TEEs: Trusted Execution Environments (TEEs) for Higher Security Data Processing (LBNL)

Sean Peisert
Staff Scientist
March 10, 2021
When property secured, anonymized, and optimized for research, administrative data can be put to work to help government programs better serve those in need.

BY JUSTINE S. HASTINGS, MARK HOWISON, TED LAWLESS, JOHN UCLES, AND PRESTON WHITE

Unlocking Data to Improve Public Policy
Covid-19 vaccines will arrive before the data sharing technology that could help track them

By CASEY ROSS / DECEMBER 2, 2020

Scientists have produced Covid-19 vaccines in record time. But the digital connectivity needed to closely track doses, side effects, and continuing infections is still lagging behind — even though the technology is now widely available.

This paradox of the pandemic was on display yesterday during a meeting hosted by the federal department of Health and Human Services. An official with the U.S. Digital Service said site visits to public health agencies around the country in recent months revealed a heavy reliance on paper documents and fax machines to collect and share data on Covid-19 tests.

WHEN APPS RULE THE ROAD

BY JANE MACFARLANE

THE PROLIFERATION OF NAVIGATION APPS IS CAUSING TRAFFIC CHAOS. IT’S TIME TO RESTORE ORDER

DURING THE 2017 WILDFIRES, THE APPS DIRECTED DRIVERS ONTO STREETS THAT WERE BEING CLOSED BY THE CITY, RIGHT INTO THE HEART OF THE FIRE.
Numerous Reasons Why Data Sharing Is Hindered

• Curation issues (e.g., preparation, description support, data quality, sensor calibration)

• Integration issues (e.g., database / data format incompatibilities)

• Regulated data (HIPAA, FISMA)

• Proprietary data (trade secrets, or $$ to produce, why share?)

• Unregulated data still containing individually private information
Numerous Reasons Why Data Sharing Is Hindered

- Curation issues (e.g., preparation, description support, data quality, sensor calibration)
- Integration issues (e.g., database / data format incompatibilities)
- Regulated data (HIPAA, FISMA)
- Proprietary data (trade secrets, or $$ to produce, why share?)
- Unregulated data still containing individually private information

Security and privacy techniques can help with some of these
Many of these data types exist

- Regulated data — biomedical data, export controlled science
- Proprietary data — power grid, materials, synthetic biology/chemistry, financial
- Unregulated / lightly regulated data still containing individually private information —
  - computer network data,
  - smart meter data,
  - smart city data,
  - vehicle / transportation location data
Some Perceived Risks with Data Sharing

- Enabling research competition
- Giving away data that cost $$ to produce
- Private data leakage / breaches
  - Accidental
  - Malicious insiders
  - External attacks
- Degrading security
- National
  - Grid
  - Automotive
  - Medical device
  - etc..
Security and Privacy Techniques Can Reduce Barriers to Sharing and/or Incentivize

Security techniques can reassure regulators and data owners by satisfying required security policies.

→Lowers risks for sharing regulated data

Privacy-preserving techniques can significantly reduce risk of exposure of raw data

→Lowers risks for private and proprietary data sharing

Security, fault tolerance, and data provenance techniques can create mechanisms to track data use.

→Incentivize data sharing by creating data marketplaces
EXISTING MODELS FOR SECURING SENSITIVE DATA
Data Exchange — Trust via Legal Agreements

Co-Located Data Sources and Compute

“MVP”

“All of Us”

NIH
Trusted Third Party

Data Sources

“MVP”

Google Cloud Platform

“All of Us”

Results
What are the problems with existing models?

- Legal agreements — what do these really protect against?
- Trusted third parties — trust for “intent” is not enough.
Trust by Attempting to Remove Data Sensitivity

Data Source

NAME: GRANT, ULYSSES S.
DATE OF DEATH: July 23, 1885

Researcher’s Compute System

“MVP”

NAME: PATIENT 24601
DATE OF DEATH: 1885

Anonymization/sanitization by:
- adding noise, (e.g., fake records)
- enforcing regularity (e.g., removing most specific aspects)
- masking (e.g., concealing / pseudonymizing)
Anonymization is increasingly easily defeated by the very techniques that are being developed for many legitimate applications of big data. In general, as the size and diversity of available data grows, the likelihood of being able to re-identify individuals.

A Precautionary Approach to Big Data Privacy

Once released to the public, data cannot be taken back. As time passes, data analytic techniques improve and additional datasets become public that can reveal information about the original data. It follows that released data will get increasingly vulnerable to re-identification—unless methods with provable privacy properties are used for the data release.

No silver bullet: De-identification still doesn’t work

Arvind Narayanan
arvindn@cs.princeton.edu

Edward W. Felten
felten@cs.princeton.edu

July 9, 2014
Trust via Physical Protections

Compute system is air-gapped and in a SCIF-like environment. Researcher goes behind a guard gate and into a locked, windowless building to access the data. Nothing goes in or out.

Data Source

“MVP”

Compute System

Results

Researcher

How many scientific researchers will work in SCIFs?
Current “Online” Model for Sensitive Data

How well does this model fit desired workflows by scientists? How usable is it? How often do researchers need to bring in new tools? Still have to trust the institution and its employees.
Secure Multiparty Computation

Data is always encrypted outside of the data source, even during computation.

Throughput for small-scale computing (AES)

Our Solution:
Hardware Trusted Execution Environments

Examples of TEEs: Intel SGX
ARM TrustZone
AMD Secure Encrypted Virtualization
RISC-V Keystone
Key Performance Findings

• AMD SEV can be used for secure scientific computing without significant performance degradation for most workloads.

Performance Analysis of Scientific Computing Workloads on Trusted Execution Environments

Ayaz Akram
UC Davis
yazakram@ucdavis.edu

Anna Giannakou
LBNL
agiannakou@lbl.gov

Venkatesh Akella
UC Davis
akella@ucdavis.edu

Jason Lowe-Power
UC Davis
jlowepower@ucdavis.edu

Sean Peisert
LBNL & UC Davis
sppeisert@lbl.gov

Proceedings of the 35th IEEE International Parallel & Distributed Processing Symposium (IPDPS), May 2021.
Differential Privacy

- Differential privacy seeks to maximize analysis accuracy of sensitive data while minimizing chances of enabling re-identification of individual entries.
- It is used by Apple and Google to collect user information (e.g., about uploaded photos) while protecting privacy.

An algorithm is $\epsilon$-differentially private if for datasets $D_1$ and $D_2$ that differ on a single element, the probability of determining if the individual record is in the dataset is less than $\epsilon$.

Accomplish DP by adding Laplace or Gaussian noise to all statistical database query responses within a bounded range.

Source: Cynthia Dwork, Microsoft Research, 2009.
Differential Privacy

Individual Health Data → Data Sources → Query → Computation

Add noise within bounded range

Validate Privacy Budget

Results
Ideal Workflow
Trusted Execution Exists Today

Chip Manufacturers:

- Intel® Software Guard Extensions (Intel® SGX)
- Intel® Trusted Execution Technology
- ARM TRUSTZONE
- AMD Secure Encrypted Virtualization (SEV)
- AMD EPYC Hardware Memory Encryption

Open Source Hardware:

- RISC-V®
- Keystone
- 9x5 HEX-Five Security
- The Linux Foundation Projects

Cloud Providers:

- Introducing Google Cloud Confidential Computing with Confidential VMs
- AWS Nitro System
Differential Privacy Exists Today

The U.S. Census Bureau Adopts Differential Privacy

John M. Abowd
United States Census Bureau
Washington, DC, USA
john.maron.abowd@census.gov

A community effort to protect genomic data sharing, collaboration and outsourcing


ngj Genomic Medicine 2, Article number: 33 (2017) Download Citation

Differential Privacy at Scale: Uber and Berkeley Collaboration

Tuesday, January 16, 2018 - 11:00 am-11:30 am

Uber Releases Open Source Project for Differential Privacy

Privacy-Enhanced and Multifunctional Health Data Aggregation under Differential Privacy Guarantees

Hao Ren, Hongwei Li, Xiaohui Liang, Shibo He, Yuanshu Dai, and Lian Zhao
Contact:
Dr. Sean Peisert
sppeisert@lbl.gov
https://www.cs.ucdavis.edu/~peisert/
https://crd.lbl.gov/sean-peisert/
https://dst.lbl.gov/security/