A cryptography game changer for biomedical research at scale

October 11 2021

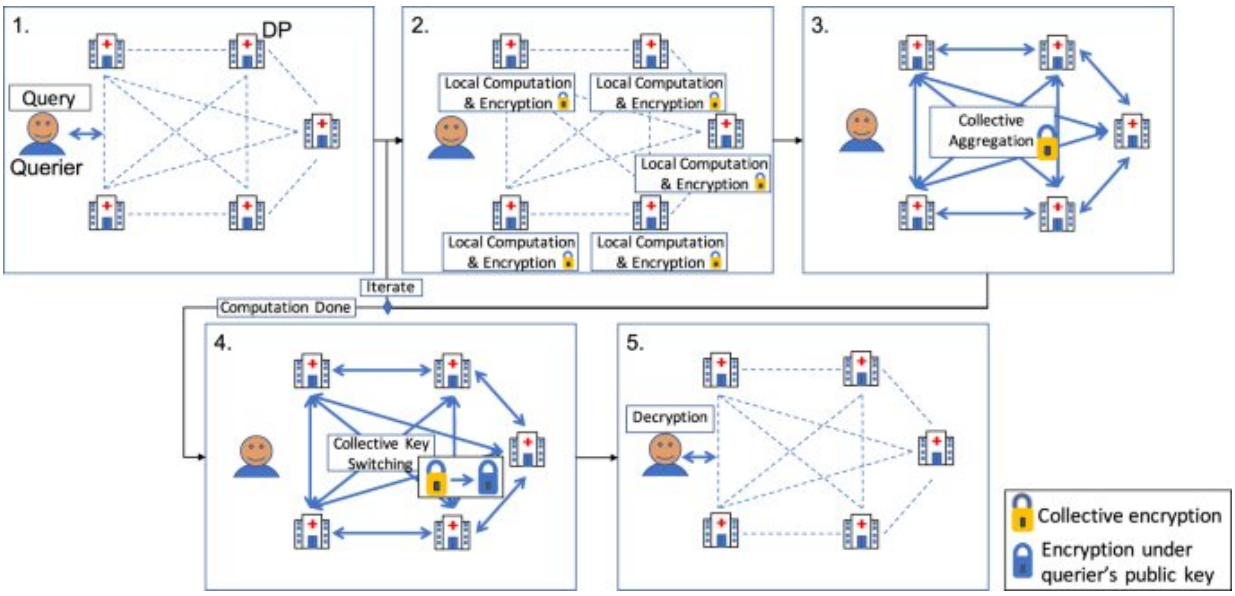

Fig. 1: System Model and FAMHE workflow. All entities are interconnected (dashed lines) and communication links at each step are shown by thick arrows. All entities (data providers (DPs) and querier) are honest but curious and do not trust each other. In 1. the querier sends the query (in clear) to all the DPs who (2.) locally compute on their cleartext data and encrypt their results with the collective public key. In 3. the DPs' encrypted local results are aggregated. For iterative tasks, this process is repeated (Iterate). In 4. the final result is then collectively switched by the DPs from the collective public key to the public key of the querier. In 5. the querier decrypts the final result. Credit: DOI: 10.1038/s41467-021-25972-y

Predictive, preventive, personalized and participatory medicine, known as P4, is the healthcare of the future. To both accelerate its adoption and maximize its potential, clinical data on large numbers of individuals must be efficiently shared between all stakeholders. However, data is hard to gather. It's siloed in individual hospitals, medical practices, and clinics around the world. Privacy risks stemming from disclosing medical data are also a serious concern, and without effective privacy preserving technologies, have become a barrier to advancing P4 medicine.

Existing approaches either provide only limited protection of patients' privacy by requiring the institutions to share intermediate results, which can in turn leak sensitive patient-level information, or they sacrifice the accuracy of results by adding noise to the data to mitigate potential leakage.

Now, researchers from EPFL's Laboratory for Data Security, working with colleagues at Lausanne University Hospital (CHUV), MIT CSAIL, and the Broad Institute of MIT and Harvard, have developed "FAMHE." This federated analytics system enables different healthcare providers to collaboratively perform statistical analyses and develop machine learning models, all without exchanging the underlying datasets. FAHME hits the sweet spot between data protection, accuracy of research results, and practical computational time—three critical dimensions in the biomedical research field.

In a paper published in *Nature Communications* on October 11, the research team says the crucial difference between FAMHE and other approaches trying to overcome the privacy and accuracy challenges is that FAMHE works at scale and it has been mathematically proven to be secure, which is a must due to the sensitivity of the data.

In two prototypical deployments, FAMHE accurately and efficiently reproduced two published, multi-centric studies that relied on data

centralization and bespoke legal contracts for data transfer centralized studies—including Kaplan-Meier survival analysis in oncology and genome-wide association studies in medical genetics. In other words, they have shown that the same scientific results could have been achieved even if the the datasets had not been transferred and centralized.

"Until now, no one has been able to reproduce studies that show that federated analytics works at scale. Our results are accurate and are obtained with a reasonable computation time. FAMHE uses multiparty homomorphic encryption, which is the ability to make computations on the data in its encrypted form across different sources without centralizing the data and without any party seeing the other parties' data" says EPFL Professor Jean-Pierre Hubaux, the study's lead senior author.

"This technology will not only revolutionize multi-site clinical research studies, but also enable and empower collaborations around sensitive data in many different fields such as insurance, financial services and cyberdefense, among others," adds EPFL senior researcher Dr. Juan Troncoso-Pastoriza.

Patient data privacy is a key concern of the Lausanne University Hospital. "Most patients are keen to share their health data for the advancement of science and medicine, but it is essential to ensure the confidentiality of such sensitive information. FAMHE makes it possible to perform secure collaborative research on patient data at an unprecedented scale," says Professor Jacques Fellay from CHUV Precision Medicine unit.

"This is a game-changer towards personalized medicine, because, as long as this kind of solution does not exist, the alternative is to set up bilateral data transfer and use agreements, but these are ad hoc and they take months of discussion to make sure the data is going to be properly

protected when this happens. FAHME provides a solution that makes it possible once and for all to agree on the toolbox to be used and then deploy it," says Prof. Bonnie Berger of MIT, CSAIL, and Broad.

"This work lays down a key foundation on which federated learning algorithms for a range of biomedical studies could be built in a scalable manner. It is exciting to think about possible future developments of tools and workflows enabled by this system to support diverse analytic needs in biomedicine," says Dr. Hyunghoon Cho at the Broad Institute.

So how fast and how far do the researchers expect this new solution to spread? "We are in advanced discussions with partners in Texas, The Netherlands, and Italy to deploy FAMHE at scale. We want this to become integrated in routine operations for medical research," says CHUV Dr. Jean Louis Raisaro, one of the senior investigators of the study.

More information: David Froelicher et al, Truly privacy-preserving federated analytics for precision medicine with multiparty homomorphic encryption, *Nature Communications* (2021). DOI: 10.1038/s41467-021-25972-y

Provided by Ecole Polytechnique Federale de Lausanne