Protecting privacy of fitness data

Thomas Marchioro [12#,], **Andrei Kazlouski**[12*]

[1]Institute of Computer Science, FORTH

[2]Computer Science Department, UOC

# Presenting author: XXX, email: marchiorot@ics.forth.gr
* Corresponding author: XXX, email: andrei@ics.forth.gr

ABSTRACT

A significant rise in popularity of sedentary and office lifestyles along with the surge of remote working have forced people to search for alternative ways to exercise and keep active. This, in turn, has prompted a substantial increase of sales for activity trackers and other wearable devices. Such devices monitor a wide variety of health and fitness parameters, including steps, sleep, heart rate, and blood pressure, etc. Users of activity trackers tend to share their fitness results online in hopes of keeping themselves motivated, receiving feedback and encouragements from the like-minded peers, and learning valuable fitness gimmicks. Furthermore, wearables and the activity data they collect have become a cornerstone of various fitness experiments and lifelogging studies. However, fitness data publishers tend to neglect sanitizing such information and making it private. In our work we investigate the reasons for such data to be protected and not to be released uncontrollably. We also study what insights can be inferred from fitness data, and the ways to mitigate those leaks.

In particular, we focus on identifying people from the fitness data they produce. In our works [1,2] we demonstrated that it is feasible to link users back to their fitness records in small-sized lifelogging datasets. We also showed that in such datasets it is possible to uniquely identify minority individuals – people who has very distinct physical attributes. Moreover, we investigate whether the adversary can actually glean these parameters. We demonstrated that it is achievable to learn the gender, height and BMI directly from the activity data produced by wearables.

Furthermore, we explore alternative ways of releasing fitness data, by sanitizing the samples and protecting personal information. This can be achieved by applying changes to the real data, or generating synthetic samples with similar data distributions.

REFERENCES

[1] Marchioro, T., Kazlouski, A., and Markatos, E. (2021). User identification from time series of fitness data. In International Conference on Security and Cryptography (SECRYPT), pages 806–811.

[2] Kazlouski, A., Marchioro, T., and Markatos, E. What your Fitbit says about you: De-anonymizing users in lifelogging datasets. *To appear* In International Conference on Security and Cryptography (SECRYPT 2022)