CoVault: A Secure Analytics Platform
Robert De Viti, Isaac Sheff, Noemi Glaeser, Baltasar Dinis, Jonathan Katz, Rodrigo Rodrigues, Bobby Bhattacherjee, Anwar Hithnawi, Deepak Garg and Peter Druschel
MPI-SWS, MPI-SP, University of Maryland, INESC-ID, ETH Zürich

Problem: Statistical queries over sensitive personal data

Raw data may be very sensitive, but (statistical) query results are usually privacy-preserving.

Example data: Personal health, mobility, activity, social contacts.

Example analytics: Epidemics, rare diseases, transportation and urban planning, finance.

Secure Analytics

CoVault is a secure analytics platform: data sources consent to the use of their sensitive data for a predefined set of analytics queries performed by a specific group of analysts, and for a limited period of time.

Key primitive: Functional Encryption (FE)

CoVault uses Functional Encryption (FE):
- a secret key allows one to learn a predefined set F of functions of encrypted cleartext m, but nothing else about m.

CoVault’s multi-party FE Primitive

Secret sharing
- To secret-share data m, the data source:
  - Generates random r
  - Computes $m \oplus r$
  - Sends r and $m \oplus r$ to two QPs.

2-party secure computation (2PC)
- To compute $f(m)$, the QPs run 2PC on the shares. The two QPs jointly compute the function on their inputs without sharing the value of their inputs!

Trusted Execution Environment (TEE)
- Only QPs that execute in TEEs and jointly implement $f$ are able to decrypt the shares.
- CoVault guarantees security as long as one TEE implementation remains secure.

Future Work

- Efficiency: Using 3PC instead of 2PC could significantly speed-up query execution.
- Applications: CoVault targets batch processing, we plan to extend it to support stream query processing.

Evaluation: Epidemic Analytics

How many encounters did a sick user have within the last 14 days?

How many unique devices did a sick user encounter within the last 14 days?

Conservative assumption: the table in input contains data of 1M users, each reporting 200 encs/day for 14 days = 28M encounter records

CoVault can run real queries even on very large datasets in the order of minutes, which is enough for the applications that we target.