Report

Written by Amalie Dyrberg Holm (amdh), Lambda Nørbjerg (emno), and Mikkel Bistrup Andersen (mbia).

GDPR Reflections

TODO: This needs to become text, not a bullet point list TODO: a_c^b

What are the potential issues in having the hospital store plaintext private data provided by patients even if they have consented to participate on the experiment and have their data processed?

- Storage limitation -> deletion when the medical experiment is done.
- Storage limitation II -> The hospital can only store the relevant information for the experiment, not extra information about the Data Subjects.
- Integrity and confidentiality -> must insure security, integrity, and confidentiality, by encryption or similar. There should be no leaks or data breaches by technical faults or staff problems.
- if the Hospital hospital stores plaintext private data, it must take appropriate Technical and Organizational measures, like MFA to view it, e3e encryption, and staff training.
- Since medical data, is a special category of data, the Hospital needs to appoint a Data Protection Officer.
- A consequence of 'The right to data portability', is that any user must be able to download their own dataset, and therefore it must exists to be downloaded. Having the Aggregation only is not enough. However, if the re-identification is impossible, this does not apply.
- There must be a Data Processing Agreement between the hospital and the security experts, if the security experts are doing processing, and not just consultancy.
- Further, there must be a Data Processing Agreement, with each of the parties in the Federated Learning with Secure Aggregation, in this case, the patients
- The hospital, and its processors, must keep a record of information on the use of data, for each process. This includes the purposes of the processing, who will receive the data, and a general description of the technical and organizational security measures.
- The data should preferably stay within the EU, encrypted or otherwise.

Would these issues be solved by removing the patients' names from their data before storing it?

- The goal of Anonymisation, is to reduce the number of direct and indirect identifiers, thus making the re-identification likelihood smaller, without significant impact to utility of the dataset.
- patients' names is one such direct identifier.

- If re-identification is next to impossible, then the data is no longer subject to GDPR regulation.
- In this case, the data is not intended to be shared with anyone in its plaintext form. The point of decentralized processing, is to make a leak less interesting and make re-identification unlikely. Should a data breach happen, it is still important to lessen re-identification.
- Replacing or Removing patients' names with numbers, UUID's, or other names, is pseudonymisation not anonymisation, and it is still considered personal data.
- Even when the patients' names have been replaced or removed, there must be measures in place, to remove other sources of indirect identification.
- With only three participants to the experiment, the re-identification likelihood, will likely stay the same regardless of Anonymisation, and of name removal.
- Most likely, there needs to be a list that maps between the patients' names, and their name replacements. Such a list would need to be kept really secret.
- Closely related to 'Data Minimization' and 'Privacy by design'

What are the remaining risks in using Federated Learning with Secure Aggregation as suggested?

- a
- b
- c

Adversarial model

describing the adversarial model you are working on,

Building blocks of the solution

describing the building blocks of your proposed solution,

Combination

how they are combined in your final solution and

Security guarantees

why they guarantee security against the adversary you describe.

Sources

- https://gdpr.eu/what-is-gdpr/
- https://www.wired.com/story/what-is-gdpr-uk-eu-legislation-compliance-summary-fines-2018/

- https://www.gdprsummary.com/gdpr-summary/
- https://www.gdprsummary.com/schrems-ii/
- https://www.gdprsummary.com/anonymization-and-gdpr/
- https://www.edps.europa.eu/system/files/2021-04/21-04-27_aepd-edps_anonymisation_en_5.pdf
- https://eprint.iacr.org/2017/281.pdf Practical Secure Aggregation for Privacy-Preserving Machine Learning
- https://www.owkin.com/blogs-case-studies/defending-against-attacks-on-federated-learning-using-secure-aggregation