Challenges of Trustable AI and Added-Value on Health
B. Séroussi et al. (Eds.)

doi:10.3233/SHTI220635

How to Publish Wearables' Data: Practical Guidelines to Protect User Privacy

Thomas MARCHIORO [a,b,1], Andrei KAZLOUSKI [a,b] and Evangelos MARKATOS [a,b]
[a] *Computer Science Department, University of Crete, Greece*
[b] *Institute of Computer Science, Foundation for Research and Technology*

Abstract. The spread of wearable fitness trackers has contributed to the increase of various fitness studies, utilizing such devices to collect data from a group of users. When these data are disclosed, the lack of sanitization can lead to severe privacy risks. In this paper, we discuss the various threats that are posed by disclosing unaltered information from wearable devices. We also dismiss common fallacies of fitness data sharing and present practical guidelines to preserve user privacy.

Keywords. Electronic Health Records, Data Privacy, Wearable Devices

1. Introduction

In lifelogging experiments the activity of a small group of people is monitored through wearable fitness trackers for a number of days. During this period, the participants provide a considerable amount of information, including surveys of microdata (e.g., age, gender, height, weight), as well as activity logs of steps, burned calories, heart rate, and more. When these data are aggregated and published, often little effort is put to really make the disclosed information "anonymous". As a result, details about the participants' activity and habits can potentially be leaked, compromising their privacy.

2. Privacy aspects of sharing data from wearables

Albeit well-known in privacy literature, sensitiveness of quasi-identifiers tends to be neglected by data publishers, likely with the intent of providing useful information about the dataset population. For example, most participants of prominent public life-logging datasets [1,2] can be de-anonymized based on quasi-identifiers, such as personal and physical characteristics. Applying standard anonymization to these attributes (e.g., using k-anonymity [3] or personalized privacy [4]) is thus a first step to protect users' privacy. However, this may not be enough to guarantee that a target is not found in the dataset. Activity logs themselves often contain information that is derived from an individual's attributes: calories, for instance, are often estimated by combining activity information with physical characteristics such as age, gender, height, and weight [5]. Thus, an ad-

[1]This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813162. The content of this paper reflects the views only of their author(s). The European Commission/ Research Executive Agency are not responsible for any use that may be made of the information it contains.

versary can leverage additional records that she possesses to find a target user in the dataset, even when personal attributes are removed [6], and glean insights about them. Such insights may include:

- Whether he has an active nightlife (on what days, what times)
- When he commutes to the workplace. Whether he is active during office hours.
- When he gets back home, the time he usually leaves the house.

Needles to say, such information is extremely sensitive, and may be utilized by a wide variety of malicious actors such as thieves, burglars, stalkers, employers, etc.

Fallacies and pitfalls In order to effectively mitigate the above-mentioned threats, the following common misconceptions must be dispelled.
Fallacy: Removing participants' identifiers (name, phone number, email address, etc.) is enough to protect their privacy.
Since fitness data themselves carry a wealth of information about the person who produce them, it might be possible to identify the users based solely on their activity information.
Fallacy: Removing physical attributes (age, weight, height, etc.) fully protects privacy.
Some of the data produced by the fitness trackers depend on physical parameters of the wearer. Therefore, it may be feasible to reconstruct those characteristics.
Pitfall: Publishing all the fitness information collected by the wearable device.
The more fitness data are published, the likelier the attacker infers sensitive insights.

Guidelines for privacy protection. Besides traditional anonymization approaches, user-specific information carried by time series of fitness records (steps, calories, etc.) must be limited via the following approaches:

- *Re-sampling/sub-sampling*, which consists in reducing the amount of samples of a time series by aggregating the data at a lower frequency. For instance, the total calories burned in a day could be published instead of 24 hourly records.
- *Reducing the granularity* in a similar fashion to the generalization principle applied for microdata. This means, e.g., binning daily steps in a range $8000 - 8500$ instead of storing a more precise value such as 8361.

Finally, fitness datasets should be published with a specific purpose in mind (e.g., showing how a physical activity intervention affected certain parameters), and all the information which is irrelevant to that purpose should be discarded.

References

[1] Thambawita V, Hicks S, Borgli H, Pettersen SA, Johansen D, Johansen H, et al.. PMData: A sports logging dataset. OSF Preprints; 2020. Available from: https://osf.io/k2apb.
[2] Open Humans Fitbit Connection. OpenHumans; 2016. Available from: https://www.openhumans.org/activity/fitbit-connection/.
[3] Sweeney L. Achieving k-anonymity privacy protection using generalization and suppression. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems. 2002;10(05):571-88.
[4] Xiao X, Tao Y. Personalized privacy preservation. In: Proceedings of the 2006 ACM SIGMOD international conference on Management of data; 2006. p. 229-40.
[5] Fitbit help;. Available from: https://help.fitbit.com/articles/en_US/Help_article/1141.
[6] Marchioro T, Kazlouski A, Markatos E. User Identification from Time Series of Fitness Data. In: Proceedings of the 18th International Conference on Security and Cryptography - SECRYPT,. INSTICC. SciTePress; 2021. p. 806-11.