Privacy-Preserving Systems for a Data-Driven World

Anwar Hithnawi
Data Driven World
Sensitive Data

Smart Homes  Genetics  Dating  Geolocation

Finance  Health  Government  Personal
DATA IS THE NEW OIL OF THE DIGITAL ECONOMY

Why Big Data Is The New Natural Resource  Forbes

How Artificial Intelligence Could Transform Medicine
Data Protection: An Age-Old Concern

- **Caesar Cipher**
  - Roman Empire
  - 50 BCE

- **Scytale**
  - Ancient Greece
  - 5 BCE

- **Cryptanalysis**
  - Arab World
  - 1400s

- **Enigma Machine**
  - 1920s

- **Information Theory**
  - 1940s

- **Modern Cryptography**
  - 1970s
Data Protection: An Age-Old Concern

- **Caesar Cipher**
  - Roman Empire
  - 50 BCE

- **Scytale**
  - Ancient Greece
  - 5 BCE

- **Cryptanalysis**
  - Arab World
  - 1400s

- **Enigma Machine**
  - 1920s

- **Information Theory**
  - 1940s

- **Modern Cryptography**
  - 1970s

- **Internet**
  - HTTP, TCP/IP
  - 1990
Data Protection: An Age-Old Concern

- **50 BCE**: Caesar Cipher
  - Roman Empire
- **5 BCE**: Scytale
  - Ancient Greece
- **1400s**: Cryptanalysis
  - Arab World
- **1920s**: Enigma Machine
- **1940s**: Information Theory
- **1970s**: Modern Cryptography
- **1990**: Internet
  - HTTP, TCP/IP

Data Protection - Individuals
Data Protection: An Age-Old Concern

- **50 BCE** Caesar Cipher, Roman Empire
- **5 BCE** Scytale, Ancient Greece
- **1400s** Cryptanalysis, Arab World
- **1920s** Enigma Machine
- **1940s** Information Theory
- **1970s** Modern Cryptography
- **1990** Internet, HTTP, TCP/IP
- **1995** HTTPS

Data Protection Individuals
Data Protection: An Age-Old Concern

- **50 BCE**: Caesar Cipher, Roman Empire
- **5 BCE**: Scytale, Ancient Greece
- **1400s**: Cryptanalysis, Arab World
- **1920s**: Enigma Machine
- **1940s**: Information Theory
- **1970s**: Modern Cryptography
- **1990**: Internet (HTTP, TCP/IP)
- **1995**: HTTPS
- **2014**: Source: Blog Scott Helme
- **2023**: Encrypted Page Loads [%]

---

Data Protection: Individuals
Securing Data: Building Blocks
Securing Data: Building Blocks

Secure Communication

Untrusted Channel

\[ \text{Encrypt} \rightarrow \text{Enc}(x) \rightarrow \text{Decrypt} \rightarrow X \]
Securing Data: Building Blocks

- Encrypt
- Decrypt
- $Enc(x)$
- Untrusted Channel
- Secure Storage
- Secure Communication

1. $X \rightarrow$ Encrypt
2. $Enc(x) \rightarrow$ Decrypt
3. $X \leftarrow$ Decrypt
4. $X \rightarrow$ Secure Communication
5. $X \leftarrow$ Secure Storage
Securing Data: Building Blocks

Encrypt $x$ to $Enc(x)$, transmitting it over an untrusted channel, and then decrypt it. Secure Storage and Secure Communication ensure data security. End-to-End Encrypted Applications guarantee confidentiality.
Securing Data: Building Blocks

End-to-End Encrypted Applications

More Applications?
Securing Data **in Use**: Modern Applications

\[ X \xrightarrow{\text{Encrypt}} Enc(x) \xrightarrow{\text{Eval}} Enc(f(x)) \xrightarrow{\text{Decrypt}} f(x) \]
Securing Data in Use: Modern Applications

\[ f(x) \xrightarrow{\text{Encrypt}} Enc(x) \xrightarrow{\text{Eval } f(\cdot)} Enc(f(x)) \xrightarrow{\text{Decrypt}} f(x) \]

X → Encrypt → Enc(x) → Untrusted Cloud → Eval \( f(\cdot) \) → Enc(f(x)) → Decrypt → Release → f(x)
Securing Data **in Use**: Modern Applications

\[ f(x) \xleftarrow{} \text{Decrypt} \quad \text{Encrypt} \quad \text{Eval } f(.) \quad \text{Decrypt} \quad \text{Release} \quad f(x) \]

\[ x \xrightarrow{} \text{Encrypt} \quad \text{Enc}(x) \quad \text{Eval } f(.) \quad \text{Enc}(f(x)) \quad \text{Release} \quad f(x) \]

**Secure Computation**
Securing Data **in Use**: Modern Applications

The process involves:
- **Encrypting** the input data $x$ to $Enc(x)$.
- **Evaluating** the function $f(x)$ in the encrypted domain.
- **Decrypting** the result $Enc(f(x))$ to $f(x)$.

This methodology enables **Secure Computation** and **Privacy-preserving Disclosure**.
End-to-End Security

data in use
secure computation

data in transit
secure communication

data at rest
secure storage
End-to-End Security

Ubiquitous Adoption

Conventional Crypto
Encryption & Digital Signature

- data at rest: secure storage
- data in transit: secure communication
- data in use: secure computation
End-to-End Security

Ubiquitous Adoption
Conventional Crypto
Encryption & Digital Signature

Just Starting
Privacy - Enhancing Technologies (PETs)
- Homomorphic Encryption
- Secure Multi-party Computation
- Zero Knowledge Proofs
- Differential Privacy

data in use
secure computation

data in transit
secure communication

data at rest
secure storage
~ 40 Years of History

- 1978: Homomorphic Encryption
- 1982: Secure Multi-party Computation
- 1989: Zero Knowledge Proofs
- 2006: Differential Privacy
~ 40 Years of History

- Homomorphic Encryption (1978)
- Secure Multi-party Computation (1982)
- Zero Knowledge Proofs (1989)
- Differential Privacy (2006)
- Big Data

- PINQ’09
- SPDM’12
- BGV’11
- BFV’12
- GSW’13
- ABY’15
- CKKS’16
- Groth’16
- SGD’16
- TFHE-rs
- Plonk’22

2023
~ 40 Years of History

- Homomorphic Encryption (1978)
- Secure Multi-party Computation (1982)
- Zero Knowledge Proofs (1989)
- Differential Privacy (2006)

Big Data: practically oriented theoretical work

- PINQ’09
- SPDZ’12
- BGV’11
- BFV’12
- GSW’13
- ABY’15
- CKKS’16
- Groth’16
- SGD’16
- TFHE-rs
- Plonk’22
~ 40 Years of History

1978: Homomorphic Encryption
1982: Secure Multi-party Computation
1989: Zero Knowledge Proofs
2006: Differential Privacy

Big Data
practically oriented theoretical work

1982: PINQ’09
1989: BFV’12
1999: BGV’11
2006: GSW’13
2010: ABY’15
2013: CKKS’16
2016: Groth’16
2016: SGD’16
2022: Plonk’22

real-word deployments

1982: Apple
1982: Microsoft
2006: Google
2015: ZAM
2015: United States Census
Theory to Practice: Barriers to Broad Adoption

Performance Gap
Practical for numerous applications but remains beyond reach for constrained use cases.

Complexity
There's a gap between the capabilities of PETs today and organizations' ability to incorporate them into applications.
Performance Gap

Practical for numerous applications but remains beyond reach for constrained use cases.

Complexity

There's a gap between the capabilities of PETs today and organizations' ability to incorporate them into applications.
Performance Gap
Fully Homomorphic Encryption
Performance Gap
Fully Homomorphic Encryption

Graph Adapted from Kristin Lauter, Talk @PriCon2020
Performance Gap
Fully Homomorphic Encryption

Performance Overhead

Graph Adapted from Kristin Lauter, Talk @PriCon2020
Performance Gap
Fully Homomorphic Encryption

Graph Adapted from Kristin Lauter, Talk @PriCon2020
Empower Constrained Environments with Encrypted Data Processing.

Approach to Efficiency

co-design

Cryptography

Application
Encrypted Data Stream Processing at Scale

[Constrained Data Sources, Large Scale, Low-Latency]

[TimeCrypt - USENIX NSDI'20]

c&o-design
Encrypted Data Stream Processing at Scale

[Constrained Data Sources, Large Scale, Low-Latency]

Symmetric HE

Time-encoded Keystream

co-design

System Performance

Throughput [ops/s]

Plain text

TimeCrypt

TimeCrypt+

ASHE

Paillier

EC-Elliptic

50k
40k
30k
20k
10k
0k
Encrypted Data Stream Processing at Scale

[Constrained Data Sources, Large Scale, Low-Latency]

System Performance

2% slowdown compared to plaintext

20x

Symmetric HE

Symmetric HE

 Plaintext

TimeCrypt

TimeCrypt+

ASHE

Paillier

EC-ElGamal

Throughput [ops/s]

co-design
Privacy-preserving, functional, and performant systems

My work aims to **build** practical systems that use cryptography to empower users and preserve their privacy.
Theory to Practice: Barriers to Broad Adoption

Advanced Cryptography → Application Demands → Complexity

Performance Gap
Theory to Practice: Barriers to Broad Adoption

Advanced Cryptography

Application Demands

Performance Gap

Complexity
My work aims to democratize access to privacy-preserving computation with new tools, systems, and abstractions.
Democratize Privacy-Preserving Computation

My work aims to democratize access to privacy-preserving computation with new tools, systems, and abstractions.

Secure Computation

FHE Compilers
IEEE S&P

HECO
USENIX Security

Differential Privacy

Cohere
IEEE S&P

Programmability

Deployments
Developing and Deploying Privacy-preserving Applications is Notoriously Hard
What does “developing these applications” entail?
Conventional Cryptography
Conventional Cryptography

App Logic

Crypto

Secure Communication

Client

send(data)

Key Agreement | Encryption | Integrity

Server

receive(data)
Conventional Cryptography

Client

Server

Key Agreement | Encryption | Integrity

send(data)

receive(data)

store(data)

load(data)

Secure Communication

Secure Storage

App

Logic

Crypto

Encryption

Integrity
Advanced Cryptography: Secure Computation

Crypto

Data Oblivious

Arithmetization

Noise

\( f \)

\( a \times b + 3 \)

\( x \times y = xy \)
Advanced Cryptography: Secure Computation

Functionality and performance depend on $f$’s representation:

• How do we express $f$
• How do we optimize $f$
Usable Fully Homomorphic Encryption

(IEEE S&P’21, USENIX Security’23)
Usable FHE

2. How can compilers address these complexities? [USENIX Security’23]
void hd(vector<bool> u, vector<bool> v)
{
    int sum = 0;
    for (int i = 0; i < v.size(); ++i)
    {
        sum += (v[i] != u[i]);
    }
}

Fully Homomorphic Encryption Programming Paradigm
```cpp
void hd(vector<bool> u, vector<bool> v)
{
    int sum;
    for(int i = 0; i < v.size(); ++i)
    {
        sum += (v[i] != u[i]);
    }
}
```

**Data Oblivious**
```c
void hd(vector<bool> u, vector<bool> v) {
    int sum;
    for (int i = 0; i < v.size(); ++i) {
        sum += (v[i] != u[i]);
    }
}
```

Data Oblivious

Arithmetization
```c
void hd(vector<bool> u, vector<bool> v)
{
  int sum;
  for(int i = 0; i < v.size(); ++i)
  {
    sum += (v[i]!=u[i]);
  }
}
```

---

**Data Oblivious**

**Arithmetization**

$x < 0 \rightarrow$ Binary Emulation
```cpp
void hd(vector<bool> u, vector<bool> v)
{
    int sum = 0;
    for (int i = 0; i < v.size(); ++i)
    {
        sum += (v[i] != u[i]);
    }
}
```
```c
void hd(vector<bool>& u, vector<bool>& v)
{
    int sum;
    for (int i = 0; i < v.size(); ++i)
    {
        sum += (v[i]!=u[i]);
    }
}
```
void hd(vector<bool> u, vector<bool> v) {
    int sum = 0;
    for (int i = 0; i < v.size(); ++i) {
        sum += (v[i] != u[i]);
    }
}

Data Oblivious
Arithmetization
Noise Management
void f(...) {
    ctxt ab = a*b + 3;
    ctxt r = ab - z*z;
    return r;
}
FHE Noise Management

```
void f(...) {
    ctxt ab = a*b + 3;
    ctxt r = ab - z*z;
    return r;
}
```

90% of runtime
void f(...) {
    ctxt ab = a*b + 3;
    ctxt r = ab - z*z;
    return r;
}
Developing FHE Applications

\[
\begin{align*}
\texttt{void } & \ f(\texttt{vec } v, \texttt{vec } u) \{
\texttt{ctxt } & \ sum = 0; \\
\text{for(int } i & \text{ = 0; } i < v.\text{size(); } ++i) \{
\texttt{sum } &= \texttt{(v[i]!=u[i]);} \\
\} \\
\text{return } & \texttt{sum;} 
\end{align*}
\]

Circuit Optimizations

- Relinearization
- Bootstrapping
- Mod Switching

Arithmetic Circuit

Developer

\(a \times b + 3\)
Developing FHE Applications

```c
void f(vec v, vec u) {
  ctxt sum = 0;
  for(int i = 0; i < v.size(); ++i) {
    sum += (v[i] != u[i]);
  }
  return sum;
}
```

Circuit Optimizations
- Relinearization
- Bootstrapping
- Mod Switching

Crypto Optimizations
- \( \mathbb{Z}[X]/(X^n + 1) \)
- FFT/NTT

Target Optimizations
- Individual Tooling
- FHE
HECO

void f(vec v, vec u)
{
    ctxt sum = 0;
    for(int i = 0; i < v.size(); ++i)
        sum += (v[i] != u[i]);
    return sum;
}

void hd(vector<bool> u, vector<bool> v)
{
    int sum = 0;
    for(int i = 0; i < v.size(); ++i)
        sum += (v[i] != u[i]);
}

Crypto Optimizations
Circuit Optimizations
Target Optimizations

FHE
ℤ[𝗆]/(𝗆ⁿ + 1)
FFT/NTT
FHE

Arith. Circuit
Batched (SIMD)

Look-Up Tables
Scalar

Mod Switching
Bootstrapping
Relinearization
HECO: Transform High-level Programs to Efficient FHE Solutions

Program Optimizations

<table>
<thead>
<tr>
<th>Arithmetization</th>
<th>Layout</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 0</td>
<td></td>
</tr>
<tr>
<td>1 0 1</td>
<td></td>
</tr>
<tr>
<td>Binary</td>
<td></td>
</tr>
<tr>
<td>Emulation</td>
<td></td>
</tr>
</tbody>
</table>

- Circuit Optimizations
- Crypto Optimizations
- Target Optimizations

- FFT/NTT
- Mod Switching
- Relinearization
- Bootstrapping

- a \times b
- 3
- Arith. Circuit
- Batched (SIMD)
- Scalar
- Look-Up Tables

Order-of-magnitude speedups via high-level transformations

Naive, HECO, Expert

Naive (non-Batched) and "Expert" synthesis-based solution

Developer
HECO: End-to-End FHE Compilation

```
void f(vec v, vec u)
{
    ctxt sum = 0;
    for(int i = 0; i < v.size(); ++i)
    {
        sum += (v[i] != u[i]);
    }
    return sum;
}
```

```
void hd(vector<s(bool)> u, vector<s(bool)> v)
{
    int sum = 0;
    for(int i = 0; i < v.size(); ++i)
    {
        sum += (v[i] != u[i]);
    }
}
```

**Program Optimizations**
- Arithmetization
- Look-Up Tables
- Arith. Circuit
- Batched (SIMD)

**Circuit Optimizations**
- Scalar
- Bootstrapping
- Mod Switching
- FFT/NTT

**Crypto Optimizations**
- \( \mathbb{Z}[X]/(X^n + 1) \)

**Target Optimizations**
- FHE

**Crypto Optimizations**
- Circuit Optimizations
- Target Optimizations
Evaluation: Effect of Batching Optimizations

![Bar chart showing the effect of batching optimizations on different operations like GxKernel, Box Blur, and RobertsCross.

HECO: Compiler for FHE

open source, automated end-to-end optimization for FHE
Democratize Privacy-Preserving Computation

My work aims to democratize access to privacy-preserving computation with new tools, systems, and abstractions.

Secure Computation
- FHE Compilers
  - IEEE S&P
- HECO
  - USENIX Security

Differential Privacy
- Cohere
  - IEEE S&P

Programmability

Deployments
Democratize Privacy-Preserving Computation

My work aims to democratize access to privacy-preserving computation with new tools, systems, and abstractions.

Secure Computation
- FHE Compilers
  - IEEE S&P
- HECO
  - USENIX Security

Differential Privacy
- Cohere
  - IEEE S&P
Differential Privacy in Large-Scale Systems

(IEEE S&P‘24)
Statistical Release

How can we release useful information without compromising privacy?

Personal Data → Analysis → Release (Privacy, Utility) → Auxiliary Data
Statistical Release

How can we release useful information without compromising privacy?
Statistical Release
How can we release useful information without compromising privacy?

Personal Data

Analysis

Privacy

Utility

Release

Auxiliary Data

Industry

Academia

Service

Users
Statistical Release
How can we release useful information without compromising privacy?

Personal Data

Analysis

Privacy

Utility

Release

Auxiliary Data

Industry → Academia

Service → Users

Government → Population
Statistical Release

How can we release useful information without compromising privacy?
Statistical Release
How can we release useful information without compromising privacy?

- **Anonymization**
  Redact Personally Identifiable Information

<table>
<thead>
<tr>
<th>Name</th>
<th>Region</th>
<th>...</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice</td>
<td>CH</td>
<td></td>
<td>100</td>
</tr>
<tr>
<td>Bob</td>
<td>DE</td>
<td></td>
<td>237</td>
</tr>
</tbody>
</table>
Statistical Release

How can we release useful information without compromising privacy?

- **Anonymization**
  Redact Personally Identifiable Information

- **Release Aggregates**

<table>
<thead>
<tr>
<th>Name</th>
<th>Region</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice</td>
<td>CH</td>
<td>100</td>
</tr>
<tr>
<td>Bob</td>
<td>DE</td>
<td>237</td>
</tr>
</tbody>
</table>
Statistical Release
How can we release useful information without compromising privacy?

- Anonymization
  Redact Personally Identifiable Information

- Release Aggregates

<table>
<thead>
<tr>
<th>Name</th>
<th>Region</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alice</td>
<td>CH</td>
<td>100</td>
</tr>
<tr>
<td>Bob</td>
<td>DE</td>
<td>237</td>
</tr>
</tbody>
</table>

Privacy Attacks
- Re-Identification
- Database Reconstruction
- Membership Inference

Auxiliary Data
Differential Privacy
Mathematical definition of privacy in the context of statistical releases
Differential Privacy
Mathematical definition of privacy in the context of statistical releases
Differential Privacy
Mathematical definition of privacy in the context of statistical releases

\[ \mathcal{M}(D_1) \]

Result A

\[ \mathcal{M}(D_2) \]

Result B
Differential Privacy

Mathematical definition of privacy in the context of statistical releases

Result A

Result B
Differential Privacy
Mathematical definition of privacy in the context of statistical releases

\[ \Pr \left[ \mathcal{M}(D_1) \in S \right] \leq e^\epsilon \cdot \Pr \left[ \mathcal{M}(D_2) \in S \right] + \delta \]
Differential Privacy
Mathematical definition of privacy in the context of statistical releases

\[ Pr [\mathcal{M}(D_1) \in \mathcal{S}] \leq e^\epsilon \cdot Pr [\mathcal{M}(D_2) \in \mathcal{S}] + \delta \]
Differential Privacy
Mathematical definition of privacy in the context of statistical releases

\[ Pr [\mathcal{M}(D_1) \in S] \leq e^{\varepsilon} \cdot Pr [\mathcal{M}(D_2) \in S] + \delta \]
Differential Privacy

Mathematical definition of privacy in the context of statistical releases

\[ Pr[\mathcal{M}(D_1) \in S] \leq e^\varepsilon \cdot Pr[\mathcal{M}(D_2) \in S] + \delta \]
Differential Privacy
Mathematical definition of privacy in the context of statistical releases

\[ Pr \left[ M(D_1) \in S \right] \leq e^\varepsilon \cdot Pr \left[ M(D_2) \in S \right] + \delta \]
Differential Privacy
Mathematical definition of privacy in the context of statistical releases

\[ Pr [M(D_1) \in S] \leq e^\varepsilon \cdot Pr [M(D_2) \in S] + \delta \]
Differential Privacy

Mathematical definition of privacy in the context of statistical releases

\[ Pr[\mathcal{M}(D_1) \in S] \leq e^\varepsilon \cdot Pr[\mathcal{M}(D_2) \in S] + \delta \]
From Theory to Practice

Calibrating Noise to Sensitivity in Private Data Analysis
Cynthia Dwork, Frank McSherry, Kobbi Nissim, Adam Smith
From Theory to Practice

Calibrating Noise to Sensitivity in Private Data Analysis
Cynthia Dwork, Frank McSherry, Kobbi Nissim, Adam Smith

Theory
6 10 13 24 46 23 50 71 80 120 163 280 400 497 671 904

arXiv keyword
Differential Privacy

Theory

Mechanism Design
DP Variants
Local Sensitivity
Composition Theorems
Synthetic Data
Local DP
DP-SGD

Calibrating Noise to Sensitivity in Private Data Analysis
Cynthia Dwork, Frank McSherry, Kobbi Nissim, Adam Smith
From Theory to Practice

Calibrating Noise to Sensitivity in Private Data Analysis
Cynthia Dwork, Frank McSherry, Kobbi Nissim, Adam Smith

OnTheMap
US Census

Real-World Applications
[Desfontaines Blog, 2021]
From Theory to Practice

Calibrating Noise to Sensitivity in Private Data Analysis
Cynthia Dwork, Frank McSherry, Kobbi Nissim, Adam Smith

Real-World Applications
OnTheMap
US Census

RAPPOR
Google

[Desfontaines Blog, 2021]
From Theory to Practice

Calibrating Noise to Sensitivity in Private Data Analysis
Cynthia Dwork, Frank McSherry, Kobbi Nissim, Adam Smith

Real-World Applications
[Desfontaines Blog, 2021]
From Theory to Practice

Calibrating Noise to Sensitivity in Private Data Analysis
Cynthia Dwork, Frank McSherry, Kobbi Nissim, Adam Smith

Real-World Applications
[Desfontaines Blog, 2021]
From Theory to Practice

Calibrating Noise to Sensitivity in Private Data Analysis
Cynthia Dwork, Frank McSherry, Kobbi Nissim, Adam Smith

Real-World Applications

[Desfontaines Blog, 2021]
From Theory to Practice

Calibrating Noise to Sensitivity in Private Data Analysis
Cynthia Dwork, Frank McSherry, Kobbi Nissim, Adam Smith

Real-World Applications
[Desfontaines Blog, 2021]
Deploying DP Applications
Deploying DP Applications

Image Dataset

pytorch

opacus

ML Model

(ε, δ)_{DP}
Deploying DP Applications

Image Dataset

Documents

pytorch

ML Model

(\varepsilon_1, \delta_1)

ML Model

(\varepsilon_2, \delta_2)
Deploying DP Applications

Image Dataset

ML Model

$(\varepsilon_1, \delta_1)$

Documents

ML Model

$(\varepsilon_2, \delta_2)$

Relational Data

SQL Analytics

$(\varepsilon_3, \delta_3)$
Deploying DP Applications

- Image Dataset
  - pytorch
  - opacus

- Documents
  - pytorch
  - opacus

- Relational Data
  - Tumult Analytics

- ML Model
  - $(\epsilon_1, \delta_1)$

- SQL Analytics
  - $(\epsilon_2, \delta_2)$

- Leakage $(\epsilon, \delta)$
Deploying DP Applications

Image Dataset

Documents

Relational Data

pytorch

opacus

ML Model

Leakage (ε₁, δ₁)

ML Model

Leakage (ε₂, δ₂)

Tumult

Analytics

SQL Analytics

Leakage (ε₃, δ₃)
Deploying DP Applications

Image Dataset
- pytorch
- opacus

Documents
- pytorch
- opacus

Relational Data
- Tumult Analytics

ML Model
- \( (\varepsilon_1, \delta_1)_\text{DP} \)

Leakage \( \varepsilon_1, \delta_1 \)

ML Model
- \( (\varepsilon_2, \delta_2)_\text{DP} \)

Leakage \( \varepsilon_2, \delta_2 \)

ML Model
- \( (\varepsilon_3, \delta_3)_\text{DP} \)

Leakage \( \varepsilon_3, \delta_3 \)

SQL Analytics
- (\varepsilon_3', \delta_3')_\text{DP}
Deploying DP Applications

Image Dataset

Documents

Relational Data

pytorch

opacus

ML Model

Leakage ($\epsilon', \delta'$)

ML Model

Leakage ($\epsilon_2, \delta_2$)

SQL Analytics

Leakage ($\epsilon_3, \delta_3$)
Deploying DP Applications

- Image Dataset
- Documents
- Relational Data
- Tumult Analytics
- pytorch-opacus

ML Model

Leakage $(\varepsilon_1, \delta_1)$

ML Model

Leakage $(\varepsilon_2, \delta_2)$

ML Model

Leakage $(\varepsilon_3, \delta_3)$
Deploying DP Applications

- Image Dataset
- Documents
- Relational Data
- Tumult Analytics
- pytorch
- opacus

Leakage \((\varepsilon_1, \delta_1)\)

ML Model \(\text{DP}^1\)

Leakage \((\varepsilon_2, \delta_2)\)

ML Model \(\text{DP}^2\)

Leakage \((\varepsilon_3, \delta_3)\)

SQL Analytics \(\text{DP}^3\)
Deploying DP Applications

Leakage $(\epsilon_1, \delta_1)$

ML Model $(\epsilon_1, \delta_1)$

Leakage $(\epsilon_2, \delta_2)$

ML Model $(\epsilon_2, \delta_2)$

Leakage $(\epsilon_3, \delta_3)$

ML Model $(\epsilon_3, \delta_3)$

SQL Analytics $(\epsilon_3, \delta_3)$

Image Dataset

Documents

Relational Data

Tumult Analytics

pytorch

opacus

pytorch

opacus

Tumult Analytics

SQL Analytics
Deploying DP Applications

Image Dataset

Documents

Relational Data

pytorch

opacus

ML Model

(\varepsilon_1, \delta_1)

\text{DP}

(\varepsilon_2, \delta_2)

\text{DP}

(\varepsilon_3, \delta_3)

\text{DP}

Leakage

(\varepsilon, \delta)

(\varepsilon, \delta)

(\varepsilon, \delta)

SQL Analytics
Deploying DP Applications

System-wide DP Guarantee

We need a system that carefully controls and allocates privacy budget across heterogeneous applications and data systems over time.
Cohere: Unified System Architecture for DP

Data Layer

DP Management Layer

Application Layer

Goal: Enforce Tight System-wide DP Guarantee

Budget Control

Fine-Grained Tracking and Coordination of Shared Global DP State

Resource Planner

Allocation of Finite Shared Privacy Resources (i.e., budget) under Complex Preferences

Documents

Image Dataset

Relational Data

ML Model

ML Model

SQL Analytics

pytorch

opacus

pytorch

opacus

Tumult Analytics

(\epsilon_1, \delta_1)_{DP_1}

(\epsilon_2, \delta_2)_{DP_2}

(\epsilon_3, \delta_3)_{DP_3}
Challenges: System-wide Privacy Guarantee
Challenges: System-wide Privacy Guarantee

1. Coordination Problem

Single Shared Privacy State
Challenges: System-wide Privacy Guarantee

1. Coordination Problem

Single Shared Privacy State

2. Composition Complexity

\[ (\varepsilon_1, \delta_1) + (\varepsilon_2, \delta_2) + (\varepsilon_3, \delta_3) \leq (\varepsilon, \delta) \text{ - DP} \]
Challenges: System-wide Privacy Guarantee

1. Coordination Problem

2. Composition Complexity

3. Scarce and Finite Resource

\[(\epsilon_1, \delta_1), (\epsilon_2, \delta_2), (\epsilon_3, \delta_3) \leq (\epsilon, \delta) - DP\]
Challenges: System-wide Privacy Guarantee

1. Coordination Problem
   - Multi-Team
   - Multi-Application
   - Multi-Library
   - Single Shared Privacy State

2. Composition Complexity
   \[ (\epsilon_1, \delta_1) + (\epsilon_2, \delta_2) + (\epsilon_3, \delta_3) \leq (\epsilon, \delta) - DP \]

3. Scarce and Finite Resource
   - Resource Allocation
   - Continuity Guarantee
   - Continuity Guarantee
   - Continuity Guarantee
   - Continuity Guarantee
   - Continuity Guarantee
   - Continuity Guarantee
Unified System Architecture for DP

Data Layer

- Documents
- Image Dataset
- Relational Data

Application Layer

- ML Model \((\varepsilon_1, \delta_1)\)_{DP}
- ML Model \((\varepsilon_2, \delta_2)\)_{DP}
- Leakage \((\varepsilon_3, \delta_3)\)
- SQL Analytics \((\varepsilon_4, \delta_4)\)_{DP}

Access Control
Unifying the Application Layer

Image Dataset

Documents

Relational Data

Application Layer

ML Model

ML Model

SQL Analytics

\((\varepsilon_1, \delta_1)_{\text{DP}}\)

\((\varepsilon_2, \delta_2)_{\text{DP}}\)

\((\varepsilon_3, \delta_3)_{\text{DP}}\)
Unifying the Application Layer

Image Dataset

Best ML Analysis

ML Model

$\epsilon_1, \delta_1$

Documents

Best ML Analysis

ML Model

$\epsilon_2, \delta_2$

Relational Data

Tumult Analytics

SQL Analytics

$\epsilon_3, \delta_3$

Application Layer
Unifying the Application Layer

Image Dataset

Best ML Analysis

ML Model

Documents

Best ML Analysis

ML Model

Relational Data

Best SQL Analysis

SQL Analytics

Application Layer
Unifying the Application Layer

- Image Dataset
  - Best ML Analysis
  - ML Model $\epsilon_1, \delta_1$

- Documents
  - Best ML Analysis
  - ML Model $\epsilon_2, \delta_2$

- Relational Data
  - Best SQL Analysis
  - SQL Analytics $\epsilon_3, \delta_3$

 Application Layer

$\epsilon, \delta$ - DP
Unifying the Application Layer

Image Dataset

Best ML Analysis

ML Model

Documents

Best ML Analysis

ML Model

Relational Data

Best SQL Analysis

SQL Analytics

Moving Beyond Local Optima

Application Layer
DP Libraries: In a Nutshell
DP Libraries: In a Nutshell

Library-specific DP Algorithm Design
Transformation | Mechanism | Sensitivity

Query Plan

DP Compiler Calibrate Noise

(\(\varepsilon, \delta\))

Noise Plan

Universal Across Libraries
Composition of Fundamental Mechanisms

- Gaussian
- Laplace
- Sparse Vector Technique
- Discrete Gaussian
- Exponential
- ...
If we can compose all fundamental mechanisms, we can support a variety of heterogeneous libraries through a unified noise plan.
Unifying the Application Layer

Best ML Analysis

ML Model

Noise Plan

Rényi DP

\[ \varepsilon(\alpha_1), \varepsilon(\alpha_2), \varepsilon(\alpha_3), \varepsilon(\alpha_4), \ldots, \varepsilon(\alpha_N) \]

\((\varepsilon, \delta)\) - DP

Best SQL Analysis

SQL Analytics

Application Layer

Image Dataset

Documents

Relational Data

Best ML Analysis

ML Model

Noise Plan

Best ML Analysis

ML Model

Noise Plan

Best SQL Analysis

SQL Analytics

ML Model

Noise Plan

[Mironov 2017]
Unifying the Application Layer

Is this the best we can do?

Assumptions: All applications are presumed to access every user.
Fine-grained Privacy Analysis

Parallel Composition

\[ \max(\varepsilon_2, \varepsilon_3) \]

[McSherry 2009]
Fine-grained Privacy Analysis

Parallel Composition

\[ \max(\varepsilon_2, \varepsilon_3) \]

[McSherry 2009]
Fine-grained Privacy Analysis

Parallel Composition

\[
\max(\varepsilon_2, \varepsilon_3)
\]

[McSherry 2009]
Fine-grained Privacy Analysis

Image Dataset
pytorch
opacus

ML Model
(ε₁)
DP

Documents
pytorch
opacus

ML Model
(ε₂)
DP

Relational Data
Tumult
Analytics

SQL Analytics
(ε₃)
DP

Block Composition

ε₁
ε₂
ε₃

[Lécuyer SOSP’19]
Fine-grained Privacy Analysis

- Image Dataset
  - pytorch
  - opacus
- Documents
  - pytorch
  - opacus
- Relational Data
  - Tumult Analytics

Block Composition

\( \varepsilon_1 \) \( \varepsilon_2 \) \( \varepsilon_3 \)

n:1 Blocks

[\text{Lécuyer SOSP’19}]

ML Model (\( \varepsilon \) DP)
ML Model (\( \varepsilon \) DP)
SQL Analytics (\( \varepsilon \) DP)
Fine-grained Privacy Analysis

- Image Dataset
  - pytorch
  - opacus
- Documents
  - pytorch
  - opacus
- Relational Data
  - Tumult Analytics

Block Composition

\[ \varepsilon_1, \varepsilon_2, \varepsilon_3 \]

\[ \max(\ldots) \]

[Lécuyer SOSP’19]
Fine-grained Privacy Analysis

Image Dataset

pytorch

ML Model

opacus

(DP)

Documents

pytorch

ML Model

opacus

(DP)

Relational Data

Tumult

ML Model

Analytics

(DP)

SQL Analytics

Block Composition

max( , , )

( )

n:1

[ ]

[Block Composition]

[Lécuyer SOSP’19]
Fine-grained Privacy Analysis

- **Image Dataset**: $\text{Image Dataset}$
- **Documents**: $\text{Documents}$
- **Relational Data**: $\text{Tumult Analytics}$

**Block Composition**

$$\max(\varepsilon_1+\varepsilon_2, \varepsilon_1+\varepsilon_3)$$

[Leçuyer SOSP’19]
Fine-grained Privacy Analysis

- Image Dataset
- Documents
- Relational Data

PyTorch/Opacus

ML Model

DP

SQL Analytics

Tumult Analytics

DP

Block Composition

\[
\text{max}(\varepsilon_1, \varepsilon_2, \varepsilon_3)
\]

\[
\varepsilon_1 + \varepsilon_2, \varepsilon_1 + \varepsilon_3, \varepsilon_2 + \varepsilon_3
\]

[Lécuyer SOSP’19]
Fine-grained Privacy Analysis

Application Layer

Image Dataset

Documents

Relational Data

pytorch

opacus

pytorch

opacus

Tumult

Analytics

ML Model

Noise Plan

ML Model

Noise Plan

ML Model

Noise Plan

Rényi DP

[\text{Mironov 2017}]

\(\varepsilon(\alpha_1), \varepsilon(\alpha_2), \varepsilon(\alpha_3), \varepsilon(\alpha_4), \ldots, \varepsilon(\alpha_N)\)

\((\varepsilon, \delta) - \text{DP}\)
Fine-grained Privacy Analysis

Application Layer

1. **Image Dataset**
   - **pytorch**
   - **opacus**
   - Noise Plan
   - ML Model

2. **Documents**
   - **pytorch**
   - **opacus**
   - Noise Plan
   - ML Model
   - - LLM

3. **Relational Data**
   - **Tumult Analytics**
   - Noise Plan
   - SQL Analytics
   - - Statistics

 duk₂, 𝛿₁kwargs
 duk₂, 𝛿₂kwargs
 duk₃, 𝛿₃kwargs
Fine-grained Privacy Analysis

- Image Dataset
  - pytorch
  - opacus
- Documents
  - pytorch
  - opacus
- Relational Data
  - Tumult Analytics
- Application Layer
  - Noise Plan
  - ML Model
  - SQL Analytics
  - Statistics
  - LLM
Fine-grained Privacy Analysis

Image Dataset
- pytorch
- opacus

Documents
- pytorch
- opacus

Relational Data
- Tumult Analytics

Application Layer
- ML Model
- Noise Plan
- LLM
- Statistics
- SQL Analytics
- Statistics
Fine-grained Privacy Analysis

- Image Dataset
  - pytorch
  - opacus
- Documents
  - pytorch
  - opacus
- Relational Data
  - Tumult Analytics
- Application Layer
  - Noise Plan
  - ML Model
  - SQL Analytics
  - LLM
  - - Statistics
Fine-grained Privacy Analysis

Partitioning Attributes
Schema must be known in advance
Fine-grained Privacy Analysis

Application Layer

- **Image Dataset**
  - pytorch
  - opacus
- **Documents**
  - pytorch
  - opacus
- **Relational Data**
  - Tumult Analytics
  - SQL Analytics
  - - Statistics

**ML Model**

**Noise Plan**

**Year of Birth**

1920

2024

Country

- **Partitioning Attributes**
  - Schema must be known in advance
Fine-grained Privacy Analysis

Image Dataset

pytorch

Documents

pytorch

opacus

ML Model

Noise Plan

Relational Data

Application Layer

Tumult Analytics

SQL Analytics

pytorch

opacus

Tumult Analytics

Fine-grained Privacy Analysis allows for a tighter Composition.

ML Model

Noise Plan

ML Model

Noise Plan

ML Model

Noise Plan

Country

Year of Birth

1920

2024

...
## Sampling: Random Subset Selection

<table>
<thead>
<tr>
<th>Image Dataset</th>
<th>Documents</th>
<th>Relational Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image Dataset" /></td>
<td><img src="image2.png" alt="Documents" /></td>
<td><img src="image3.png" alt="Relational Data" /></td>
</tr>
</tbody>
</table>

- **Image Dataset**
  - `pytorch`
  - `opacus`

- **Documents**
  - `pytorch`
  - `opacus`

- **Relational Data**
  - `Tumult Analytics`

**ML Model**
- `Noise Plan`
- `ML Model`
- `LLM`

**Application Layer**

<table>
<thead>
<tr>
<th>Country</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>🇺🇸 USA</td>
<td>...</td>
</tr>
<tr>
<td>🇨🇦 Canada</td>
<td>...</td>
</tr>
<tr>
<td>🇪🇸 Spain</td>
<td>...</td>
</tr>
<tr>
<td>🇨🇭 Switzerland</td>
<td>...</td>
</tr>
<tr>
<td>🇳🇴 Norway</td>
<td>...</td>
</tr>
<tr>
<td>🇨ielding Germany</td>
<td>...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year of Birth</th>
<th>1920</th>
<th>...</th>
<th>2024</th>
</tr>
</thead>
<tbody>
<tr>
<td>1930</td>
<td></td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td></td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>2020</td>
<td></td>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

**Noise Plan**
- `Statistics`
- `SQL Analytics`

**Notes:**
- `2024` Year of Birth
- `1920` Year of Birth
- `...` Indicates omitted data for visualization.
Sampling: Random Subset Selection

- Image Dataset
  - ML Model
  - Noise Plan
- Documents
  - ML Model
  - Noise Plan
- Relational Data
  - ML Model
  - Noise Plan

Amplification via Subsampling

Population

ML Model
- LLM

SQL Analytics
- Statistics

Application Layer
Scarce andFinite Resource

Image Dataset

ML Model

pytorch

opacus

Documents

pytorch

opacus

ML Model

- LLM

Relational Data

Tumult

Analytics

SQL Analytics

- Statistics

Application Layer

Rényi DP

\[ \varepsilon(\alpha_1) \varepsilon(\alpha_2) \varepsilon(\alpha_3) \ldots \varepsilon(\alpha_N) \]

Privacy Requirements

Noise Plan

Data Requirements

\[ \sigma( ) \]

Partitioning Attributes

Subsampling

Management Layer

Year of Birth

1920

2024

Country

Relational Data

- Statistics

ML Model

- LLM

Tumult

Analytics

Relational Data

Image Dataset

Documents

pytorch

opacus

SQL Analytics

- Statistics

Application Layer
Scarce and Finite Resource

Application Layer

Privacy Leakage

ML Training

ML Training

SQL

Management Layer

Year of Birth

1920

2024

Users

Country

1920

1920

...
Scarce and Finite Resource

Application Layer

Management Layer

Privacy Leakage

Year of Birth
1920
... 2024

Users

ML Training

SQL

ML Training

Country

1920
...
Scarce and Finite Resource

Application Layer

Management Layer

Privacy Leakage

Year of Birth
1920
2024

Country

ML Training
SQL
ML Training
ML Training
Scarce and Finite Resource

Application Layer

Management Layer

Privacy Leakage

Year of Birth

Country

Time

ML Training

SQL
Scarce and Finite Resource

Resource Allocation under Finite Privacy Budget
Continuity under a Finite Budget
Ensuring Sustained Budget Allocation Over Time

Resetting Budget

→


Continuity under a Finite Budget
Ensuring Sustained Budget Allocation Over Time

Resetting Budget

DP Violation
Continuity under a Finite Budget
Ensuring Sustained Budget Allocation Over Time

Resetting Budget

DP Violation

User Rotation

retired groups

active groups

future groups
Continuity under a Finite Budget
Ensuring Sustained Budget Allocation Over Time

User Rotation

Resetting Budget

DP Violation

Biased Set of Active Users
Continuity under a Finite Budget
Ensuring Sustained Budget Allocation Over Time

Resetting Budget → DP Violation

User Rotation
- Retired groups
- Active groups
- Future groups

Biased Set of Active Users
Budget Guarantees with Unlocking
Continuity under a Finite Budget
Ensuring Sustained Budget Allocation Over Time

User Rotation

Resetting Budget

DP Violation

Biased Set of Active Users
Budget Guarantees with Unlocking

Retired groups
Active groups
Future groups
Privacy Resource Allocation

Potential Applications
Privacy Resource Allocation

Optimize the number of Applications
Privacy Resource Allocation

Optimize the number of Applications
Privacy Resource Allocation

Optimize Privacy Cost Relative to Error
Privacy Resource Allocation

![Diagram showing the relationship between privacy cost and relative error, with a note to optimize for utility.](image-url)
Privacy Resource Allocation

Potential Applications

Multidimensional Knapsack Problem

Objective:

\[ \max \sum_{i \in \text{Apps}} \text{Utility}_i \cdot y_i \]

\[ y_i = 1 \quad \text{if application } i \text{ is allocated, else 0} \]
Privacy Resource Allocation

Potential Applications

Available Blocks

Multidimensional Knapsack Problem

Objective:
\[ \max \sum_{i \in \text{Apps}} Utility_i \cdot y_i \]

Budget Constraints:
\[ \sum_{i \in \text{Apps}} \epsilon_{ij} \cdot y_i \leq \text{Budget}_j \quad \forall j \in \text{Blocks} \]

Privacy cost of application $i$ for block $j$
* for simplicity we show the cost in $\epsilon$-DP rather than RDP
Privacy Resource Allocation

Multidimensional Knapsack Problem

Objective:
\[
\max \sum_{i \in \text{Apps}} \text{Utility}_i \times y_i
\]

Budget Constraints:
\[
\sum_{i \in \text{Apps}} \varepsilon_{ij} \times y_i \leq \text{Budget}_j \quad \forall j \in \text{Blocks}
\]

Privacy cost of application $i$ for block $j$

* for simplicity we show the cost in $\varepsilon$-DP rather than RDP
### Resource Allocation: Taming the Complexity

#### Request 1

<table>
<thead>
<tr>
<th>Year of Birth</th>
<th>Country</th>
<th>Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>1925</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1926</td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2024</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Resource Allocation: Taming the Complexity

<table>
<thead>
<tr>
<th>Year of Birth</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>1925</td>
<td></td>
</tr>
<tr>
<td>1926</td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td></td>
</tr>
<tr>
<td>2024</td>
<td></td>
</tr>
</tbody>
</table>

Request 1

- All

Request 2

- D-A-CH

Groups
Resource Allocation: Taming the Complexity

Request 1

- All

Request 2

- D-A-CH

Request 3

- Europe Gen Z

Year of Birth

Country

Groups

...
Resource Allocation: Taming the Complexity

Request 1
- All

Request 2
- D-A-CH

Request 3
- Europe Gen Z

Request 4
- France
Resource Allocation: Taming the Complexity

Request 1
- All

1925
1926

Request 2
- D-A-CH

1997

Request 3
- Europe Gen Z

2012

Request 4
- France

Year of Birth

Country

Contending Costs

- R1
- R1 + R2
- R1 + R3
- R1 + R2 + R3
- R1 + R4
- R1+ R3 + R4

Groups

171
Resource Allocation: Taming the Complexity

Request 1
- All
- 1925
- 1926

Request 2
- D-A-CH
- 1997

Request 3
- Europe Gen Z
- 2012

Request 4
- France
- 2024

Country
- Year of Birth
- Contending Costs
- Budgets

Dimensionality Reduction

Application History

Groups

R1
R1 + R2
R1 + R3
R1 + R2 + R3
R1 + R4
R1+ R3 + R4
Resource Allocation: Taming the Complexity

```
Request 1
  All
  1925
  1926
  ...

Request 2
  D-A-CH
  1997
  ...

Request 3
  Europe Gen Z
  2012

Request 4
  France
  2024

Problem Size depends only on Requests rather than the domain size of the partitioning attributes
```

Contending Costs

- R1
- R1 + R2
- R1 + R3
- R1 + R2 + R3
- R4 + R4

Budgets
Evaluation Scenario

Weekly Allocations

... 40 Weeks

~500 Requests / Round
$\varepsilon \in [0.01, 0.75]$ $\delta = 10^{-9}$

Total Budget
$\varepsilon = 3$ $\delta = 10^{-7}$
Evaluation Scenario

Weekly Allocations

\( \varepsilon \in [0.01, 0.75] \quad \delta = 10^{-9} \)

\( \varepsilon = 3 \quad \delta = 10^{-7} \)

\(~500 \text{ Requests / Round}~

Total Budget

Baseline

PrivateKube

[ Luo et al. OSDI’21 ]

Fixed Coarse-Grained Privacy Analysis
Workload: Mixture of Analytics and ML Tasks

Allocation Algorithms
- Upper bound (ILP)
- FCFS
- DPF
- DPK
- Knapsack Approximation Algorithms

Utility [%]

PrivateKube  Ours
Workload: Mixture of Analytics and ML Tasks

Allocation Algorithms
- Upper bound (ILP)
- FCFS
- DPF
- DPK

Knapsack Approximation Algorithms

Utility [%]
PrivateKube  Ours
Workload: Mixture of Analytics and ML Tasks

Allocation Algorithms
- Upper bound (ILP)
- FCFS
- DPF
- DPK

Knapsack Approximation Algorithms

Utility [%]

PrivateKube  Ours

9x w/o PAs
Workload: Predicate Counting Queries

\[
\text{SELECT Count(*) FROM x WHERE } \Phi \quad \text{(Only Gaussian Mechanism)}
\]
Differential Privacy

Theory – System-wide DP Guarantee
Cross-framework Compatibility and Efficient Privacy Analysis

Resource Allocation
Distributing Budget across various Applications

System Continuity
Ensuring Sustained Budget Allocation Over Time

Practice

ppslab/cohere
My work aims to democratize access to privacy-preserving computation with new tools, systems, and abstractions.
My work aims to build practical systems that use cryptography to empower users and preserve their privacy.
Looking Forward
Democratize Privacy-Preserving Computation

Privacy-Preserving Systems Designs
Democratize Privacy-Preserving Computation

Hybrid Compilation

Secure Computation on Heterogeneous Hardware

Privacy-Preserving Systems Designs
Democratize Privacy-Preserving Computation

Hybrid Compilation

Secure Computation on Heterogeneous Hardware

Privacy-Preserving Systems Designs

End-to-End Privacy

Privacy-Transparency Dichotomy
End-to-End Privacy
Secure Computation
Homomorphic Encryption | Secure Multi-party Computation

E2E Security
**Secure Computation**

Homomorphic Encryption | Secure Multi-party Computation

1. Encrypt
2. Eval $f(\cdot)$
3. Decrypt

**Personal**

**E2E Security**

**Releasing Data**

Differential Privacy | Anonymization

1. Release
2. Database

**Privacy-enhanced view of the data**

**Public**
End-to-End Privacy Platform

Homomorphic Encryption | Secure Multi-party Computation | Zero Knowledge Proofs | Differential Privacy

- Encrypt
- Decrypt
- Eval $f(\cdot)$
- Enc
- Release

Personal → Encrypt → Enc( ) → Eval $f(\cdot)$ → Enc( ) → Release → Public

Privacy-enhanced view of the data

E2E Privacy

Data Subject → E2E Privacy → Data Use
Privacy-Transparency Dichotomy
Privacy-Transparency Dichotomy

[ Holding Secrets Accountable: Auditing Private ML Algorithms ]
Privacy-Preserving Machine Learning

Client → Private Inference → Service Provider → Private Training

- Privacy, Liability
- Privacy, Model Stealing, Model Evasion
- Privacy, Regulation, Competition
Verifiable Claims and Accountability in PPML
My work aims to **build** practical systems that use cryptography to empower users and preserve their privacy.