Optimized, direct sale of privacy in personal data marketplaces

Javier Parra-Arnau*

Department of Computer Science and Mathematics, Universitat Rovira i Virgili (URV), E-08034 Tarragona, Spain

ABSTRACT

Very recently, we are witnessing the emergence of a number of start-ups that enables individuals to sell their private data directly to brokers and businesses. While this new paradigm may shift the balance of power between individuals and companies that harvest and mine data, it raises some practical, fundamental questions for users of these services: how they should decide which data must be vended and which data protected, and what a good deal is. In this work, we investigate a mechanism that aims at helping users address these questions. The investigated mechanism relies on a hard-privacy model and allows users to share partial or complete profile data with broker and data-mining companies in exchange for an economic reward. The theoretical analysis of the trade-off between privacy and money posed by such mechanism is the object of this work. We adopt a generic measure of privacy although part of our analysis focuses on some important examples of Bregman divergences. We find a parametric solution to the problem of optimal exchange of privacy for money, and obtain a closed-form expression and characterize the trade-off between profile-disclosure risk and economic reward for several interesting cases. Finally, we evaluate experimentally how our approach could contribute to privacy protection in a real-world data-brokerage scenario.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

Over the last recent years, much attention has been paid to government surveillance, and the indiscriminate collection and storage of tremendous amounts of information in the name of national security. However, what most people are not aware of is that a more serious and subtle threat to their privacy is posed by hundreds of companies they have probably never heard of, in the name of commerce.

They are called *data brokers*, and they gather, analyze and package massive amounts of sensitive personal information, which they sell as a product to each other, to advertising companies or marketers, often without our knowledge or consent. A substantial chunk of this is the kind of harmless consumer marketing that has been going on for years. Nevertheless, what has recently changed is the amount and nature of the data being extracted from the Internet and the rapid growth of a tremendously profitable industry that operates with no control whatsoever. Our habits, preferences or interests, our friends, personal data such as date of birth, number of children or home address, and even our daily movements, are some examples of the personal information we are giving up without being aware it is being collected, stored and finally sold to a wide range of companies.
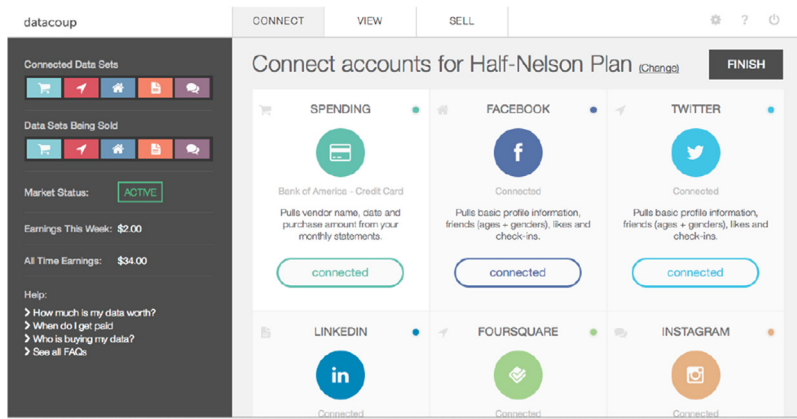

E-mail address: javier.parra@urv.cat

Fig. 1. Screenshot of Datacoup which allows users to earn money by sharing their personal data.
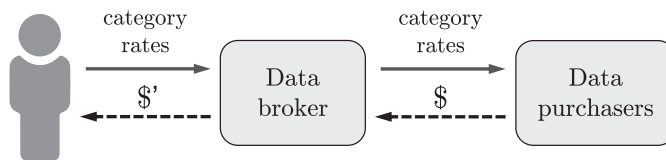

Fig. 2. Conceptual depiction of the data-purchasing model assumed in this work. In this model, users first send the data broker their category rates, that is, the money they would like to be paid for completely exposing their actual interests in each of the categories of a profile. For example, a user might be willing to reveal their purchasing habits on clothing for 7 dollars. Based on the rates chosen for each category, data buyers decide then whether to pay the user for learning their profile and gaining access to the underlying data. Finally, depending on the offer made, the disclosure may range from portions of their profile to the complete actual profile.

A vast majority of the population understands that this is part of an unwritten contract whereby they get content and services free in return for letting advertisers track their behavior; this is the current barker economy that, for example, currently sustains the Web. But while a significant part of the population finds this tracking invasive, there are people who do not give a toss about being mined for data [3].

Very recently we are witnessing the emergence of a number of start-ups that hope to exploit this by buying access to our social-networks accounts and banking data. One such company is Datacoup, which lets users connect their apps and services via APIs in order to sell their data. Datacoup, and similar start-ups, however, do not provide raw data to potential data purchasers, among others, retailers, marketers, insurance companies and banks. Rather, they typically build a profile that gives these companies an overview of a user's data.

The emergence of these start-ups is expected to provide a win-win situation both for users and data buyers. On the one hand, users will receive payments, discounts or various rewards from purchasing companies, which will take advantage of the notion that users are receiving a poor deal when they trade personal data in for access to "free" services. On the other hand, companies will earn more money because the quality of the data these start-ups will offer to them will be much greater than that currently provided by traditional brokers —the problem with the current data brokers is often the stale and inaccurate data [51].

Undoubtedly, the creation of a marketplace in personal data will represent a significant shift in the balance of power between individuals and companies that gather and mine data. According to some recent studies, this is a shift people would be willing to embrace. Just over half of the 9 000 people surveyed worldwide said they would share data about themselves with companies in exchange for cash [5]. A separate survey has found that 42 percent of more than a thousand 13-17-year-olds in the U.K. would rather accept cash for their personal data than earn money from a job [6]. Lastly, it was reported in [4] that 56% of the consumers surveyed would be willing to give up personal data provided that they received some kind of economic compensation.

The possibility that individuals may vend their private data *directly* to businesses and retailers will be one step closer with the emergence of companies like Datacoup. For many, this can have a liberating effect. It permeates the opaque data-exchange process with a new transparency, and empowers online users to decide what to sell and what to retain. However, the prospect of people selling data directly to brokers poses a myriad of new problems for their owners. How should they manage the sale of their data? How should they decide which elements must be offered up for mining and which ones protected? What is a good deal?