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

Sean Peisert

Staff Scientist

March 10, 2021

BERKELEY LAB

INTELLECTUAL PROPERTY OFFICE

DOI:10.1145/3335150

When properly 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

BERKELEY LAB

IP INTELLECTUAL PROPERTY OFFICE

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.

Casey Ross. Covid-19 vaccines will arrive before the data sharing technology that could help track them. *Stat+*, Dec. 2, 2020. https://www.statnews.com/2020/12/02/covid19-vaccines-interoperability-data-hospitals/

BERKELEY LAB

IP INTELLECTUAL PROPERTY OFFICE

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.

BERKELEY LAB

IPO INTELLECTUAL PROPERTY OFFICE

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

BERKELEY LAB

IP INTELLECTUAL PROPERTY OFFICE

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

BERKELEY LAB

INTELLECTUAL PROPERTY OFFICE

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..

BERKELEY LAB

INTELLECTUAL PROPERTY OFFICE

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

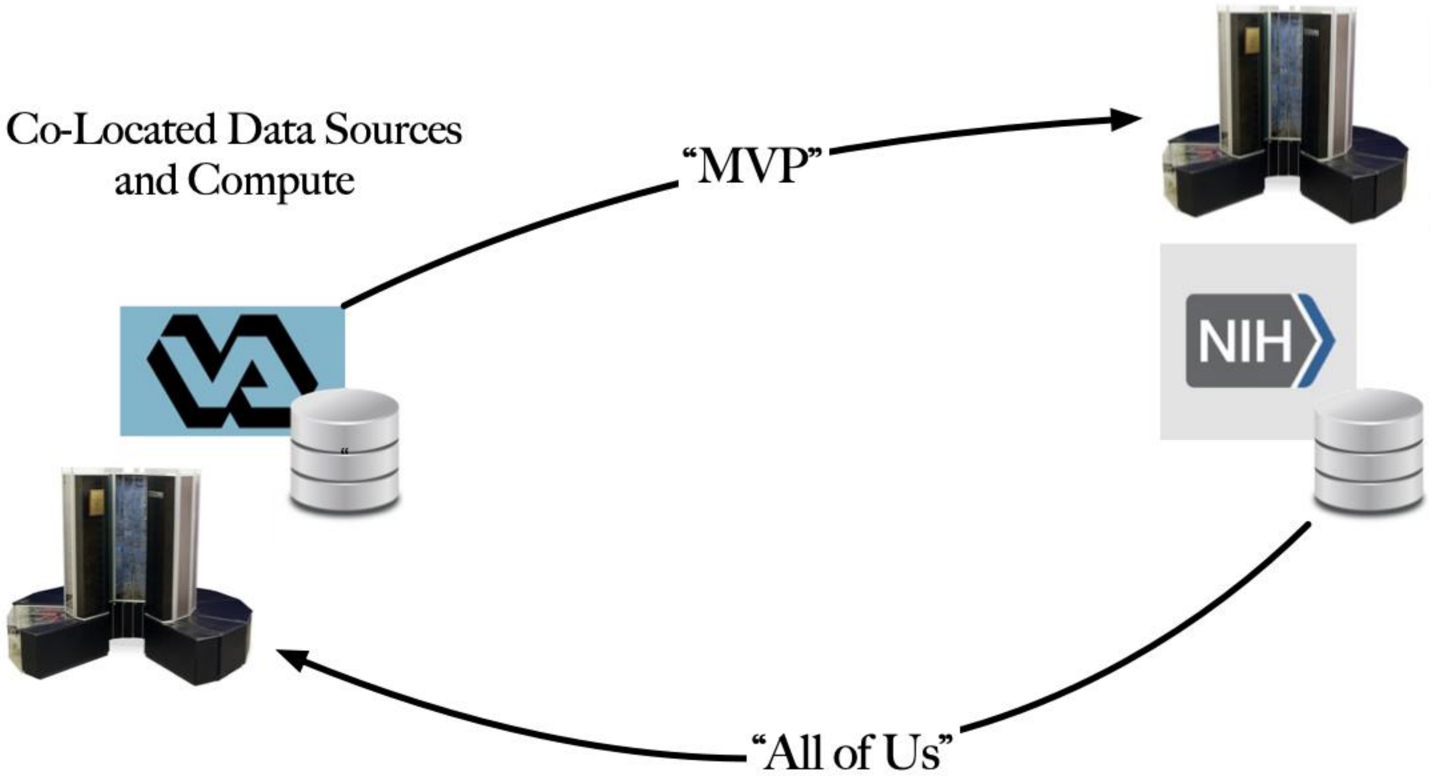
→Lowers risks for private and proprietary data sharing

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

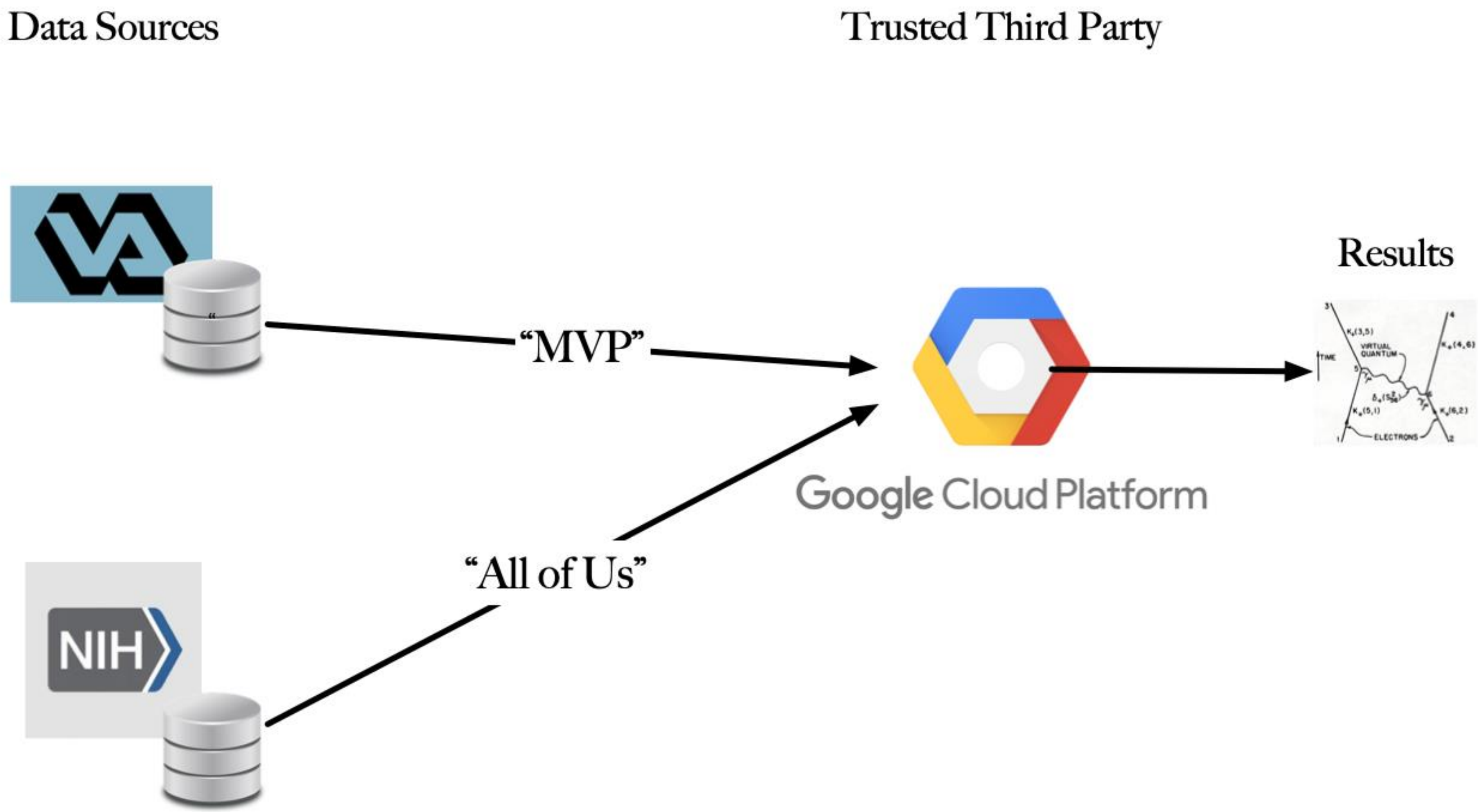
→Incentivize data sharing by creating data marketplaces

EXISTING MODELS FOR SECURING SENSITIVE DATA

Data Exchange — Trust via Legal Agreements



Trusted Third Party



BERKELEY LAB

IP
INTELLECTUAL PROPERTY OFFICE

What are the problems with existing models?

- Legal agreements — what do these really protect against?
- Trusted third parties — trust for "intent" is not enough.

RYAN SINGEL SECURITY 10.01.09 08:05 AM

PROBE TARGETS ARCHIVES' HANDLING OF DATA ON 70 MILLION VETS

The New York Times

Facebook Says Cambridge Analytica Harvested Data of Up to 87 Million Users

The New York Times

Facebook Security Breach Exposes Accounts of 50 Million Users

The New York Times

Millions of Anthem Customers Targeted in Cyberattack

tions ☰ The Washington Post
Democracy Dies in Darkness

The Switch

145 million Social Security numbers, 99 million addresses and more: Every type of personal data Equifax lost to hackers, by the numbers

Sections ☰ The Washington Post
Democracy Dies in Darkness

Federal Insider

Hacks of OPM databases compromised 22.1 million people, federal authorities say

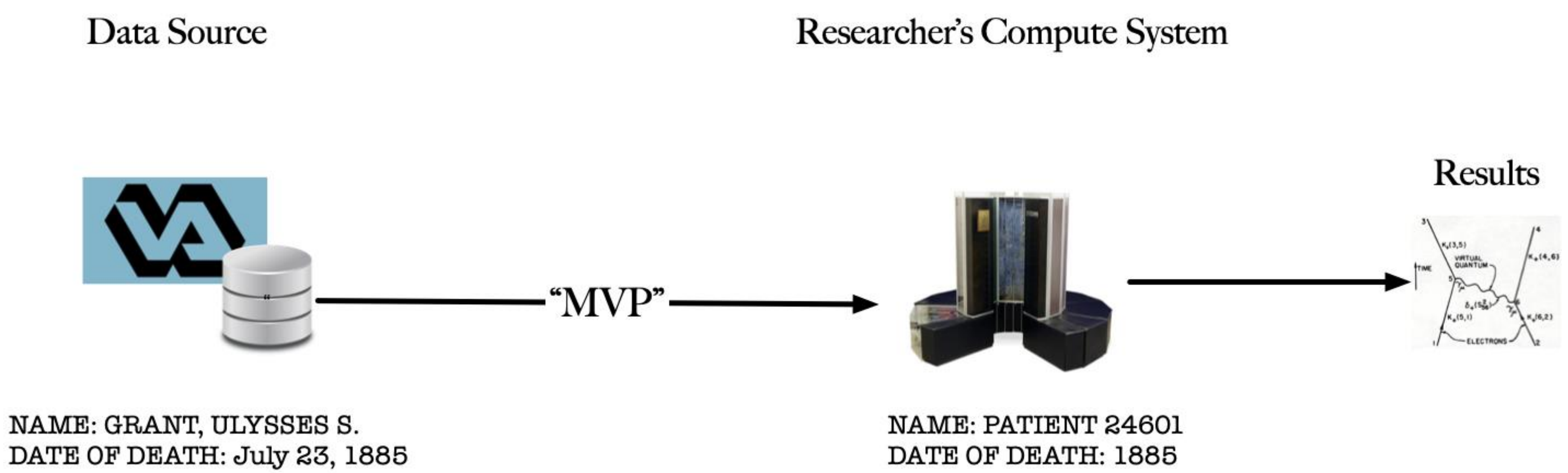
By Ellen Nakashima
July 9, 2015

Most Read Politics

U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB
Lawrence Berkeley National Laboratory

Trust by Attempting to Remove Data Sensitivity



Anonymization/sanitization by: adding noise, (e.g., fake records)
enforcing regularity (e.g., removing most specific aspects)
masking (e.g., concealing / pseudonymizing)

BERKELEY LAB

IP INTELLECTUAL PROPERTY OFFICE

What about “anonymization”?

The New York Times

A Face Is Exposed for AOL Searcher No. 4417749

By MICHAEL BARBARO and TOM ZELLER Jr. AUG. 9, 2006

REPORT TO THE PRESIDENT

BIG DATA AND PRIVACY: A TECHNOLOGICAL PERSPECTIVE

Executive Office of the President

President’s Council of Advisors on Science and Technology

May 2014

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.

ars TECHNICA BIZ & IT TECH SCIENCE POLICY CARS GAMING & CULTURE FOR

POLICY —

“Anonymized” data really isn’t—and here’s why not

Companies continue to store and sometimes release vast databases of " ...

NATE ANDERSON - 9/8/2009, 4:25 AM

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

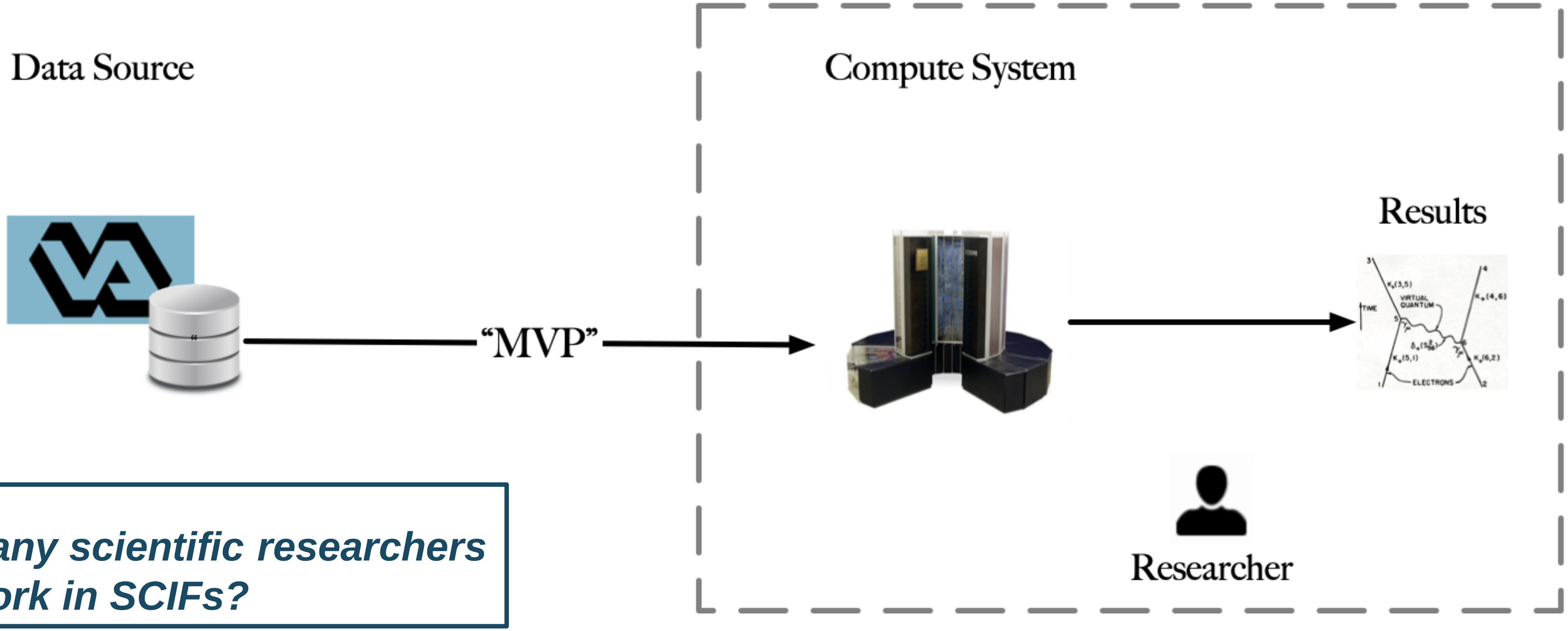
Arvind Narayanan
arvindn@cs.princeton.edu

Edward W. Felten
felten@cs.princeton.edu

July 9, 2014

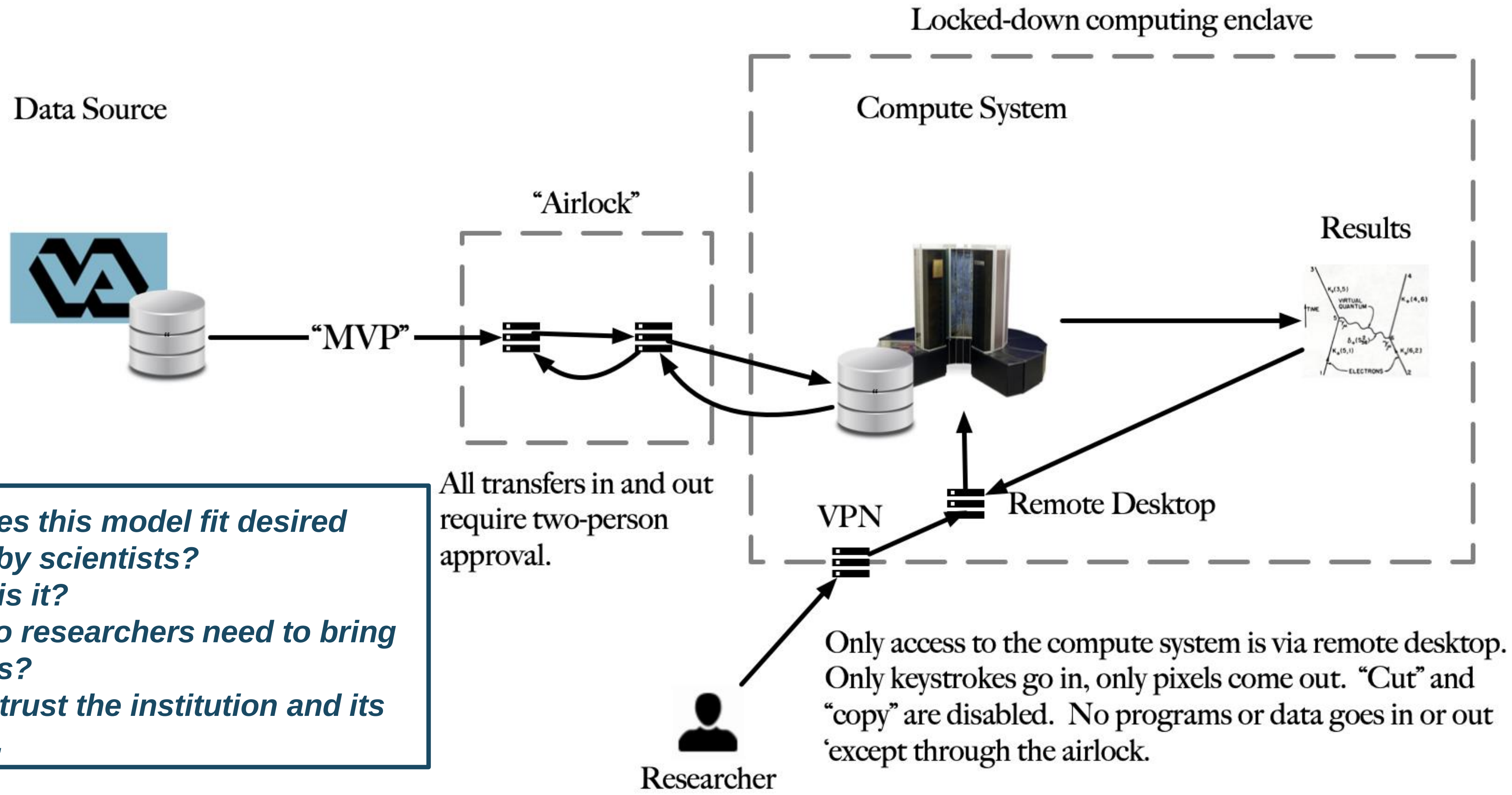
Trust via Physical Protections

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.



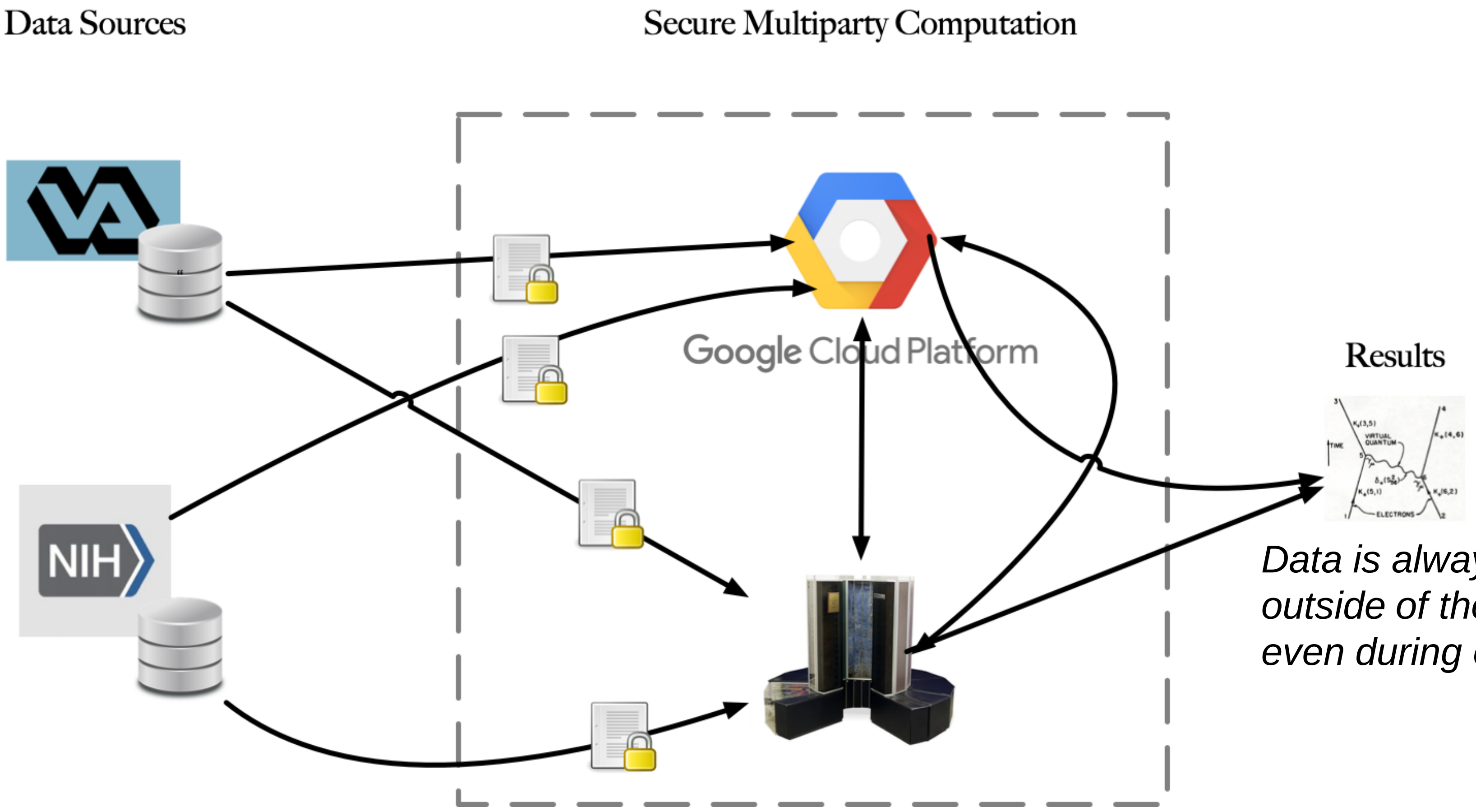
How many scientific researchers will work in SCIFs?

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.

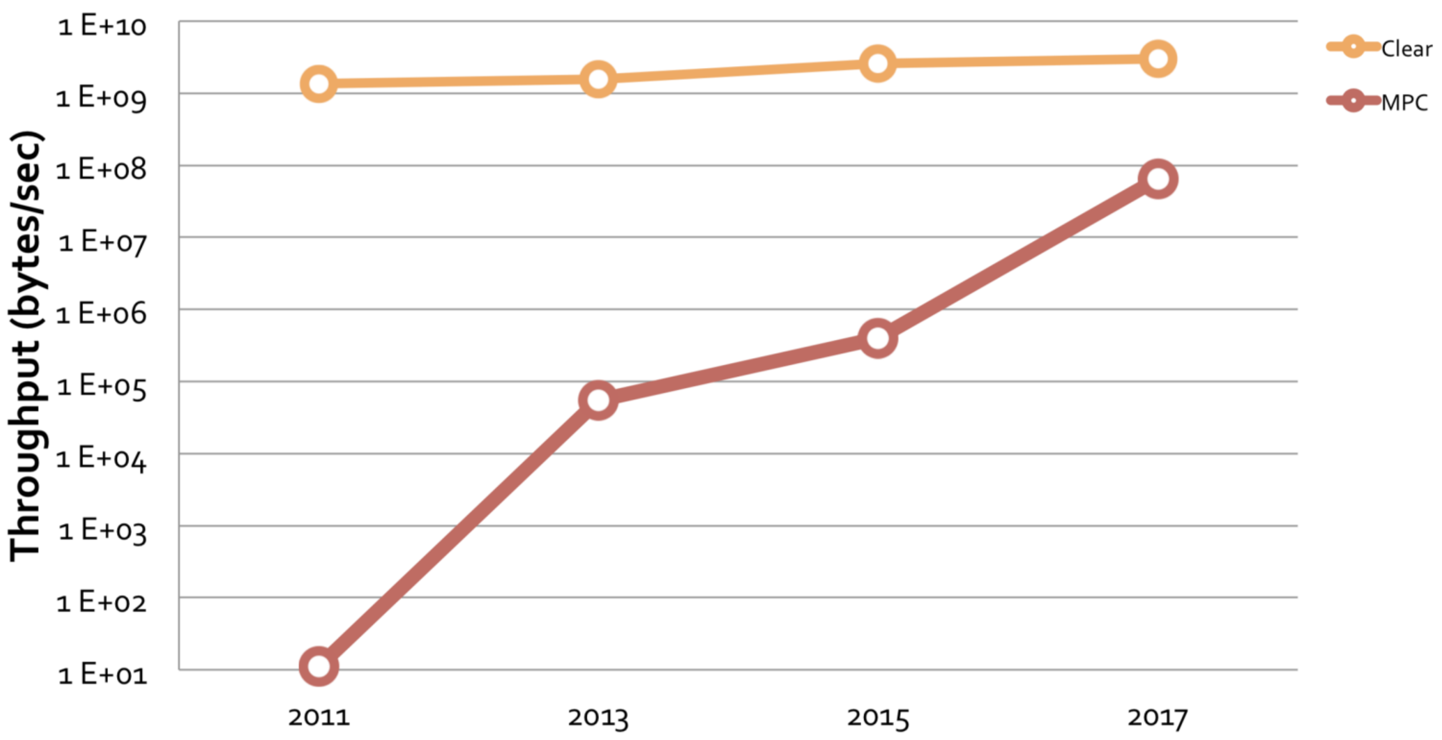
Secure Multiparty Computation



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

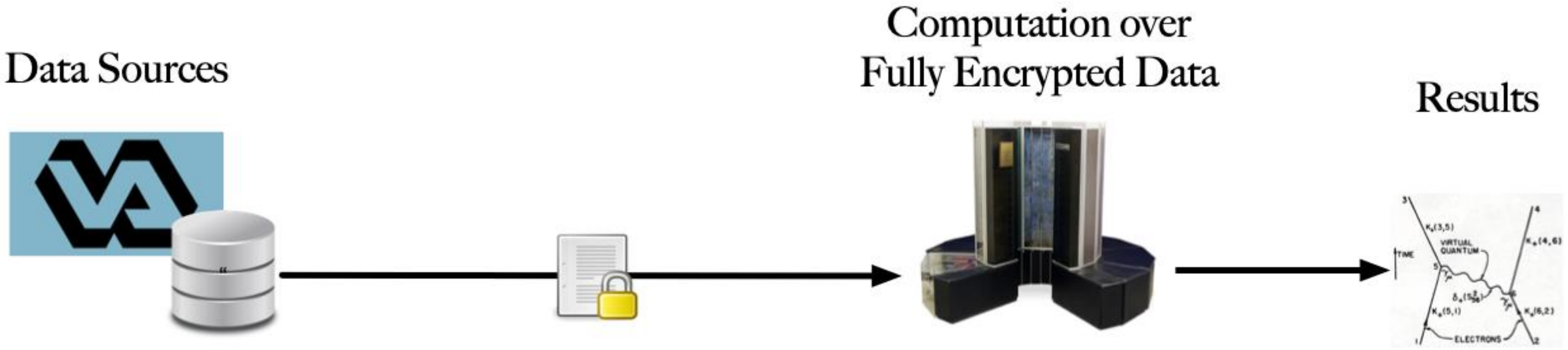
Source: diagram inspired by Mayank Varia and Andrei Lapets, "Trustworthy Computing for Scientific Workflows," Trusted CI Webinar, July 23 2018.

Throughput for small-scale computing (AES)



Source: Mayank Varia and Andrei Lapets, “Trustworthy Computing for Scientific Workflows,” Trusted CI Webinar, July 23 2018.

Our Solution:
Hardware Trusted Execution Environments



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

BERKELEY LAB

INTELLECTUAL PROPERTY OFFICE

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	Anna Giannakou	Venkatesh Akella	Jason Lowe-Power	Sean Peisert
UC Davis	*LBNL*	*UC Davis*	*UC Davis*	*LBNL & UC Davis*
yazakram@ucdavis.edu	agiannakou@lbl.gov	akella@ucdavis.edu	jlowepower@ucdavis.edu	sppeisert@lbl.gov

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

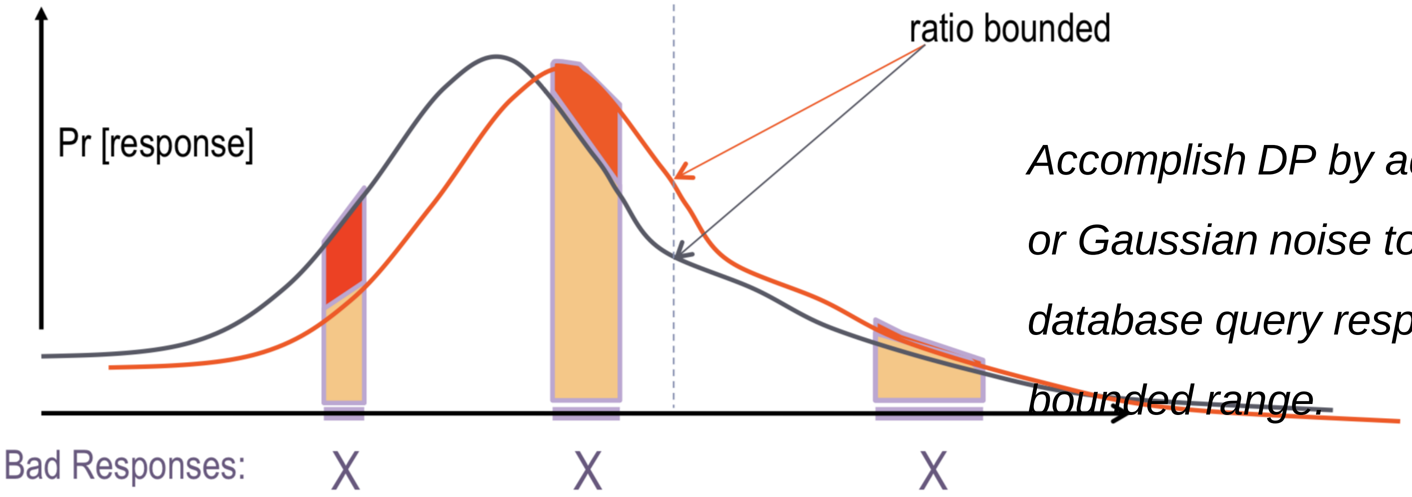
BERKELEY LAB

IP INTELLECTUAL PROPERTY OFFICE

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.

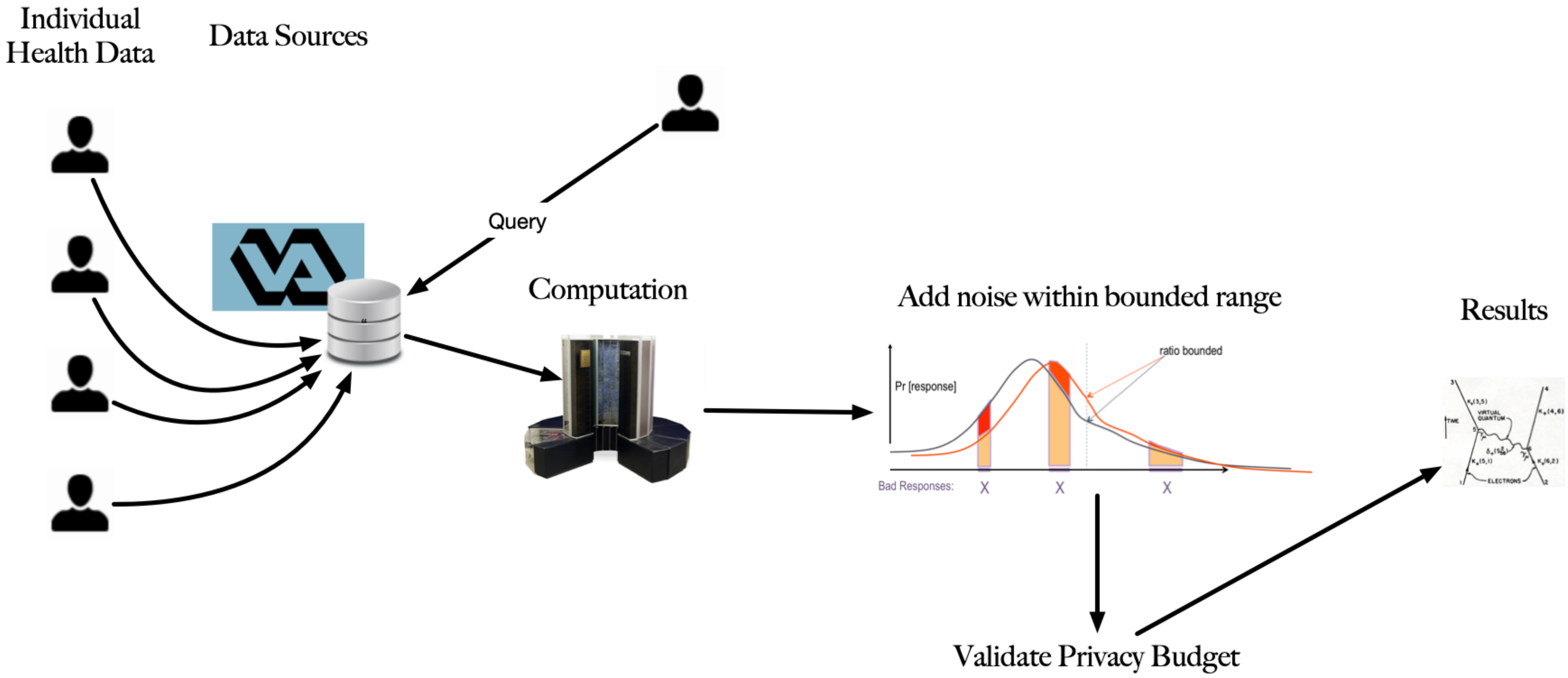
An algorithm *is ϵ-differentially private if for datasets D_1 and D_2 that differ on a single element, the probability of determining if the individual record is in the dataset is less than ϵ.*



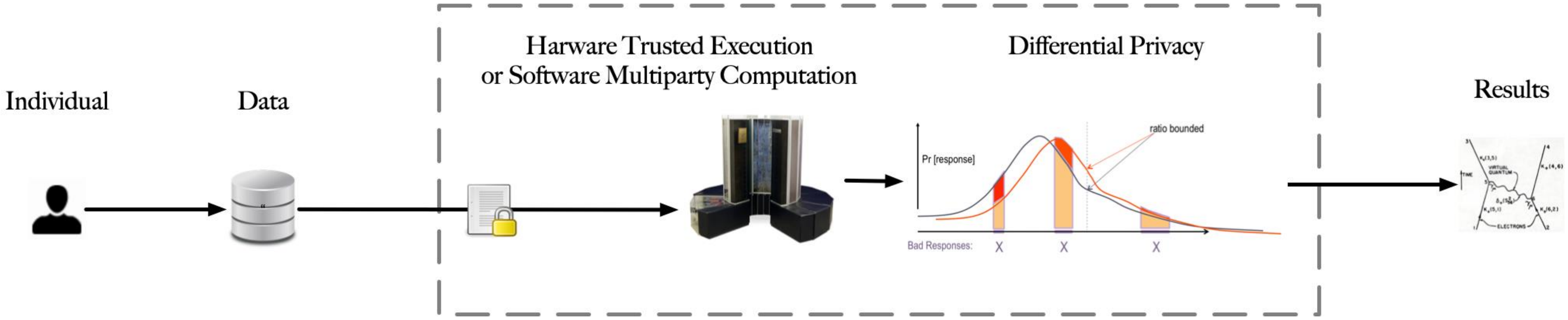
Accomplish DP by adding Laplace or Gaussian noise to all statistical database query responses within a bounded range.

Source: Cynthia Dwork, Microsoft Research, 2009.

Differential Privacy



Ideal Workflow



Trusted Execution Exists Today

Chip Manufacturers:

Intel® Software Guard Extensions (Intel® SGX)

Intel® Trusted Execution Technology

Hardware-based Technology for Enhancing Server Platform Security

arm TRUSTZONE

AMD

Secure Technology

AMD Secure Encrypted Virtualization (SEV)

AMD EPYC Hardware Memory Encryption

Open Source Hardware:

RISC-V®

Keystone

Open-source Secure Hardware Enclave

0x5 HEX–Five Security

THE LINUX FOUNDATION PROJECTS

CONFIDENTIAL COMPUTING CONSORTIUM

Cloud Providers:

Introducing Google Cloud Confidential Computing with Confidential VMs

AWS Nitro System

Differential Privacy Exists Today

Machine Learning Journal

Learning with Privacy at Scale

Vol. 1, Issue 8 • December 2017
by Differential Privacy Team

The U.S. Census Bureau Adopts Differential Privacy

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

Google AI Blog

The latest news from Google AI

Federated Learning: Collaborative Machine Learning without Centralized Training Data

Thursday, April 6, 2017

Posted by Brendan McMahan and Daniel Ramage, Research Scientists

HARVARD UNIVERSITY

OpenDP

building an open-source suite of tools for deploying differential privacy

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

Shuang Wang, Xiaoqian Jiang, Haixu Tang, Xiaofeng Wang, Diyue Bu, Knox Carey, Stephanie OM Dyke, Dov Fox, Chao Jiang, Kristin Lauter, Bradley Malin, Heidi Sofia, Amalio Telenti, Lei Wang, Wenhao Wang & Lucila Ohno-Machado

npj 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

ANDY GREENBERG SECURITY 07.13.17 10:02 AM

UBER'S NEW TOOL LETS ITS STAFF KNOW *LESS* ABOUT YOU

UBER Uber Security Follow
Jul 13, 2017 · 4 min read

Uber Releases Open Source Project for Differential Privacy

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

Hao Ren,[1] Hongwei Li,[1,2,*] Xiaohui Liang,[3] Shibo He,[4] Yuanshun Dai,[1] and Lian Zhao[5]

BERKELEY LAB

IP INTELLECTUAL PROPERTY OFFICE

Contact:
Dr. Sean Peisert
sppeisert@lbl.gov
https://www.cs.ucdavis.edu/~peisert/
https://crd.lbl.gov/sean-peisert/
https://dst.lbl.gov/security/

U.S. DEPARTMENT OF ENERGY

UNIVERSITY OF CALIFORNIA

BERKELEY LAB
Lawrence Berkeley National Laboratory