Online Privacy: Risks, Challenges, and New Trends

Esma Aïmeur(✉)

Département d'informatique et de recherche opérationnelle,
Faculté des arts et des sciences, Université de Montréal, Montréal, Canada
aimeur@iro.umontreal.ca

1 Summary

Being on the Internet implies constantly sharing information, personal or not. Nowadays, preserving privacy is not an easy feat: technology is growing too fast, leaving legislation far behind and the level of security awareness is insufficient. Websites and Internet services are collecting personal data with or without the knowledge or consent of users. Not only does new technology readily provide an abundance of methods for organizations to gather and store information, people are also willingly sharing data with increasing frequency, exposing their intimate lives on social media websites. Online data brokers, search engines, data aggregators, geolocation services and many other actors on the web are monetizing our online presence for their own various purposes. Similarly, current technologies including digital devices such as smartphones, tablets, cloud computing/SaaS, big data, BYOD are posing serious problems for individuals and businesses alike. Data loss is now a common event and the consequences are exceedingly damaging. Although there are means at our disposal to limit or at least acknowledge how and what we're sharing, most do not avail themselves of these tools and so the current situation remains unacceptable. Many privacy enhancing technologies (PETs) have been available for some time, but are not effective enough to prevent re-identification and identity theft.

In this tutorial, we propose how to address various issues inherent to Internet data collection and voluntary information disclosure – the Achilles' heel of privacy. We emphasize the problems and challenges facing privacy nowadays and conclude with some recommendations and best practices.

2 Goals

The goals of this tutorial are: (i) present the different facets of online privacy; (ii) review the various risks and technologies for preserving privacy (iii) provide recommendations and best practices.

3 Outline

Online Privacy. We introduce the ten principles of privacy and present various problems that make it difficult to preserve privacy.

J. Lopez et al. (Eds.): CRiSIS 2014, LNCS 8924, pp. 263–266, 2015.
DOI: 10.1007/978-3-319-17127-2_20

Information Sought. There are diverse types of information that can be obtained: Identifying information (name, age, gender, address, phone number, mother's maiden name, social insurance/security number, personal identification number (PIN), income, occupation, marital status, place of residence, etc.); Buying patterns (stores visited on a regular basis, accounts, assets, liabilities, etc.); Navigation habits (websites visited, frequency of visits, pseudonyms used on forums, acquaintances on the net, etc.); Lifestyle (hobbies, social networks, travel behaviour, vacation periods, etc.); Sensitive data such as employment, medical or criminal records; or Biological information (blood group, genetic code, fingerprints).

Internet Data Collection Techniques

Social Media. By their very nature, those websites aggregate, classify and collect various data about our preferences (*Likes, Shares, Re-tweets,* etc.), our opinions, and what we follow. As they try to mimic our day-to-day life, social networks can provide better insight about how we shop, how we judge products and services and how we share our preferences to marketers and companies.

Online Data Brokers. There are websites such as Abika.com or USSearch.com that, for a fee (sometimes for free), let anyone search for a name in order to retrieve all personal information about him or her available in a multitude of public records. Possible data include the person's name, address, date of birth, marital status, age of children, list of relatives, mortgage information, bankruptcy history and even sensitive information such as Social Security numbers, voting records or court records.

Search Engines. Search tools such as 123people.com, Whozat.com, Pipl.com, Peekyou.com, PeopleSearch.net, Peoplefinder.com, AnyWho, Yasni.com, are also good sources of information for administrators. They are free real-time people search tools that look into nearly every corner of the web to provide and gather information. There are also social network aggregator web sites such as Lifehacker.com, Spokeo.com, Spoke.com and Intelius.com, which collect data from various, online and offline sources (phone directories, social networks, etc.) and have large databases which they may unknowingly sell to malicious people.

Geolocation. Most of today's mobile phones are equipped with a Global Positioning System (GPS) chip, allowing people to know where they are located at any instant. Not only does the GPS user have access to his location, so do applications residing on the device. This is now known as geolocation. Aside from the well-known map functionality made possible by this technology, there are many interesting applications such as FourSquare. They are used to indicate your location to friends and, conversely, see their location. Similarly, Twitter has the option of attaching the user's location to its tweets.

Background Check. As we're adding content each and every day on social network, writing in blogs, and commenting on websites, we are often unaware on how much of that information is freely available for anybody to see. This provides a tremendous source of information for future employers and background check firms. Since a company wants to minimize hiring risks, it will certainly refrain from hiring someone

having tasteless pictures of him wandering online or expressing dangerous opinions whenever she or he gets the chance.

Online Conversations. Online conversation is clearly one of the main uses of the Internet: the rise of many webchat applications over the past few years (Facebook Chat, Google Chat, etc.) is a clear sign. IRC, even though we hear less and less about it, is still very active (approximately 400 000 simultaneously connected users on the top 100 servers at the time of writing) is still one of the most popular decentralized platforms. Since its creation, the use of SSL (secure socket layer) was slowly introduced for client-to-server connections, increasing the difficulty to eavesdrop in on a conversation.

Mobile Phones and Applications. Two main operating systems are now shaping the mobile market: Android, pushed by Google, and Apple iOS. Emerging from those two platforms are hundreds of thousands applications, or "apps", available from Google Play or Apple's AppStore. As some applications are tightly coupled with the users' personal data, some can go as far as requiring full access to the SMS information or the users' own address book. Most mobile platforms use two-way synchronization in order to keep the users' information up-to-date. This requires a considerable amount of trust from the user base, since the information shared can be very intimate and diverse: contacts, emails, passwords, credit card number, browser history, WIFI hotspots, WPA keys etc. The mobile-to-computer experience is narrowing the gap, allowing users to share all the information collected from one browser to another.

Big Data. Big data analytics offers powerful opportunities to access insights from new and existing data sets. It is driving data collection to become ubiquitous and permanent. Our digital traces can be used for different purposes including new ways of discriminating.

Privacy Enhanced Technologies (PETs). In order to ensure privacy and anonymity online, some technologies have been made available to the public.

PETs for Anonymization. For anonymous Communication Techniques, various technologies such as *Hordes, Crowds, Anonymizer1*, and private authentication protocols for mobile scenarios, have been proposed to keep users anonymous. Tor is a well known circuit-based low-latency anonymous communication service that addresses perfect forward secrecy, congestion control, directory servers, integrity checking, configurable exit policies, and has a practical design for location-hidden services via rendezvous points. Some ISPs and servers won't allow Tor for the time being.

PETs for Identity Management. There are several approaches in this area. In particular, Liberty Alliance's *federated approach7, OpenID8 authentication* (a decentralized approach), *Identity Metasystem Architecture* and *Generic Bootstrapping Architecture (GBA)* (focused telecommunication). More specifically, there are credential systems that allow authentication (and authorization) without identification by providing only the Personally Identifiable Information (PII) necessary for the transaction or a proof of entitlement.

PETs for Data Processing. In the field of data mining, various methods have been proposed to minimize access to users' privacy: *additive data perturbation; multiplicative data perturbation; data anonymization; secure multi-party computation; privacy preserving multivariate statistical analysis; probabilistic automata; privacy preserving formal methods; sampling-based methods; k anonymization classification; privacy in graph data; and statistical disclosure control.*

Re-identification. Despite all the tools cited above, it is possible to re-identify people—that is, to determine the exact identity of a person by gathering various pieces of information disseminated across the web. Different techniques exist such as: *Linkage, Inference, Homogeneity, Graphs and Machine Learning.*

This process has recently gained popularity following the emergence of personal databases on the Internet. As a consequence, there is no absolute protection, even with Privacy Enhancing Technologies.

Identity Theft. According to the definition given by the OCDE, "identity theft occurs when a party acquires, transfers, possesses, or uses personal information of a natural or legal person in an unauthorized manner, with the intent to commit, or in connection with, fraud or other crimes". People must understand that identity theft not only affects people using their credit card or debit card, it also includes people who use their name, their Social Insurance/Security, Number, online passwords and even their address.

The Right to Erasure. Another hot topic is the right to permanently remove our online presence and how the different services (search engines, social networks, etc.) should behave. Currently, it is very handy to sanitize our online presence as many services are keeping an artificial presence, Facebook being the main culprit here. Since anybody is able to tag any picture (even if the tag isn't linked to a precise profile, it can be used by facial recognition software), a bad picture can therefore be easily be retrieved.

Recommendations. To minimize harm we provide recommendations and best practices for each of the following aspects: technology, social behaviour, ethics, legislation etc. We conclude with the new challenges that face privacy nowadays.