Quantified Self and the Privacy Challenge

Hamed Haddadi
Queen Mary University of London

Ian Brown
Oxford Internet Institute

ABSTRACT

The increasing availability of personal activity monitors, tracking devices, wearable recording devices, and associated smartphone apps has given rise to a wave of *Quantified Self* individuals and applications. The data from these apps and sensors are usually collected by associated apps and uploaded to the software developers for feedback to individual and their selected partners. In this paper we highlight the privacy risks associated with this practice, demonstrating the ease with which an app provider can infer individuals co-location and joint activities without having access to specific location data. We highlight a number of potential solution to this challenge in order to minimise the privacy leakage from these applications.

Author Keywords

Quantified Self; Privacy ; Data Mining

INTRODUCTION

The *Quantified Self* (QS) phenomena is currently at the centre of consumer interest and industry attention in the wearable tech industry. Many individuals are interested in understanding their own activity, emotions [6], sleep, and health patterns, while *experimenting* with their own bodies.[1] There are an increasing number of QS sensors and complementary smart phone applications are available on the market today: sleep quality monitors, heart rate monitors, personal video recording devices such as Google Glass, skin conductance measurement sensors, accelerometers, pedometers and step counters, to mention a few. The new range of smartphones are in fact designed to cater for continuous activity tracking.

One obstacle to adoption of activity trackers has been the loss of consumers' interest after a period of use. Hence many developers have recently focused on correlating physical activity with other user-provided data such as calorie intake and mood, in order to draw more appealing inferences to visual feedback to the users. These include life logging apps (e.g., Saga[2]) or location-mining apps (e.g., GoogleNow) which heavily rely on continuos reporting of the users' location, all leading to major privacy concerns for the users and those associated with them. In addition to increasing number of high precision sensors, the associated smartphone apps ask for an increasingly larger number of permissions in order to look for more signals and data to feed into their pattern matching algorithms.

Availability of more data sources and feedback to the user is essential in usefulness and accuracy involved with the personal data ecosystem [7]. In [10], Watson discusses some societal aspects of use of personal data for QS applications. Continuous reportage of health data to doctors and emergency services can have benefits for researchers. However, with addition of personal data from a variety of linked sources, privacy issues start to emerge which need to be identified, presented to individuals, and addressed through technology and regulatory mechanisms [1].

In this paper we discuss the privacy threats posed by the use of activity monitors. We discuss the ease of detecting whether two individuals' have been spending time together, *without* the need for location information. Using aggregate activity data collected by a popular fitness tracker, we demonstrate the similarity between the activity log of individuals who have been spending time together. This enables the aggregator to be able to easily identify the location and mobility patterns of an individual who has not opted to share their location with third parties. We will then discuss potential mitigation strategies for respecting and preserving individuals' location privacy, while enabling them to enjoy the personal benefits of QS devices.

PRIVACY RISKS

We use the data available from the Jawbone UP API[3] from 4 individuals with known activities and overall 147 hours spent together during 2 short travel periods. These activities include walking, running, commuting, and exercise periods. Our data includes the step count, activity times, inactive periods, and distance covered available from the pedometer and accelerometer in the Jawbone dataset, in addition to inactive times. We treat the step count as a time-series signal for our analysis of the data from individuals. Figure 1 displays a sample of the data from two adults spending a day together in a ski resort. When looking at the step-count data from the two individuals using Kolmogorov-Smirnov distance[4] we see that the hourly step count time series data for individuals spending time together in nearly all cases displays less than 5% difference, with increasing confidence the longer time the individuals spend together. Using time-series correlations [8] would yield an even higher accuracy by removing the transient time-lag in activities such as walking up the same set of

[1] `http://www.economist.com/node/21548493`
[2] `http://getsaga.com/`
[3] `https://jawbone.com/up/developer`
[4] See [3] for an explanation on the reasons for use of KS-distance.

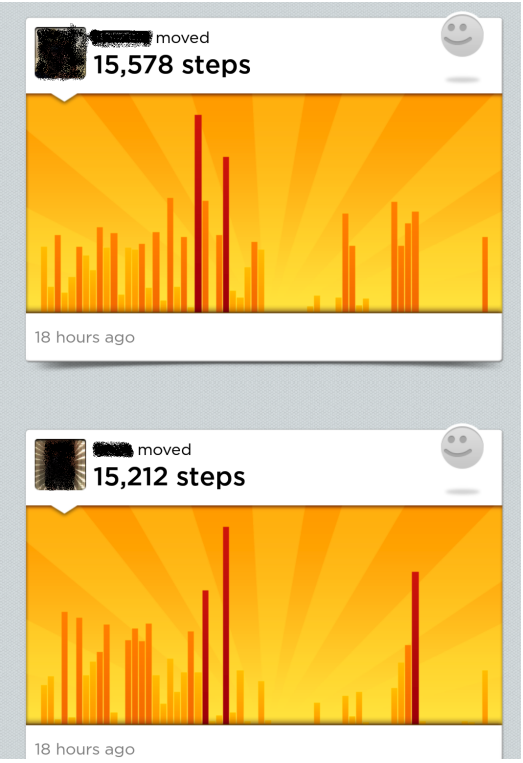

Figure 1. Activity data from two participants

stairs or walking through the same terrain within a few seconds of each other. Hence the current cloud-based data collection methods enable easy identification of co-located individuals,[5] or infer their specific type of activity, if only one of them chooses to share their location with the app provider.

DISCUSSIONS & RESEARCH CHALLENGES

Considering the ease of identification of individuals using a small number of personal information pieces [2, 9] and the issues of data ownership and ethics, we need to take active steps to enable the individuals' rights and privacy in the wearable tech industry. We are developing a client side platform for e-health and QS applications to provide a thorough feedback mechanism to different interest groups relevant to an individual, e.g., on a personal level, at a community level, and to health practitioners, using the Privacy Analytics framework [5].

Creation of a successful personal data ecosystem relies on cooperation between service providers, users, and regulators. An important challenge in the wearable tech landscape is to consider the rights of individual being tracked by these devices, e.g., the individuals appearing in Google Glass videos in public. Currently the individuals have no way of getting engaged in the data collection and tracking process. A simple method could rely on continuous broadcast of a *Do-Not-Track* beacon from smart devices carried by individuals who prefer not to be subjected to image recognition by wearable cameras. Naturally, respecting this beacon and requirement depends on the regulatory enforcement and the device providers conforming with these requests. This approach is similar to the Do-Not-Track initiative on the web,[6] though relying on the local broadcast of the signal, successful reception by the tracking device, and interrupting the recording process. Indeed the level of intrusion of Google Glass may justify a requirement for an opt-in approach instead.

In ongoing research we are investigating the feasibility of this form of broadcast for signalling privacy preferences as well as privacy-preserving location-based advertising [4]. In related work, we are developing a framework for the tracking services (such as Google and Facebook) to inform the individual about the identity and location of the data requesting party, hence reducing information asymmetry

[5]perhaps in a similar manner to the NSA CO-TRAVELER program `http://apps.washingtonpost.com/g/page/world/how-the-nsa-is-tracking-people-right-now/634/`

[6]`http://donottrack.us/`

REFERENCES

1. Brown, I., Brown, L., and Korff, D. Using nhs patient data for research without consent. *Law, Innovation and Technology 2*, 2 (2010-12-01T00:00:00), 219–258.
2. Domenico, M. D., Lima, A., and Musolesi, M. Interdependence and predictability of human mobility and social interactions. In *Nokia Mobile Data Challenge* (June 2012).
3. Haddadi, H., Fay, D., Jamakovic, A., Maennel, O., Moore, A. W., Mortier, R., Rio, M., and Uhlig, S. Beyond node degree: evaluating as topology models. *arXiv preprint arXiv:0807.2023* (2008).
4. Haddadi, H., Hui, P., and Brown, I. Mobiad: private and scalable mobile advertising. In *Proceedings of the fifth ACM International Workshop on Mobility in the Evolving Internet Architecture*, MobiArch '10, ACM (New York, NY, USA, 2010), 33–38.
5. Haddadi, H., Mortier, R., Hand, S., Brown, I., Yoneki, E., McAuley, D., and Crowcroft, J. Privacy analytics. *SIGCOMM Comput. Commun. Rev. 42*, 2 (Mar. 2012), 94–98.
6. Lathia, N., Pejovic, V., Rachuri, K., Mascolo, C., Musolesi, M., and Rentfrow, P. Smartphones for large-scale behavior change interventions. *Pervasive Computing, IEEE 12*, 3 (July 2013), 66–73.
7. Mortier, R., Haddadi, H., Henderson, T., McAuley, D., and Crowcroft, J. Challenges & opportunities in human-data interaction. In *DE2013: Open Digital* (Salford, UK, 2013).
8. Rojkova, V., and Kantardzic, M. M. Analysis of inter-domain traffic correlations: Random matrix theory approach. *CoRR abs/0706.2520* (2007).
9. Sweeney, L. Simple demographics often identify people uniquely. *Health (San Francisco)* (2000), 1–34.
10. Watson, S. M. *Living with Data: Personal Data Uses of the Quantified Self*. Oxford Internet Institute Masters Thesis, 2013.