# International Workshop on Privacy Engineering

http://www.ieee-security.org/TC/SPW2017/IWPE

*May 25th, 2017 at the Fairmont, San Jose, CA*

*Co-located with 38th IEEE Symposium on Security and Privacy*

| | |
|---|---|
| 09:00-09:15 | **Welcome, introductions and opening remarks** |
| 09:15–10:15 | **The Diffix Framework: Revisiting Noise, Again**<br>Invited talk by Paul Francis (Max Planck Institute for Software Systems)<br>For over 40 years, the holy grail of database anonymization is a system that allows a wide variety of statistical queries with minimal answer distortion, places no limits on the number of queries, is easy to configure, and gives strong protection of individual user data. This keynote presents Diffix, a database anonymization system that promises to finally bring us within reach of that goal. Diffix adds noise to query responses, but "*fixes*" the noise to the response so that repeated instances of the same response produce the same noise. While this addresses the problem of averaging attacks, it opens the system to "*difference attacks*" which can reveal individual user data merely through the fact that two responses differ. Diffix proactively examines queries and responses to defend against difference attacks. This talk presents the design of Diffix, gives a demo of a commercial-quality implementation, and discusses shortcomings and next steps.<br> <br>***Paul Francis** is a tenured faculty at the Max Planck Institute for Software Systems in Germany. Paul has held research positions at Cornell University, ACIRI, NTT Software Labs, Bellcore, and MITRE, and was Chief Scientist at two Silicon Valley startups. In the past, Paul's research centered around routing and addressing problems in the Internet and P2P networks. Paul's innovations include NAT, shared-tree multicast, the first P2P multicast system, the first DHT (as part of landmark routing), and Virtual Aggregation. More recently, Paul's research has focused on Internet privacy with a focus on anonymized analytics. Paul is cofounder of the startup Aircloak.* |
| Coffee Break | |
| 10:45–11:35 | **Session 1: Privacy Engineering Evaluation**<br>**Privacy Impact Assessments in practice: Outcome of a descriptive field research in the Netherlands**<br>Jeroen van Puijenbroek and Jaap-Henk Hoepman<br>*Privacy by design is not only important from an economic perspective but also from a legal one. The upcoming European General Data Protection Regulation makes privacy by design and default mandatory. One concrete step an organization can take towards privacy by design is to perform a so-called privacy impact assessment. To verify the assumption that the outcome of the privacy impact assessment leads to sufficient and adequate input for designing privacy-friendly products and systems that comply to privacy regulations and social norms regarding privacy we performed a descriptive field study in the Netherlands. In this paper, we present the results of this field study. Our main results are the following. When performing a privacy impact assessment, organisations use the organization itself as a focal point, instead of the data subjects whose data is being processed. A consequence of this focus is that the outcome of the privacy impact assessment will lead, at best, to a product or system that is compliant with data protection regulation. It will not lead to a product or system that is privacy friendly, or one that takes into account social norms regarding the processing of personal information. Another significant result is that the* |

| | |
|---|---|
| | *interviewed data protection officers perceive the process of determining privacy risks, based on the gathered information about a specific product or system, as vague. Moreover, the proposed countermeasures tend to address the effect rather than the cause of a privacy risk. Further research is needed to develop a more rigorous and transparent process for determining privacy risks that can be used by organisations.*<br><br>**Assessing Privacy in Social Media Aggregators**<br>Gaurav Misra, Jose M. Such and Lauren Gill<br>*Social Media Aggregator (SMA) applications present a platform enabling users to manage multiple Social Networking Sites (SNS) in one convenient application, which results in a unique concentration of data from several SNS accounts in addition to the user's mobile phone data available to them. We describe a three-step methodology to assess how privacy is considered in these applications: 1) We inspect the mobile data and social media data; 2) we study any privacy policies and their compliance with respect to distributor's vetting policies; and 3) we perform a qualitative assessment of traceability between privacy policies and the actual transparency and control mechanisms offered to users by the apps' interfaces. We then present the results we obtained for 13 popular SMAs from 3 app stores, showing a variation in data accessed by the individual applications, an absence of privacy policies for 5 of the SMAs evaluated, and a lack of traceability between privacy policies and transparency and control of interface operations. After this, we report our experiences using the methodology and the lessons learned, together with potential future work to improve the methodology and its potential to also assess privacy in other mobile applications that also connect with social media.* |
| 11:35-12:25 | **Session 2: Privacy Engineering Case Studies**<br>**Battery Status Not Included: Assessing Privacy in Web Standards**<br>Lukasz Olejnik, Steven Englehardt and Arvind Narayanan<br>*The standardization process is core to the development of the open web. Until 2013, the process rarely included privacy review and had no formal privacy requirements. But today the importance of privacy engineering has become apparent to standards bodies such as the W3C as well as to browser vendors. Standards groups now have guidelines for privacy assessments, and are including privacy reviews in many new specifications. However, the standards community does not yet have much practical experience in assessing privacy.*<br>*In this paper we systematically analyze the W3C Battery Status API to help inform future privacy assessments. We begin by reviewing its evolution — the initial specification, which only cursorily addressed privacy, the discovery of surprising privacy vulnerabilities as well as actual misuse in the wild, followed by the removal of the API from major browser engines, an unprecedented move. Next, we analyze web measurement data from late 2016 and confirm that the majority of scripts used the API for fingerprinting. Finally, we draw lessons from this affair and make recommendations for improving privacy engineering of web standards.*<br><br>**Statistical Detection of Downloaders in Freenet**<br>Brian Levine, Marc Liberatore, Brian Lynn and Matthew Wright<br>*Images posted to file-sharing networks without a person's permission can remain available indefinitely. When the image is sexually explicit and involves a child, the scale of this* |

| | |
|---|---|
| | *privacy violation grows tremendously worse and can have repercussions for the victim's whole life. Providing investigators with tools that can identify the perpetrators of child pornography (CP) trafficking is critical to addressing these violations. Investigators are interested in identifying these perpetrators on Freenet, which supports the anonymous publication and retrieval of data and is widely used for CP trafficking. We confirmed that 70,000 manifests posted to public forums dedicated to child sexual abuse contained tens of thousands of known CP images including infants and toddlers. About 35% of traffic on Freenet was for these specific manifests. In this paper, we propose and evaluate a novel approach for investigating these privacy violations. In particular, our approach aims to distinguish whether a neighboring peer is the actual requester of a file or just forwarding the requests for other peers. Our method requires only a single peer that passively analyzes the traffic it is sent by a neighbor. We derive a Bayesian framework that models the observer's decision for whether the neighbor is the downloader, and we show why the sum traffic from downloaders relayed by the neighbor is not a significant source of false positives. We validate our model in simulation, finding near perfect results, and we validate our approach by applying it to real CP-related manifests and actual packet data from Freenet, for which we find a false positive rate of about 2%. Given these results, we argue that our method is an effective investigative method for addressing privacy violations resulting from CP published on Freenet.* |
| 12:25-12:30 | **Best Paper Award Ceremony** |
| Lunch | |

| | |
|---|---|
| 13:30–14:30 | **The Price of Privacy in the Cloud or The Economic Consequences of Mr. Snowden**<br>Invited talk by Simon Wilkie (Microsoft)<br>Cloud computing involves distributed and shared use of computing facilities in a network allowing end users new flexibility and lower sunk costs. As a result, the cloud computing market has exhibited phenomenal growth. However, Edward Snowden's revelations of the NSA's spying program in 2013 degraded the privacy reputation of the US-based cloud service providers. We examine the economic impact of the Snowden revelations using a panel dataset of global cloud revenues across service types and vendors. We assume that the Snowden revelations are a negative demand-shock “treatment" for US-based providers, and regard non-US-based cloud providers as the control group. We find the revelations decreased the growth rate of revenues of US providers by 11% from Q3 2013 to Q4 2014. The expected losses to the US cloud industry are at least $18 billion, providing a benchmark “price of privacy.”<br>We then evaluate how users and cloud service providers changed their behavior using Microsoft's free trial database and 18 online service providers' privacy policies. Following the demand shock there is an increase in privacy protection and a significant price war in the US. Thus, firms' strategic reactions to the demand shock led to lower prices with a higher quality of privacy protection which in the long run increased US firms market share.<br> <br>***Simon Wilkie** is a Professor of Economics and of Communications Law at the USC Law School. He is currently on leave from the University of Southern California Economics department and is serving as the Senior Economist at Microsoft. His research focuses on game theory, its application to business strategy, economic and regulatory policy design, and the economics of the communications industries. His most recent research is on the wholesale telecommunications market and the concept diversity in media markets.* |
| 14:30–15:15 | **Session 3: Privacy Engineering Methodologies**<br>**Addressing Early Life Cycle Privacy Risk: Applying System-Theoretic Early Concept Analysis and Model-Based Systems Engineering to Privacy**<br>Stuart Shapiro<br>*This paper adapts System-Theoretic Early Concept Analysis (STECA), an instrumental safety risk management technique, for privacy to better identify and address privacy risks early in the engineering process. The technique, STECAPriv, aims to infer a nominal functional privacy control structure based on a conceptual system description and privacy-related system behavioral constraints. Model-based systems engineering (MBSE) is employed in conjunction with STECA-Priv to validate the projected control structure and to identify privacy risks in the form of constraint violations. To illustrate STECAPriv as supported by MBSE, it is applied to the simplified example of a smart television.*<br>**A Metamodel for Privacy Engineering Methods**<br>Yod-Samuel Martín García and Jose M. Del Alamo<br>*Engineering privacy in information systems requires systematic methods to capture and address privacy issues throughout the development process. However, the diversity of both* |

| | |
|---|---|
| | *privacy and engineering approaches, together with the specific context and scope of each project, have spawned a plethora of privacy engineering methods. Method engineering can help to cope with this landscape, as it allows describing existing methods in terms of a limited variety of method elements (and eventually enable their recombination into new, customized methods). This paper applies method engineering to introduce a privacy engineering metamodel, whose applicability is illustrated with a set of popular privacy engineering method elements, and a widely recognized privacy engineering method.* |
| Coffee Break | |
| 15:45–16:30 | **Session 4: Privacy Engineering in the domain of Advertisement**<br>**Privacy Broker: Message Oriented Middleware to implement Privacy Controls in Service Oriented Architectures (Industry paper)**<br>Narasimha Raghavan Veeraragavan and Karen Lees<br>*Schibsted is a global media and classified ads conglomerate with more than 200 million unique users per month, operating mainly from Europe. The company is currently being transformed away from traditional paper media and siloed sites towards a unified global media giant. As part of this transformation, Schibsted needs to collect a wide variety of datasets such as profile, behavior, location, payment and communication messages about the user in order to provide personalized content and target advertisements to the end users.*<br>*With the new EU General Data Protection Regulations (GDPR) taking effect from May 25 2018, each user using our products has a right to decide how his/her datasets should be governed or used in our products. To this end, we are building some privacy controls for the end users. These privacy controls are realized by a message-oriented middleware.*<br>*In this paper, we present a case study of design of a centralized topic based pub/sub style of middleware towards implementing the privacy controls in Schibsted's ecosystem of services.*<br><br>**Engineering Privacy and Protest: a Case Study of AdNauseam**<br>Daniel C Howe and Helen Nissenbaum<br>*The strategy of obfuscation has been broadly applied— in search, location tracking, private communication, anonymity— and, as such, has been recognized as an important element of the privacy engineer's toolbox. However, there remains a need for clearly articulated case studies describing not only the engineering of obfuscation mechanisms but, further, providing a critical appraisal of obfuscation's fit for specific socio-technical applications. This is the aim of our paper, which presents our experiences designing and implementing AdNauseam, an open-source browser extension that leverages obfuscation to counter tracking by online advertisers.* |
| 16:30-17:30 | **Panel: Privacy Engineering and Practice**<br>Anupam Datta (Moderator), Saikat Guha (Microsoft), Sid Stamm (Rose-Hulman Institute of Technology), Aleksandra Korolova (University of Southern California), Ilya Mironov (Google)<br>This panel will present the current developments in different companies, and how they have managed to engineer privacy in their information systems.<br><br>***Saikat Guha*** *is a researcher at Microsoft Research India.* |

| | |
|---|---|
| | ***Sid Stamm*** *is Associate Professor of Computer Science and Software Engineering at Rose-Hulman Institute of Technology*<br>***Aleksandra Korolova*** *is a WiSE Gabilan Assistant Professor of Computer Science at Univeristy of Southern California.*<br>***Ilya Mironov*** *is a Staff Research Scientist working in cryptography and privacy at Google.* |
| 17:30–17:45 | **Wrap-up and concluding remarks** |