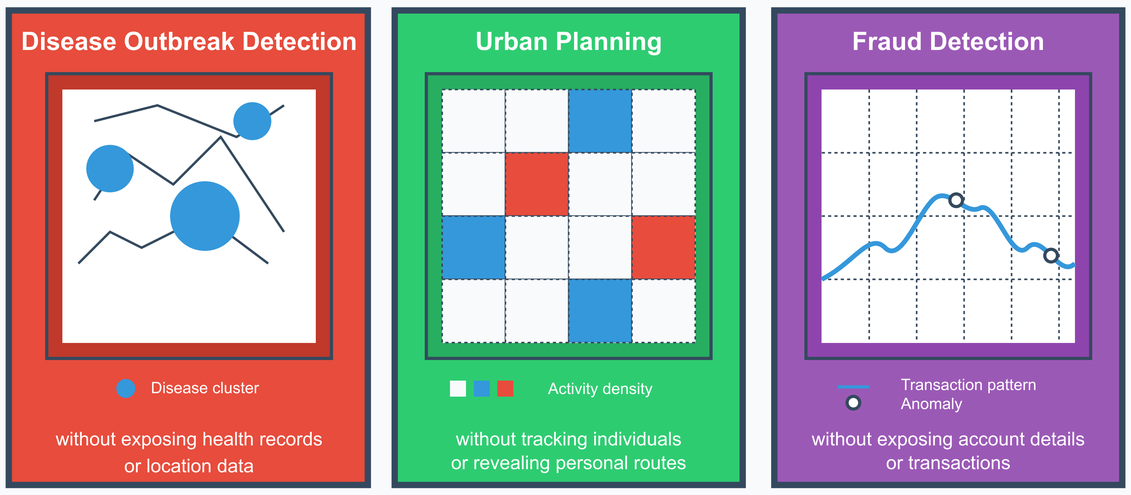
CoVault: Secure, Scalable Analytics of Personal Data

Roberta De Viti[1], Isaac Sheff[1], Noemi Glaeser[24], Baltasar Dinis[3], Rodrigo Rodrigues[3],
Bobby Bhattacharjee[4], Anwar Hithnawi[5], Deepak Garg[1], Peter Druschel[1]
[1]MPI-SWS | [2]MPI-SP | [3]IST (ULisboa), INESC-ID | [4]UMD | [5]ETH Zürich

1. Secure Analytics



Society can benefit from large-scale analysis of personal data
Opportunity: Improvement of human life and billions in savings through data-driven innovations
Barrier: Privacy, trust, and scale issues

Challenge: Given the high sensitivity of this data, a system must provide:

- Confidentiality
- Scalability
- Integrity of Results
- Selective Forward Consent

2. Current designs

TEE-Based Systems

TEE

The TEE HW vendor is a **single root of trust** (possible compromise)

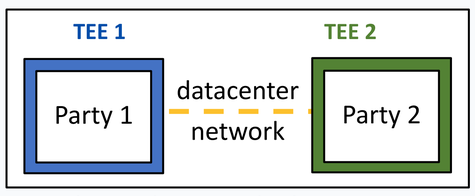
MPC-Based Systems

Geographically separated datacenters

Trust is distributed among parties, which are geo-separated for non-collusion assumptions to hold

3. CoVault design

Unlocks datacenter-scale MPC



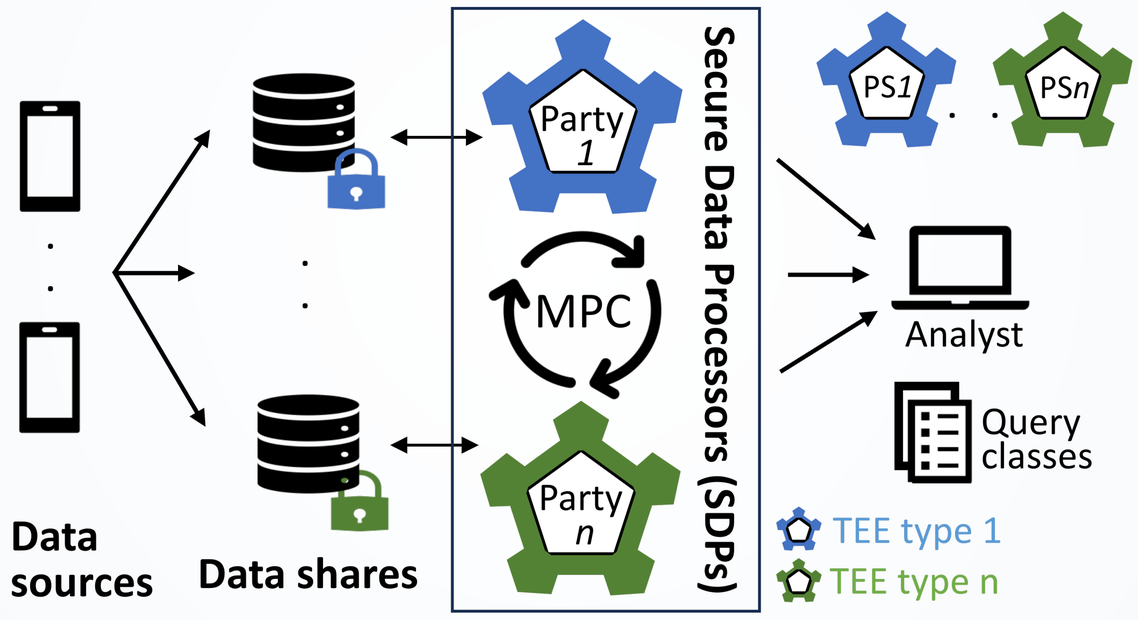
Colocated in a single datacenter

Key: Trust distributed among **diverse TEE HW vendors**

Diverse TEE vendors ► Non-collusion without geo-separation ► Colocation enabled ► High bandwidth b/w parties

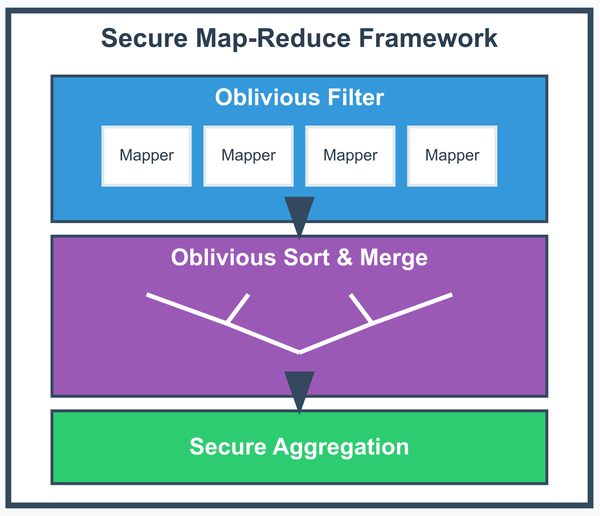
4. CoVault architecture

Server-aided MPC + Diverse TEEs enable colocation and scaling



5. Horizontal scaling

Beyond the bandwidth bottleneck: Core parallelism

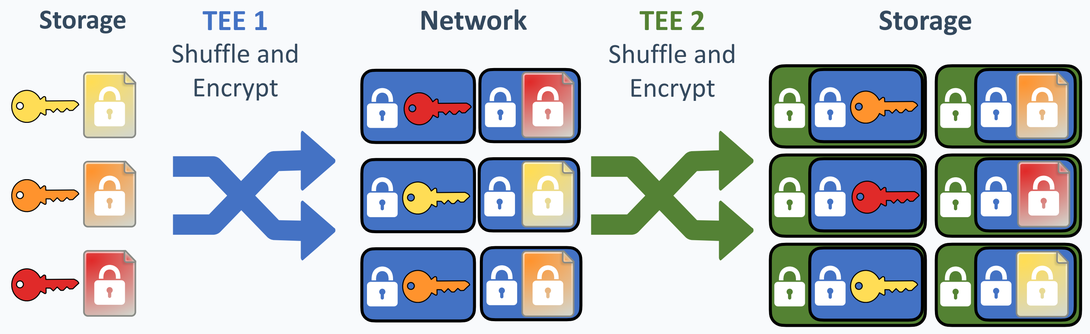


Map-Reduce Primitives in MPC:
Filter, Sort, Merge, Compact

Side-Channel Mitigation:
Pad variable-length communication

6. Oblivious Data Retrieval

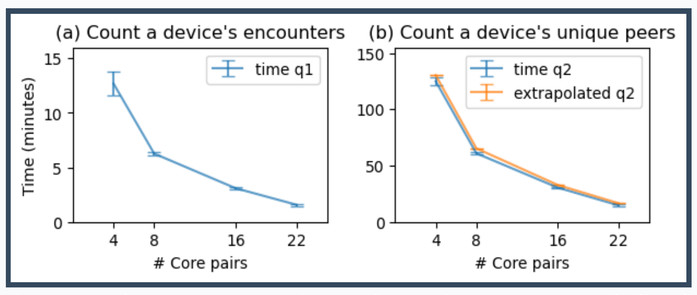
Hides patterns of random array accesses with private indexes



Faster than ORAM, with constant query time

7. Epidemic Analytics Scenario

Query Performance with Millions of Records



Minute-scale query execution on large datasets, enough for the applications that we target

Setup: Intel TDX and AMD SEV-SNP TEEs in Google Cloud

Extrapolation: from Millions to Billions of Records

1.5k	1k	30k
core pairs for continuous ingestion	core pairs for query execution	citizens per core pair

Assumption: a person has 200 encounters/day, so a country the size of Germany (80M people) generates a total of 11.85B records/day

Scan me!