Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/cose

Computers & Security

A taxonomy for privacy enhancing technologies

Johannes Heurix [a,*], Peter Zimmermann [a], Thomas Neubauer [b], Stefan Fenz [b]

[a] Vienna University of Technology, Institute of Software Technology and Interactive Systems, Favoritenstrasse 9-11, 1040 Vienna, Austria

[b] SBA Research, Favoritenstrasse 16, 1040 Vienna, Austria

ARTICLE INFO

Article history:
Received 21 December 2014
Received in revised form
29 March 2015
Accepted 11 May 2015
Available online 20 May 2015

Keywords:
Privacy
Taxonomy
Privacy-enhancing technologies
Survey
Classification
Categorization

ABSTRACT

Privacy-enhancing technologies (PETs) belong to a class of technical measures which aim at preserving the privacy of individuals or groups of individuals. Numerous PETs have been proposed for all kinds of purposes, but are difficult to be compared with each other. The challenge here lies in the fact that information privacy is a comprehensive concept with solutions being diverse, with different focus and aims. As existing taxonomies cover information security-related aspects, while neglecting privacy-specific properties, this work aims at filling this gap by describing a universal taxonomy of PETs where the taxonomy aspects are selected such that they allow the categorization of PETs in different dimensions and properties to cover a wide area of privacy (e.g., user privacy, data privacy). It provides the reader with a tool for the systematic comparison of different PETs. This helps in identifying limitations of existing PETs, complementary technologies, and potential research directions. To demonstrate its applicability, the proposed taxonomy is applied to a set of key technologies covering different disciplines such as data anonymization, privacy-preserving data querying, communication protection, and identity hiding.

1. Introduction

Privacy is a notion known to virtually everybody, yet it is surprisingly difficult to define. Privacy is a complex and multi-dimensional concept and has been perceived as a legal, philosophical, or even technical term. Historically, the definition of an individual's privacy as the "right to be let alone", as it was phrased by the US Supreme Court in 1834, became famous. Privacy is also recognized as a human right in Article 8 of the European Convention for the Protection of Human Rights and Fundamental Freedoms (Council of Europe (1987)). A more specific notion is information privacy that specifically addresses personal information of an individual and the disclosure of this personal information, or in other words "the claim of individuals, groups or institutions to determine for themselves when, how, and to what extend information about them is communicated to others" (Westin, 1967). The emergence and advances of information processing technologies, the shift from an analogous to a digital data-centric world, and the trend towards collecting and storing personal data about everyone have led to understanding that the preservation of privacy is, in these days, more important than ever.

* *Corresponding author*. Tel.: +43 (1) 588 01 188 118.
E-mail address: johannes.heurix@tuwien.ac.at (J. Heurix).
http://dx.doi.org/10.1016/j.cose.2015.05.002

Privacy is not only a personal issue but a security issue in general, as one has to trust that the parties that are entrusted with the personal information protect this data adequately. Several serious security incidents such as the data breach at Heartland Payment Systems in 2008 or the infamous Sony hack of 2011 (Grocer, June 2011) clearly underpin this circumstance. More recently, Sony once again has been the target of an attack in 2014 which was even more dramatic than the 2011 incident as it demonstrated the existence of critical vulnerabilities and bad practices such as the existence of folders actually named 'passwords'. The stolen data included sensitive internal documents such as unreleased scripts, but also very personal employee-specific data such as salaries, internal memos, and even information about the medical conditions of employees and their family members (Rosenblatt, December 13 2014). Further security breaches reported in 2014 include incidents at Target, Home Depot, Michaels, and Apple. But not only consumer-related companies have been compromised. Still, financial institutions remain a major target for hackers as well, as demonstrated by the attack on JPMorgan Chase in 2014 (Silver-Greenberg et al., 2014). More and more, companies have stopped following the general rule of data minimization, i.e., to collect and store only as much information as necessary for the proper execution of their business processes. In fact, some of the largest Internet-centered companies, such as Google and Facebook, generate their revenues by collecting, processing, and selling as much personal data of their users as possible. Although thoughts of privacy were largely neglected by the average user for a long time, controversial actions such as the large-scale field-mapping for Street View, the introduction of automated facial recognition or the Timeline feature on Facebook (Dwyer, Fall 2011), have sparked the interest and the resistance of the general public, and people are beginning to raise their awareness for the preservation of their privacy.

Companies often argue that sensitive data records are sufficiently anonymized before being processed for marketing or other reasons. However, it has been shown that there is a clear risk of re-identification when considering meta data or background knowledge. A recent study (de Montjoye et al., 2015) involving the investigation of anonymized credit card records of 1.1 million people has proven that it is possible to uniquely re-identify 90% of the individuals when having knowledge of only four spatiotemporal points, i.e., where these individuals have been at which point in time. Depending on whether transactions of the same individual are linked together (which usually is the case for those anonymized data sets), once individuals have been re-identified, their complete set of transactions can thus be traced. This proves that simple anonymization is usually not sufficient but more sophisticated approaches are required.

In order to protect an individual's privacy, several legal acts were introduced, including the Directive 95/46/EC by the European Union (European Union, 1995) and its national derivatives or the US Privacy Act of 1974 (US Congress, 1974). However, legal regulations can be amended or overridden when deemed necessary with severe consequences as demonstrated in the reduction of civil rights by the introduction of the USA Patriot Act (US Congress, 2001). The introduction of more or less laws-conforming practices such as the unconditional data retention and other surveillance operations (e.g., Tempora (MacAskill et al., June 21 2013)) prove that legal regulations are less effective than expected and that they were often plainly ignored by governmental institutions themselves. Therefore, legal acts must be complemented with robust and sound technological solutions. However, privacy is not only compromised by enforced surveillance: Ubiquitous computing also plays a big role, especially with the advent of the Internet of Things where traditional computers are no longer required for data transfer, but also people, animals, or simple everyday objects are assigned unique identifiers and have the ability to communicate without computers. The increased address range of IPv6 makes it possible that even objects such as tooth brushes, thermostats, or refrigerators are granted access to the Internet. This interconnectivity produces a huge amount of additional data which, when combined with other data sources, may be more effective for surveillance than traditional CCTV. The big difference however is that while enforced surveillance is generally perceived as invasive and is thus disapproved, connecting all sorts of things to the Internet is often regarded as a big boon and is therefore widely accepted, although it is known how dangerous this can be.

So called privacy-enhancing technologies (PETs) aim at protecting the individual's privacy by the use of technical means. Their goal is to protect user identities by providing anonymity, pseudonymity, unlinkability, and unobservability of users as well as of data subjects (Fischer-Hübner, 2001), (Pfitzmann and Hansen, 2008). In the last decade, numerous PETs were proposed dealing with network traffic anonymization (e.g., TOR (Dingledine et al., 2004)), identity management (e.g., IDEMIX (Camenisch and Van Herreweghen, 2002)), or anonymous data storage (e.g., Free Haven (Dingledine et al., 2001)) based on different building blocks, including cryptographic primitives or the separation of information. As privacy is a many-faceted concept, PETs can target all different aspects of information privacy, which makes the classification of PETs a non-trivial task. Several taxonomies and classifications such as the excellent taxonomy of dependable and secure computing by Avizienis et al. (Avizienis et al., 2004) were proposed, but they largely deal with security-related issues and do not focus on privacy, or concentrate on specific aspects of privacy only (e.g., (Alvarez and Petrovic, 2003; Padayachee, 2012; Hansman and Hunt, 2005; Kjaerland, 2006; Shameli-Sendi et al., 2014)).

This work creates a comprehensive taxonomy of PETs that allows the categorization of PETs in different dimensions and aspects to cover a wide area of privacy (e.g., user privacy, data privacy). In particular, the taxonomy is designed to (i) categorize privacy-related aspects of PETs which are neglected in security-focused categorizations, as well as (ii) identify common aspects of PETs in order to cover a wide range of PETs from different disciplines. The overall goal is to provide the research community in the privacy area with a useful tool for the systematic comparison of existing PETs which makes it possible to identify the limitations of existing approaches and research gaps.