
Senior Director, Associate General Counsel, Adobe Systems Incorporated

Neil Richards
Professor of Law, Washington University Law School

Shirley Rooker
President, Call for Action

Russell Schrader
Chief Privacy Officer and Associate General Counsel – Global Enterprise Risk, Visa Inc.

Paul Schwartz
Professor of Law, University of California-Berkeley School of Law

Scott Shipman
Senior Counsel, Global Privacy Practices, eBay

Meredith Sidewater
Senior Vice President and General Counsel, Lexis Nexis Risk Solutions

Al Silipigni
Vice President, Privacy Officer, American Express Company

Daniel Solove
Professor of Law, George Washington University Law School

Tim Sparapani
Director of Public Policy, Facebook

Greg Stuart
CEO, Mobile Marketing Association

Peter Swire
Professor, Ohio State University Moritz College of Law

Omar Tawakol
CEO, BlueKai

Omer Tene
Associate Professor, College of Management School of Law, Rishon Le Zion, Israel

Anne Toth
Chief Trust Officer, Yahoo! Inc.

Steven Vine
Chief Privacy Officer, Datran Media

Michael Zimmer
Assistant Professor in the School of Information Studies, University of Wisconsin-Milwaukee

Table of Contents

*Recipients of the IAPP award for best papers at the 2011 Privacy Law Scholars Conference

Out of respect for copyright law and for ease of reference, this compilation is a digest of the papers selected by the Future of Privacy Forum Advisory Board and does not contain the full text. The selected papers in full text are available through the referenced links

Accountability as the Basis for Regulating Privacy: Can Information Security Regulations Inform Privacy Policy?

Mary J. Culnan

Full paper available at: www.futureofprivacy.org/the-privacy-papers/

Executive Summary

The predominate approach to consumer privacy regulation in the US is grounded in two principles: notice and choice. Firms provide notice describing their information practices while choice provides consumers with limited rights to opt out of certain uses of their personal information. Companies that break these promises may be subject to an FTC investigation. This paper argues that this current approach to regulating privacy is not effective and needs to be revisited. The current approach places too much burden on the individual, frequently deals with harm only after the fact, and has failed to motivate organizations to proactively prevent privacy or security incidents resulting from their information processing activities. As an alternative, the paper proposes augmenting the current approach with new regulations based on accountability where firms are delegated responsibility to develop risk management programs for privacy tailored to their individual circumstances.

The paper analyzes the requirements of three information security laws (GLB Safeguards Rule, HIPAA Security Rule and the Massachusetts Standards for the Protection of Personal Information) which require organizations to adopt comprehensive security programs against the elements of accountability and concludes that these laws provide a starting point for designing a new privacy regulatory regime. Based on this analysis, the paper describes what a sample privacy program might look like including the types of evidence that could be maintained to demonstrate compliance. An accountability analysis of three recent FTC enforcement actions illustrates how this approach might work in practice.

While current security laws provide a good starting point, privacy also raises new implementation challenges that will need to be addressed including the absence of standards for "reasonable privacy," identifying the types of records organizations need to maintain to document their compliance with the regulations, and how firms with different contexts should operationalize fair information principles. The paper concludes by reviewing arguments in favor of the more flexible delegation approach to privacy regulation which is based on the assumption that firms have superior information and expertise to develop solutions that will lead to the desired results compared with the traditional "command and control" compliance model.

Author

Mary J. Culnan is the Slade Professor of Management and Information Technology at Bentley University in Waltham, Massachusetts. Her current research is focusing on accountability as a public policy approach to privacy. She authored the 1999 Georgetown Information Privacy Policy Survey which was used by the FTC to make policy recommendations about online privacy to Congress, and co-authored the Future of Privacy Forum's "icon study" related to online behavioral advertising. She has testified before the U.S. Congress, the Massachusetts House and Senate and other government agencies on a range of privacy issues. Currently she serves as a member of the GAO's Executive Committee on Management and Information Technology. She also served as a Commissioner on the President's Commission on Critical Infrastructure Protection. She holds a Ph.D. in information systems from UCLA

Against Notice Skepticism

Ryan Calo

Full paper available at: www.futureofprivacy.org/the-privacy-papers/

Executive Summary

This is a work-in-progress that explores how design might help resuscitate notice in the context of privacy---and possibly elsewhere. The paper describes why notice has failed and even backfired as a regulatory strategy in privacy. In recognition of the potential benefits of notice over government-mandated restrictions on information, the paper identifies errors officials may be making in deploying notice strategies. The first is that privacy policies are the only form of information strategy that could work in privacy. The second is that notice must be textual, verbal, or its symbolic equivalent. Companies are using innovative ways to convey information that do not rely primarily on lengthy documents and the law should encourage these practices. Recent studies in human-computer interaction suggest even more radical and potentially effective forms of consumer communication.

(Draft)

Author

Ryan Calo is the director for privacy and robotics at the Stanford Law School Center for Internet and Society. Prior to joining the law school in 2008, Calo worked as an associate in the Washington, D.C. office of Covington & Burling, LLP and clerked for Judge R. Guy Cole Jr. on the U.S. Court of Appeals for the Sixth Circuit. Calo's work on privacy and robotics has appeared in the Wall Street Journal, the New York Times, and other major news outlets.

The Case For Online Obscurity

Woodrow Hartzog* and Frederic Stutzman**

Full paper available at: www.futureofprivacy.org/the-privacy-papers/

Executive Summary

On the Internet, obscure information has a minimal risk of being discovered or understood by unintended recipients. Empirical research demonstrates that Internet users rely on obscurity perhaps more than anything else to protect their privacy. Users routinely hide information by making it invisible to search engines, using pseudonyms and multiple profiles, and taking advantage of privacy settings.

Yet, online obscurity has been largely ignored by courts and lawmakers. In this article, we argue that obscurity is a critical component of online privacy, but it has not been embraced by courts and lawmakers because it has never been adequately defined or conceptualized. To that end, this article develops the first clear definition of online obscurity.

Empirical Support for Online Obscurity

The choice to disclose online is the product of a complex and highly contextual decision process, where risks are weighed against the potential reward of disclosure. It is normal to expect obscurity in everyday life. When we stroll down the street, we do not expect to be identified by all passers-by; indeed, we expect to be obscure in the eye of these observers. With the rise of peer-produced online content, it is now just as clear that our expectation of obscurity transfers online.

Empirical research demonstrates that individuals exert control over the information disclosed online by limiting the audience of the disclosure, by bounding the meaning of the disclosure, and by adapting the disclosure to a particular website. In social network sites, where the use of anonymity would violate norms and limit benefits attained from site use, individuals strategically develop techniques that effectively produce obscurity in disclosure. Interacting with both rules and norms, obscurity is flexibly – and reflexively – created in sites that we would consider highly identified. Even in remarkable, anonymous contexts such as Facebook, individuals rely on obscurity as an important aspect of managing both identity and privacy.

Contrary to the powerful popular discourse that argues that individuals online have essentially different privacy and notoriety goals, we demonstrate that online obscurity is a crucial aspect of privacy for Internet users. Through obfuscation techniques and other normative practices, it is clear that obscurity is both desired and expected online.

*Assistant Professor of Law, Cumberland School of Law at Samford University; Affiliate Scholar, Center for Internet and Society at Stanford Law School.

**Postdoctoral Fellow, H. John Heinz III College, Carnegie Mellon University.

The Specter of Obscurity in Online Privacy Law

Because courts and lawmakers have failed to develop online obscurity as a concept, the law in a number of online privacy disputes remains difficult to square with the expectations of Internet users. For example, if a blogger limits access to her website to those who have a password, are her posts considered public or private? How should courts classify pseudonymous postings that are invisible to search engines but could have been accessed by anyone in possession of the URL? If a website introduces facial recognition technology as a way to search faces in photos, have they broken any promises of privacy to users who previously uploaded photos and may have relied on the fact those photos were not searchable?

Courts have not explicitly embraced the concept of online obscurity, but its existence is hard to ignore in a number of disputes. Judicial support for the analog version of online obscurity – practical obscurity – has laid the foundation for the recognition of online obscurity. Courts already rely upon obfuscation features like passwords, privacy settings, encryption, and code to limit search visibility. However, without a clear conceptualization of online obscurity, courts consistently reach one conclusion – the unfettered ability of any hypothetical individual to find and access information on a website renders that information "public," or ineligible for privacy protection.

Courts also have a problematic tendency to rely upon passwords to define what information is public – that is, the password-protection of information is a critical test of that information's intended and expected publicity. This is another important reason a workable definition of online obscurity is needed.

Lawmakers have also implicitly recognized the value of obscurity, but their failure to embrace obscurity as a concept has resulted in criticism that their laws fail to protect "privacy"—meaning secrets—or that they protect information that is not private at all. If lawmakers were to clarify that they were seeking to protect the obscurity of information, these laws might be perceived and implemented differently. For example, the Drivers Privacy Protection Act prohibits the disclosure of information about any individual obtained by the DMV in a motor vehicle record. Of course, much of the information protected by this statute, such as home address, height, and hair color, is hardly secret, or even private. But the law implicitly protects whatever obscurity the information exists in by restricting access to it.

A Proposed Definition and Framework

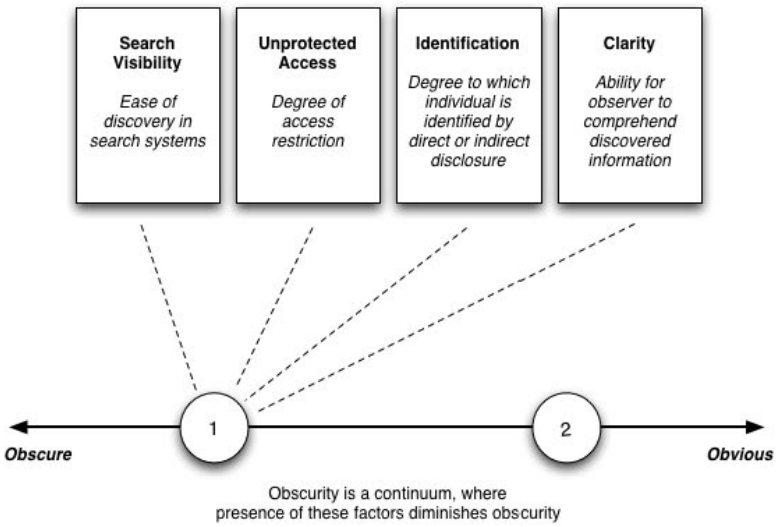
We conceive of online obscurity as a form of everyday obfuscation. Thus, we think the proper metaphor is the key and lock; to understand encountered information (i.e., release the lock), one must possess context (the key or keys) that renders the information un-obscure. This metaphor is likely better suited to online disputes given the judicial reliance on the digital version of the key: the password. In essence, we are simply proposing that there is more than one key that can lock information. Indeed, many kinds of keys and locks, each with varying strengths, exist, and considered cumulatively, fall along a spectrum that will allow courts to make a more nuanced analysis of online information on a scale of obscurity.

To that end, we propose the following definition: Information is obscure online if it exists in a context missing one or more key factors that are essential to discovery or comprehension. We have identified four of these factors as part of a non-exhaustive and flexible list: 1) search visibility 2) unprotected access 3) identification 4) clarity.

The presence of these factors diminishes obscurity, and their absence enhances it. Thus, in determining whether information is obscure online, courts should consider whether any of these factors were present in their determination. Information that is entirely un-obscure is completely obvious, and vice versa. Courts should engage in a case-by-case analysis of the factors, examining each one individually, then as a whole to determine the degree of online obscurity. Figure 1 depicts how this conceptualization would work in two different scenarios:

Figure 1

Factors Determining Online Obscurity



Obscurity is a continuum, where presence of these factors diminishes obscurity

Scenario 1 is a blog that is visible only to invited users and is not searchable by general search engines like Google. It is close to being completely obscure because it is missing two of the most important factors for finding and understanding information: search visibility and unprotected access. Scenario 2 is a Twitter account that uses only a first name and a blurry photo to identify the poster. While this information is more obvious than the information in Scenario 1 because it is freely searchable and accessible, it is still obscure because only certain Internet users would be able to identify the poster of the content or completely comprehend any idiosyncratic posts.

Potential Application of Online Obscurity

This framework could be applied as an analytical tool or as part of an obligation. Obscurity could be relied upon as a continuum to help determine if information is eligible for privacy protections. Obscurity could be used as a protective remedy by courts and lawmakers; instead of forcing websites to remove sensitive information, a compromise could be some form of mandated obscurity. Finally, obscurity could serve as part of an agreement. Internet users bound to a "duty to maintain obscurity" would be allowed to further disclose information, so long as they kept the information generally as obscure as they received it.

The conceptualization and proposed implementations of online obscurity in this article are meant to be introductions, not the final word. Much more research and analysis is required to fully explore how online obscurity might be utilized in the law.

Authors

Woodrow Hartzog is an Assistant Professor at the Cumberland School of Law at Samford University. He is also a Junior Affiliate Scholar at the Center for Internet and Society at Stanford Law School. His research focuses on privacy, human-computer interaction, online communication, and electronic agreements. He holds a Ph.D. in mass communication from the University of North Carolina at Chapel Hill, an LL.M. in intellectual property from the George Washington University Law School and a J.D. from Samford University. He previously worked as an attorney in private practice and at the United States Patent and Trademark Office. He also served as a clerk for the Electronic Privacy Information Center.

Fred Stutzman is a postdoctoral fellow at Carnegie Mellon University, where he works with Alessandro Acquisti. In 2011, he graduated from the University of North Carolina at Chapel Hill, where he was advised by Gary Marchionini. Fred's research focuses on privacy in social computing, where he explores the economics of privacy choice, and designs systems and policies that produce positive privacy outcomes. He is also interested in pro-social outcomes of social media use, particularly social media use during life transition. In addition to his academic work, Fred is the co-founder of ClaimID.com, founder of EPS, Inc. (distributor of the productivity software Freedom and Anti-Social), and consultant to select organizations.

Dispelling the Myths Surrounding De-Identification: Anonymization Remains a Strong Tool for Protecting Privacy

Ann Cavoukian and Khaled El Emam

Full paper available at: www.futureofprivacy.org/the-privacy-papers/

Executive Summary

Recent studies have put into question the value of de-identifying personal information as an essential tool to protect privacy. Repeated claims regarding the ease of re-identification may lead to the mistaken impression that it is futile to de-identify personal information. Furthermore, these assertions may drastically reduce the availability of de-identified information for potentially beneficial secondary purposes, such as much-needed health research.

This paper aims to dispel this myth. The authors illustrate the enormous value of the de-identification of personal information as an essential tool that should be routinely used to minimize risks, particularly in the context of health information. This paper demonstrates the possibility of solving the traditional zero-sum paradigm pitting data quality against privacy. As long as proper de-identification techniques and re-identification risk measurement procedures are used, re-identification remains a relatively difficult task in actual practice. It is thus possible to achieve a high degree of privacy, while at the same time preserving a high level of data quality. Maximizing both privacy and data quality enables a shift from a zero-sum paradigm to a positive-sum paradigm, a key principle of Privacy by Design.

While de-identification of information is not a perfect tool, it continues to be a valuable and effective mechanism for protecting personal information, in conjunction with additional safeguards. The objective of this paper is to shatter the myth that de-identification is not a strong tool to protect privacy - it is. The authors urge organizations that collect, use and disclose personal information to continue to de-identify personal data, in a comprehensive and responsible manner, as part of an overall risk assessment framework.

Authors

Dr. Ann Cavoukian is recognized as one of the leading privacy experts in the world. Noted for her seminal work on Privacy Enhancing Technologies (PETs) in 1995, her concept of *Privacy by Design* seeks to proactively embed privacy into the design specifications of information technology and accountable business practices, thereby achieving the strongest protection possible. In October, 2010, regulators from around the world gathered at the annual assembly of International Data Protection and Privacy Commissioners in Jerusalem, Israel, and unanimously passed a landmark Resolution recognizing *Privacy by Design* as an essential component of fundamental privacy protection. This was followed by the U.S. Federal Trade Commission's inclusion of *Privacy by Design* as one of its three recommended practices for protecting online privacy – a major validation of its significance.

An avowed believer in the role that technology can play in the protection of privacy, Dr. Cavoukian's leadership has seen her office develop a number of tools and procedures to ensure that privacy is strongly protected, not only in Canada, but around the world. She has been involved in numerous international committees focused on privacy, security, technology and business, and endeavours to focus on strengthening consumer confidence and trust in emerging technology applications.

Dr. Cavoukian serves as the Chair of the Identity, Privacy and Security Institute at the University of Toronto, Canada. She is also a member of several Boards including, the European Biometrics Forum, Future of Privacy Forum, RIM Council, and has been conferred as a Distinguished Fellow of the Ponemon Institute. Dr. Cavoukian was honoured with the prestigious *Kristian Beckman Award* in 2011 for her pioneering work on *Privacy by Design* and privacy protection in modern international environments. In the same year, Dr. Cavoukian was also named by *Intelligent Utility* Magazine as one of the Top 11 Movers and Shakers for the Global Smart Grid industry, received the SC Canada Privacy Professional of the Year Award and was honoured by the University of Alberta Information Access and Protection of Privacy Program for her positive contribution to the field of privacy.

Khaled El Emam, PhD, is an Associate Professor at the University of Ottawa, Faculty of Medicine. He is a Canada Research Chair in Electronic Health Information at the University of Ottawa. Previously Khaled was a Senior Research Officer at the National Research Council of Canada, and prior to that he was head of the Quantitative Methods Group at the Fraunhofer Institute in Kaiserslautern, Germany. In 2003 and 2004, he was ranked as the top systems and software engineering scholar worldwide by the Journal of Systems and Software based on his research on measurement and quality evaluation and improvement, and ranked second in 2002 and 2005. He holds a Ph.D. from the Department of Electrical and Electronics, King's College, at the University of London (UK). His lab's web site is: http://www.ehealthinformation.ca/.

The Failure of Online Social Network Privacy Settings

Michelle Madejski, Maritza Johnson and Steven Bellovin

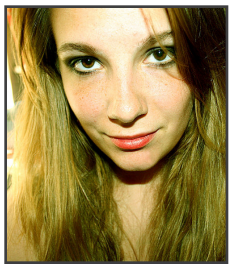
Full paper available at: www.futureofprivacy.org/the-privacy-papers/

Executive Summary

Increasingly, people are sharing sensitive personal information via online social networks (OSN). While such networks do permit users to control what they share with whom, access control policies are notoriously difficult to configure correctly; this raises the question of whether OSN users' privacy settings match their sharing intentions.

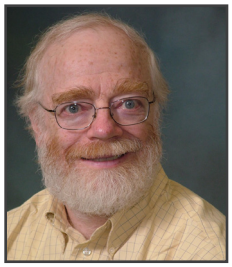
We present the results of an empirical evaluation that measures privacy attitudes and intentions and compares these against the privacy settings on Facebook. Our results indicate a serious mismatch: every one of the 65 participants in our study confirmed that at least one of the identified violations was in fact a sharing violation. In other words, OSN users' privacy settings are incorrect. Furthermore, a majority of the participants report that they cannot or will not fix such errors. We conclude that the current approach to privacy settings is flawed and cannot be fixed; a fundamentally different approach is needed. We present recommendations to ameliorate the current problems, as well as provide suggestions for future research.

Authors



Michelle Madejski graduated from Columbia University in 2010 with an undergraduate degree in Computer Science from the School of Engineering and Applied Science. Her interests include user privacy, web development, and the role of technology in society. Her undergraduate-originated work regarding user privacy in online social networks has received a Best Paper award and has been featured in numerous press including Gawker, MSNBC Today, Chicago Tribune, Huffington Post. Michelle is currently an engineer at Boeing in Seattle and anticipates attending a PhD program in the Fall of 2012.

Maritza Johnson is a Ph.D. candidate in the Department of Computer Science at Columbia University. Her research interests include computer security and human factors. Maritza was a recipient of the AT&T Labs Fellowship. Prior to studying at Columbia, Maritza received a B.A. in Computer Science from the University of San Diego.



Steven M. Bellovin is a professor of computer science at Columbia University, where he does research on networks, security, and especially why the two don't get along. He joined the faculty in 2005 after many years at Bell Labs and AT&T Labs Research, where he was an AT&T Fellow. He received a BA degree from Columbia University, and an MS and PhD in Computer Science from the University of North Carolina at Chapel Hill. While a graduate student, he helped create Netnews; for this, he and the other perpetrators were given the 1995 Usenix Lifetime Achievement Award (The Flame). He is a member of the National Academy of Engineering and is serving on the Computer Science and Telecommunications.

The PII Problem:
Privacy and a New Concept of Personally Identifiable Information

Paul M. Schwartz and Daniel J. Solove

(Forthcoming 86 NYU Law Review – (2010))

Full paper available at: www.futureofprivacy.org/the-privacy-papers/

Executive Summary

Personally identifiable information (PII) is one of the most central concepts in information privacy regulation. The scope of privacy laws typically turns on whether PII is involved. The basic assumption behind the applicable laws is that if PII is not involved, then there can be no privacy harm. At the same time, there is no uniform definition of PII in information privacy law. Moreover, computer science has shown that the very concept of PII can be highly malleable.

To demonstrate the policy implications of the failure of the current definitions of PII, this Article examines current practices of behavioral marketing. In their use of targeted technologies, companies direct offerings to specific consumers based on information collected about their characteristics, preferences, and behavior. Behavioral marketing also has enormous implications for public health due to food marketing to youth. Over the last three decades, the extent of obesity among minors has risen dramatically throughout the U.S. Experts also point to a detrimental effect of the marketing of food products to minors. Yet, the present regulatory regime for information privacy with PII as the cornerstone has proven incapable of an adequate response to behavioral marketing.

This Article proceeds in four steps.

A Typology of PII

First, this Article develops a typology of PII that shows three basic approaches in United States law to defining this term.

- The "tautological" approach defines PII as any information that identifies a person. The Video Privacy Protection Act demonstrates this model. The problem with the tautological approach is that it fails to define PII or explain how it is to be singled out.
- The "non-public" approach defines PII by what it is *not* rather than what it is. The non-public approach says that PII is all that is not aggregate data because such information does not identify a person. The Gramm-Leach Bliley Act epitomizes this approach. The problem with the non-public approach is that it does not map onto whether the information is in fact identifiable.
- The "specific-types" approach lists specific types of data that constitute PII. If the information in question falls into the enumerated group, it then becomes a kind of statutory "per se" PII. The Massachusetts Breach Notification Statute, California's Song-Beverly Credit Card Act, and Children's Online Privacy Protection Act illustrate this model. This approach is flawed because technology can broaden the kinds of information that constitute PII.

A Critique of PII

Second, this Article discusses defects in the existing distinction between PII and non-PII. The line between PII and non-PII is not fixed but depends upon factors including changes in technology and the specific context of data processing. For example, whether or not a search query is PII cannot be determined in the abstract.

The Example of Behavioral Marketing

Third, this Article uses behavioral marketing, with a special emphasis on food marketing to children, as a test case for demonstrating the flaws in the current definitions of PII. Individuals can now be tracked across different websites or digital media. Moreover, online advertising networks follow people around the Web. Advertising networks place tracking files on people's computers, which allow the company to gather information about behavior and preferences.

Marketers today engage in a pinpoint process that focuses on ever-smaller groups of people. Instead of companies selling ads for specific websites, advertisers now seek to buy access to people who fit a certain pattern. Information that is collected is packaged into profiles, which are then sold on stock-market-like exchanges. Yet, in behavioral marketing, companies generally do not track individuals through use of their names. Instead they use software to build personal profiles that exclude this item but that contain a wealth of details about the individual. Typically, the firms associate these personal profiles with a single alphanumerical code placed on an individual's computer.

Thus, behavioral marketing occurs without identifying (in the traditional sense) a specific individual. While advertising networks may not know people's names, identification of individuals is nonetheless possible in many cases. For example, enough pieces of information linked to a single person—even in the absence of a name, Social Security Number, or financial information—will permit identification of the individual. Such identification of seeming non-PII is a genuine possibility. Nonetheless, online companies have attempted to short-circuit the discussion of privacy harms and necessary legal reforms by simply asserting that they do not collect PII.

Policy Proposals

This Article concludes by developing an approach to redefining PII based on the different levels of risk to individuals. In our model of PII 2.0, information refers to (1) an identified, (2) identifiable, or (3) non-identifiable person. The continuum runs from actually being identified to no risk of identification, and our three categories divide up this spectrum and provide three different regimes of regulation.

Information refers to an *identified* person when it singles out a specific individual from others. Put differently, a person has been identified when her identity is ascertained.

In the middle of the risk continuum, information refers to an *identifiable* individual when a specific identification, while possible, is not a significantly probable event. In other words, an individual is identifiable when there is some non-remote possibility of future identification. The risk level is moderate to low. This information should be treated differently than an important sub-category of nominally identifiable information, where a linkage to a specific person has not yet been made, but where such a connection is more likely. When there is a significant risk of identification, the non-identified data should be treated the same as identified data.

At the other end of the risk continuum, *non-identifiable* information carries only a remote risk of identification. Such data cannot be said to be relatable to a person taking account of the means reasonably likely to be used for identification. In certain kinds of data sets, for example, the original sample is so large that other information will not enable the identification of individuals. An example would be high-level information about the population of the U.S., China, and Japan, and their relative access to telecommunications.

In our reconceptualized notion of PII, the key is to think about identification in terms of risk. Our model, PII 2.0, conceives of identifiability as a continuum of risk rather than as a simple dichotomy. A clear way to demonstrate the functioning of this new approach is by considering the applicability of FIPs. When information refers to an *identified* person, all of the FIPs generally should apply.

As for the category of *identifiable*, it is not appropriate to treat such information as fully equivalent to identified. The information does not yet refer to a specific person and may never do so. While some protections are in order because there is a risk of linkage to a specific individual, full notice, access, and correction rights should not be granted to an affected individual simply because identifiable data about her are processed. For one thing, the law's creation of such interests would decrease rather than increase privacy by requiring that all such data be associated with a specific person. This connection would be necessary to allow an individual to exercise her rights of notice, access, and correction. In this fashion, the law would promote a vicious circle that could transform all identifiable data into identified data.

Moreover, limits on information use, data minimization, and restrictions on information disclosure should not be applied across-the-board to identifiable information. Such limits would be disproportionate to risks from data use and also cripple socially productive uses of analytics that did not raise significant risks of harms to individuals. At the same time, some FIPs should apply to identifiable data. The key obligations concern data security, transparency, and data quality.

Thus, one benefit of PII 2.0 is that it tailors FIPs to whether information is identified or identifiable. A further benefit of PII 2.0 is that it creates an incentive for companies to keep information in the least identifiable form. The payoff is that the company, by making information identifiable or non-identifiable, will benefit from FIPs that become easier to meet as it moves along this continuum *away* from identified information.

Authors

Paul M. Schwartz is Professor of Law at the University of California-Berkeley Law School and a Director of the Berkeley Center for Law & Technology. A leading international expert on informational privacy and information law, he has published widely on these topics. In this country, his articles and essays have appeared in periodicals such as the Harvard Law Review, Yale Law Journal, Stanford Law Review, and N.Y.U. Law Review. With Daniel Solove, he is a co-author of the casebook, *Information Privacy Law* (Aspen, 4th ed., forthcoming 2012) and the handbook, *Privacy Law Fundamentals* (IAPP, 2011).

Professor Schwartz has provided advice and testimony to numerous governmental bodies in the United States and Europe. He has also assisted numerous corporations in the United States and abroad with information privacy issues. He belongs to the American Law Institute and is a member of the Editorial Board of International Data Privacy Law and the International Journal of Law and Information Technology.

Professor Schwartz received a J.D. degree from Yale Law School, where he was a Senior Editor on the Yale Law Journal, and a B.A. degree from Brown University. His home page is www.paulschwartz.net.

Daniel J. Solove is the John Marshall Harlan Research Professor of Law at the George Washington University Law School. He is also Senior Policy Advisor at Hogan Lovells. Additionally, he is the founder of TeachPrivacy, http://teachprivacy.com, a company that helps schools develop a comprehensive privacy program.

One of the world's leading experts in privacy law, Solove is the author of numerous books, including Nothing to Hide: The False Tradeoff Between Privacy and Security (Yale 2011), Understanding Privacy (Harvard 2008), The Future of Reputation: Gossip and Rumor in the Information Age (Yale 2007) (winner of the 2007 McGannon Award), and The Digital Person: Technology and Privacy in the Information Age (NYU 2004).

Professor Solove is also an author of Privacy Law Fundamentals (IAPP 2011), a short guide to information privacy law. Additionally, he is also the author of a textbook, Information Privacy Law with Aspen Publishing Co. now in its third edition, with co-author Paul Schwartz. He is the author of several other textbooks, including Privacy and the Media (1st edition, Aspen Publishing Co. 2009) and Privacy, Information, and Technology (2nd edition, Aspen Publishing Co. 2009). He has published more than 40 articles and essays, which have appeared in leading law reviews.

Solove has testified before Congress and has been involved as an expert and consultant in a number of high-profile privacy cases. He has been interviewed and featured in several hundred media broadcasts and articles. He blogs at http://www.concurringopinions.com. More information about his work can be found at: http://www.danielsolove.com.

Privacy Papers of Notable Mention

To View the Following Papers Visit: www.futureofprivacy.org/the-privacy-papers/

"Flash Cookies and Privacy II: Now with HTML5 and ETag Respawning"
By: Chris Hoofnagle, Mika Ayenson, Deitrich James Wambach, Ashkan Soltani and Nathan Good

"Regulating Privacy by Design"
By: Ira S. Rubinstein


FPF
THE FUTURE OF
PRIVACY FORUM

This report was designed with the environment and cost-effectiveness in mind. It is printed on recovered fiber paper that has no ozone layer threatening emissions and generates no detectable amounts of sulfur, chlorine, nitrogen, or dioxide gases when properly incinerated.

The Future of Privacy Forum (FPF) is a Washington, DC based think tank that seeks to advance responsible data practices. The forum is led by Internet privacy experts Jules Polonetsky and Christopher Wolf and includes an advisory board comprised of leading figures from industry, academia, law and advocacy groups. FPF was launched in November 2008, and is supported by Adobe, American Express, AOL, AT&T, Bank of America, Bering Media, BlueKai, BrightTag, Comcast, comScore, Datran Media, Dell, Deloitte, DoubleVerify, eBay, Evidon, Facebook, General Electric, Google, Intel, Intelius, Intuit, Jumptap, LexisNexis, LinkedIn, Lockheed Martin, Lotame, Microsoft, Neustar, News Corporation, The Nielsen Company, Privacy Choice, Proctor & Gamble, Qualcomm, Quantcast, Reputation.com, Sprint, Stroz Friedberg, Time Warner Cable, TruEffect, TRUSTe, Verizon, Visa, Vodafone, Yahoo! and Zynga.

To learn more about FPF, please visit www.futureofprivacy.org