Prof. Christian Cachin
Orestis Alpos, Luca Zanolini

Exercise 9

9.1 Large-scale differentially private data analysis (10pt)

References to deployments of differentially private machine learning by Google and Microsoft were given in class. Apple has rolled out d.p. analytics capabilities in its operating systems iOS and macOS since 2017. (These features were also broadly publicized.) Apple's stated goal is to better understand how people use its devices and services.

i. Read Sections 1–3 and 6–7 of the paper "Learning with privacy at scale," which describes Apple's algorithms:
   `https://docs-assets.developer.apple.com/ml-research/papers/learning-with-privacy-at-scale.pdf`
   There is also a blog post with some of the same material:
   `https://machinelearning.apple.com/research/learning-with-privacy-at-scale`
ii. Optionally, if you are interested to learn about the *count min sketch* and the *private count mean sketch* algorithms[1], watch a 45-min tutorial by Abhradeep Guha Thakurta, one of the creators of Apple's d.p. analytics system:
   - Differential Privacy: From Theory to Deployment. Invited talk at the 26th Usenix Security Symposium, 2017.
     `https://www.usenix.org/conference/usenixsecurity17/technical-sessions/presentation/thakurta`

With this background you are asked to design a d.p. data analysis method for a large Swiss company or institution, which has a million or more customers. Its app is installed in 100'000s of mobile devices.

Answer the following in max. one page of text overall:

a) Describe your scenario, i.e., the company or institution, the type of data it wants to analyze, and the goals of this analysis. What assumptions do you need to roll out your method?
b) Which aspects can you adopt from the solutions used by Google or Apple?
c) What differs from these deployments?

[1] They are described in Sections 4–5 of the paper, but much more technically.