

I am pleased to nominate the paper “Privacy-Preserving and Efficient Verification of the Outcome in Genome-Wide Association Studies” (accepted and presented at PETS 2022) for the Best Scientific Cybersecurity Paper Award.

This paper proposes a framework that efficiently verifies the correctness of the aggregate statistics obtained as a result of a genome-wide association study (GWAS) conducted by a researcher while protecting individuals’ privacy in the researcher’s dataset. GWAS is a popular method for identifying genetic variations (mutations) that are associated with a particular phenotype (disease). The researcher publishes the workflow of the conducted study, its output, and associated metadata. For researchers, showing that discovered associations are correctly computed and the results are reproducible is of immense importance, especially if they are planning to use the research findings in their study (e.g., personalized medicine). Computational errors might occur during GWAS, and the researcher may unintentionally provide wrong results as the output of the research (GWAS). It is trivial to verify the correctness of the research findings if the input dataset is provided. However, the input dataset might not always be released as it may contain sensitive information about individuals. GWAS studies include highly sensitive datasets that contain genomic and phenotypic information of individuals that participate in the study. Thus, in the proposed framework, the researcher keeps the research dataset private while providing, as part of the metadata, a partial noisy dataset (that achieves local differential privacy). To check the correctness of the workflow output, a verifier makes use of the workflow, its metadata, and results of another GWAS (conducted using publicly available datasets) to distinguish between correct statistics and incorrect ones. Our results on real genomic data show that the proposed framework can correctly classify all the correct statistics that are highly associated with the considered phenotype and all the incorrect statistics that imply a significant overselling of the real outcome (e.g., the researcher unintentionally reports stronger associations than the original ones).

Overall, this paper tackles an important problem and shows that the correctness of GWAS statistics can be efficiently verified with high confidence in a privacy-preserving way. We believe that this work will be a valuable step towards providing provenance in a privacy-preserving way while providing guarantees to the users about the correctness of the results.