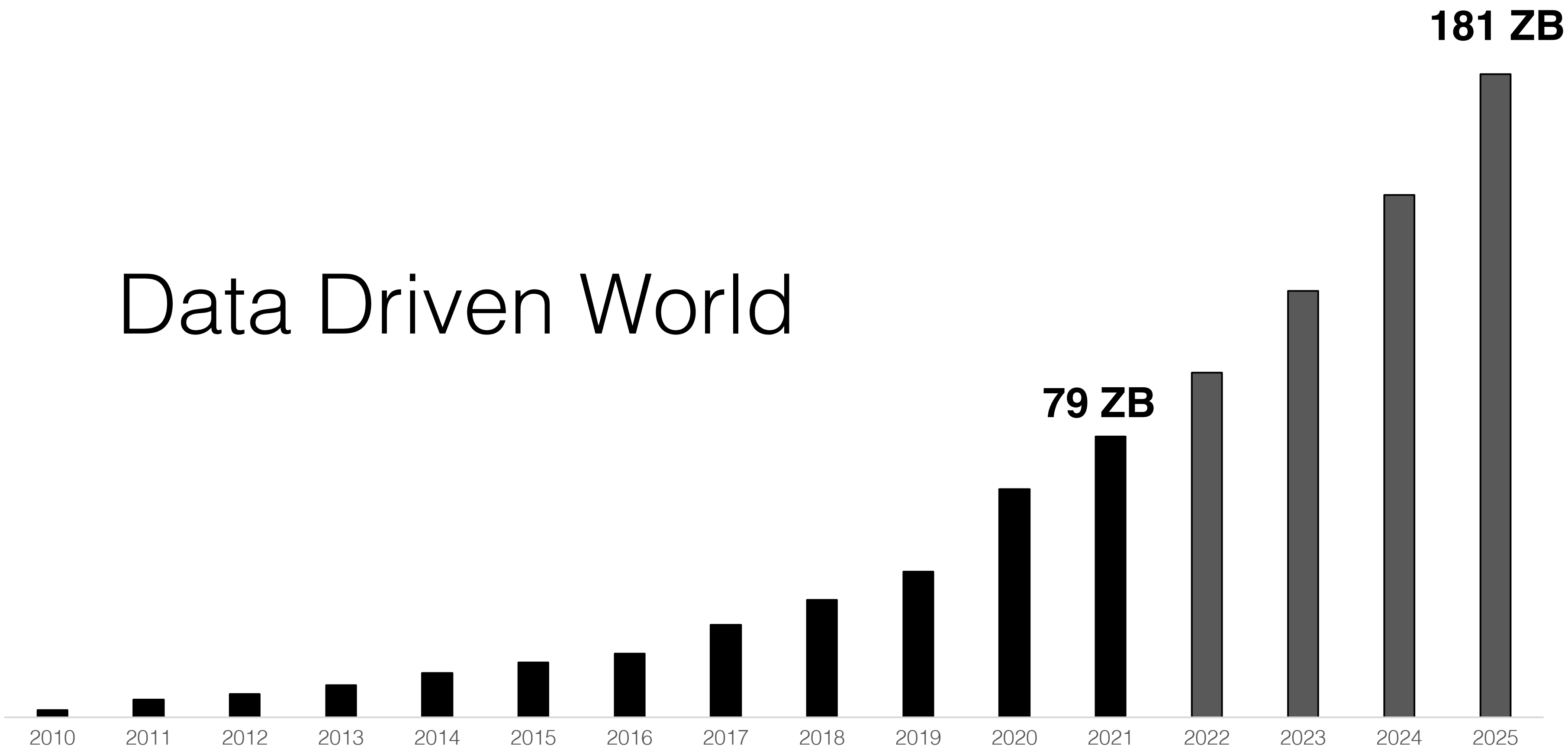
System Designs for End-to-End Privacy

Functionality, Performance, & Accessibility

Anwar Hithnawi

ETH zürich

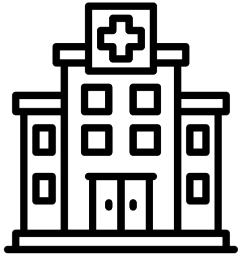
Based on joint work with: Lukas Burkhalter, Alexander Viand, Nicolas Kuechler, and Hidde Lycklama.



Data Driven World
79 ZB
181 ZB
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025

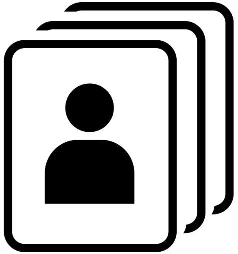
Sensitive Data

Finance



Health

Government



Personal

Morgan Stanley settles personal data breach lawsuit for $60 million

Data Breaches Keep Happening. So Why Don't You Do Something?

Capital One Data Breach Compromises Data of Over 100 Million

All 3 Billion Yahoo Accounts Were Affected by 2013 Attack

The Government Uses 'Near Perfect Surveillance' Data on Americans

Congressional hearings are urgently needed to address location tracking.

By THE EDITORIAL BOARD

Grindr and OkCupid Spread Personal Details, Study Says

Norwegian research raises questions about whether certain of sharing of information violate data privacy laws in Europ the United States.

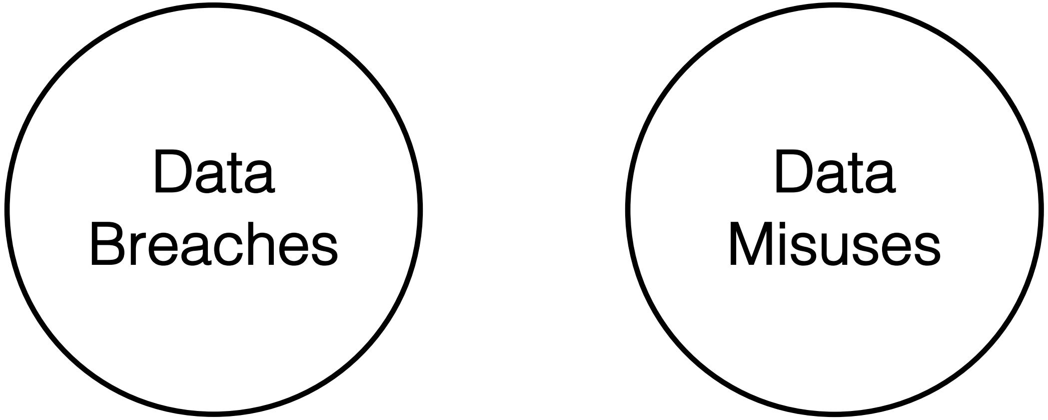
You Should Be Freaking Out About Privacy

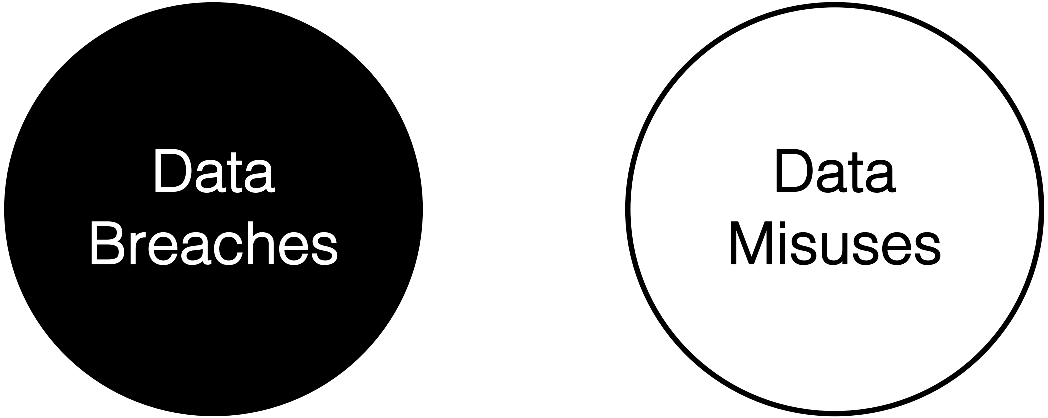
Nothing to hide, nothing to fear? Think again.

Technology

Data broker shared billions of location records with District during pandemic

The bulk sales of location data have fueled a debate over public health and privacy

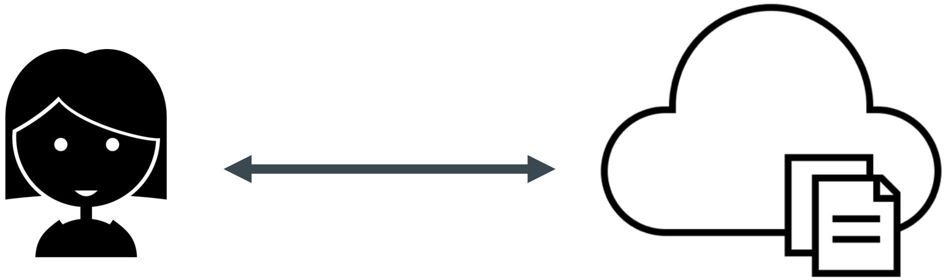

Data Breaches
Data Misuses


Data Breaches
Data Misuses

~ **1.245** Billion

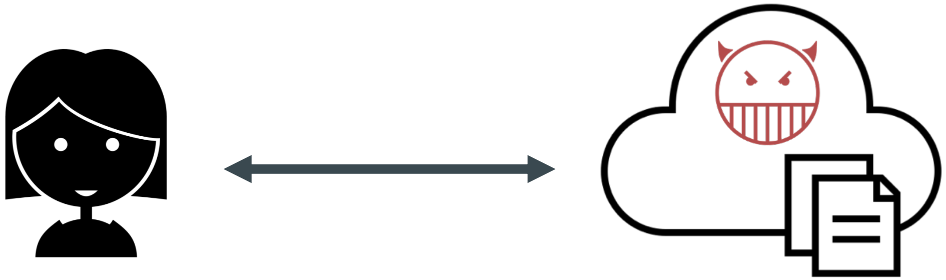
The number of data records **stolen** in 2020

Equifax	Uber	Twitter	Facebook
143,000,000	57,000,000	330,000,000	533,000,000
2017	2017	2018	2019



> “Where the sensitive information is concentrated, that is where the spies will go. This is just a fact of life.”

former NSA official Ken Silva.



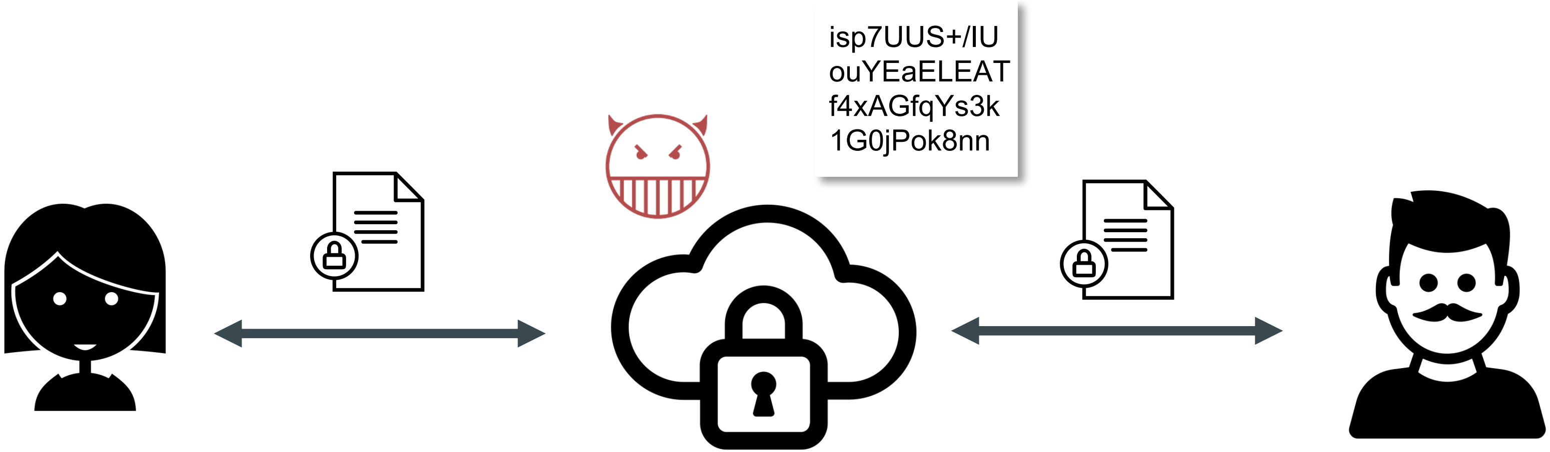
> Where the sensitive information is concentrated, that is where the spies will go. This is just a fact of life.

former NSA official Ken Silva.

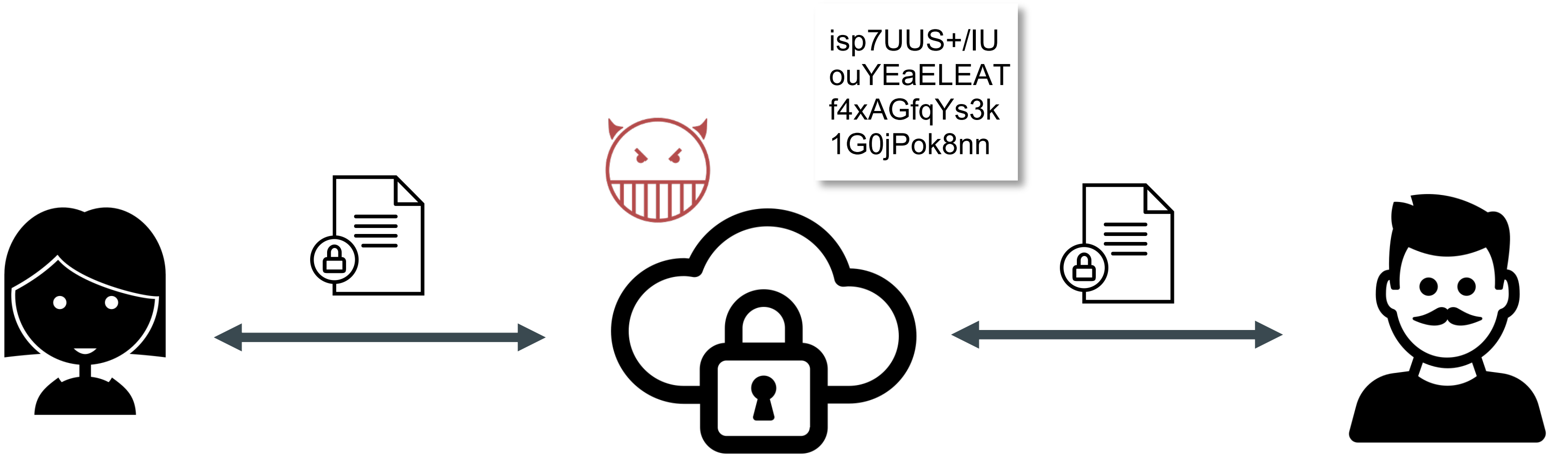
Software Vulnerabilities | Insider Threats | Physical Attacks

End-to-End Encrypted Systems

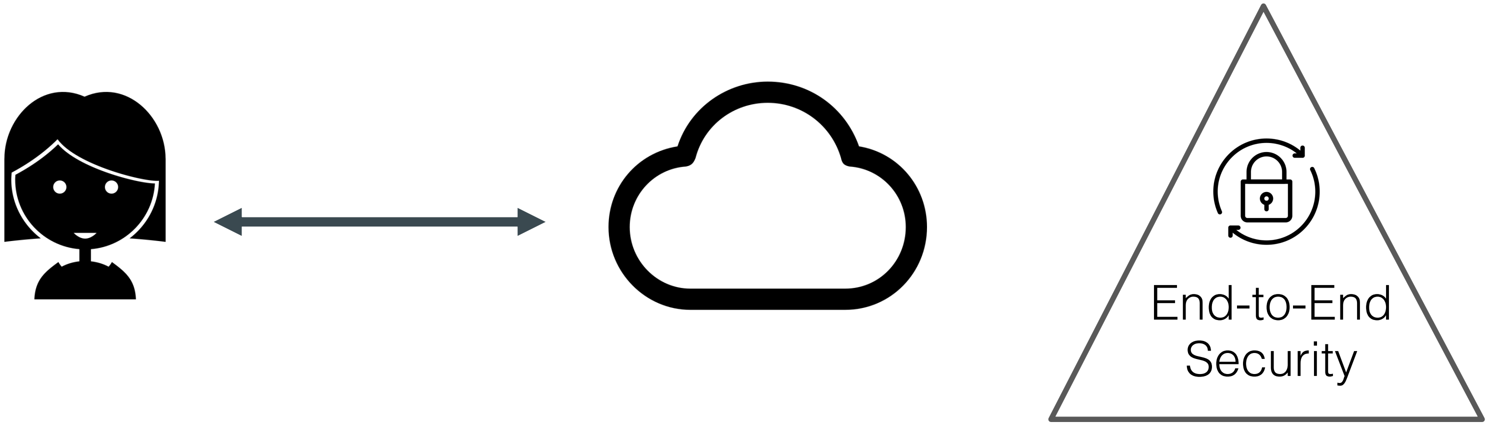
End-to-End Encrypted Systems



End-to-End Encrypted Systems

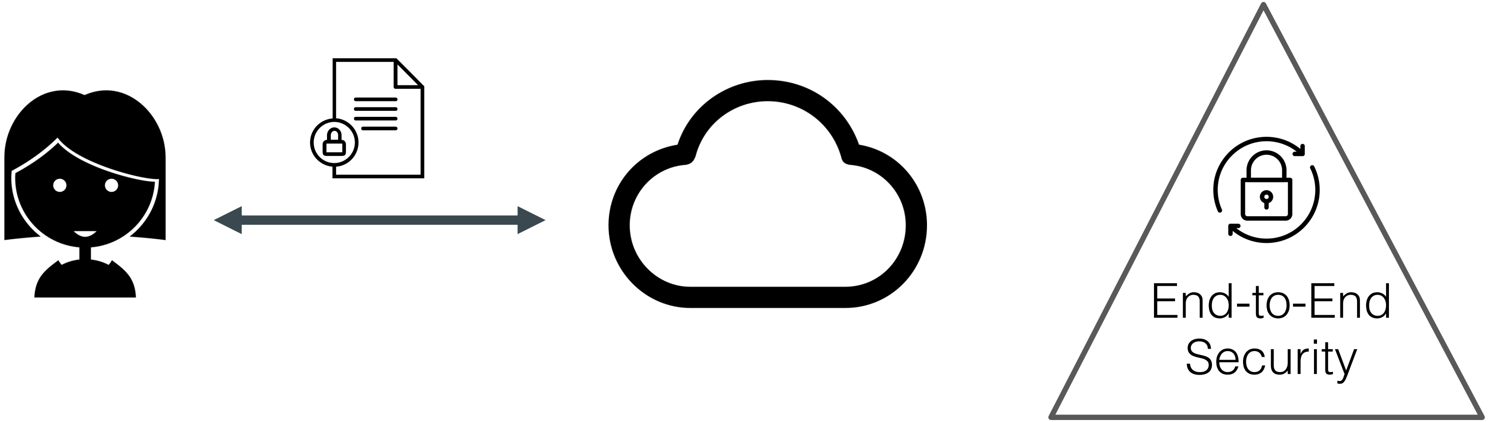


More Applications?



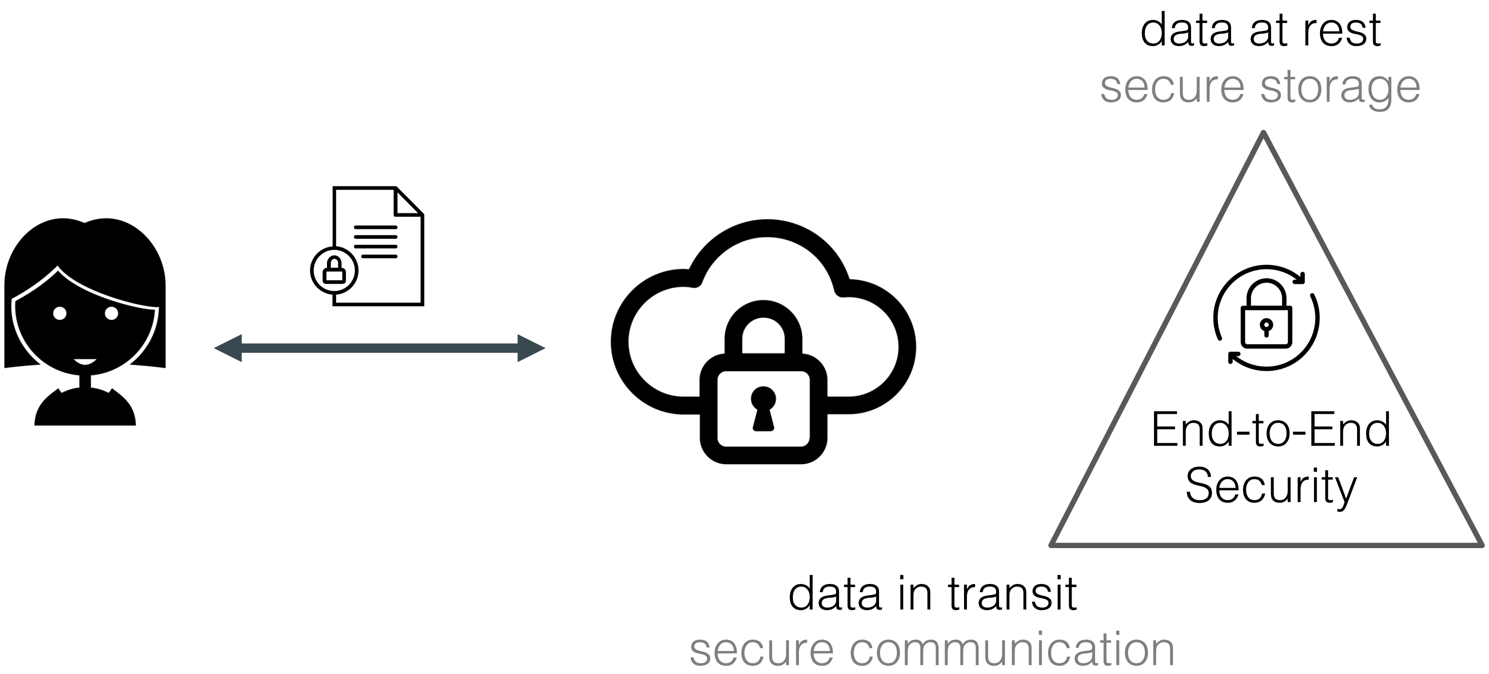
data in transit

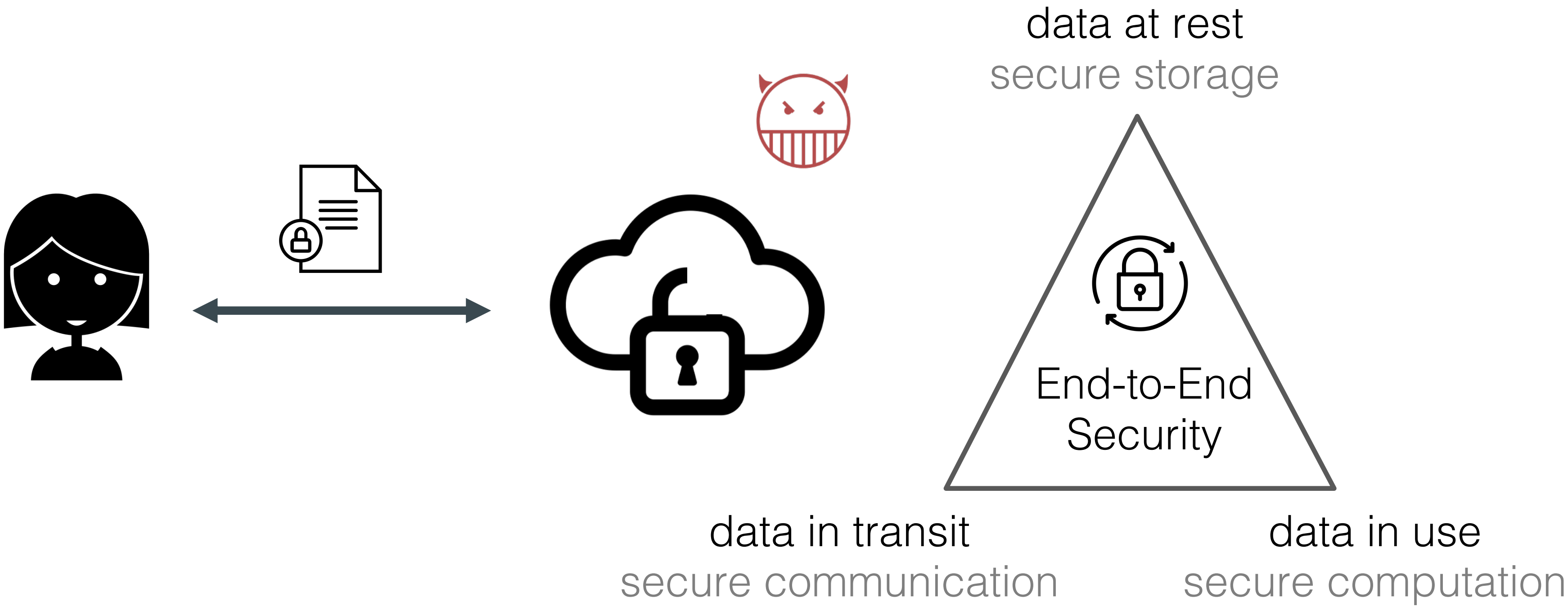
secure communication



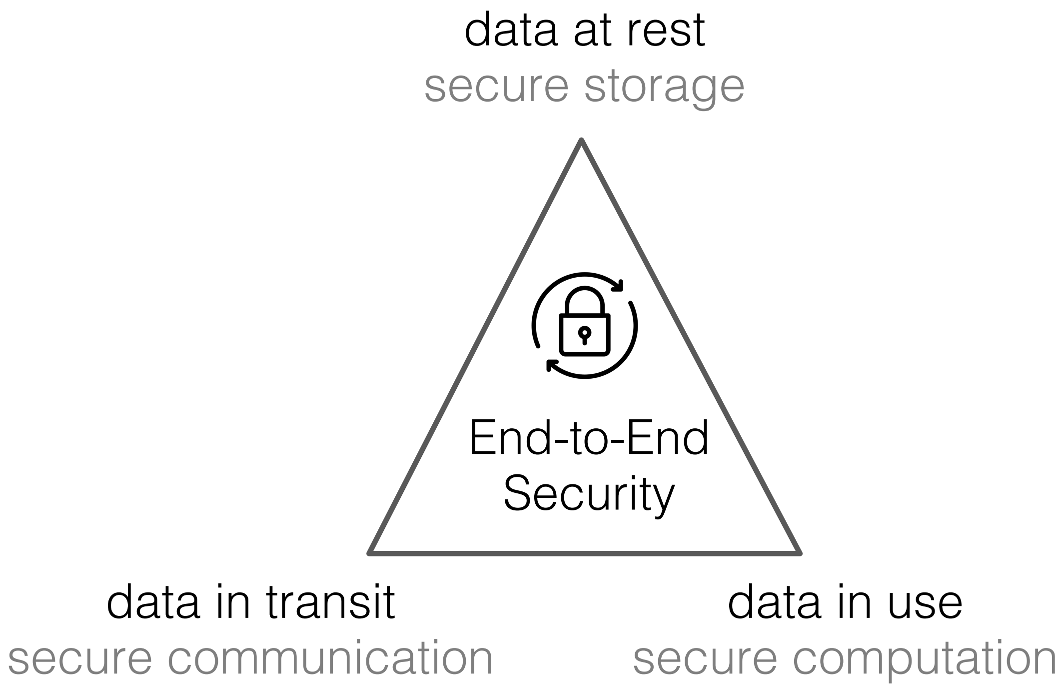
data in transit

secure communication


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


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

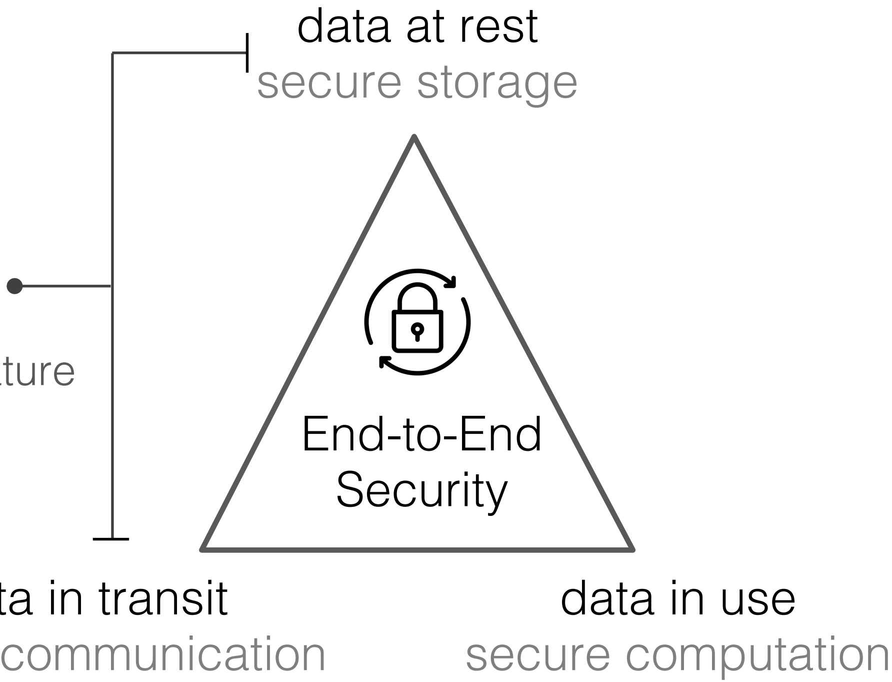
Modern Cryptography



Modern Cryptography

Conventional Crypto

Encryption & Digital Signature

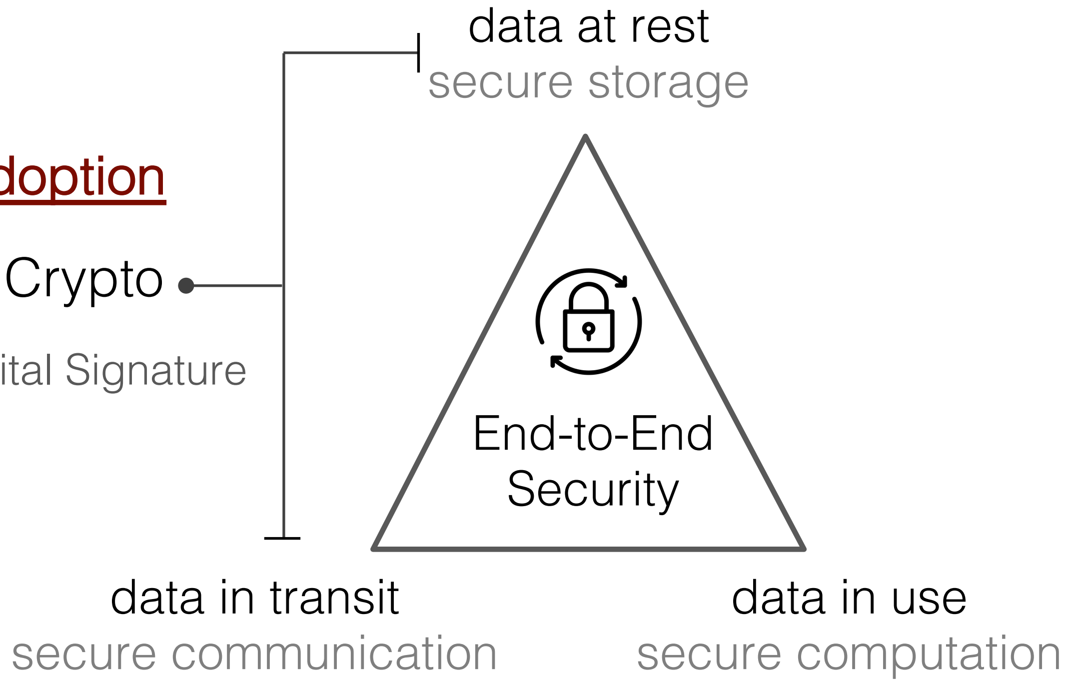


Modern Cryptography

Ubiquitous Adoption

Conventional Crypto

Encryption & Digital Signature



Modern Cryptography

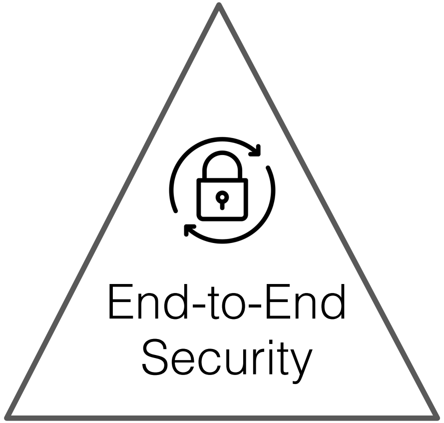
data at rest

secure storage

<u>Ubiquitous Adoption</u>

Conventional Crypto

Encryption & Digital Signature



Advanced Crypto

- Homomorphic Encryption
- Secure Multi-party Computation
- Zero Knowledge Proofs

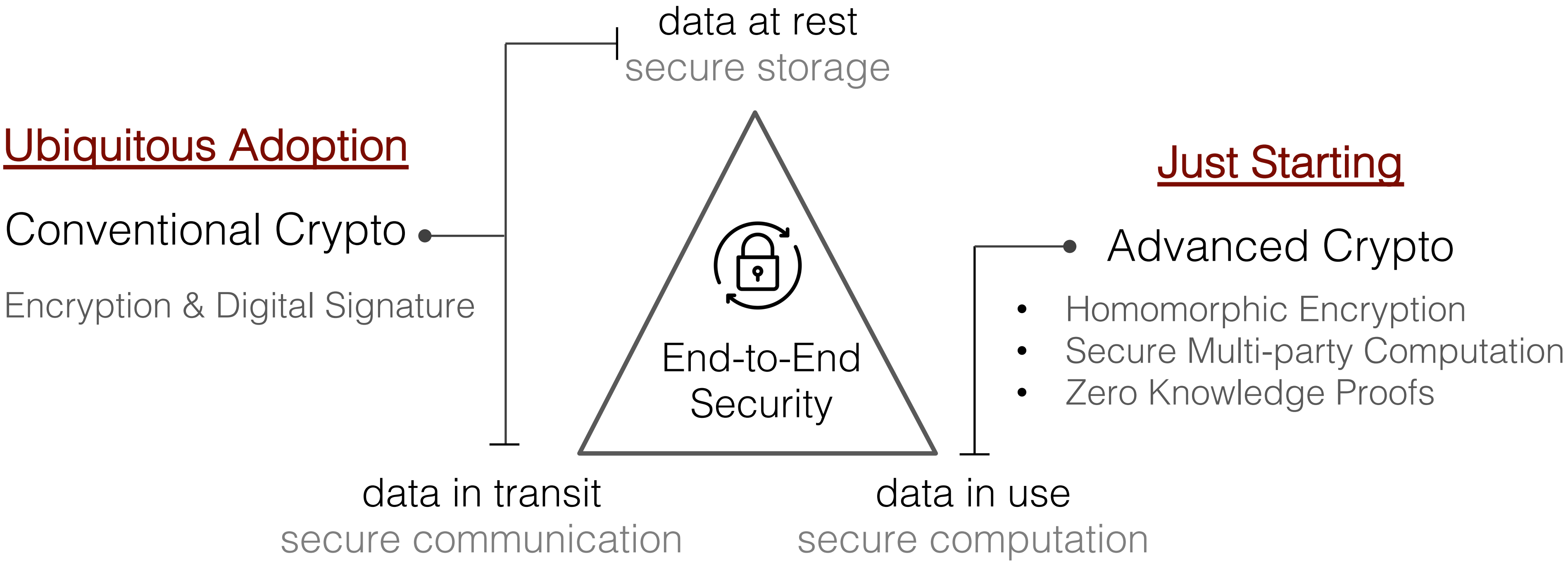
data in transit

secure communication

data in use

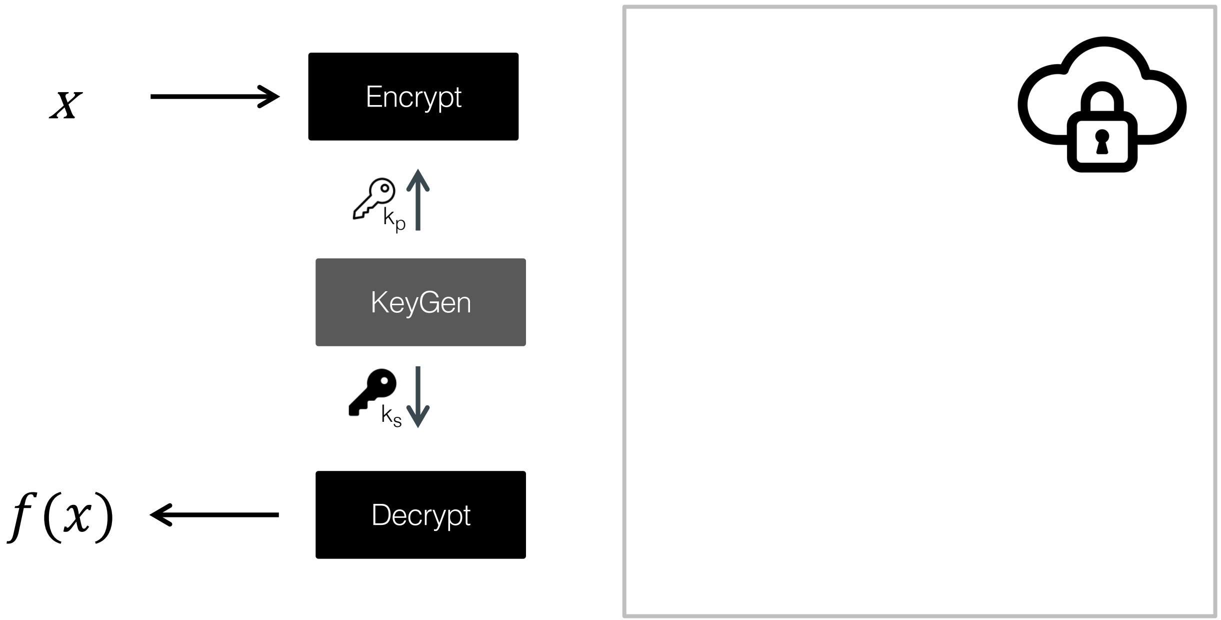
secure computation

Modern Cryptography



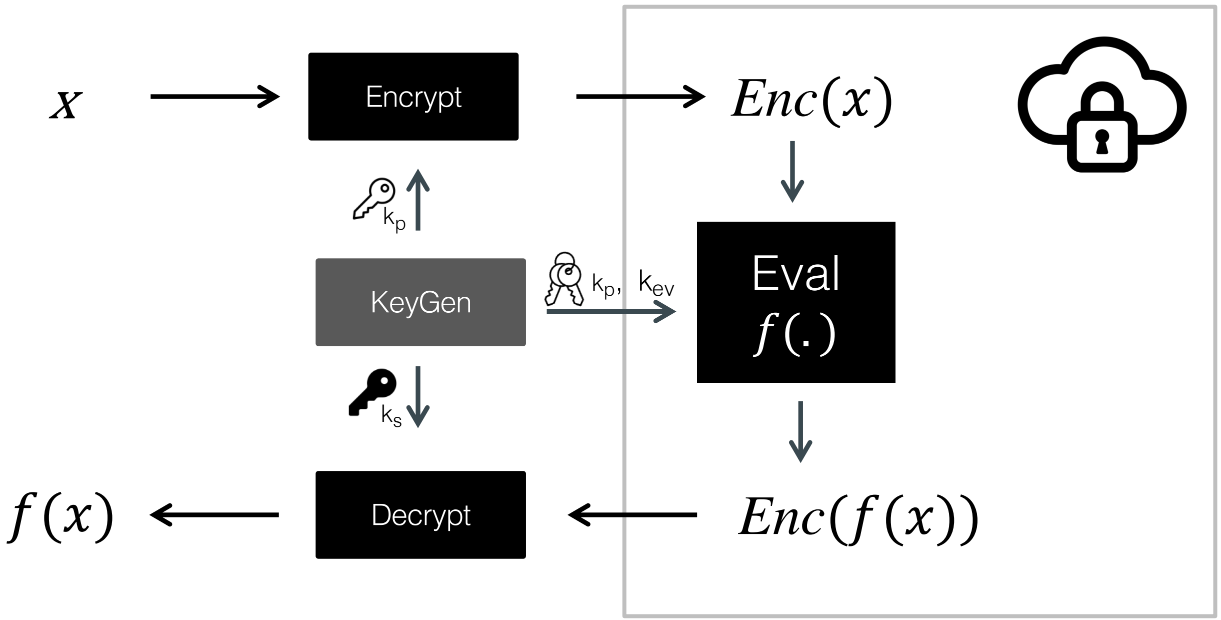
Fully Homomorphic Encryption

Enables computation on encrypted data



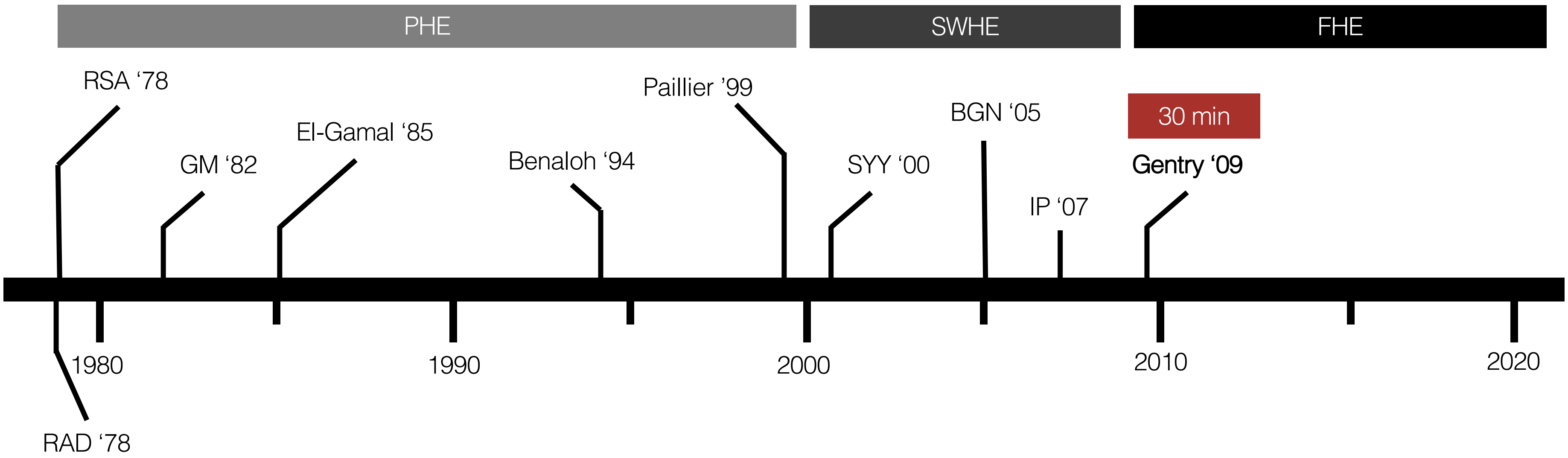
Fully Homomorphic Encryption

Enables computation on encrypted data

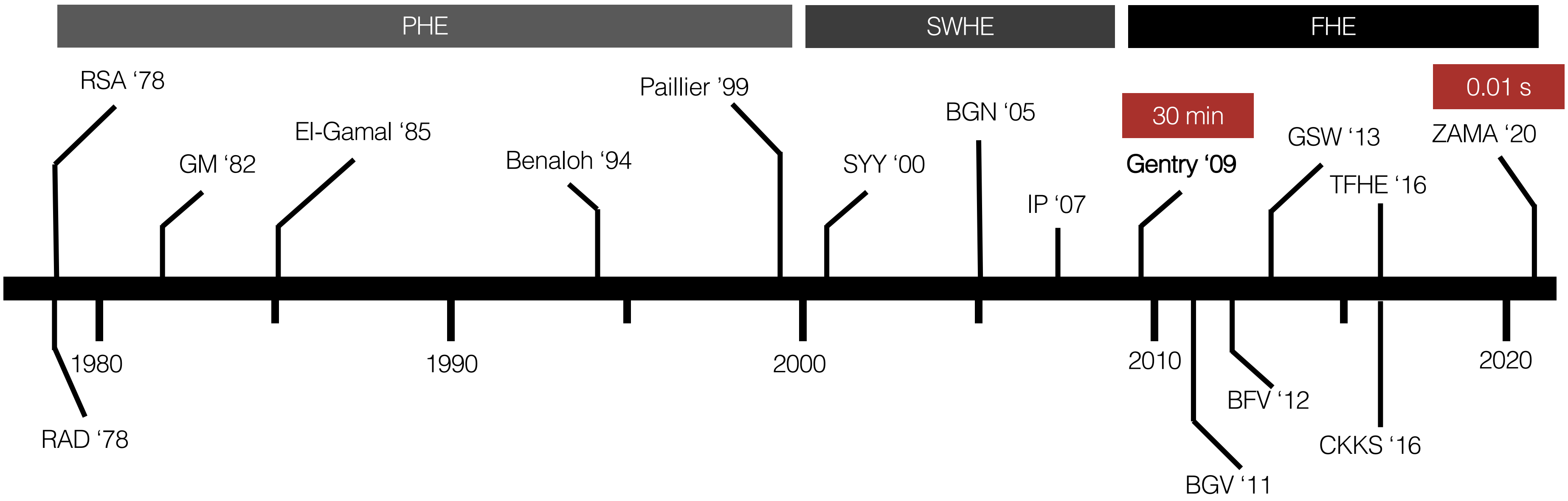


Delegate the processing of data without giving away access to it

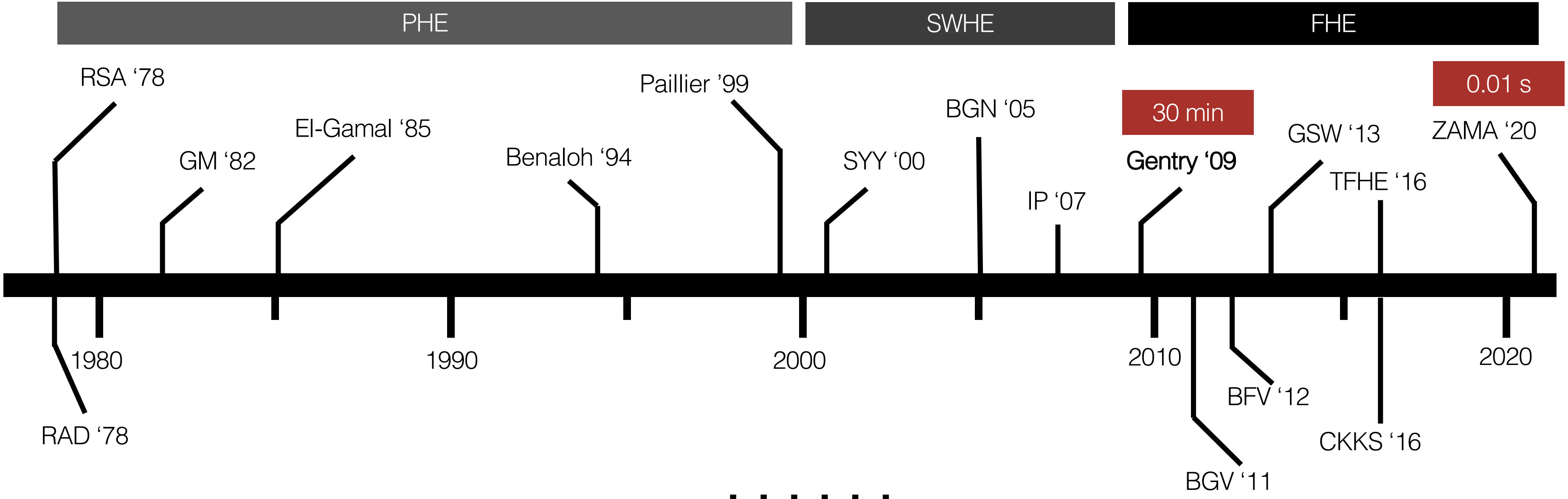
40 Years of FHE History



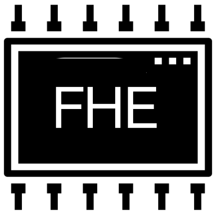
40 Years of FHE History



40 Years of FHE History



FHE Hardware Accelerators



FHE is not yet practical for many applications

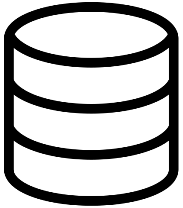
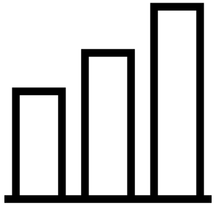
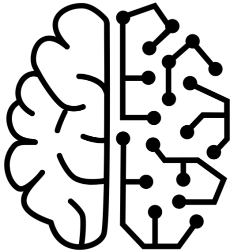
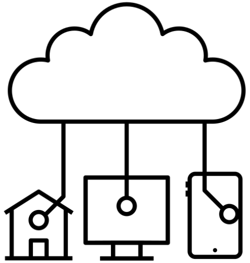
<u>but</u> will soon be practical for a wider set of applications…

Performance gap of modern applications ?

FHE is not yet practical for many applications
<u>but</u> will soon be practical for a wider set of applications…

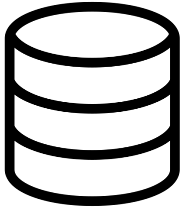
Facilitate FHE use in real world deployment ?

Building Encrypted Data Processing Systems

DBMS	Analytics	Machine Learning	Streaming	Internet of Things	
					…
CryptDB	Seabed	CryptoNets	TimeCrypt	Talos	
Blind Seer	Seanat	Helen	Zeph	Pilatus	
Arx	Conclave	RoFL	Waldo	Kryptein	
…	…	…	…	…	

Building Encrypted Data Processing Systems

DBMS



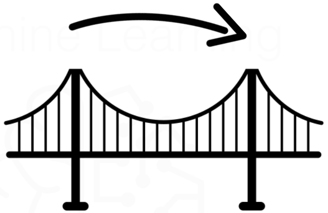
CryptDB

Blind Seer

Arx

...

Advanced Cryptography



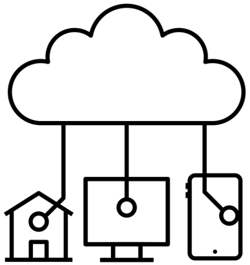
Modern Applications

...

...

...

Internet of Things



...

Talos

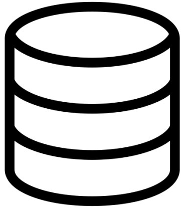
Pilatus

Kryptein

...

Building Encrypted Data Processing Systems

DBMS



CryptDB

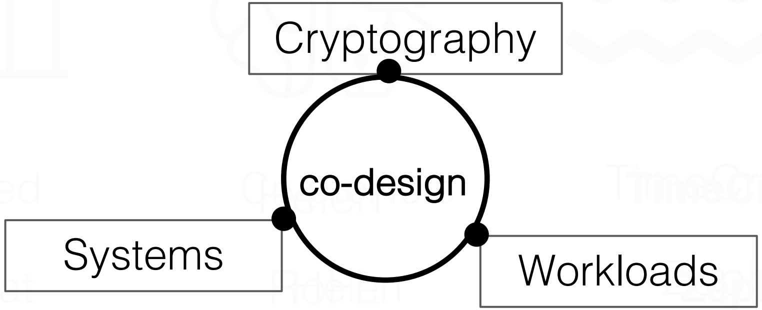
Blind Seer

Arx

…

Advanced Cryptography

Modern Applications

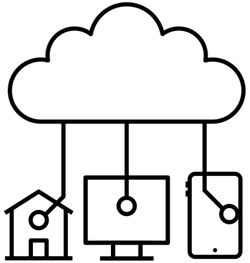


…

…

…

Internet of Things



…

Talos

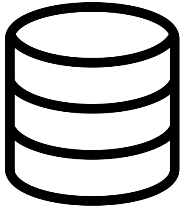
Pilatus

Kryptein

…

Building Encrypted Data Processing Systems

DBMS



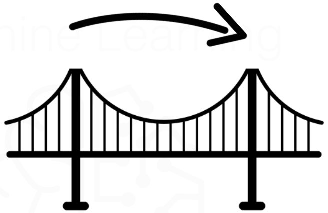
CryptDB

Blind Seer

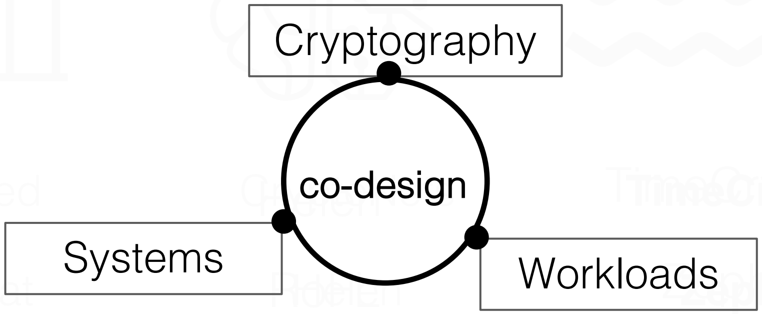
Arx

...

Advanced Cryptography



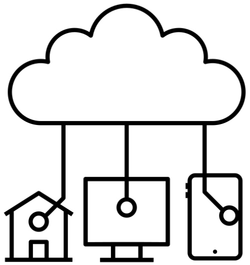
Modern Applications



- Functionality
- Performance
- Leakage

... ...

Internet of Things



...

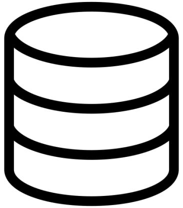
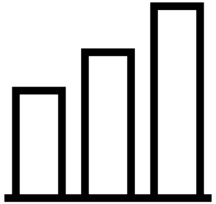
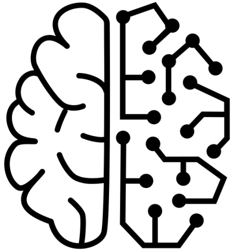
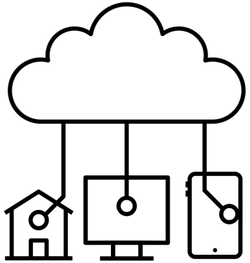
Talos

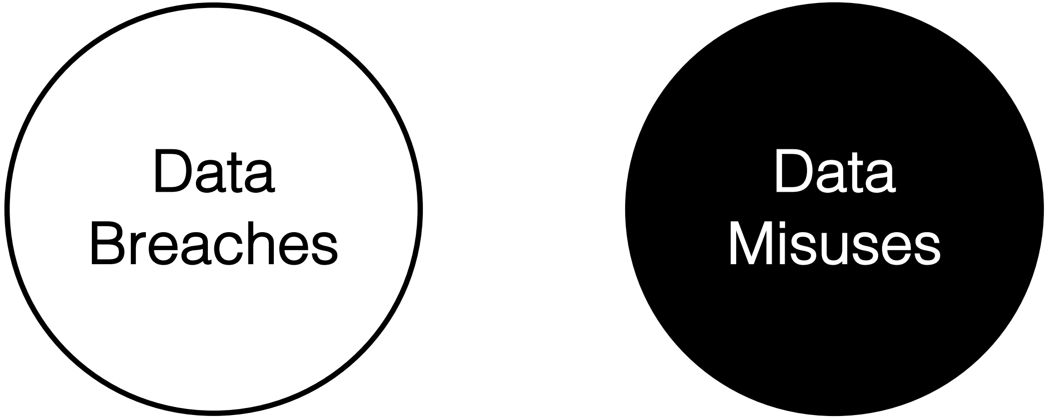
Pilatus

Kryptein

...

Building Encrypted Data Processing Systems

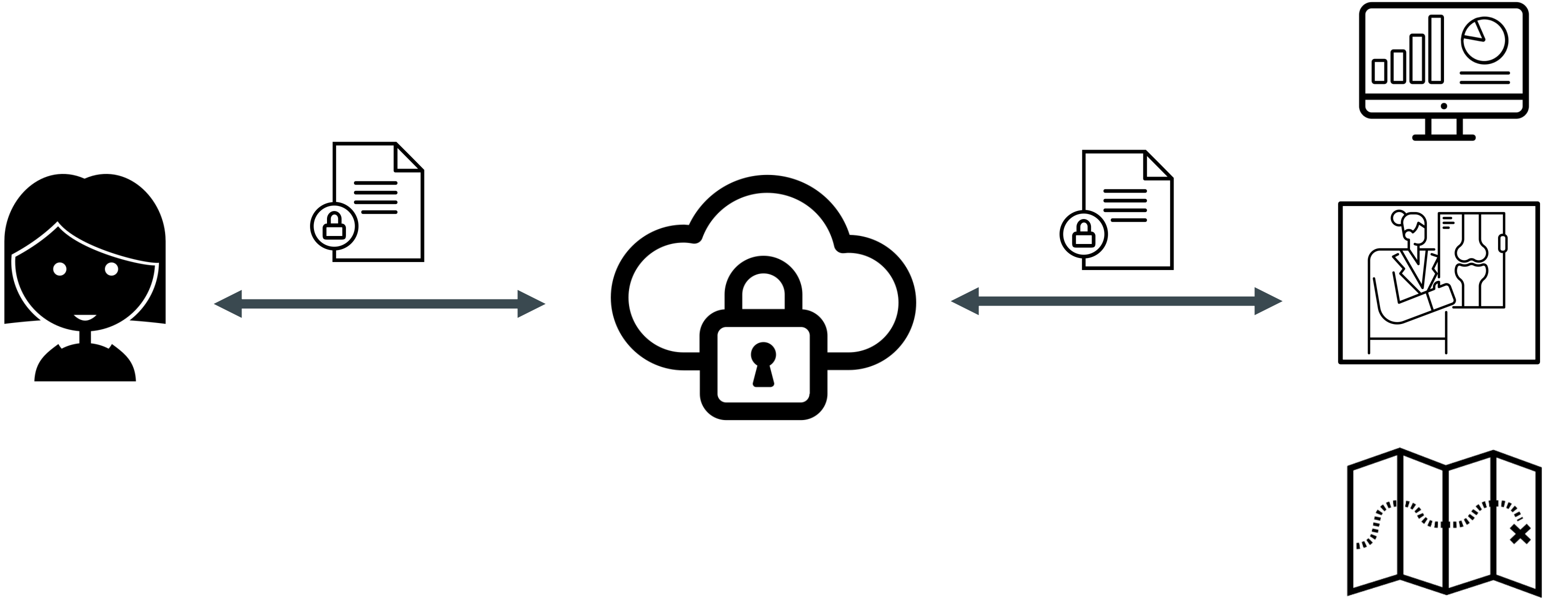
DBMS	Analytics	Machine Learning	Streaming	Internet of Things	
					…
CryptDB	Seabed	CryptoNets	**TimeCrypt**	**Talos**	
Blind Seer	Senat	Helen	**Zeph**	**Pilatus**	
Arx	Conclave	**RoFL**	Waldo	Kryptein	
…	…	…	…	…	


Data
Breaches
Data
Misuses

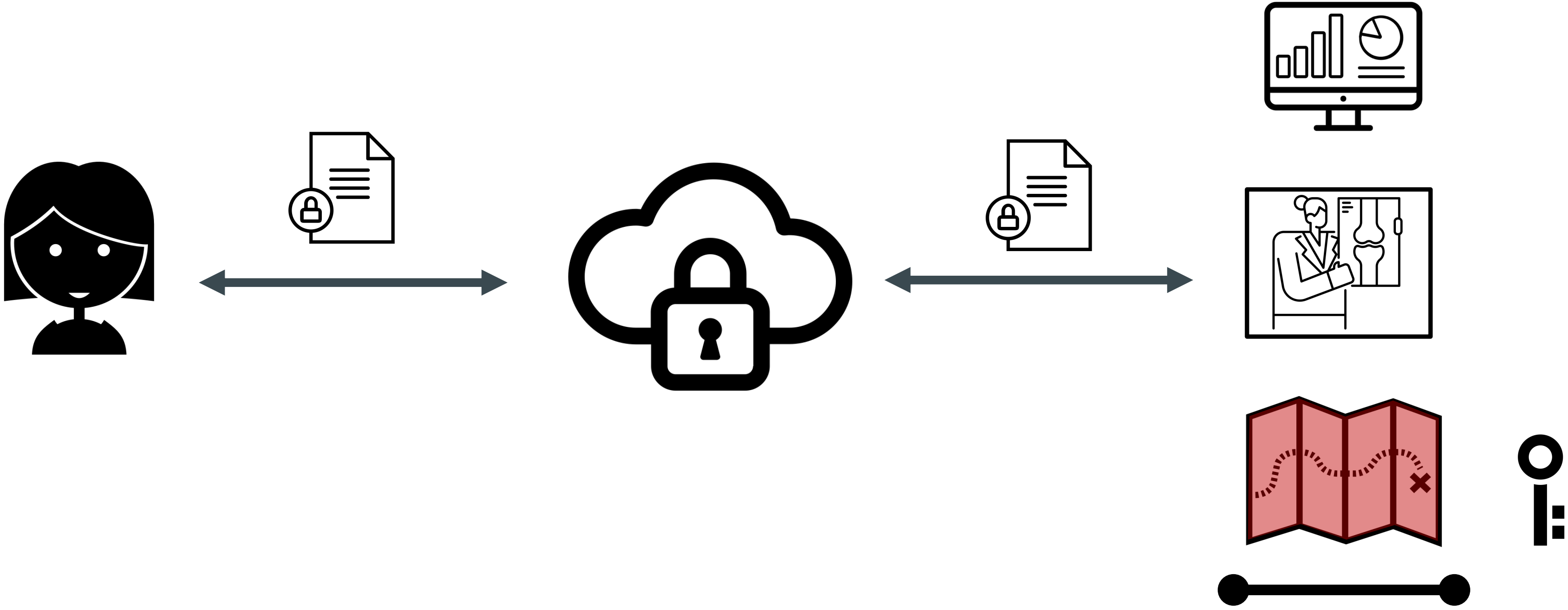
End-to-End Encrypted Systems

End-to-End Encrypted Systems

End-to-End Encrypted Systems



End-to-End Encrypted Systems



Privacy protection means more than securing the data …

Data Misuse

use of data for purposes that
the user did not agree to

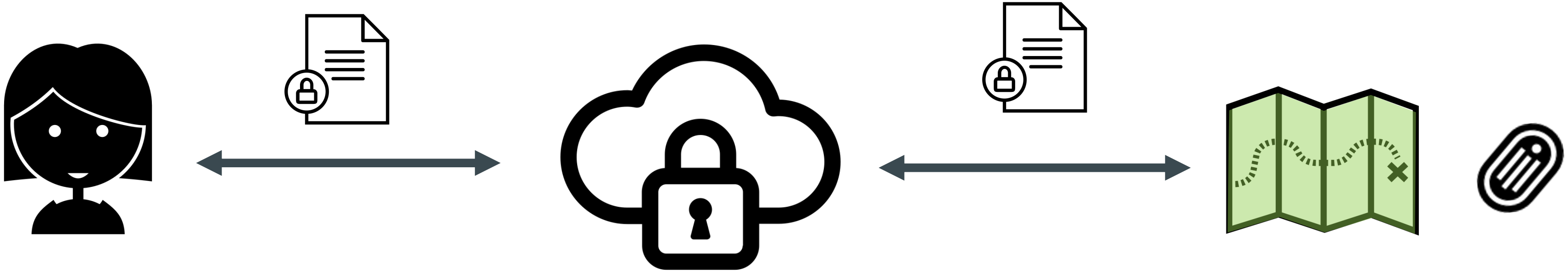
End-to-End Encrypted Systems

Fundamental issue: unrestricted views of the data

End-to-End Encrypted Systems → End-to-End Privacy

privacy enhanced

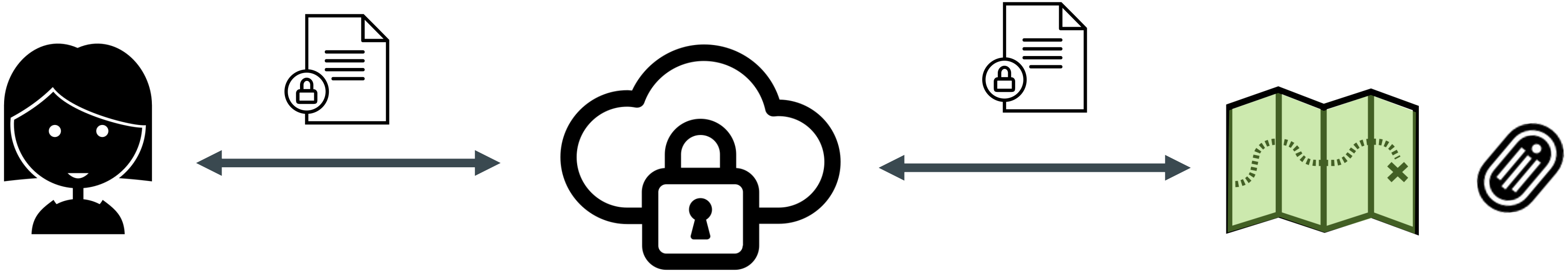
Fundamental issue: ~~unrestricted~~ views of the data



End-to-End Encrypted Systems → End-to-End Privacy

privacy enhanced

Fundamental issue: ~~unrestricted~~ views of the data



Data Minimization

Purpose Limitation

Goal:

End-to-End Security → End-to-End Privacy

Goal:

End-to-End Security → End-to-End Privacy

data confidentiality

unauthorized parties

Goal:

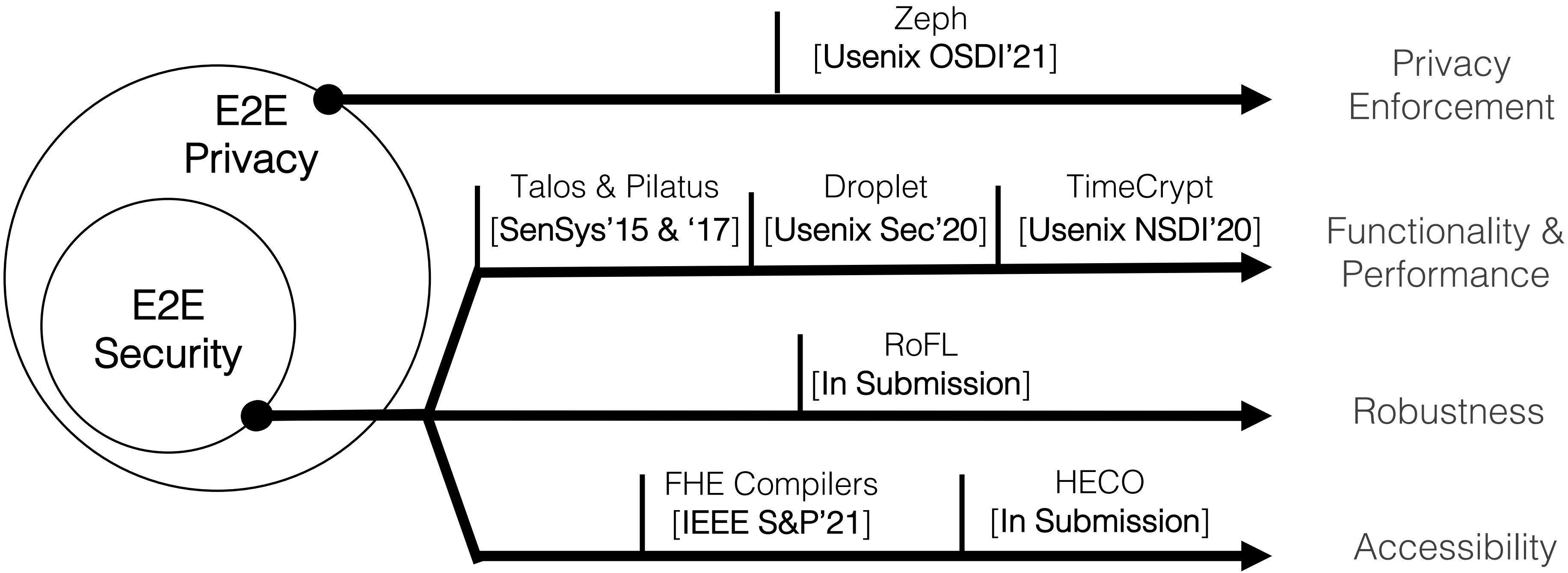
End-to-End Security → End-to-End Privacy

data confidentiality
unauthorized parties

&

strong privacy guarantees
authorized parties

My Research: Building practical systems that use cryptography to empower users and preserve their privacy & tools to democratize cryptography



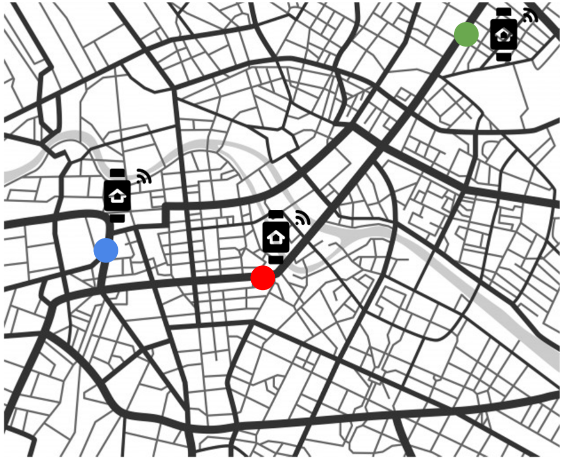
Zeph (Usenix OSDI '21)

User-centric Model
for Privacy

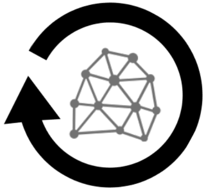
Cryptographically
Enforces Privacy

One of Many Scenarios

"Raw Location Data"



Privacy Transformation



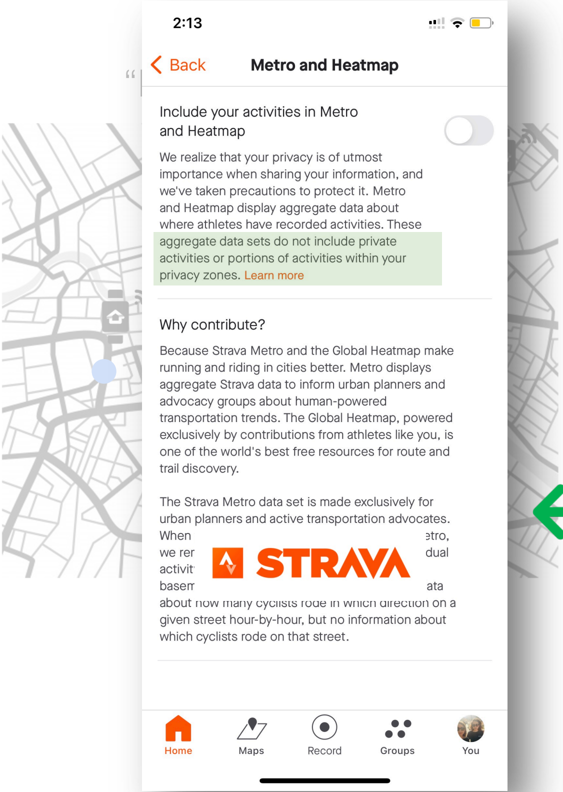
Policy

- Temporal
- Spatial
- Population

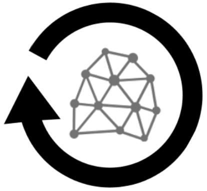
"Daily popular running tracks"

Day 6

One of Many Scenarios



Privacy Transformation



Policy

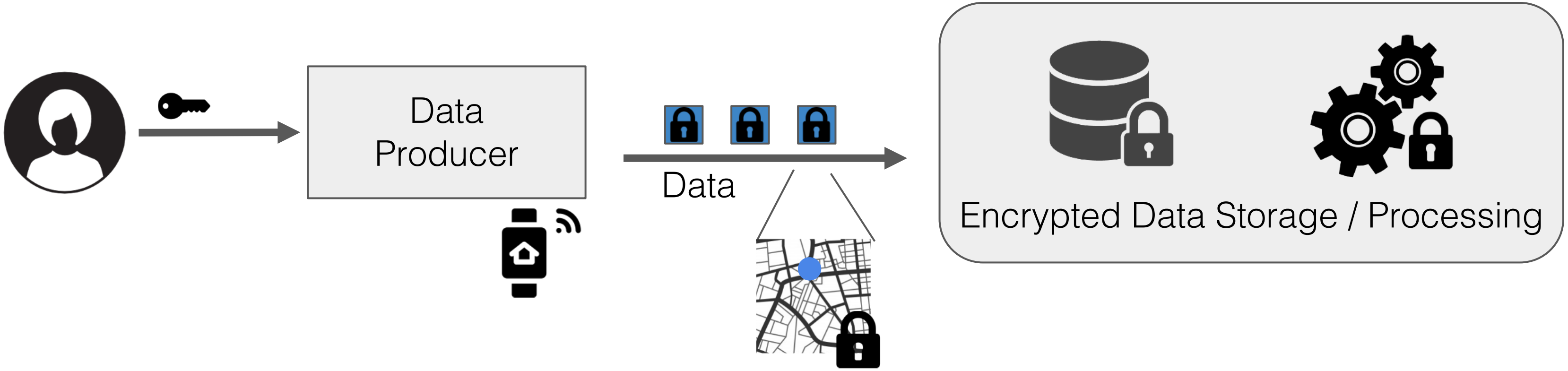
- Temporal
- Spatial
- Population

“Daily popular running tracks”

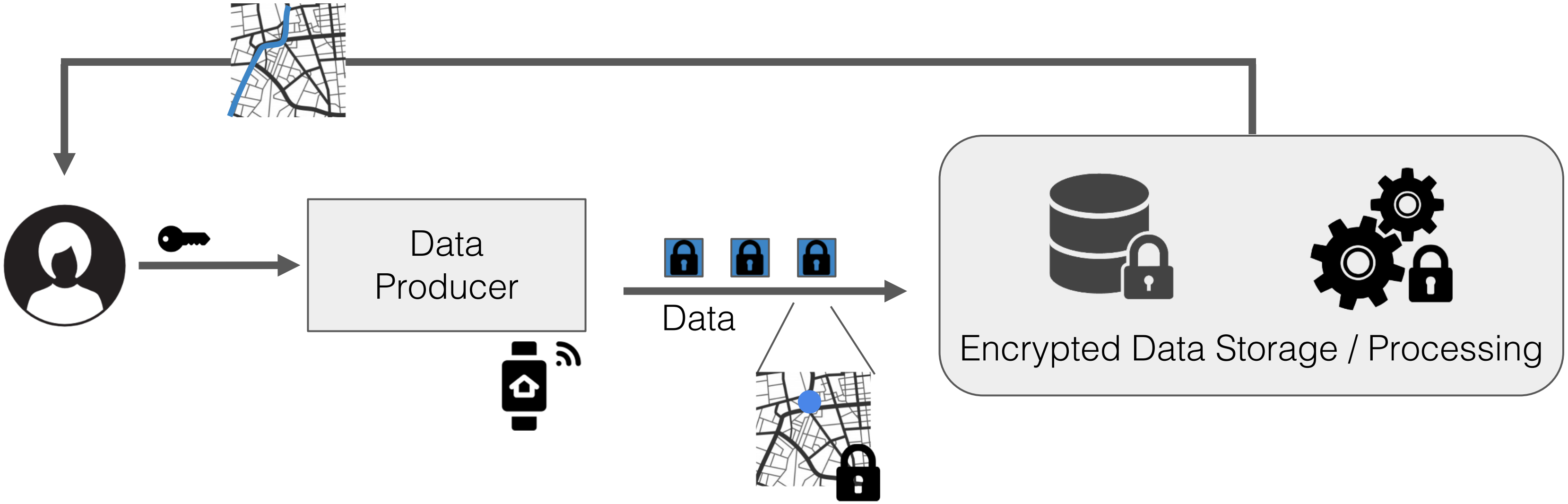
Day 6

✗ End-to-End Security ✗ Enforcement

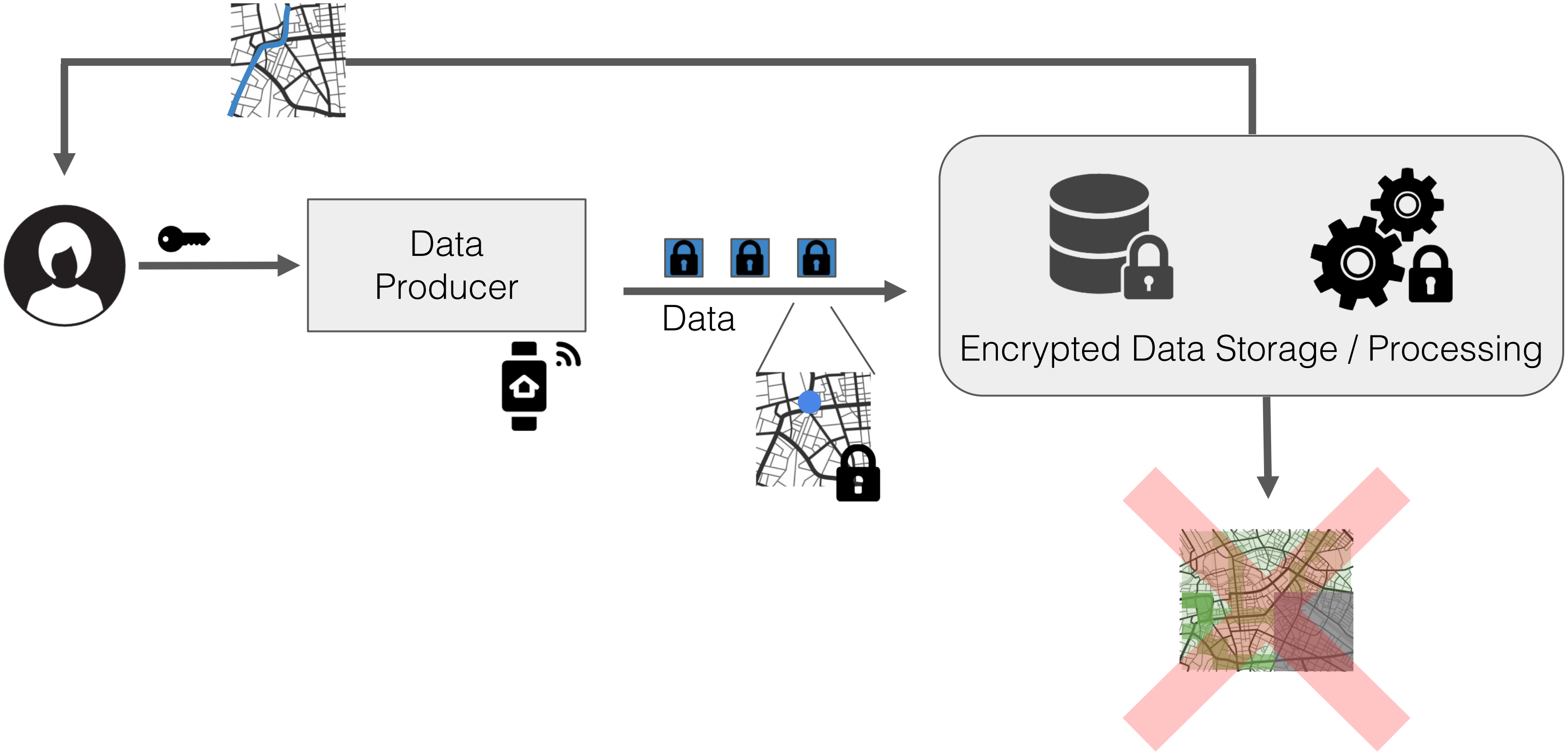
Existing End-to-End Encrypted Streaming Pipeline



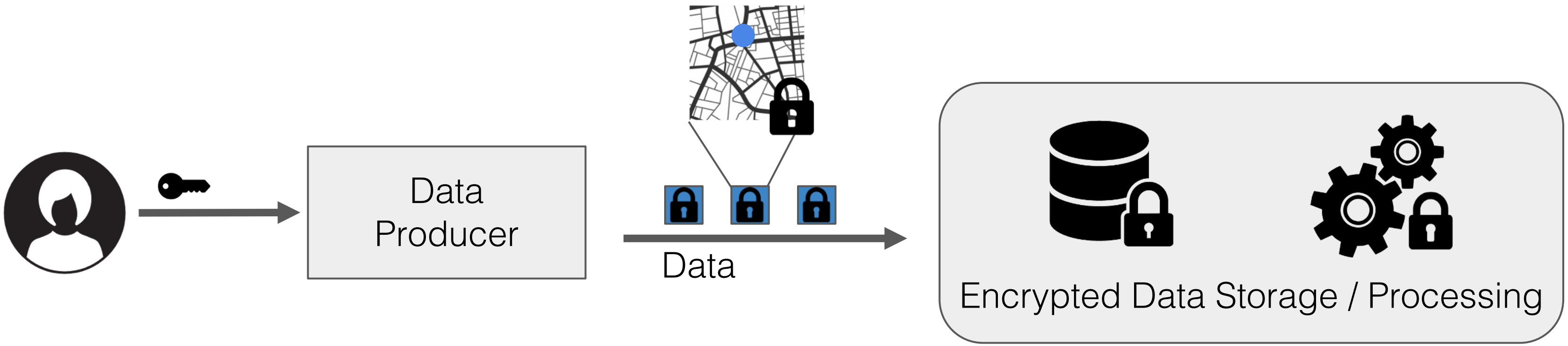
Existing End-to-End Encrypted Streaming Pipeline



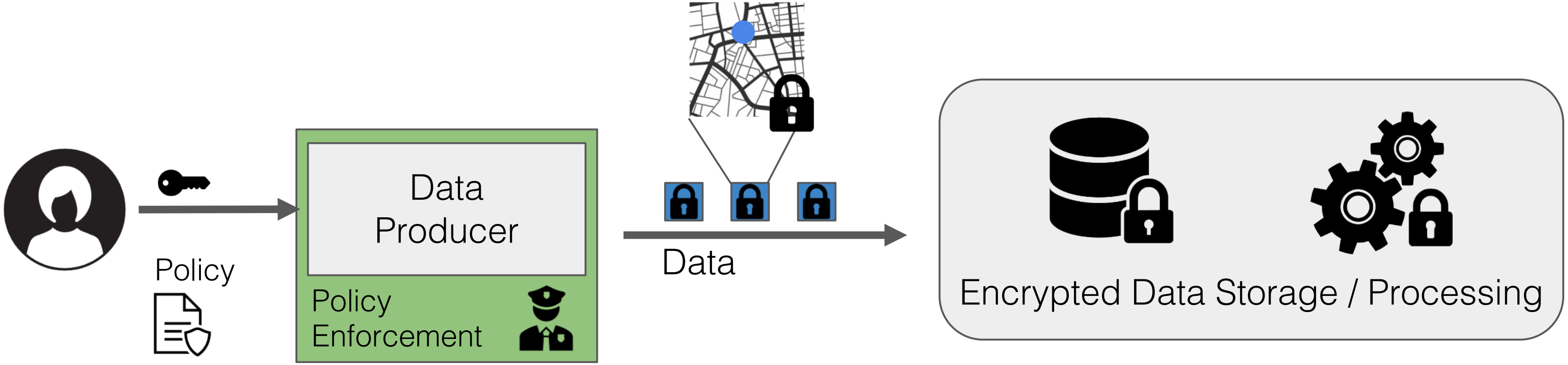
Existing End-to-End Encrypted Streaming Pipeline



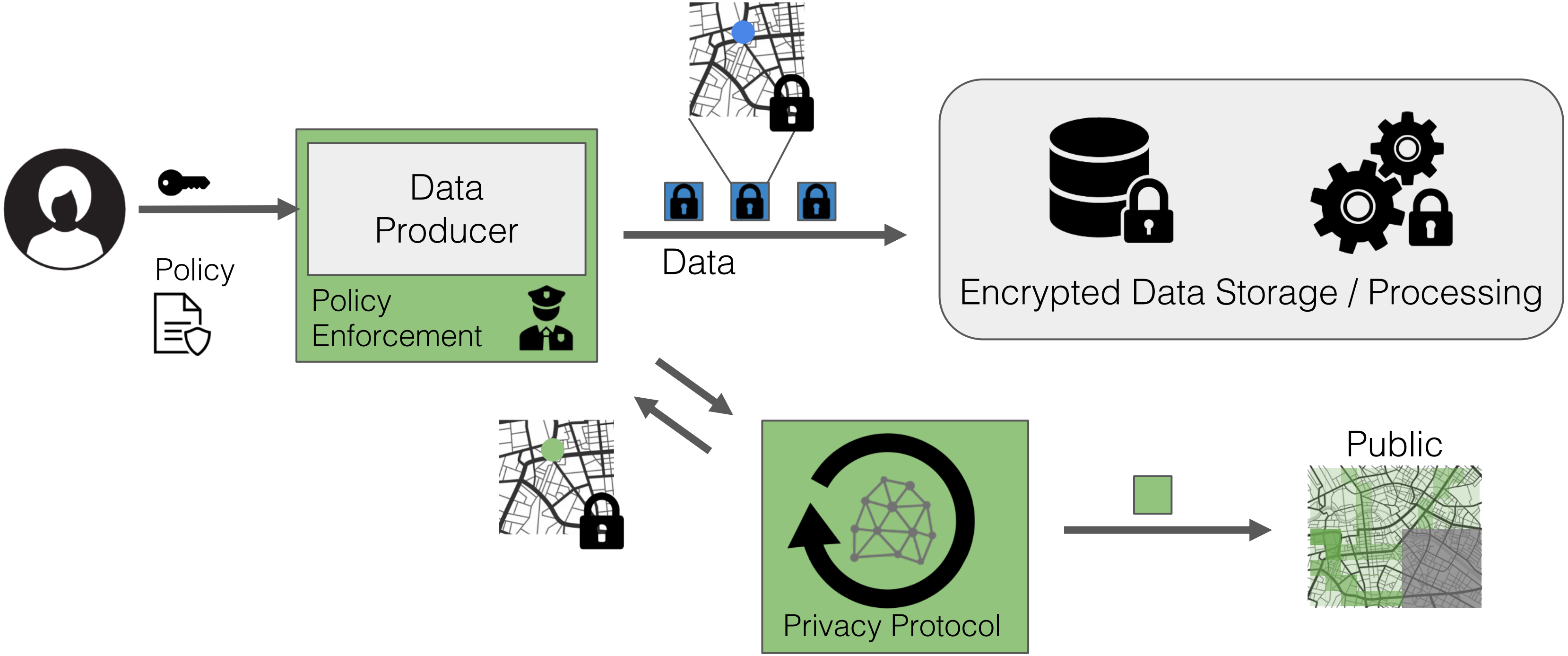
Integrate Privacy Controls into Existing Pipelines



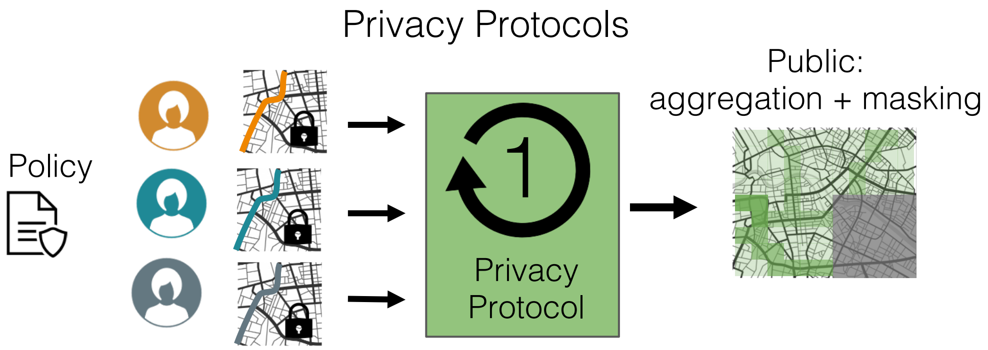
Integrate Privacy Controls into Existing Pipelines



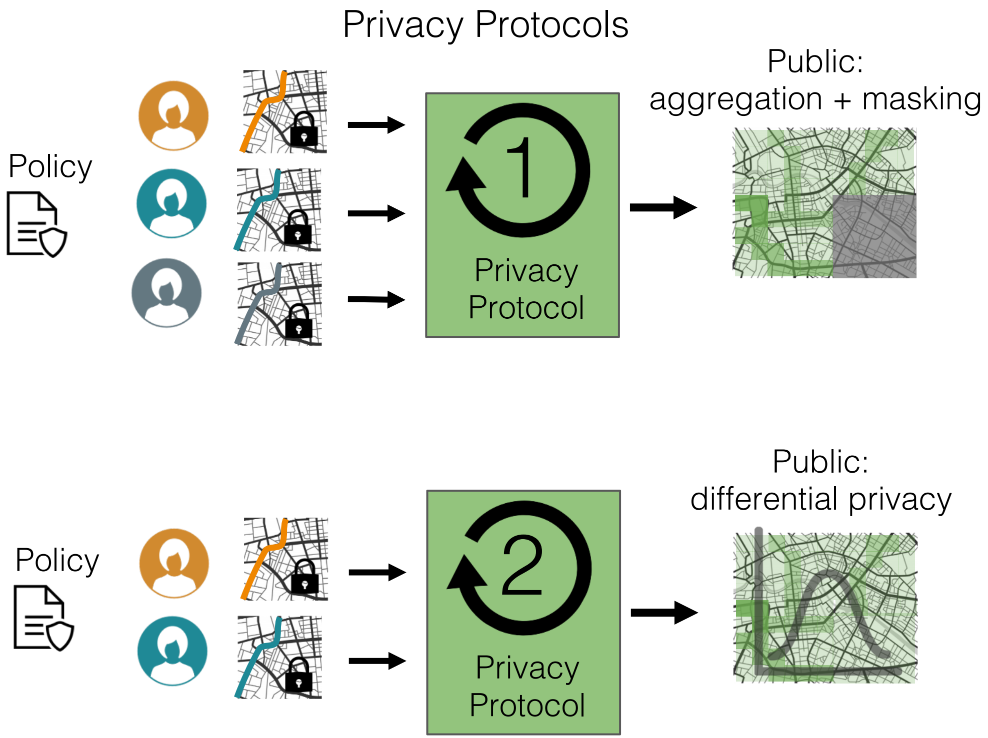
Integrate Privacy Controls into Existing Pipelines

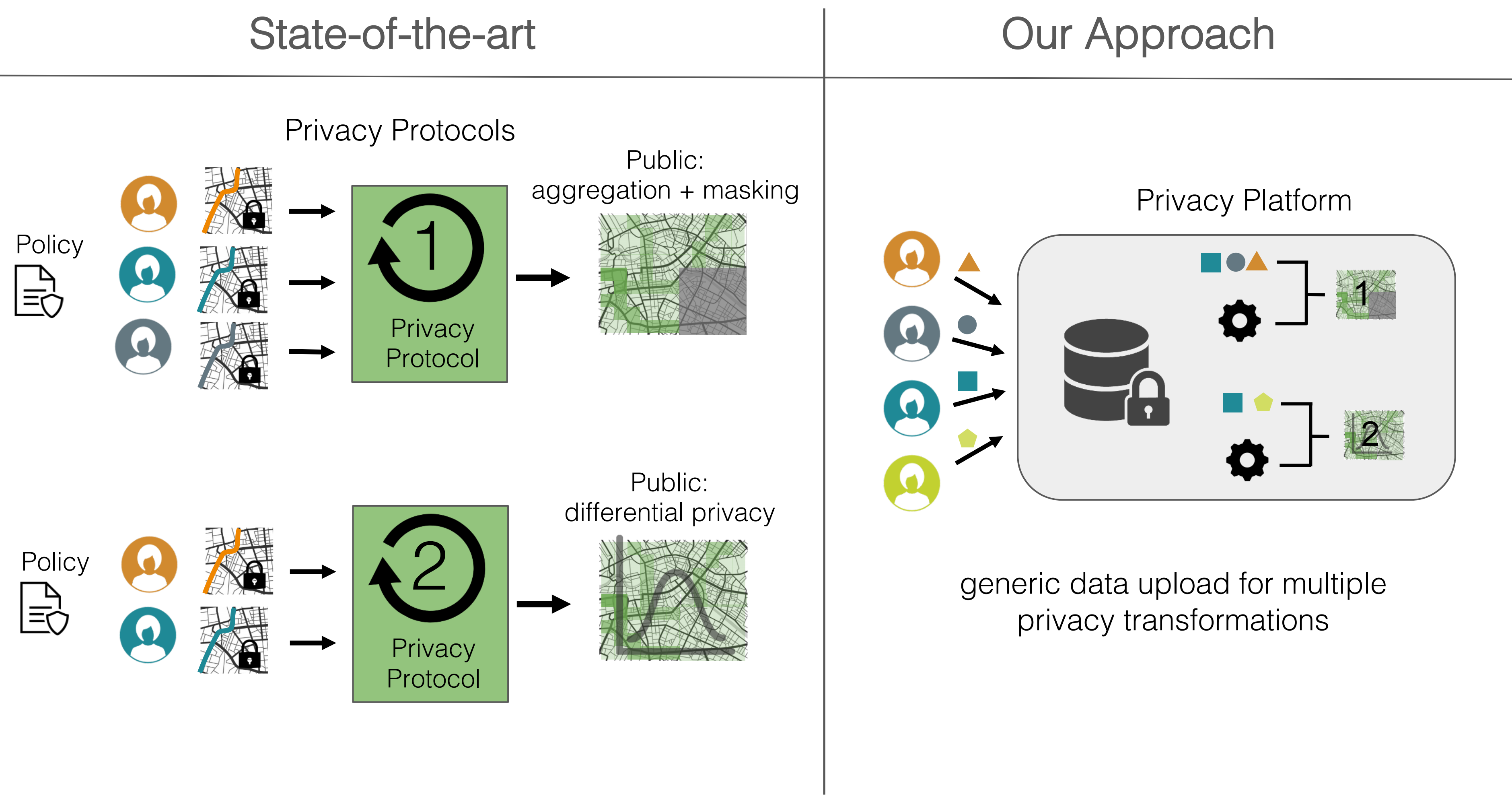


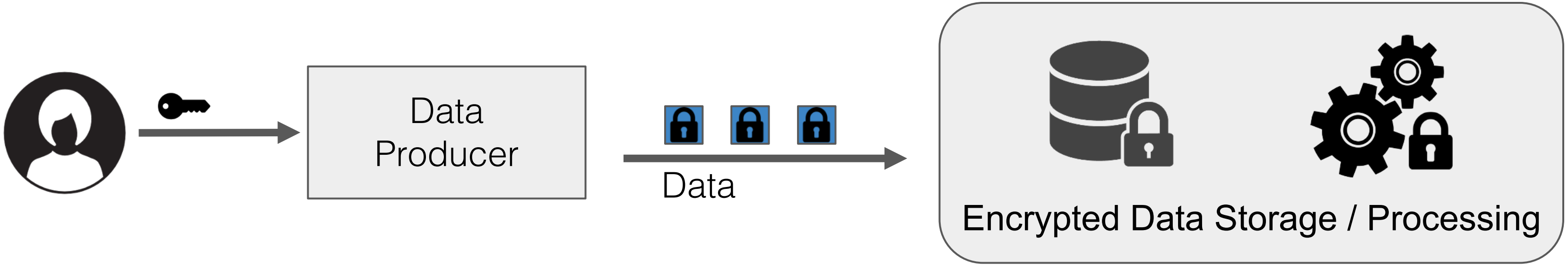
State-of-the-art



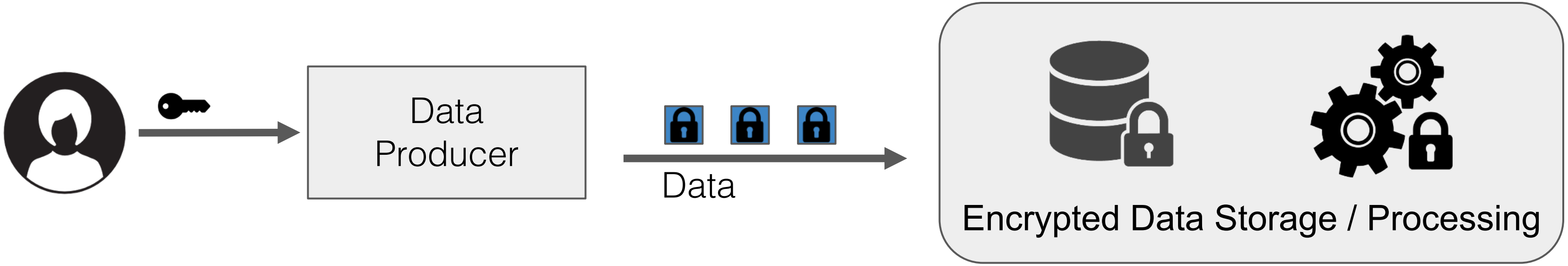
State-of-the-art

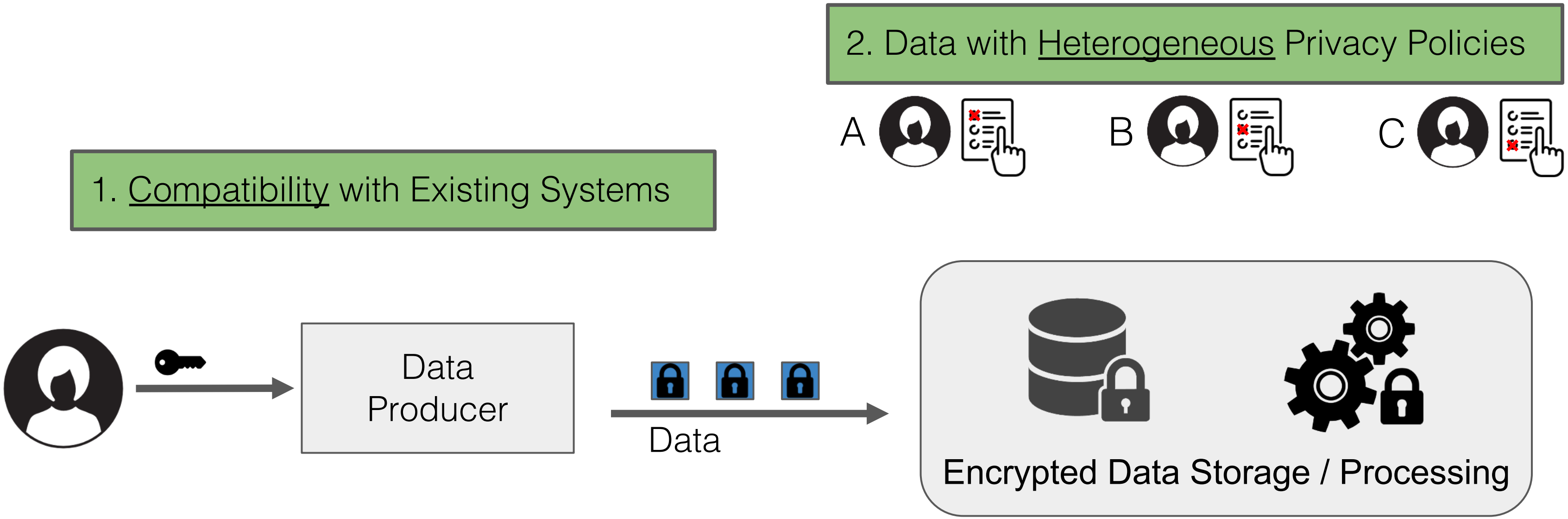



State-of-the-art
Our Approach
Privacy Protocols
Policy
Public:
aggregation + masking
1
Privacy Protocol
Policy
Public:
differential privacy
2
Privacy Protocol
Privacy Platform
1
2
generic data upload for multiple privacy transformations

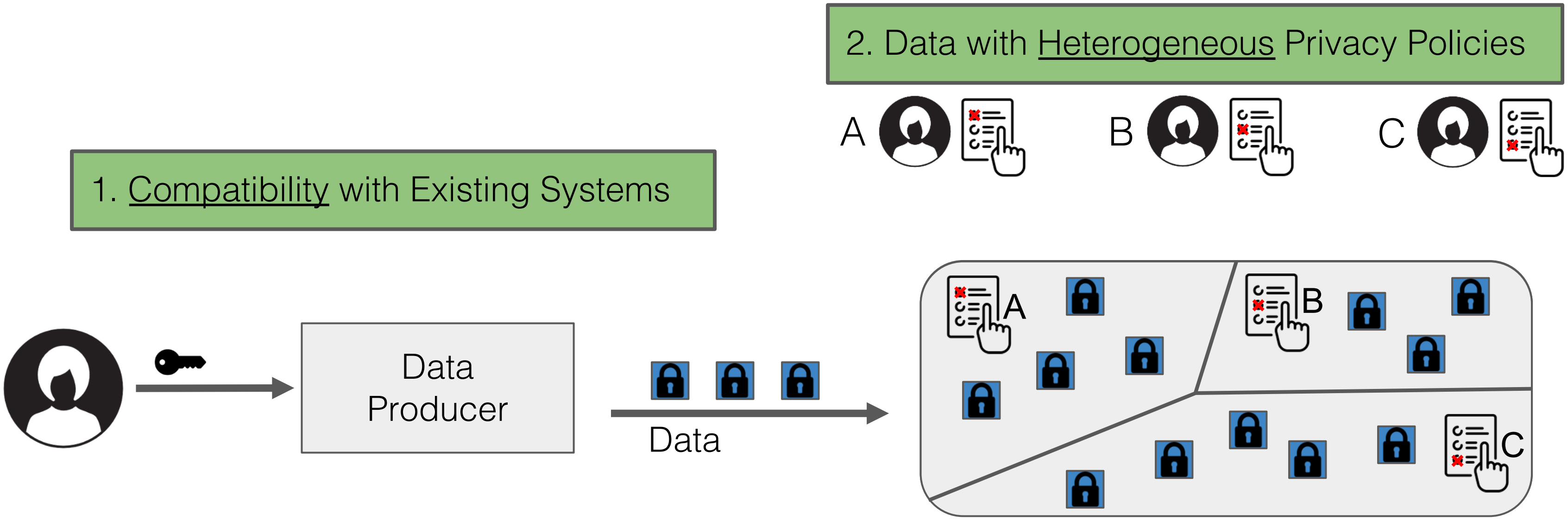

Data
Producer
Data
Encrypted Data Storage / Processing

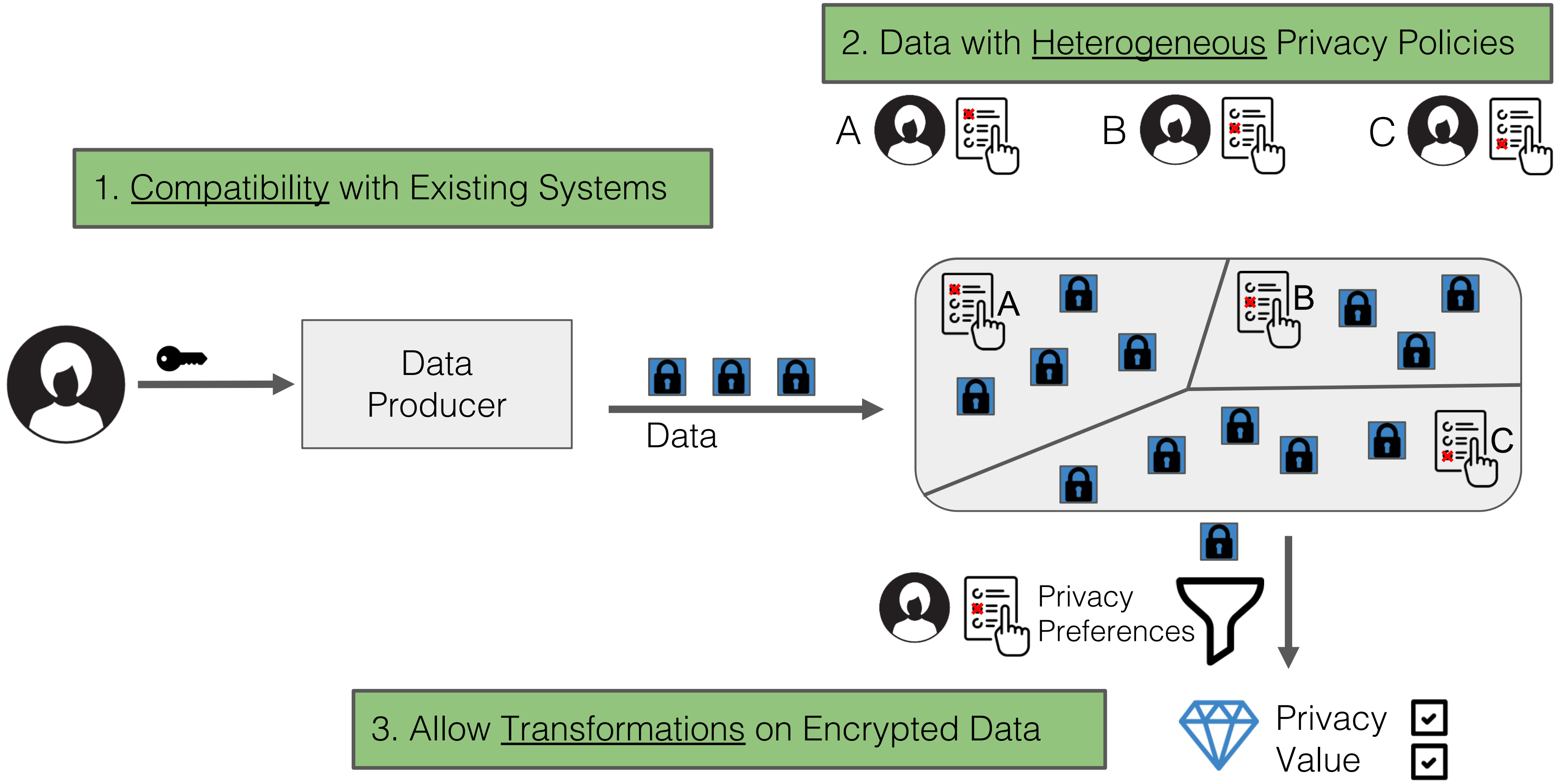
1. Compatibility with Existing Systems





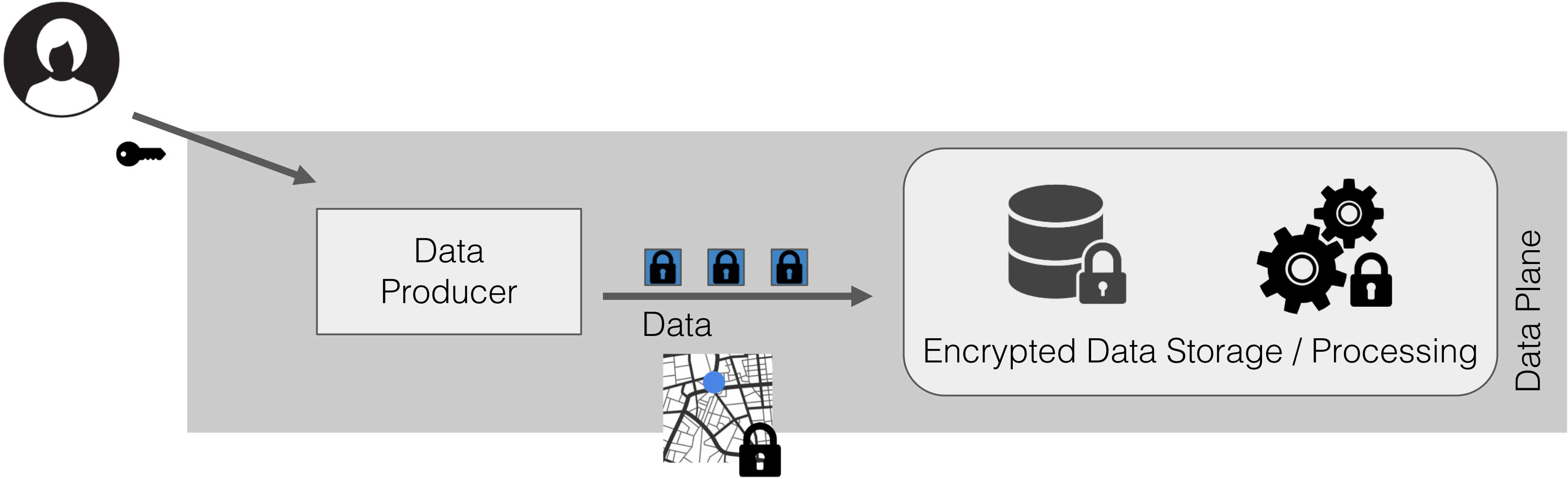
2. Data with Heterogeneous Privacy Policies
A
B
C
1. Compatibility with Existing Systems
Data Producer
Data
Encrypted Data Storage / Processing


2. Data with Heterogeneous Privacy Policies
A
B
C
1. Compatibility with Existing Systems
Data
Producer
Data
A
B
C

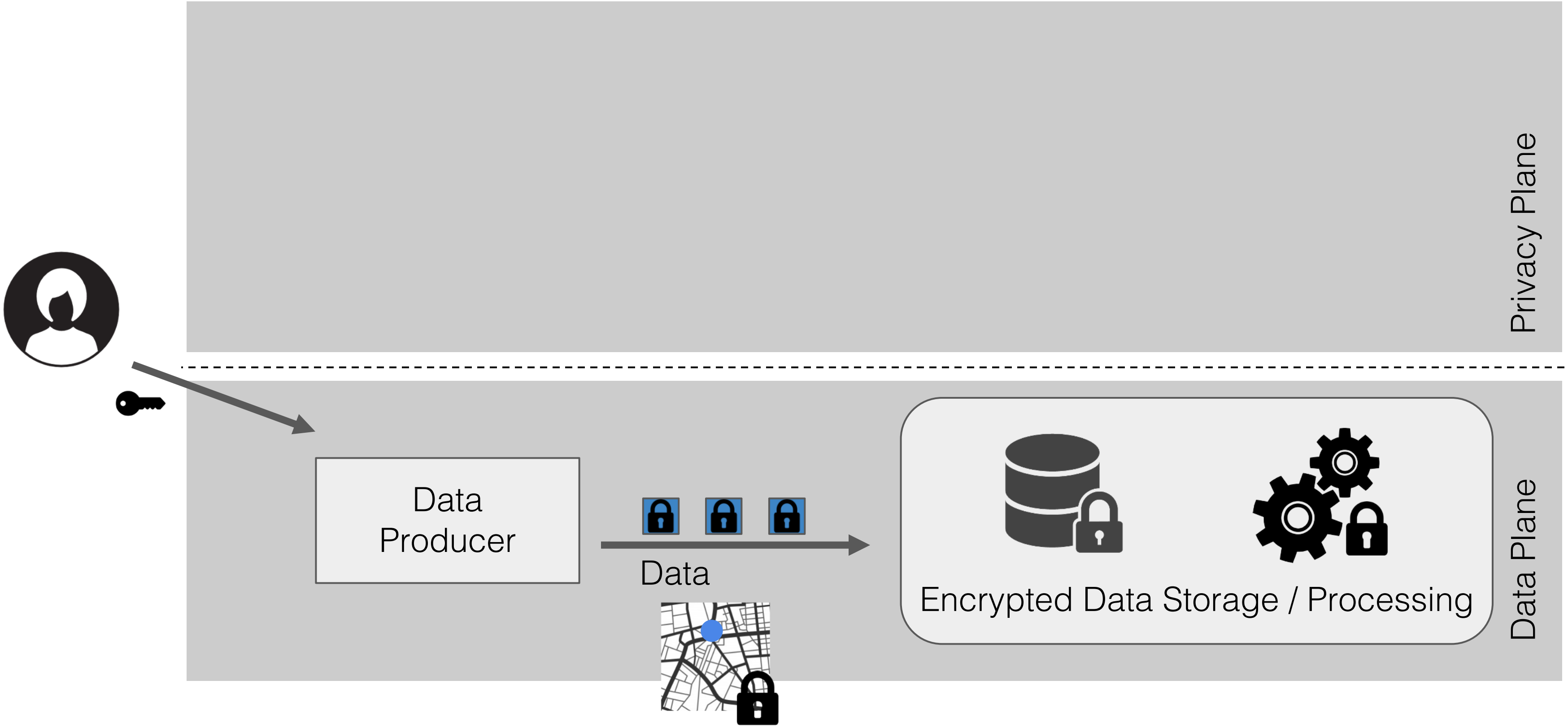


2. Data with Heterogeneous Privacy Policies
A
B
C
1. Compatibility with Existing Systems
Data Producer
Data
A
B
C
Privacy Preferences
Privacy
Value
3. Allow Transformations on Encrypted Data

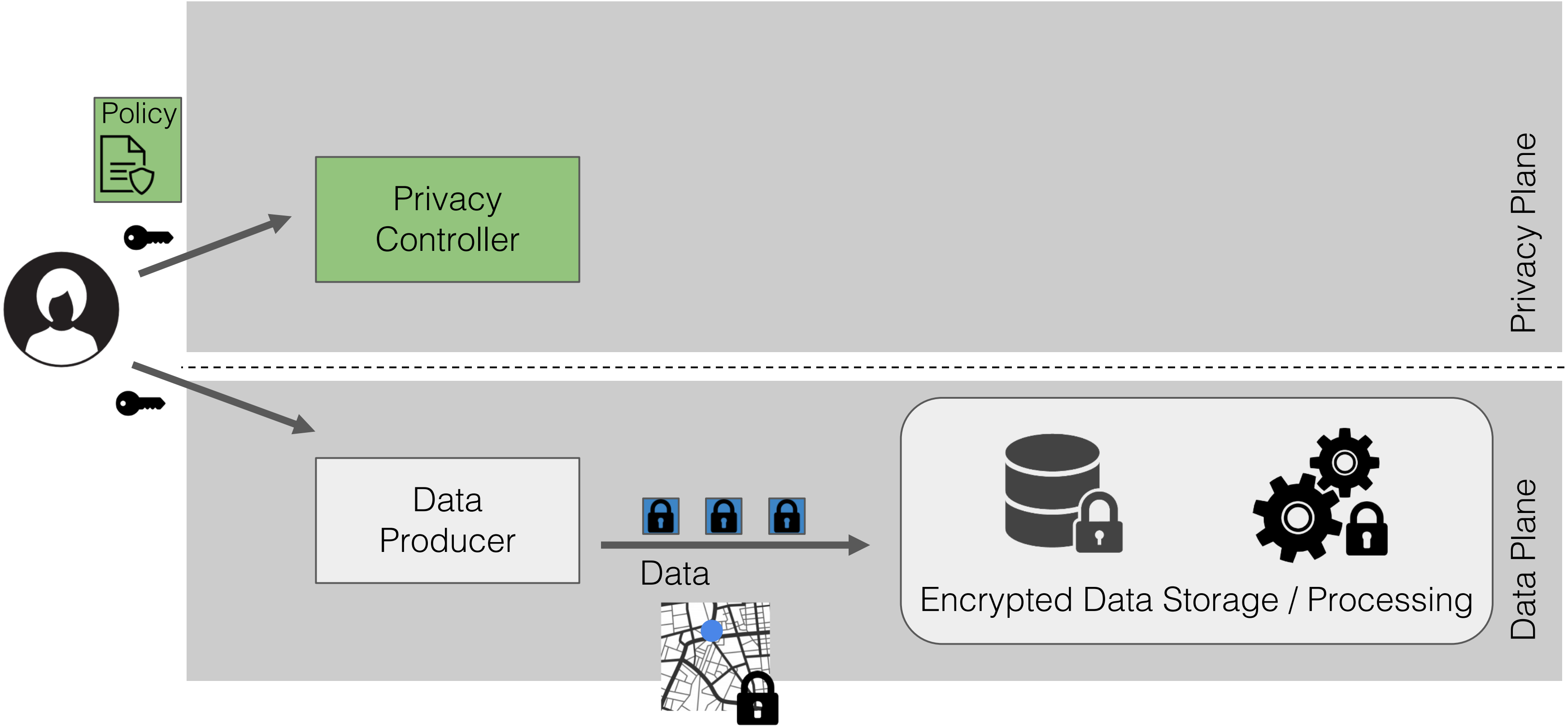
Zeph's Approach



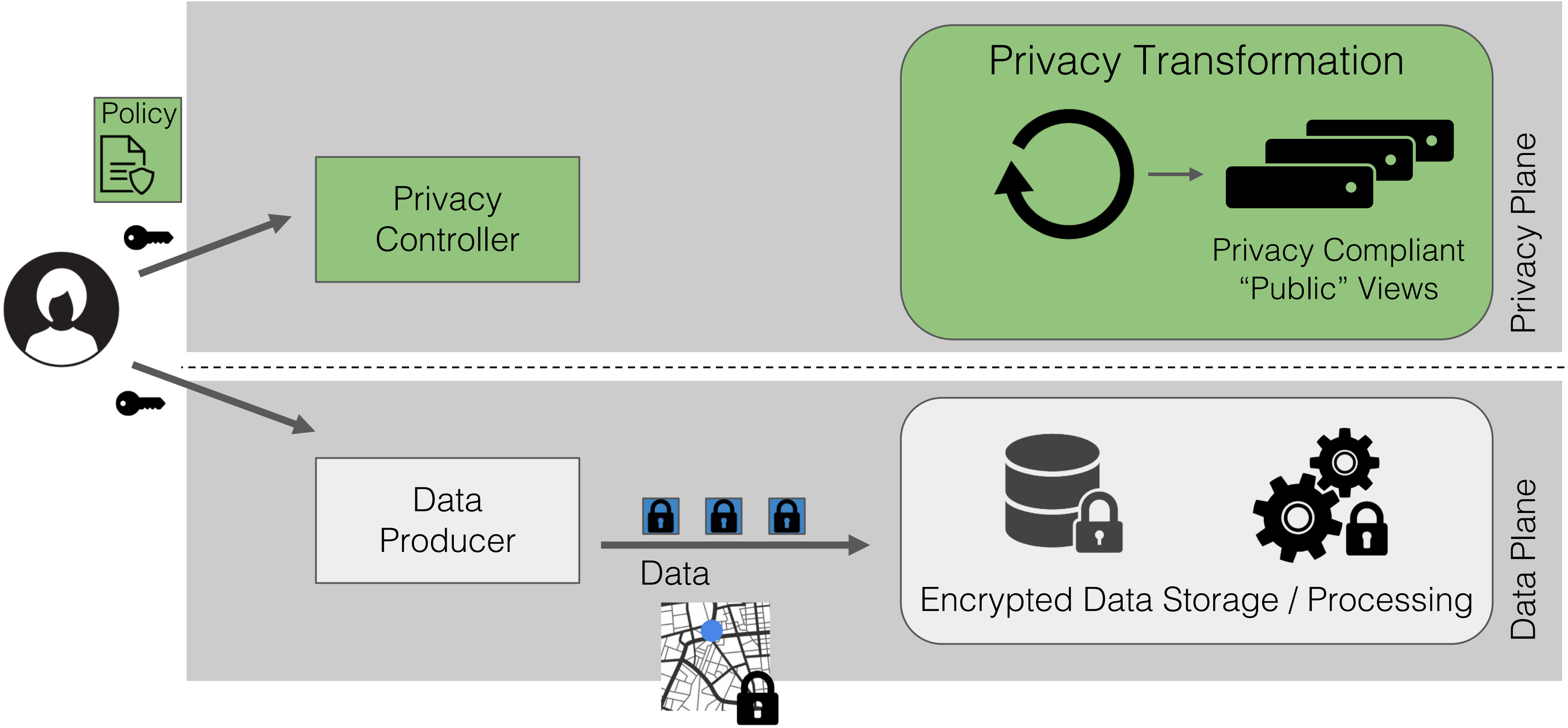
Zeph's Approach



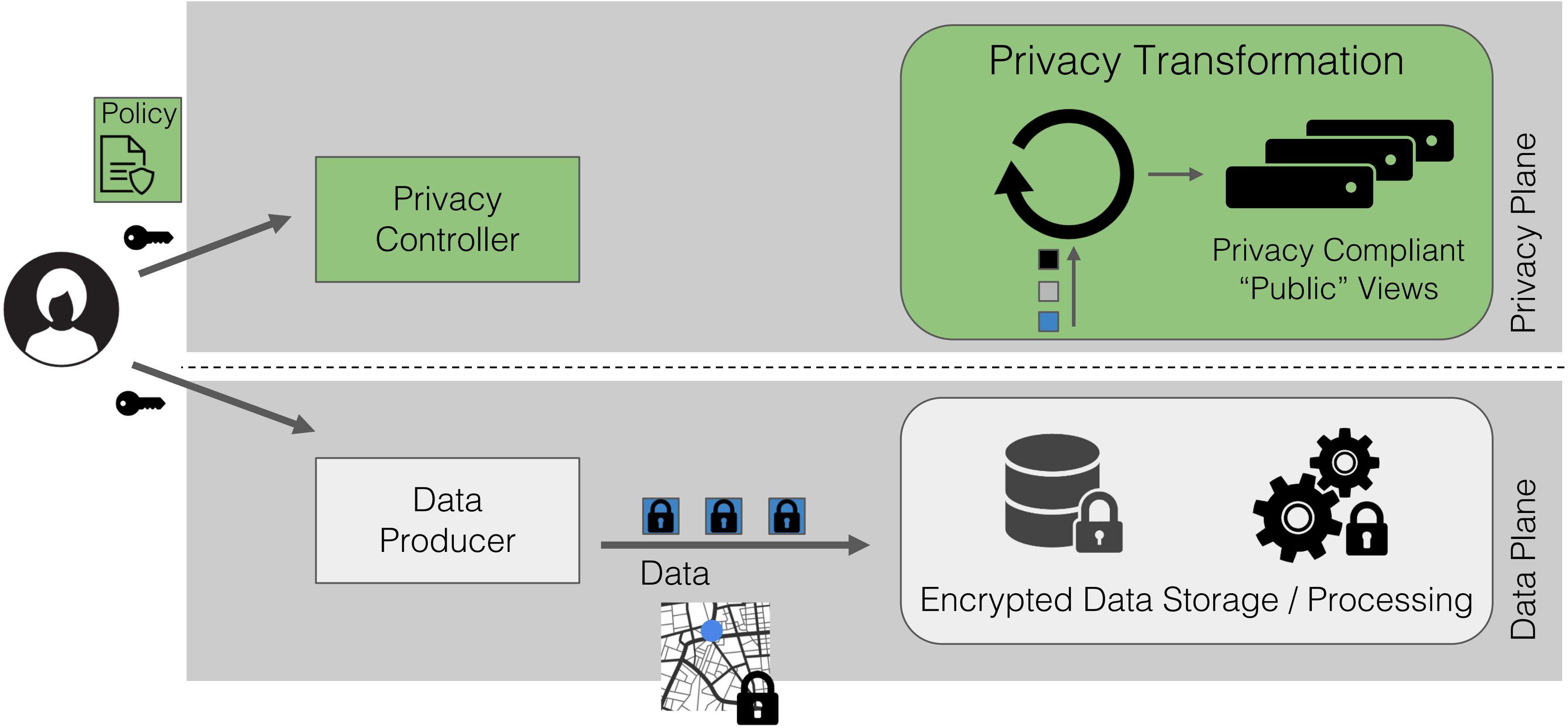
Zeph's Approach



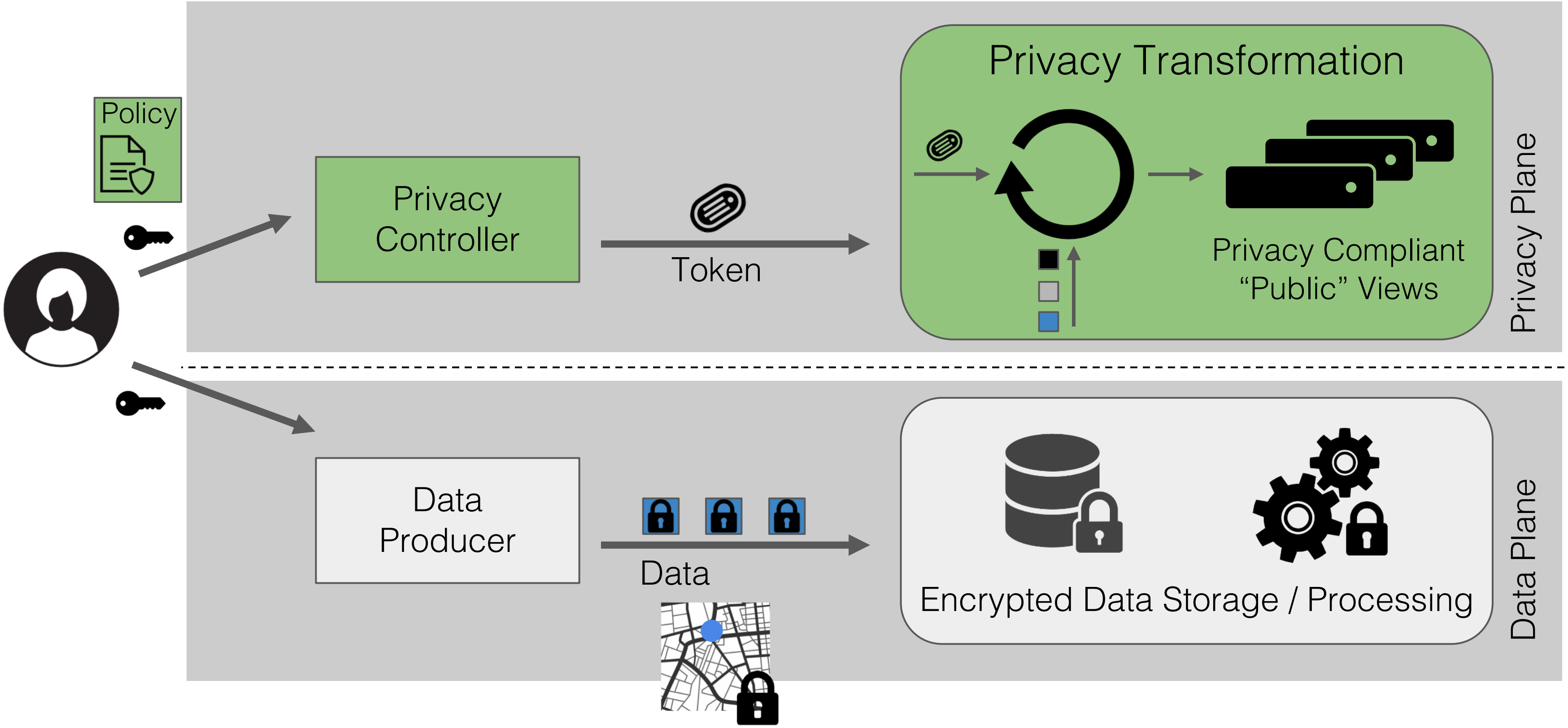
Zeph's Approach



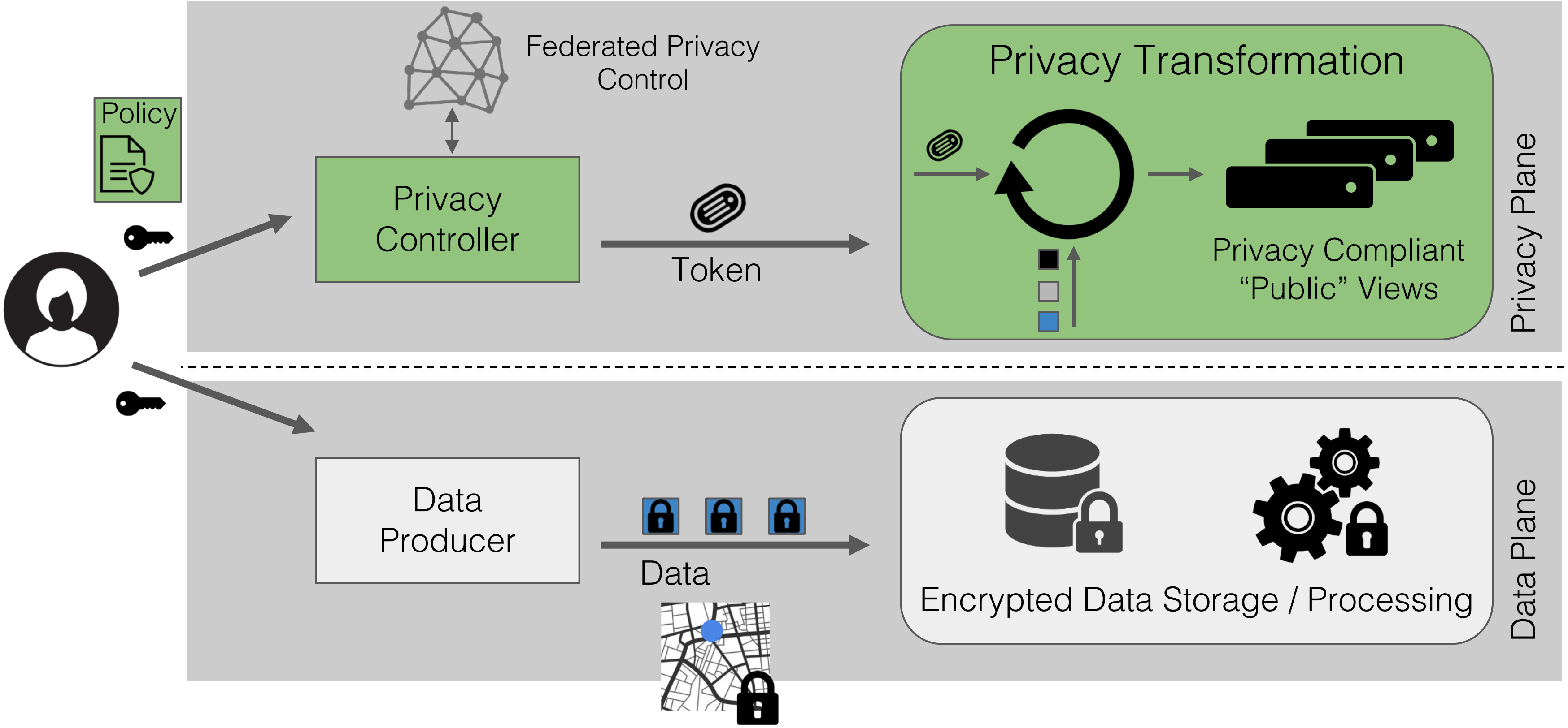
Zeph's Approach



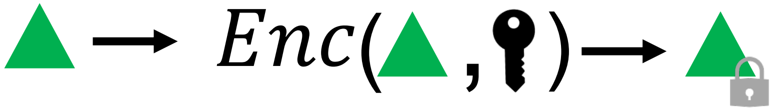
Zeph's Approach



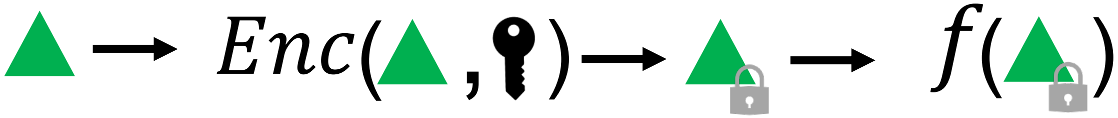
Zeph's Approach



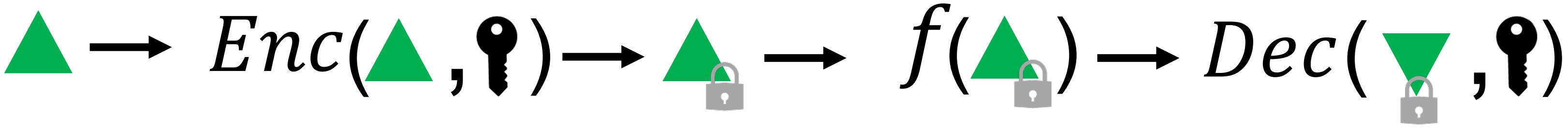
Cryptographic Privacy Tokens



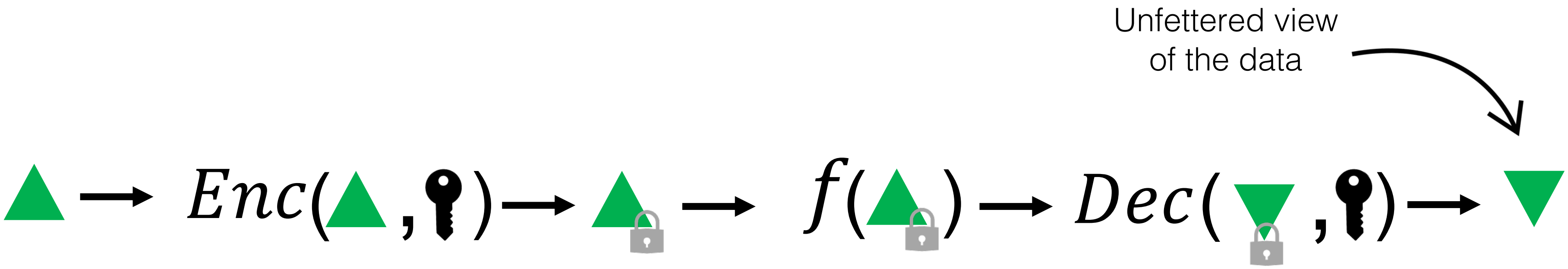
Cryptographic Privacy Tokens



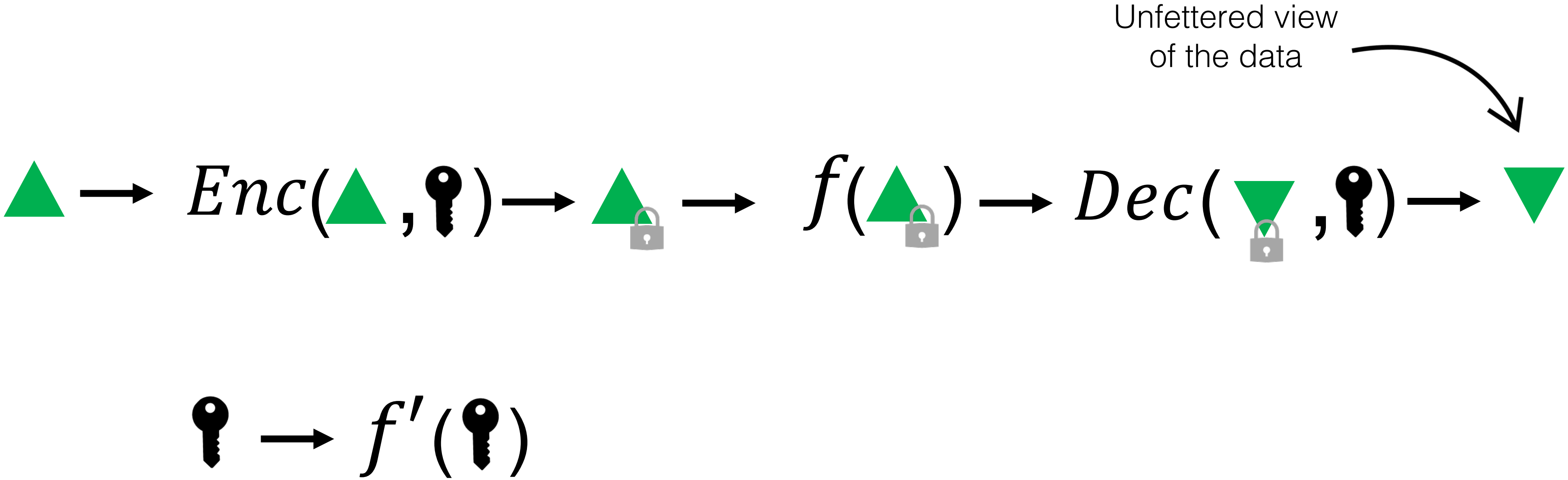
Cryptographic Privacy Tokens



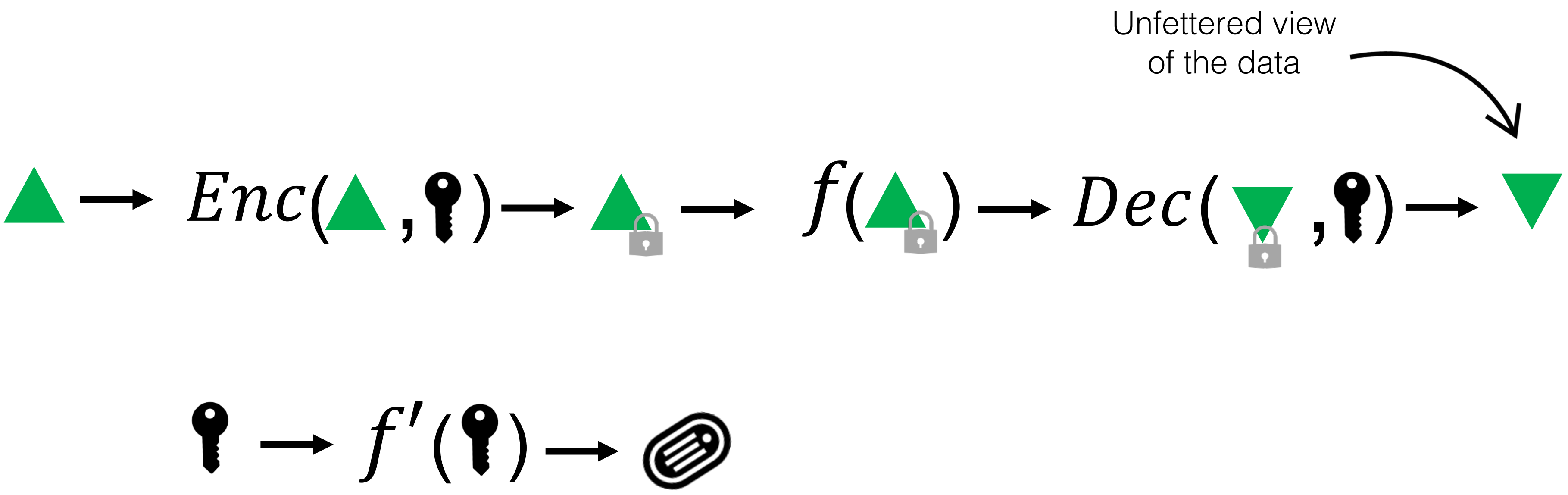
Cryptographic Privacy Tokens



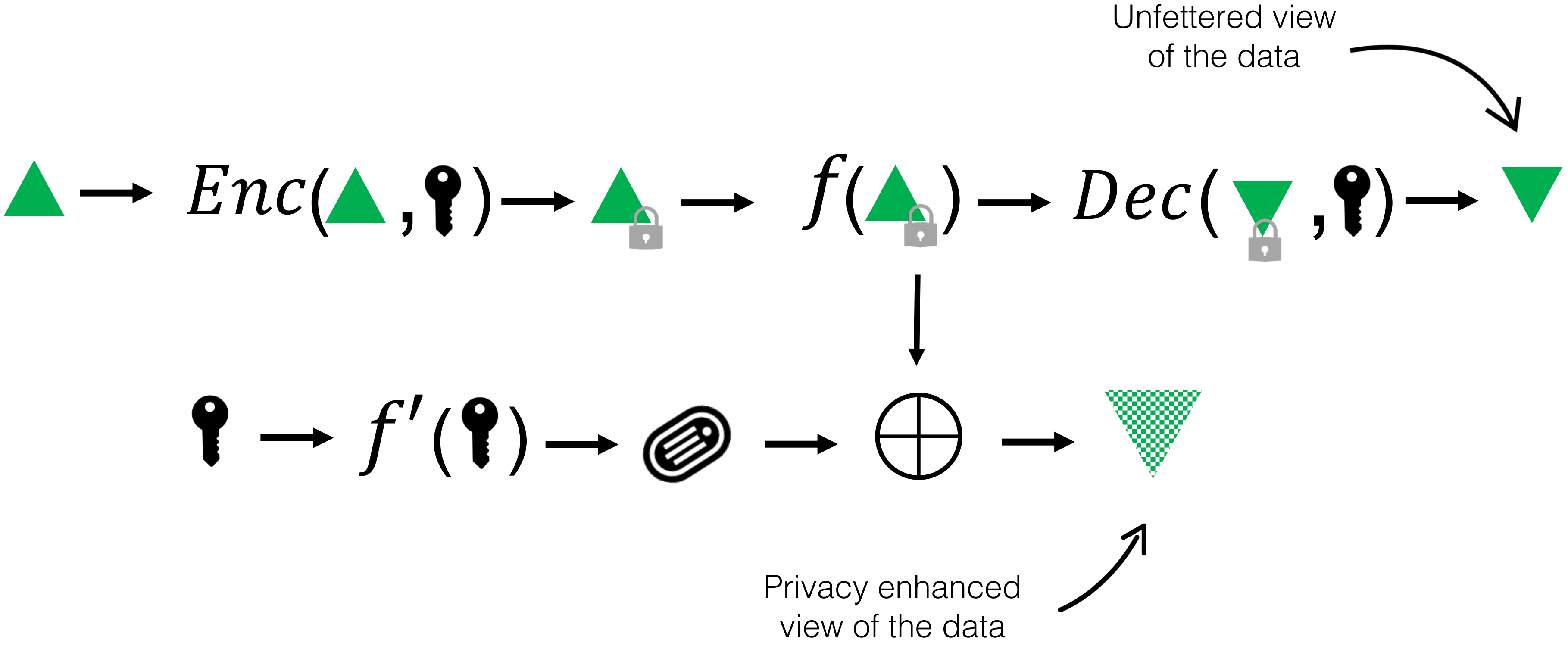
Cryptographic Privacy Tokens



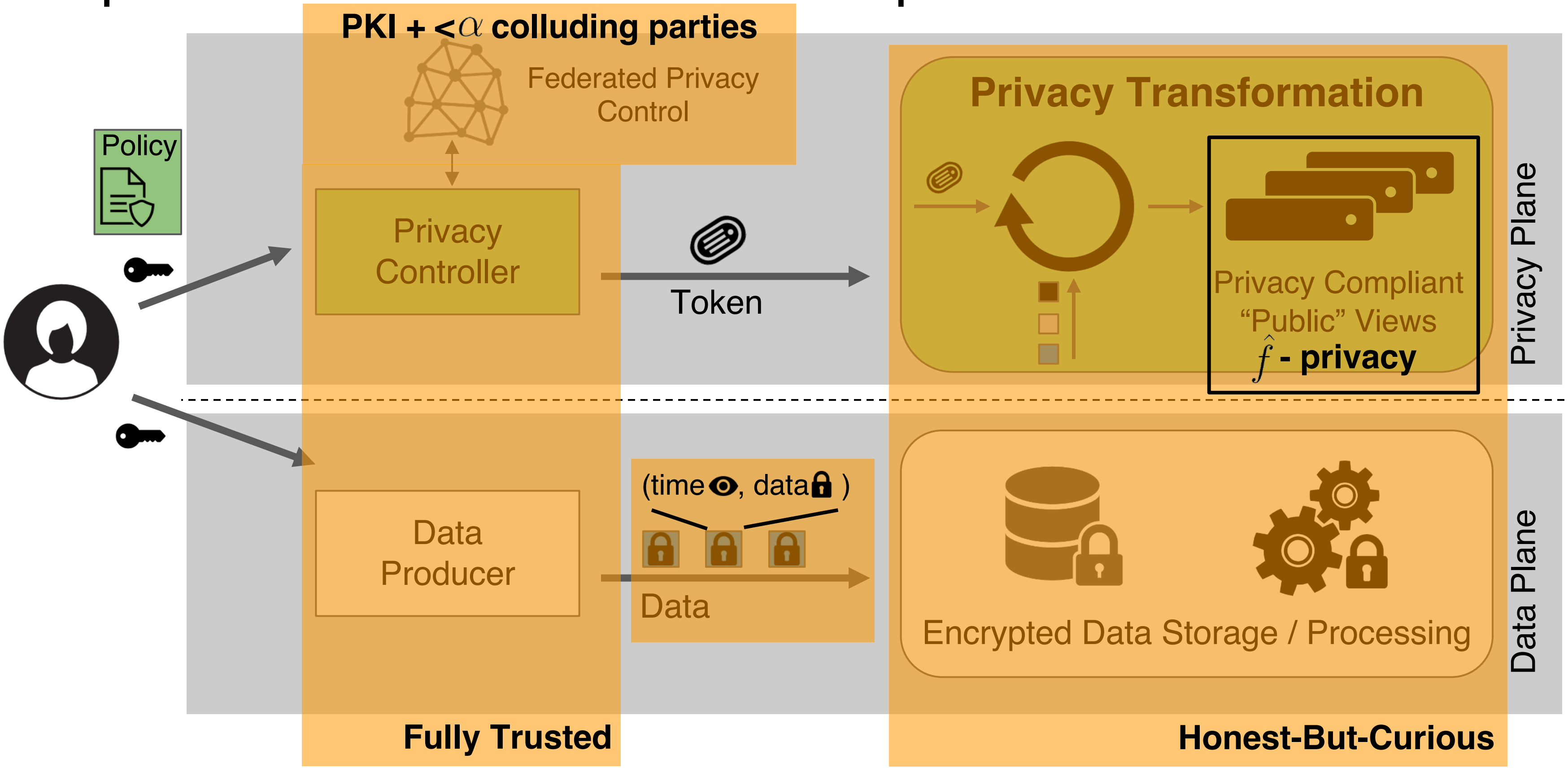
Cryptographic Privacy Tokens



Cryptographic Privacy Tokens



Zeph's Threat Model and Assumptions



Privacy Transformations

Existing Privacy Transformations

“Practical” Privacy Tools

- Data Masking
  - Pseudonymization
  - Tokenization
  - Perturbation
  - Redaction
  - Date Shifting
- Data Generalization
  - Bucketing
  - Time Resolution
  - Population

Formal Privacy Models

Existing Privacy Transformations

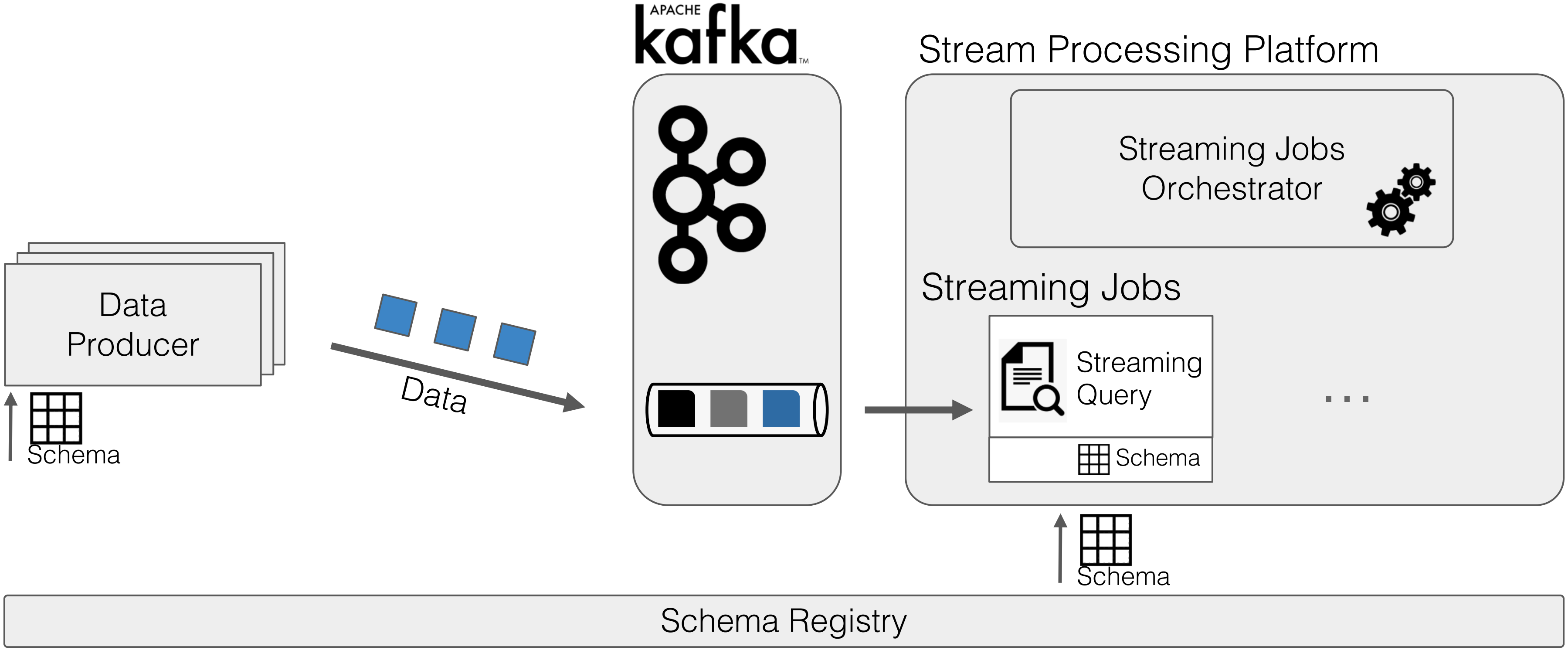
"Practical" Privacy Tools

- Data Masking
  - Pseudonymization
  - Tokenization
  - Perturbation
  - Redaction
  - Date Shifting
- Data Generalization
  - Bucketing
  - Time Resolution
  - Population

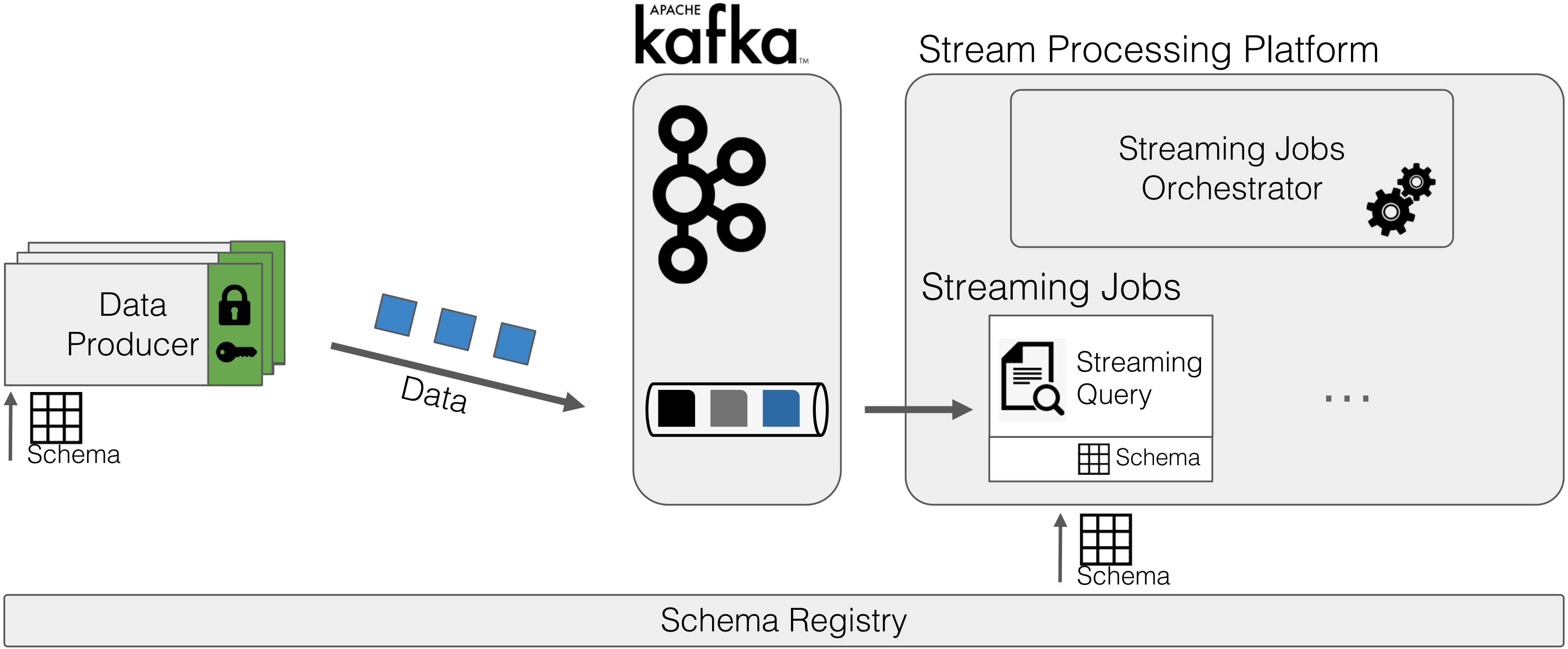
Formal Privacy Models

- Differential Privacy DP
  - Additive Noise Mechanism
  - Others (e.g., ME, SVT)
- k - Anonymity

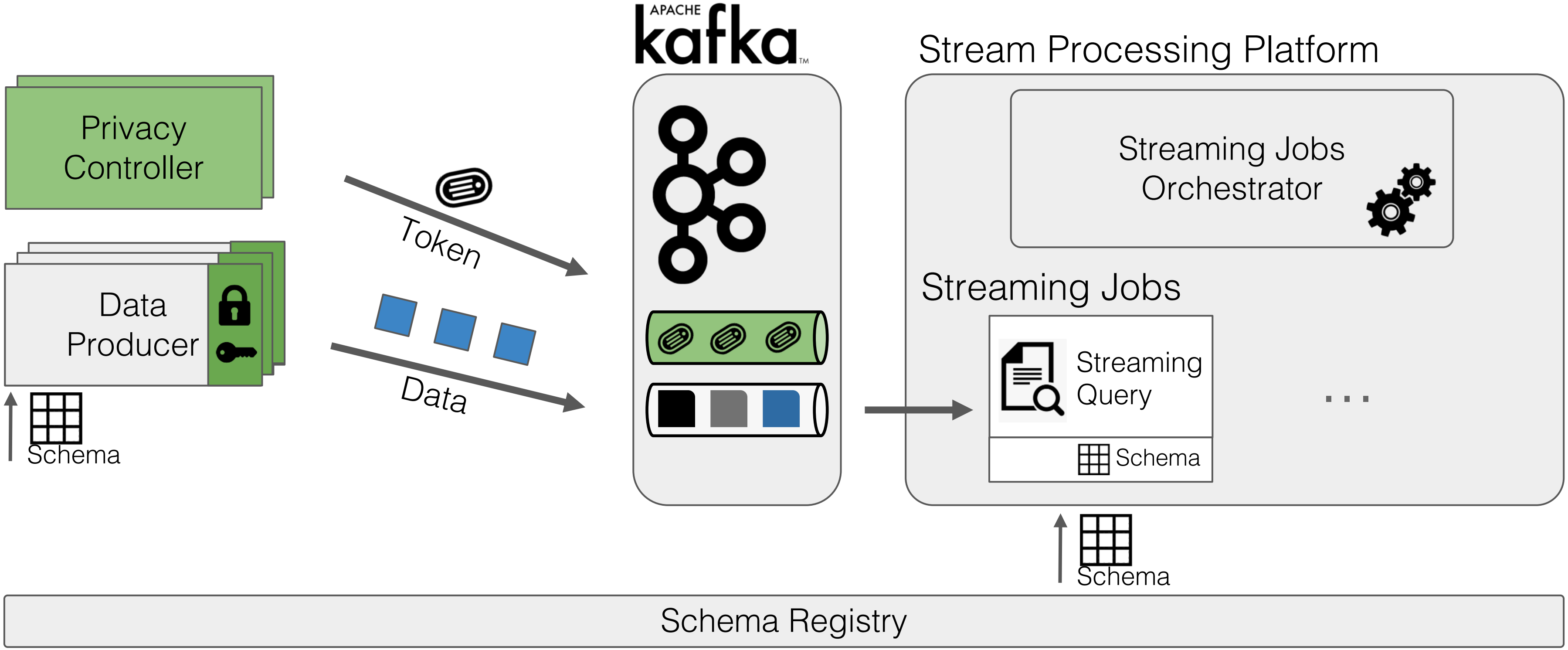
How Zeph augments existing System Designs



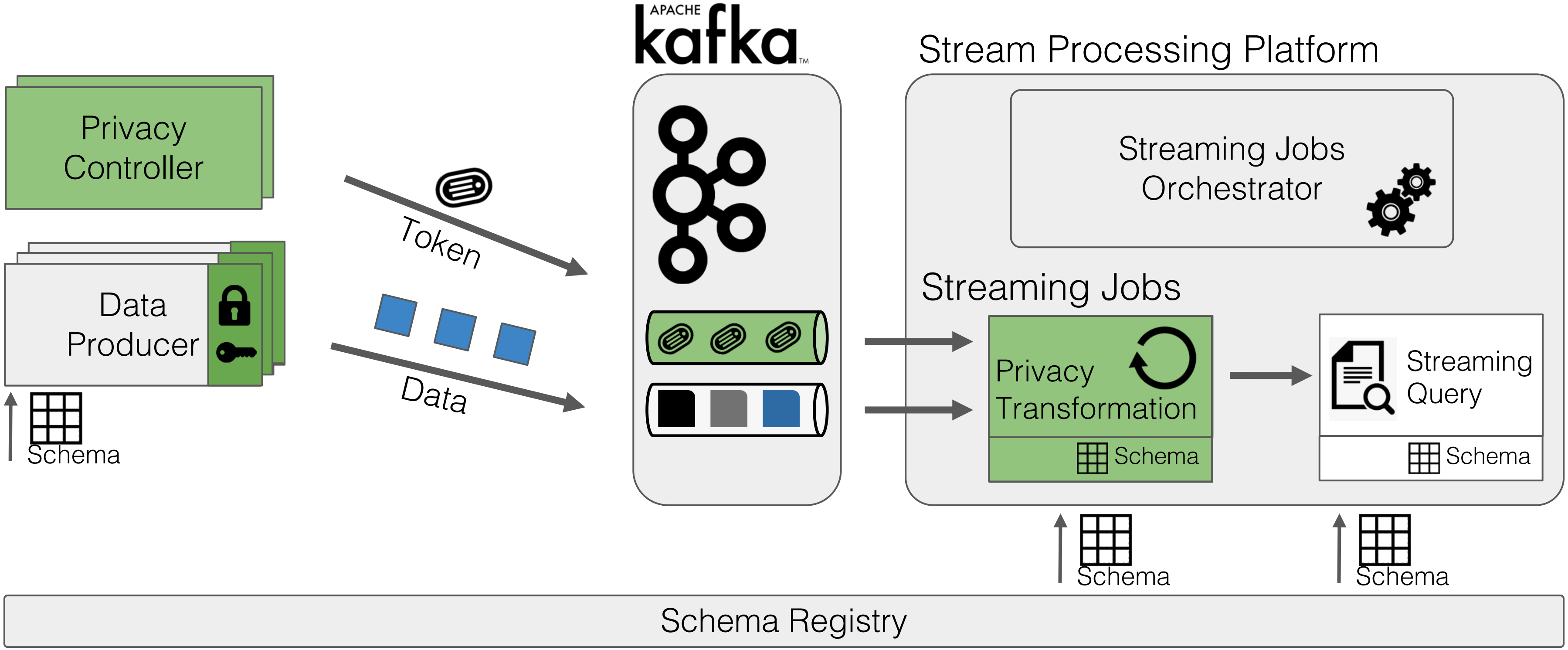
How Zeph augments existing System Designs



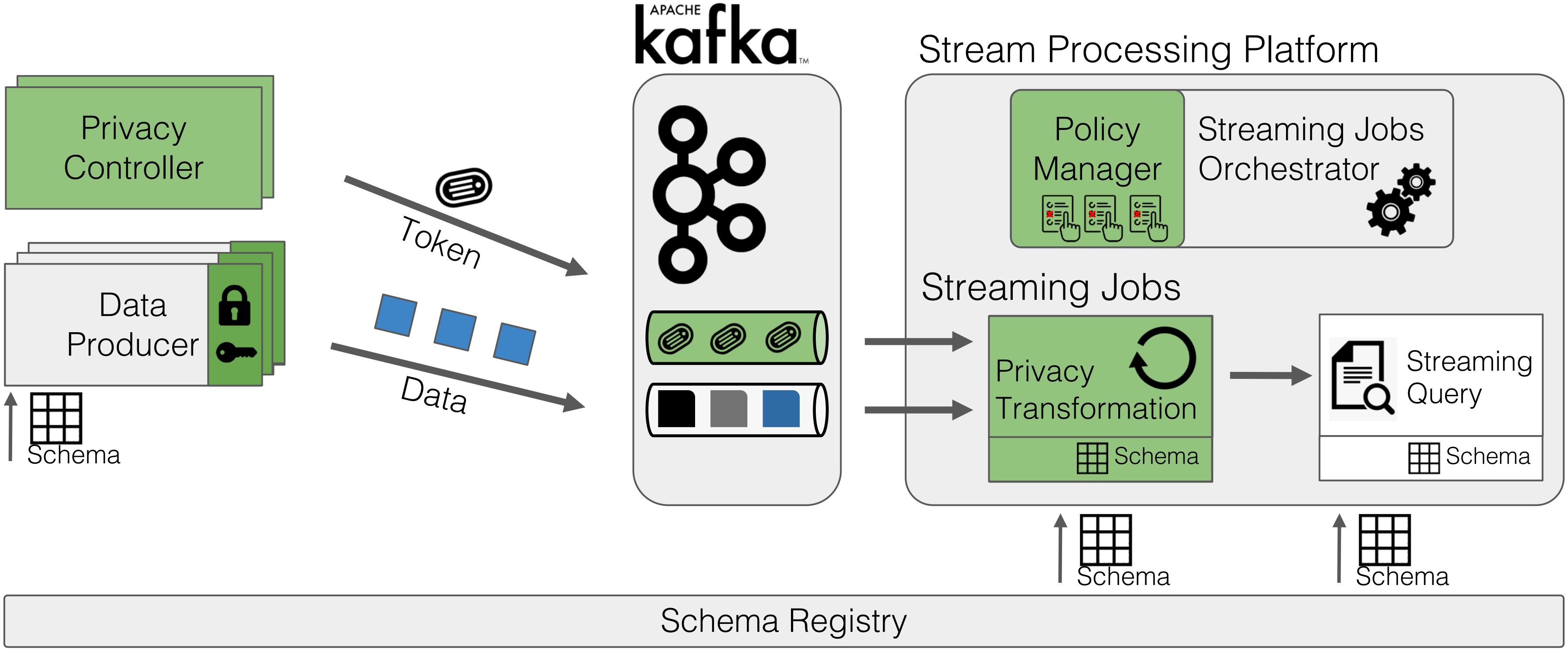
How Zeph augments existing System Designs



How Zeph augments existing System Designs



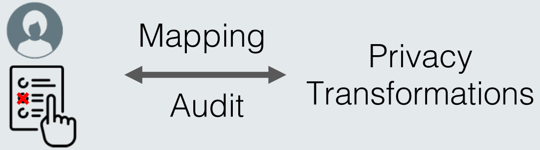
How Zeph augments existing System Designs



Contributions

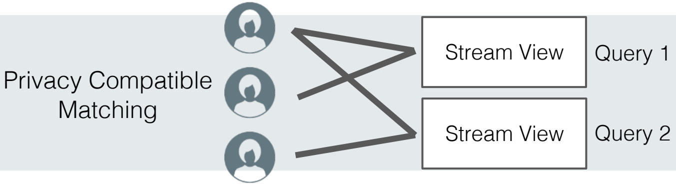
User-centric privacy

Keep End-User Control Simple



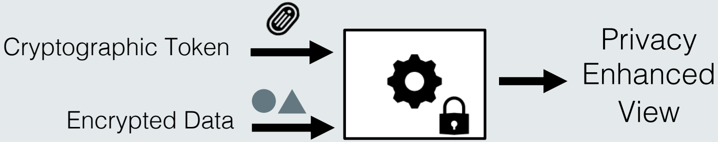
Privacy Orchestration

Organize Privacy Transformations



Cryptographic Enforcement

Cryptographic Privacy Tokens



Contributions

User-centric privacy

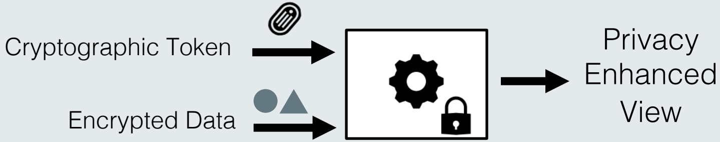
Keep End-User Control Simple

Privacy Orchestration

Organize Privacy Transformations

Cryptographic Enforcement

Cryptographic Privacy Tokens



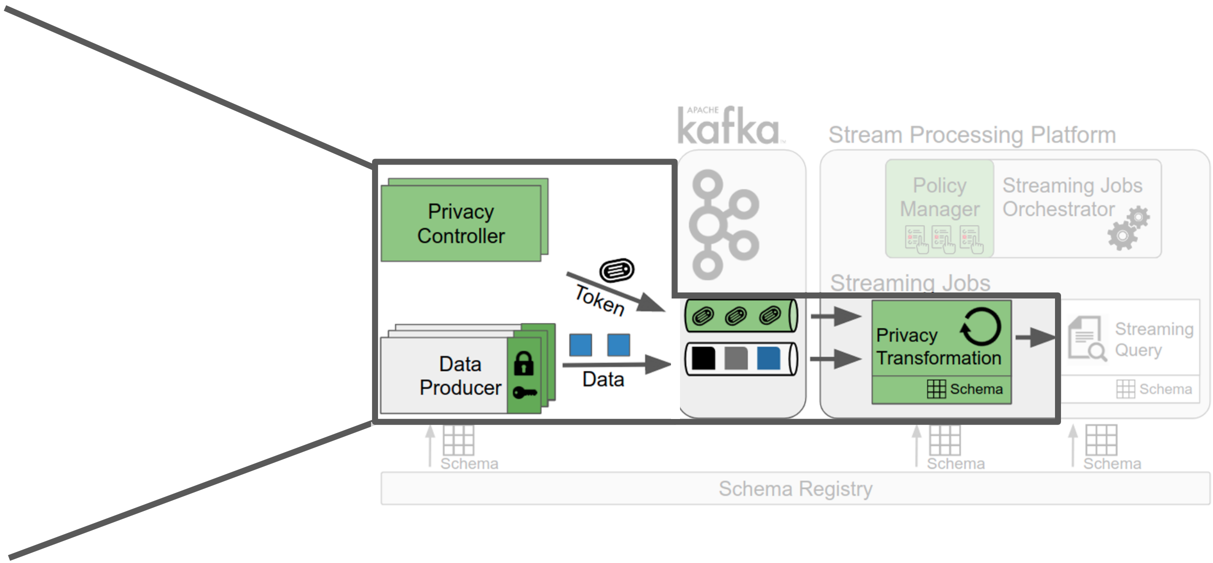
Cryptographic Enforcement of Privacy

1) **Confidentiality** of data

2) Transformation **Authorization** by Privacy Controller

3) **Compute** transformation on confidential data

4) Privacy Controller is **efficient** and **independent** of data



Cryptographic Enforcement of Privacy

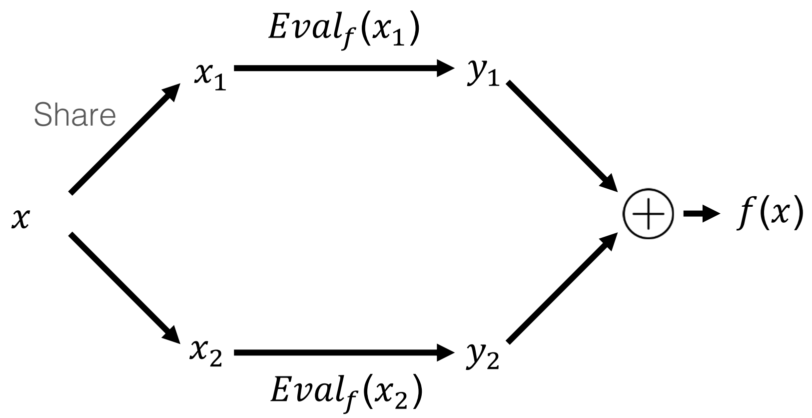
1) **Confidentiality** of data

2) Transformation **Authorization** by Privacy Controller

3) **Compute** transformation on confidential data

4) Privacy Controller is **efficient** and **independent** of data

Additive Homomorphic Secret Sharing



Cryptographic Enforcement of Privacy

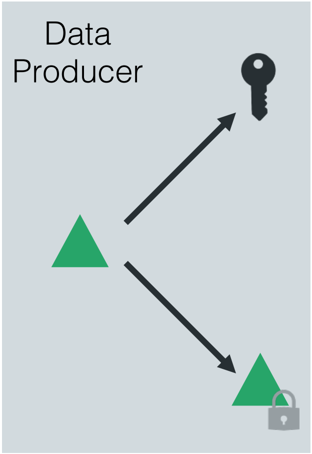
1) **Confidentiality** of data

2) Transformation **Authorization** by Privacy Controller

3) **Compute** transformation on confidential data

4) Privacy Controller is **efficient** and **independent** of data

Additive Homomorphic Secret Sharing



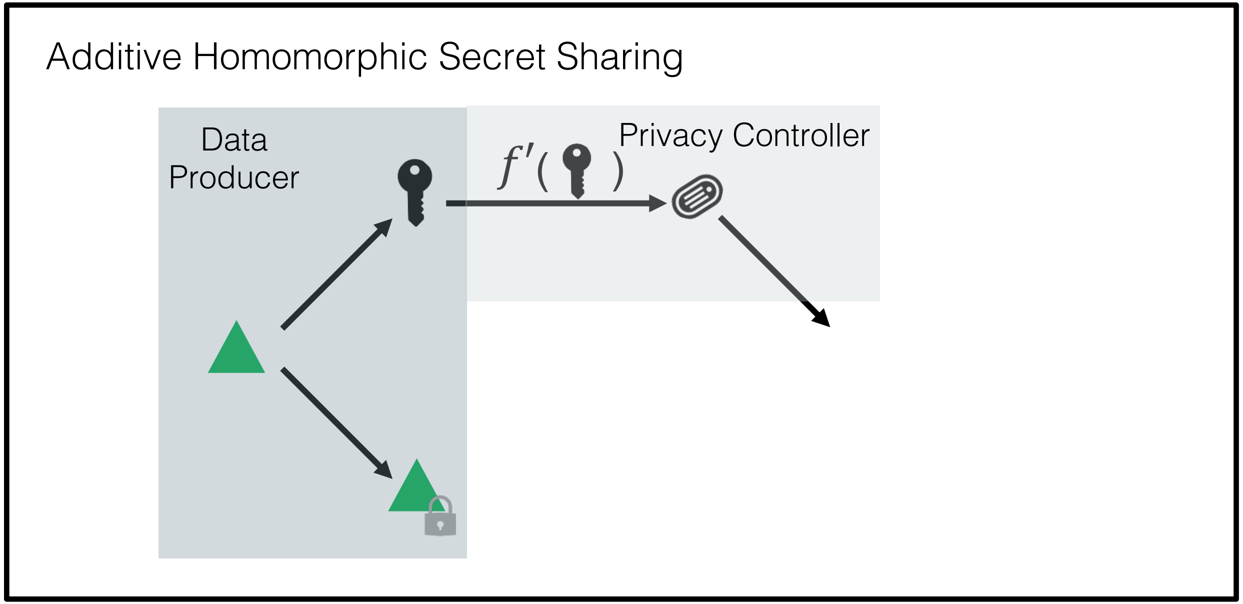
Cryptographic Enforcement of Privacy

1) **Confidentiality** of data

2) Transformation **Authorization** by Privacy Controller

3) **Compute** transformation on confidential data

4) Privacy Controller is **efficient** and **independent** of data



Cryptographic Enforcement of Privacy

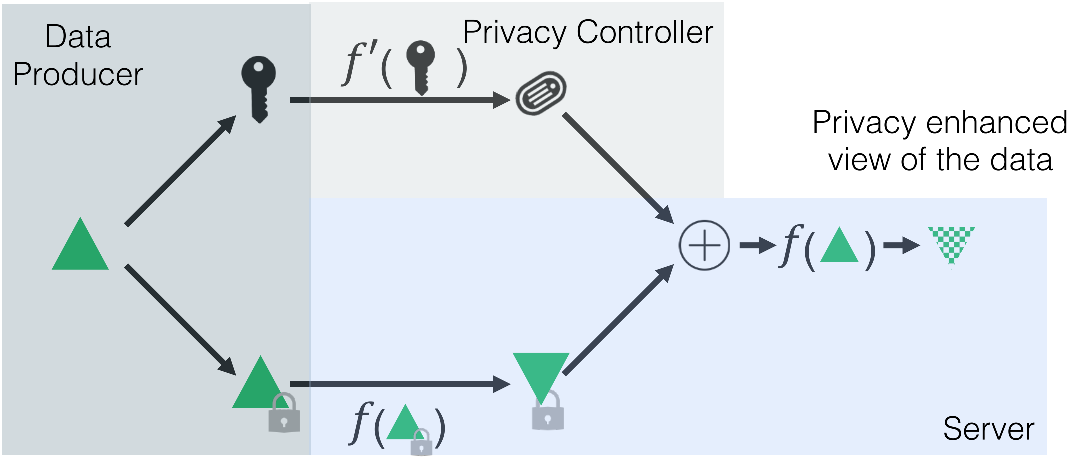
1) **Confidentiality** of data

2) Transformation **Authorization** by Privacy Controller

3) **Compute** transformation on confidential data

4) Privacy Controller is **efficient** and **independent** of data

Additive Homomorphic Secret Sharing



Privacy Transformations

$T'(\cdot) = \sum$

"Practical" Privacy Tools

- **Data Masking**
  - Pseudonymization
  - Tokenization
  - Perturbation
  - Redaction
  - Date Shifting
- **Data Generalization**
  - Bucketing
  - Time Resolution
  - Population

Formal Privacy Models

- **Differential Privacy** DP
  - Additive Noise Mechanism
  - Others (e.g., SVT)
- **k - Anonymity**

Privacy Transformations

T'(·) = Σ

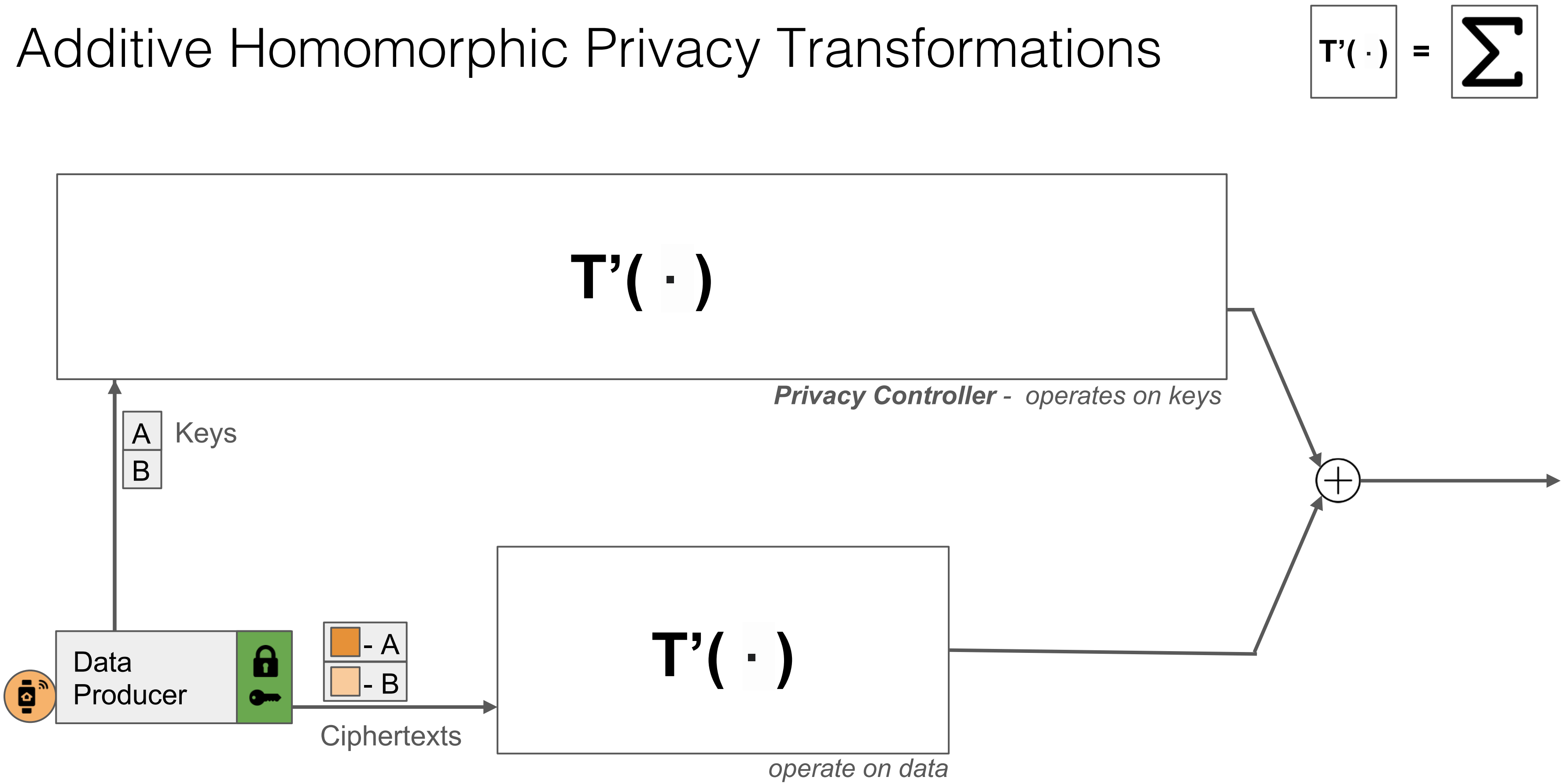
"Practical" Privacy Tools

- Data Masking
  - Pseudonymization ✗
  - Tokenization ✗
  - Perturbation ✓
  - Redaction ✓
  - Date Shifting ✓
- Data Generalization
  - Bucketing ✓
  - Time Resolution ✓
  - Population ✓

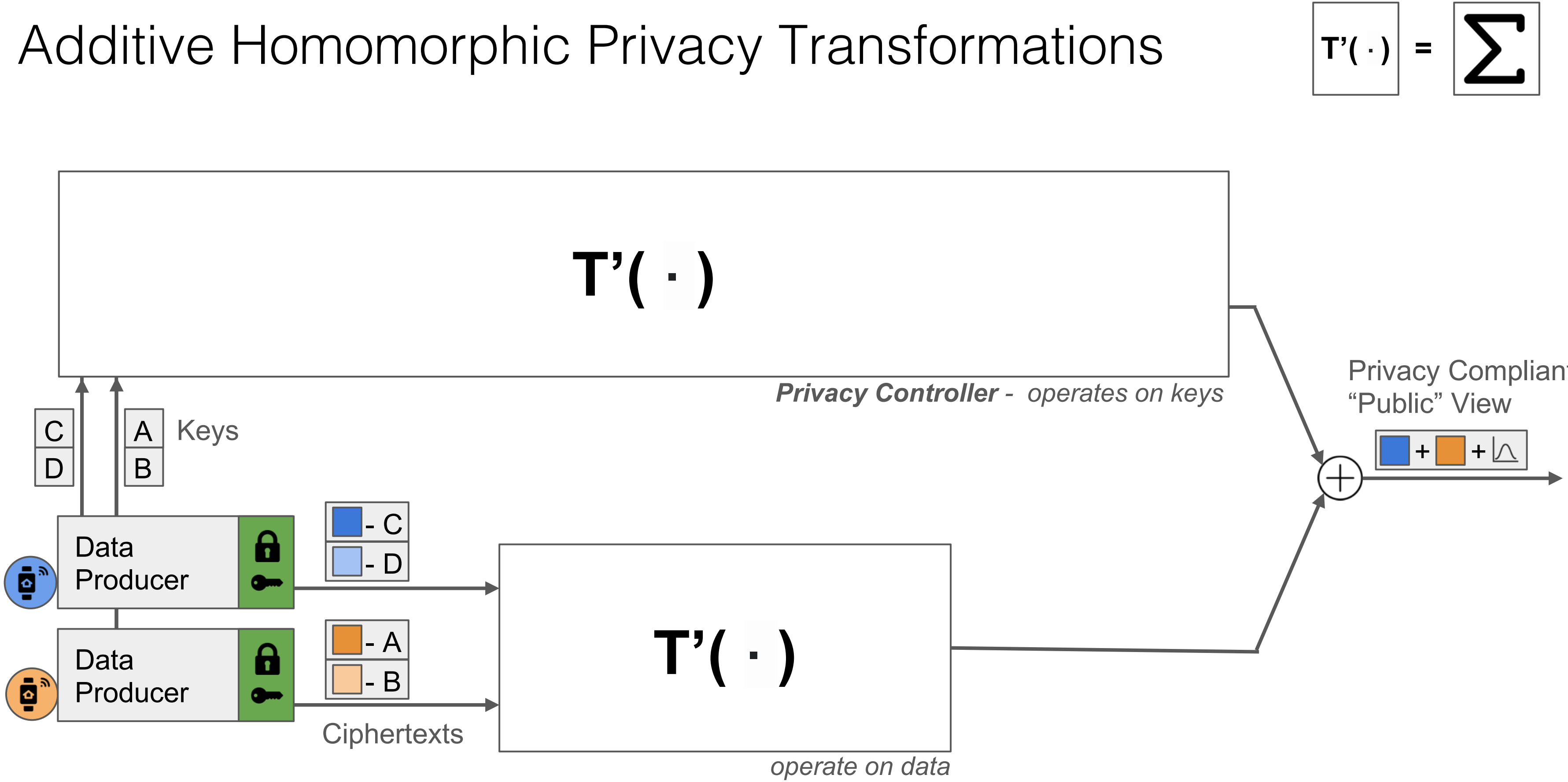
Formal Privacy Models

- Differential Privacy **DP**
  - Additive Noise Mechanism ✓
  - Others (e.g., SVT) ✗
- k - Anonymity ✗

Additive Homomorphic Privacy Transformations

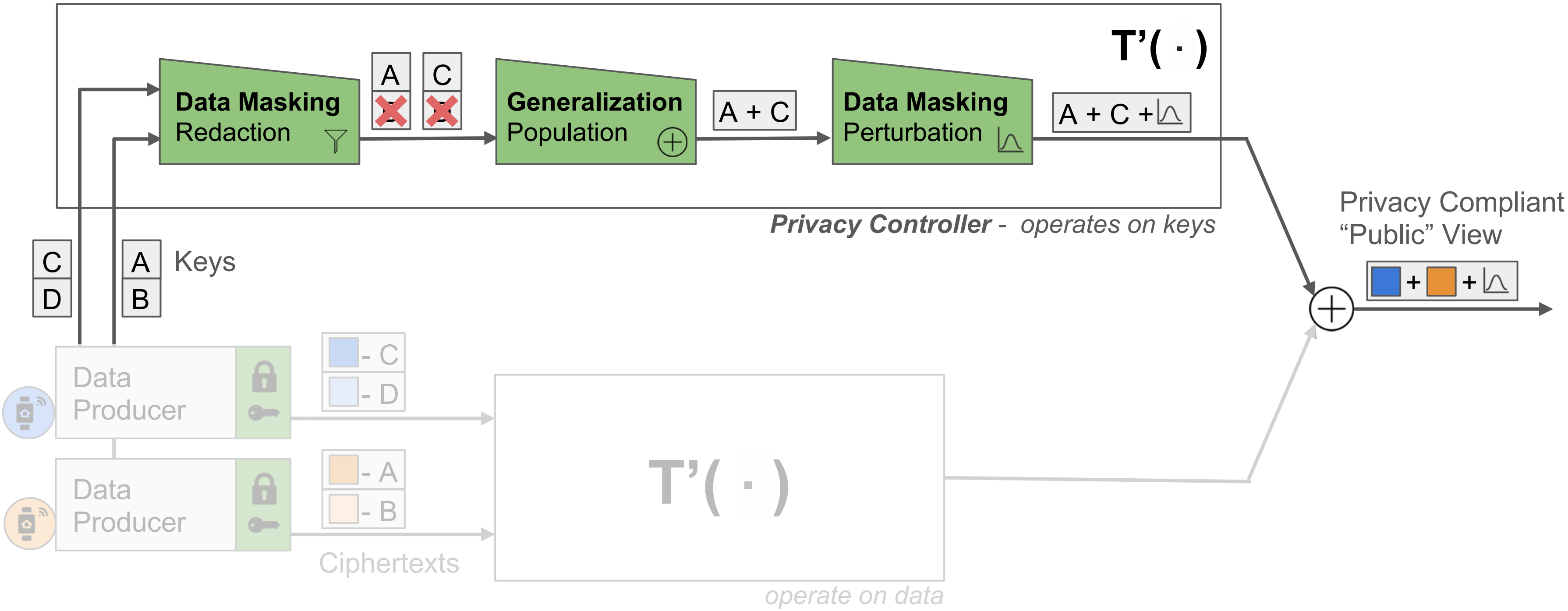


Additive Homomorphic Privacy Transformations

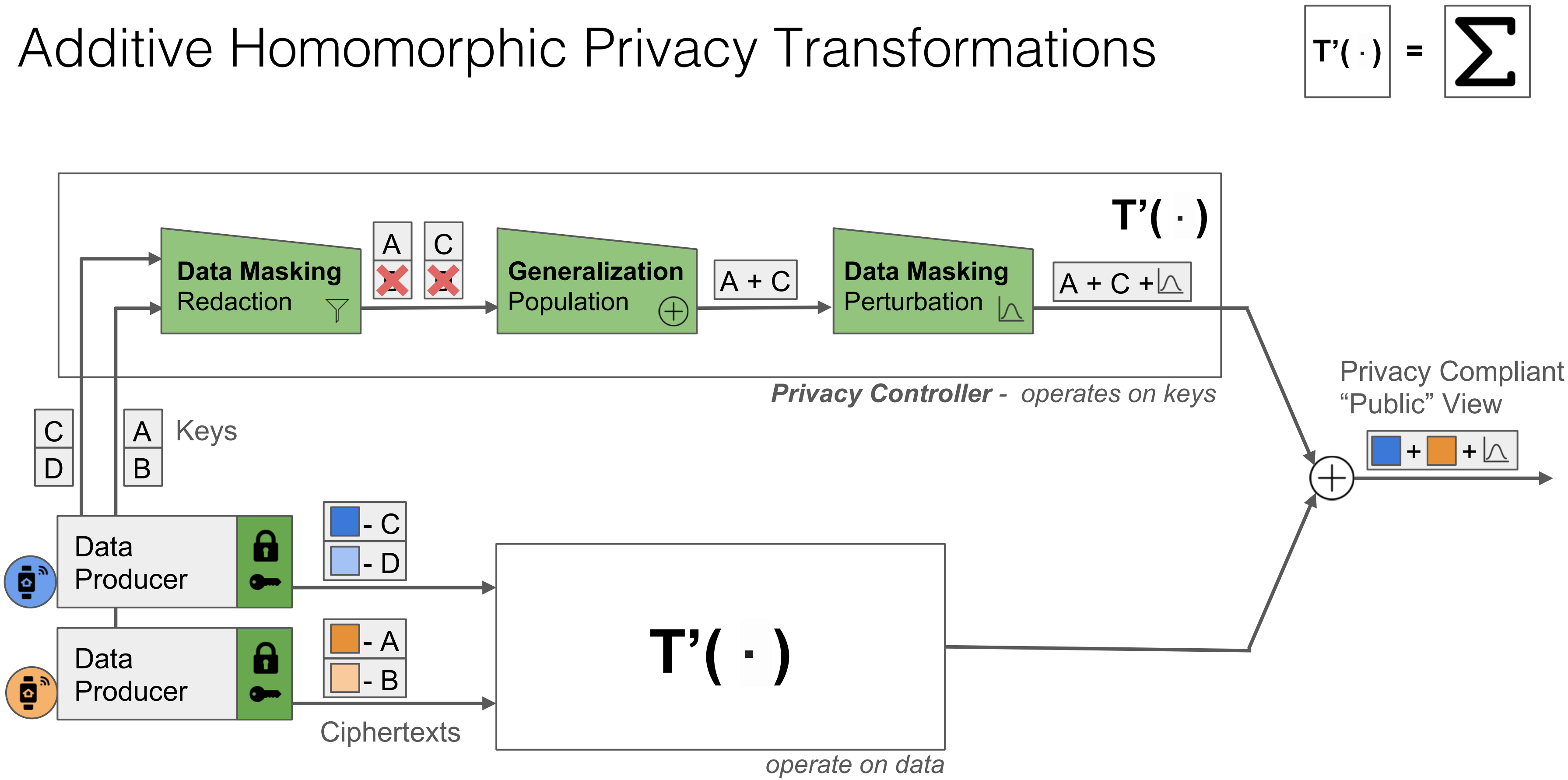


Additive Homomorphic Privacy Transformations

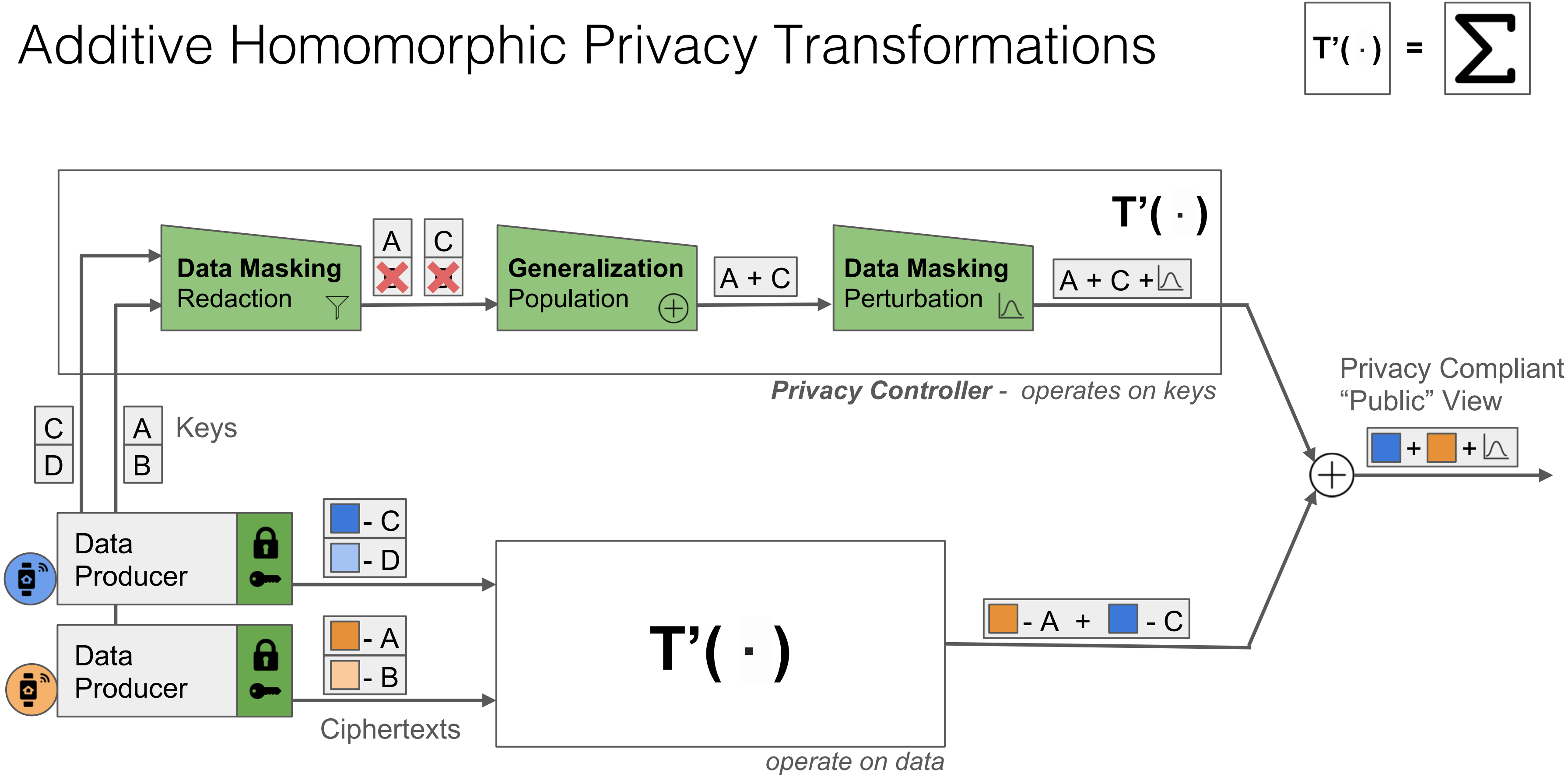
$\boxed{T'(\cdot)} = \boxed{\sum}$



Additive Homomorphic Privacy Transformations

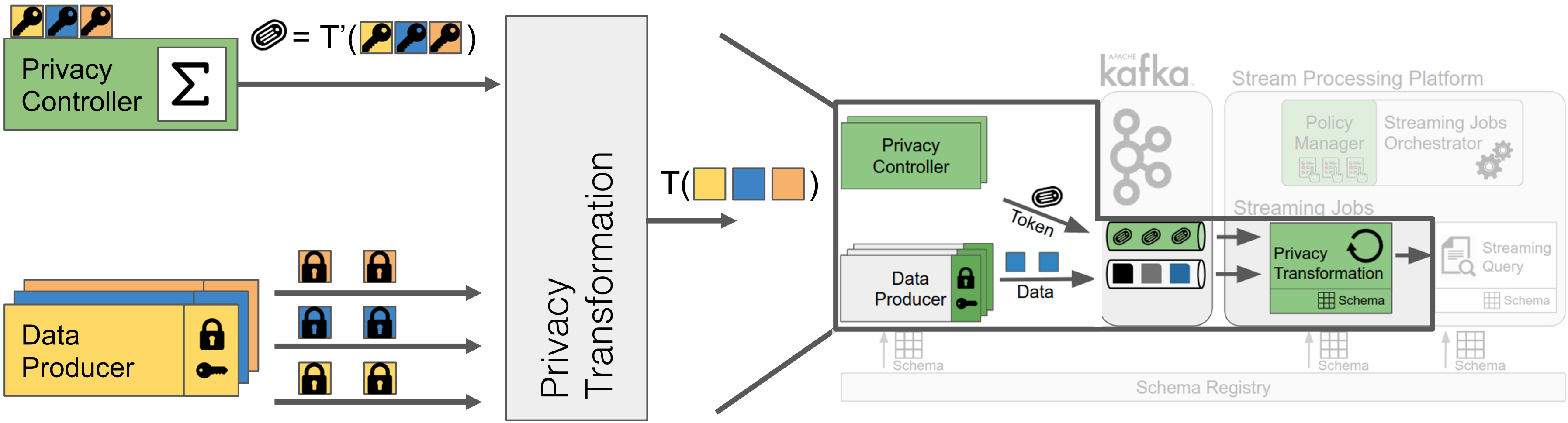


Additive Homomorphic Privacy Transformations



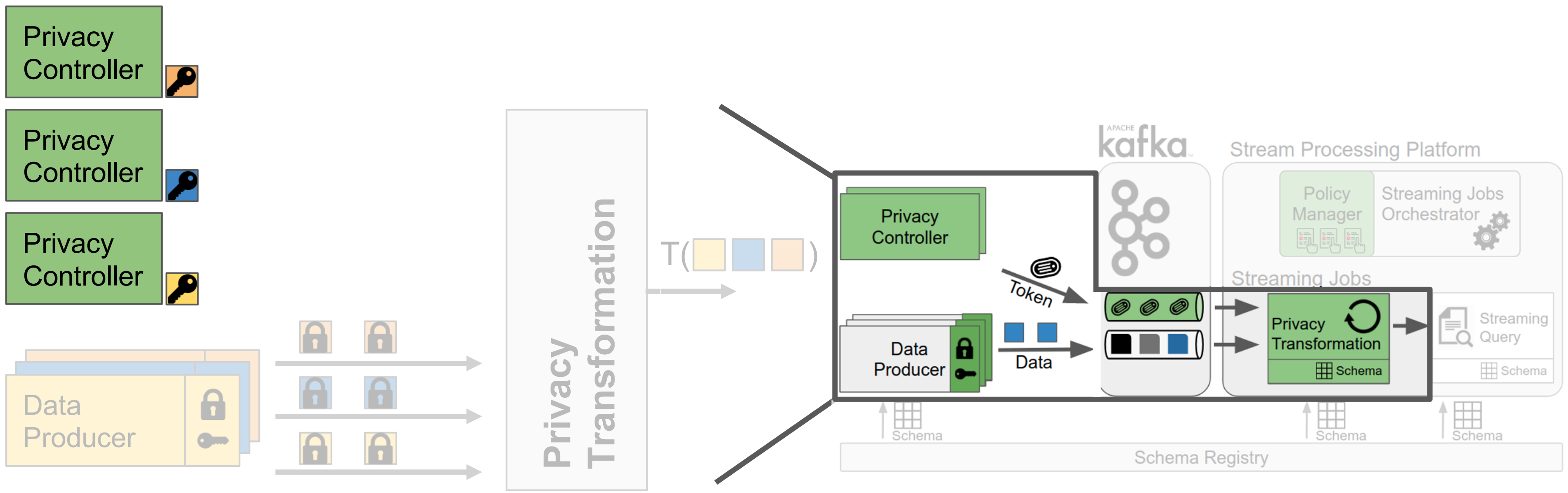
Enable Federated Privacy Control

"multiple Data Producers - one Privacy Controller"



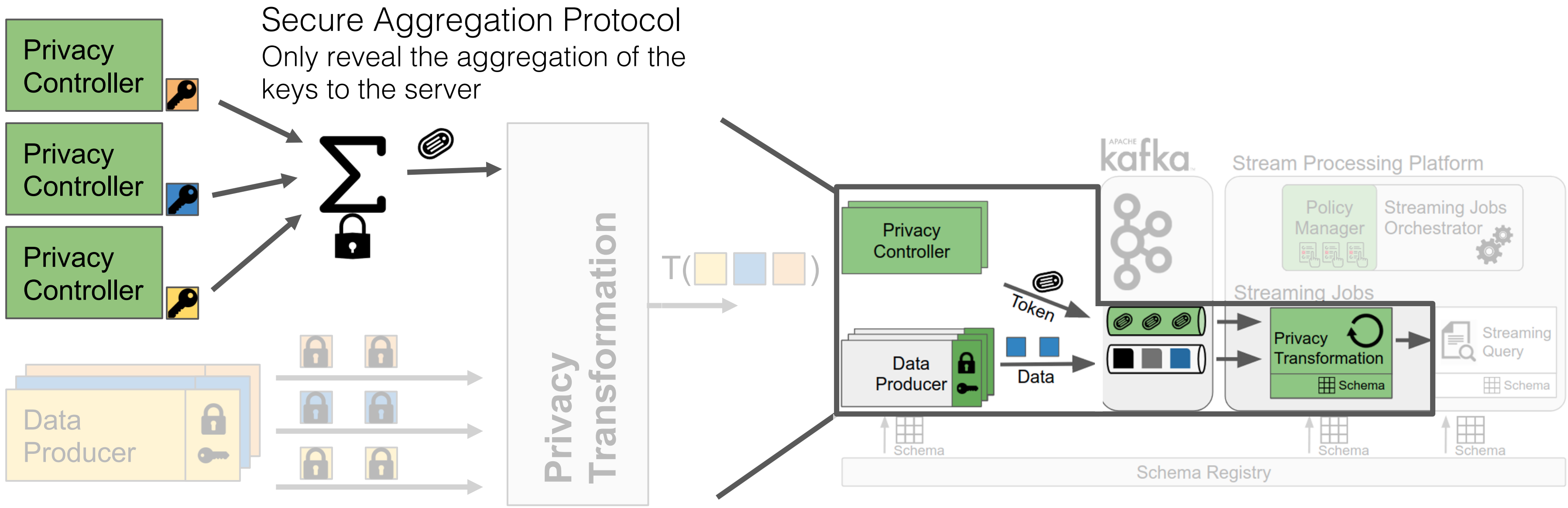
Enable Federated Privacy Control

“multiple Data Producers - multiple Privacy Controllers”



Enable Federated Privacy Control

"multiple Data Producers - multiple Privacy Controllers"



Zeph Implementation and Evaluation

ANSIBLE aws

Data Producer & Privacy Controller

Java

The Rust Programming Language

Privacy Transformation

APACHE kafka streams

Zeph Implementation and Evaluation

ANSIBLE aws

Data Producer & Privacy Controller

Java

The Rust Programming Language

Privacy Transformation

APACHE kafka® streams

Fitness App

Website Analytics

Smart Car

Zeph Implementation and Evaluation

ANSIBLE aws

Data Producer & Privacy Controller

Java

The Rust Programming Language

Privacy Transformation

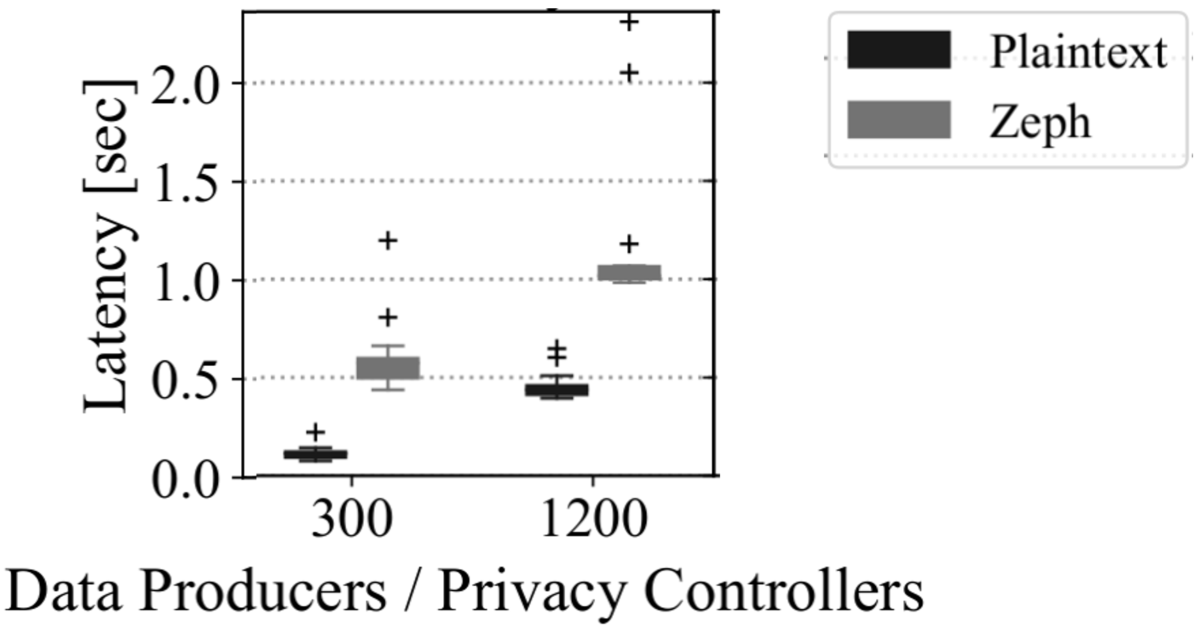
APACHE kafka streams

Fitness App

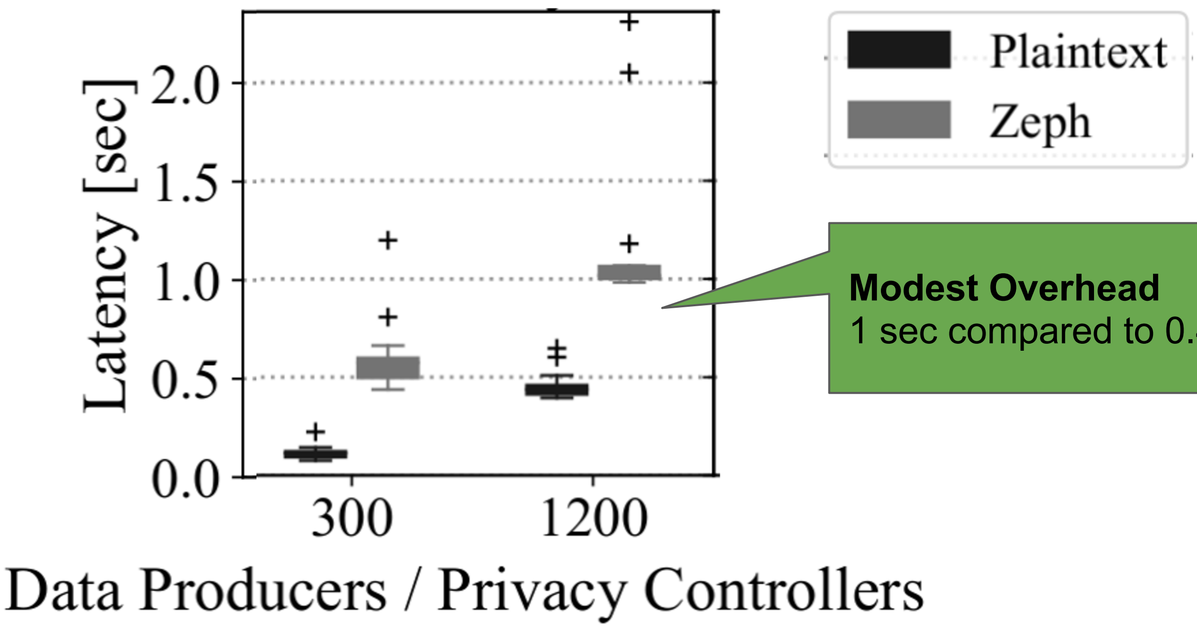
Website Analytics

Smart Car

Web Analytics: End-to-End Benchmark

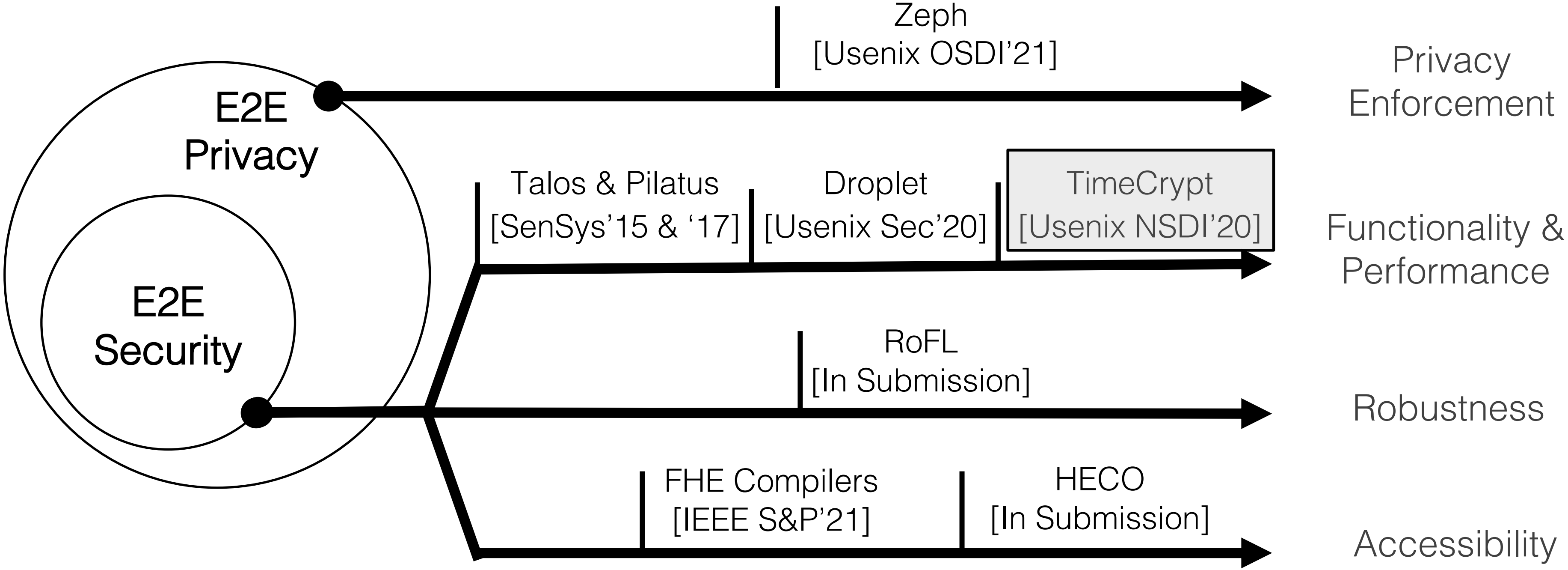


Web Analytics: End-to-End Benchmark



Modest Overhead
1 sec compared to 0.46 sec

My Research: Building practical systems that use cryptography to empower users and preserve their privacy & tools to democratize cryptography



TimeCrypt (Usenix NSDI ‘20)

Encrypted Time Series Database

Can we enable encrypted data processing for time series workloads?

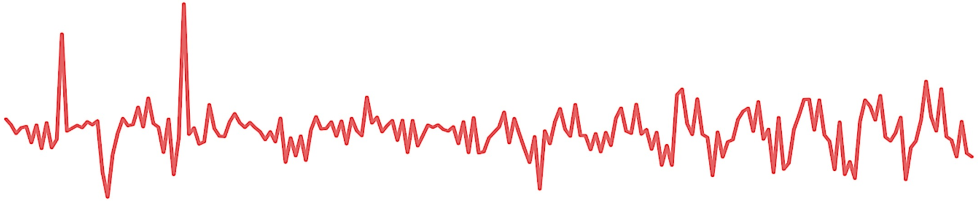
Can we enable encrypted data processing for time series workloads?

Large Scale

Low-Latency

Time Series Data

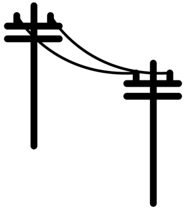
is Emerging Everywhere



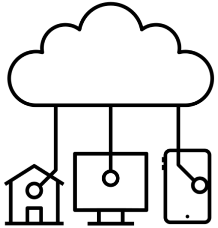
Time-ordered observations of a quantitative characteristic of an individual or phenomenon taken at successive points in time.

Financial

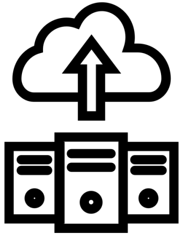
Health Monitoring



Smart Grid

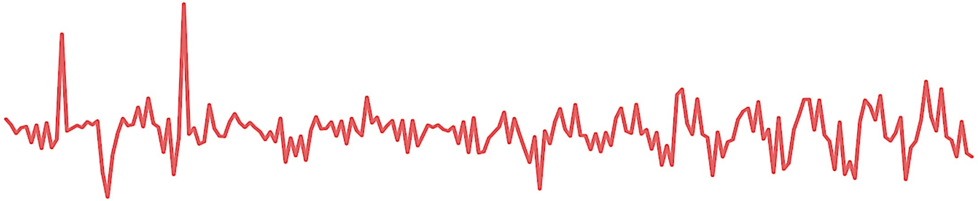


Internet of Things



DevOps/Telemetry

Time Series Data
is Emerging Everywhere

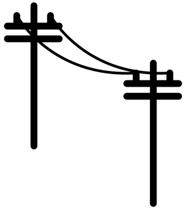


Time-ordered observations of a quantitative characteristic of an individual or phenomenon taken at successive points in time.

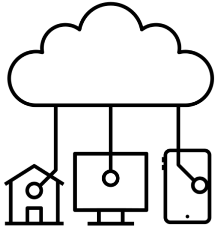
high **resolution sensitive** data!

Financial

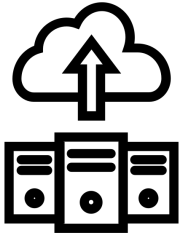
Health Monitoring



Smart Grid



Internet of Things



DevOps/Telemetry

Time Series Databases

Time Series Workloads

- Primarily INSERTS to recent time interval
- Statistical queries over time ranges
- Single writer

Performance Requirements

- High throughput writes
- Data compaction (aging out data)
- Scale with data volume and velocity

Prometheus

druid

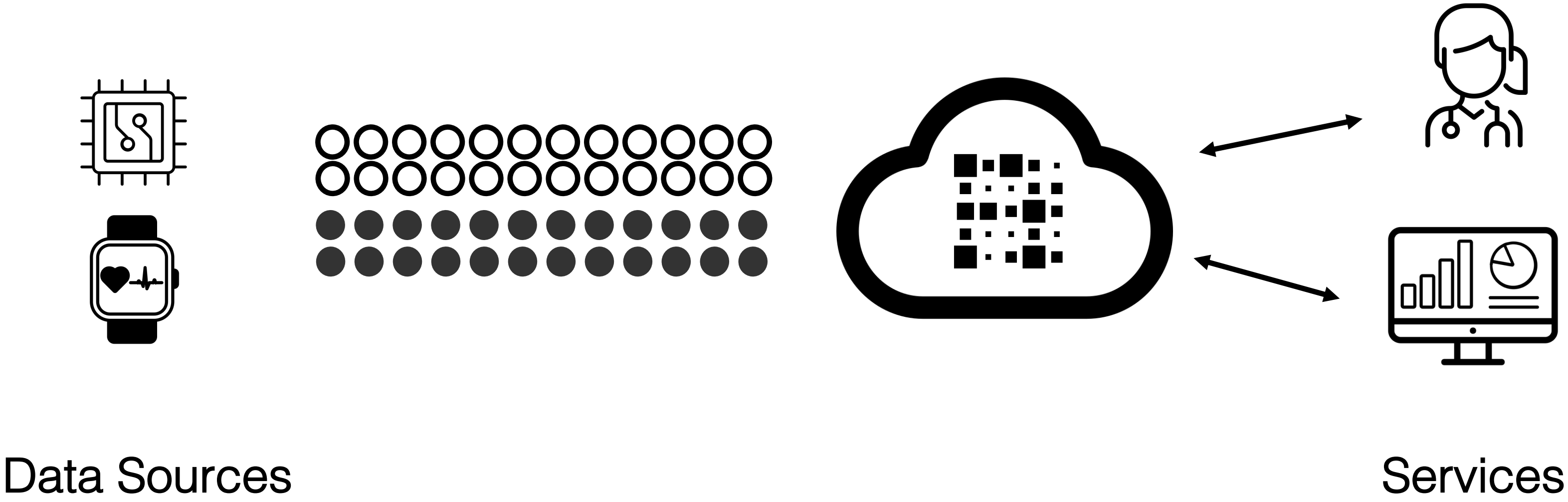
TIMESCALE

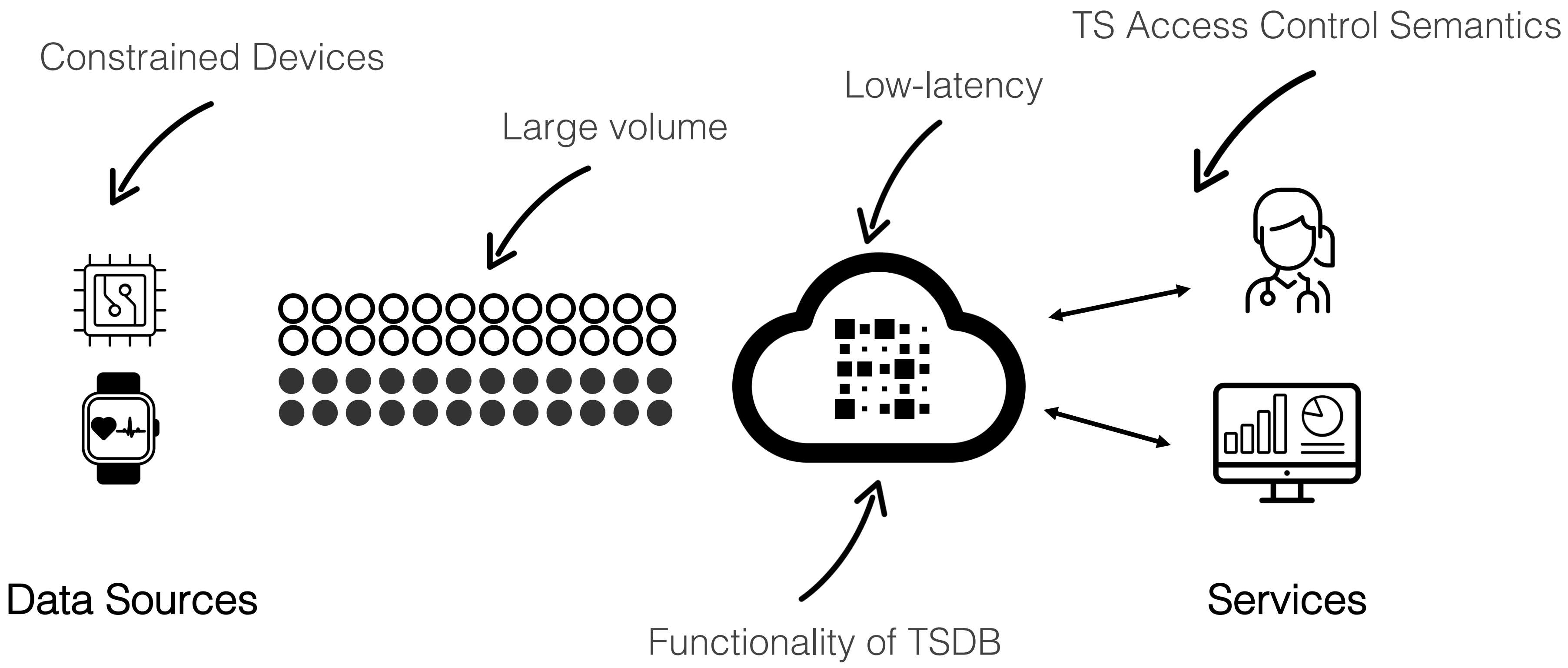
influxdata

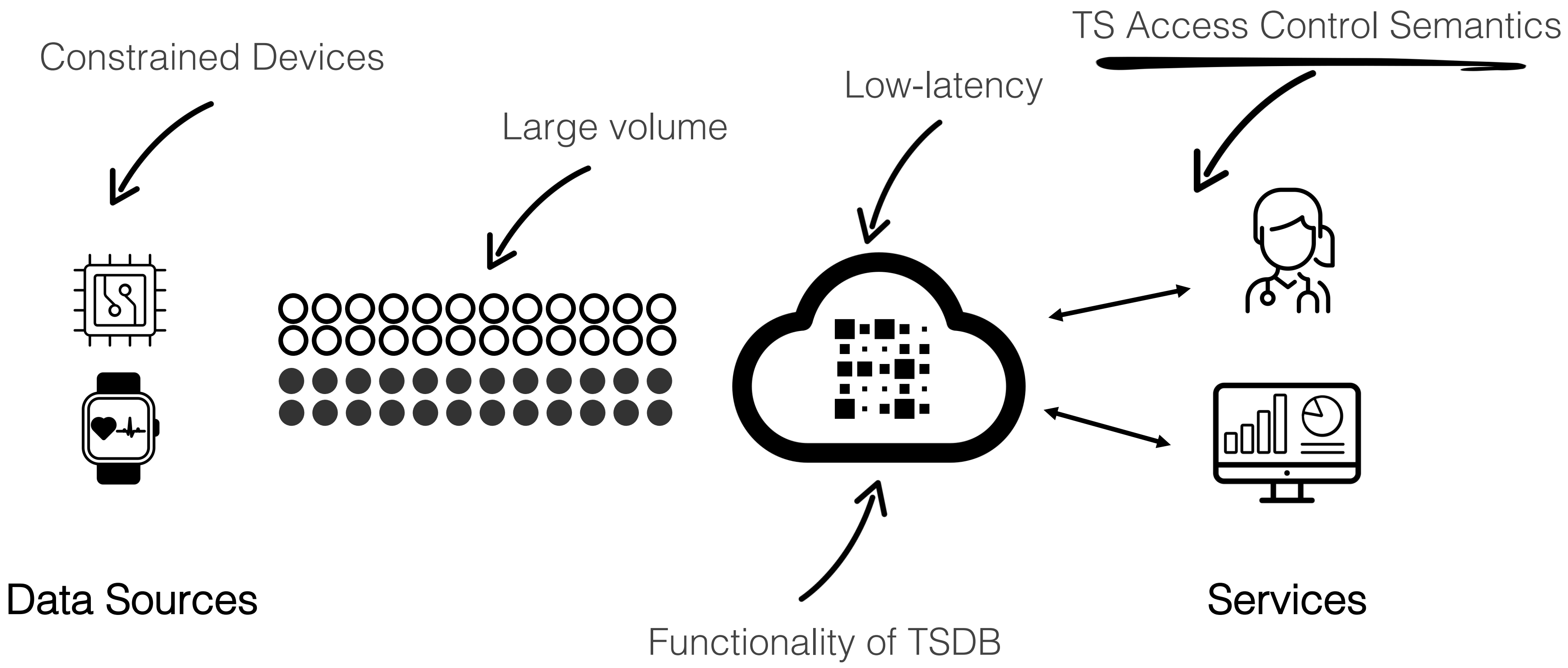
SiriDB

riakTS

KairosDB

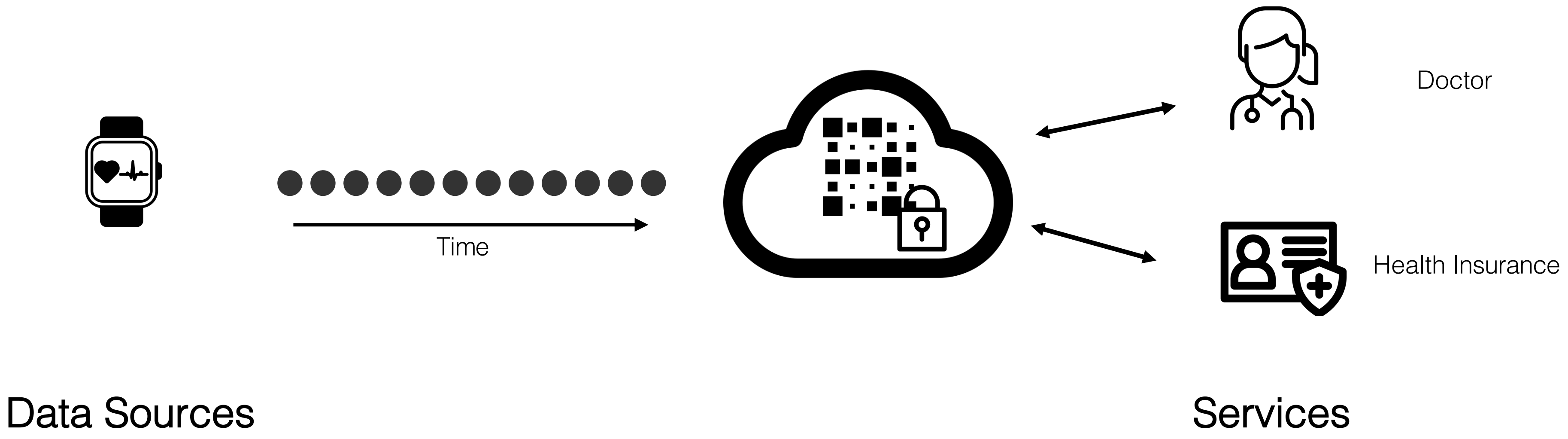

Data Sources
Services


Constrained Devices
Large volume
Low-latency
TS Access Control Semantics
Data Sources
Services
Functionality of TSDB

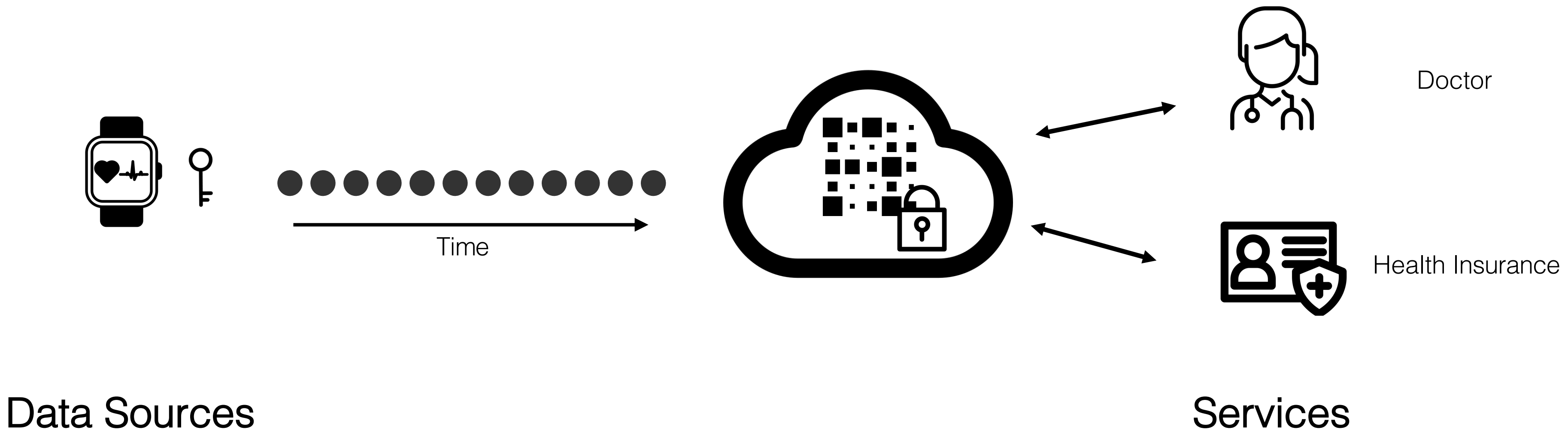


TS Access Control Semantics
Constrained Devices
Low-latency
Large volume
Data Sources
Services
Functionality of TSDB

Cryptographic Access Control

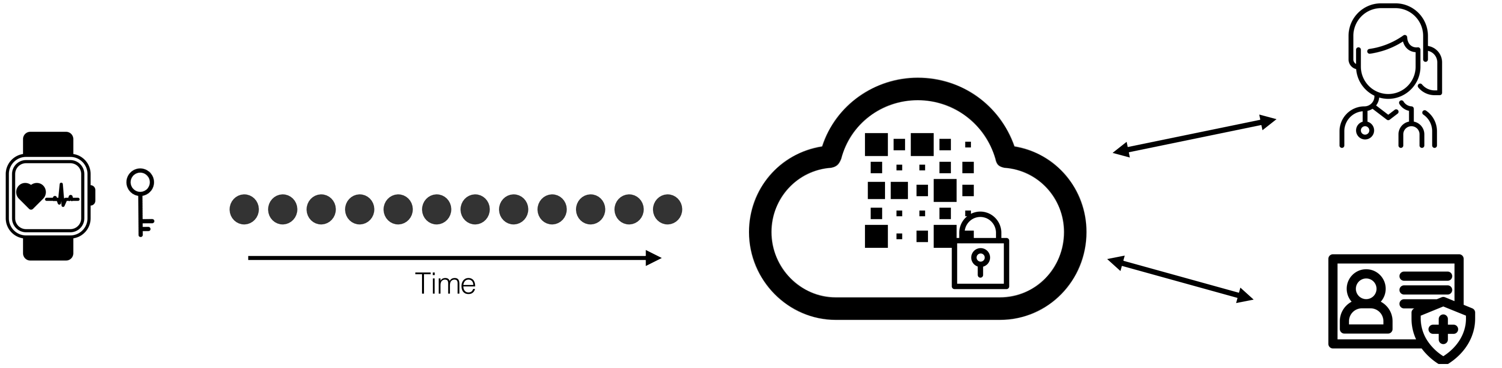


Cryptographic Access Control



Cryptographic Access Control

The services **learned** more than they **needed**!

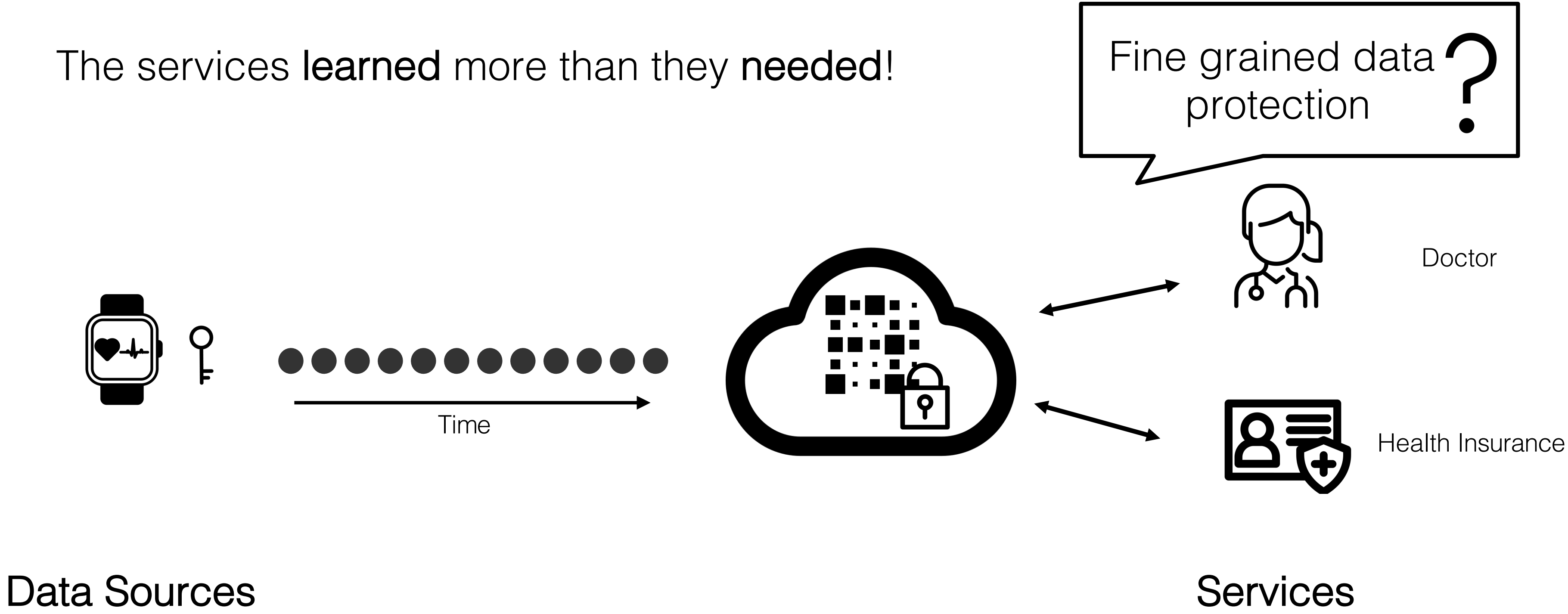


Data Sources

Services

Cryptographic Access Control

The services **learned** more than they **needed**!



Cryptographic Access Control

Support **time series** access control semantics

- Time
- Resolution
- Attribute

Data Sources

Services

Cryptographic Access Control

Support **time series** access control semantics

- Time
- Resolution
- Attribute

Cryptographic Access Control

Support **time series** access control semantics

- Time
- Resolution
- Attribute

TimeCrypt

Encrypted Time Series Database that enables scalable computation over large volumes of encrypted time series data.

Functionalities	Statistical queries and lifecycle operations over encrypted data
Cryptographic Access Control	Fine-grained access policies over time, resolution, and attributes
Security	Data Secrecy -- Homomorphic Encryption Function Integrity -- Homomorphic MACs
Efficiency	Interactive queries over large scale data

Large-Scale Challenges

Lots of Data:

Supporting “big data” computations

Fine-grained Access Control:

Scalability as data and the number of data consumers grow

E.g., Partial Homomorphic Encryption [Paillier]

E.g., Attribute Based Encryption [KP-ABE]

Large-Scale Challenges

Lots of Data:

Supporting “big data” computations

Fine-grained Access Control:

Scalability as data and the number of data consumers grow

E.g., Partial Homomorphic Encryption [Paillier]

E.g., Attribute Based Encryption [KP-ABE]

Solution that supports both fine-grained access control and computations over large-scale encrypted data

Building Blocks

Efficiency	Additive Symmetric Homomorphic Encryption [Castelluccia]

Building Blocks

Efficiency

Additive Symmetric Homomorphic Encryption [Castelluccia]

Key stream: $k_0, k_1, k_2, k_3, k_4, k_5, \ldots$

●●●●

Δ

Building Blocks

Efficiency

Additive Symmetric Homomorphic Encryption [Castelluccia]

Key stream: $k_0, k_1, k_2, k_3, k_4, k_5, \ldots$

●●●● Δ

$\boxed{m_0 + k_0}$ $\boxed{m_1 + k_1}$ $\boxed{m_2 + k_2}$ $\cdots$ modulo M

$\longrightarrow$ Time

Building Blocks

Efficiency

Additive Symmetric Homomorphic Encryption [Castelluccia]

Key stream: $k_0, k_1, k_2, k_3, k_4, k_5, \ldots$

●●●● Δ

$\underbrace{\boxed{m_0 + k_0} \quad \boxed{m_1 + k_1}}_{+}$ $\quad \boxed{m_2 + k_2} \quad \cdots$ modulo M

Time

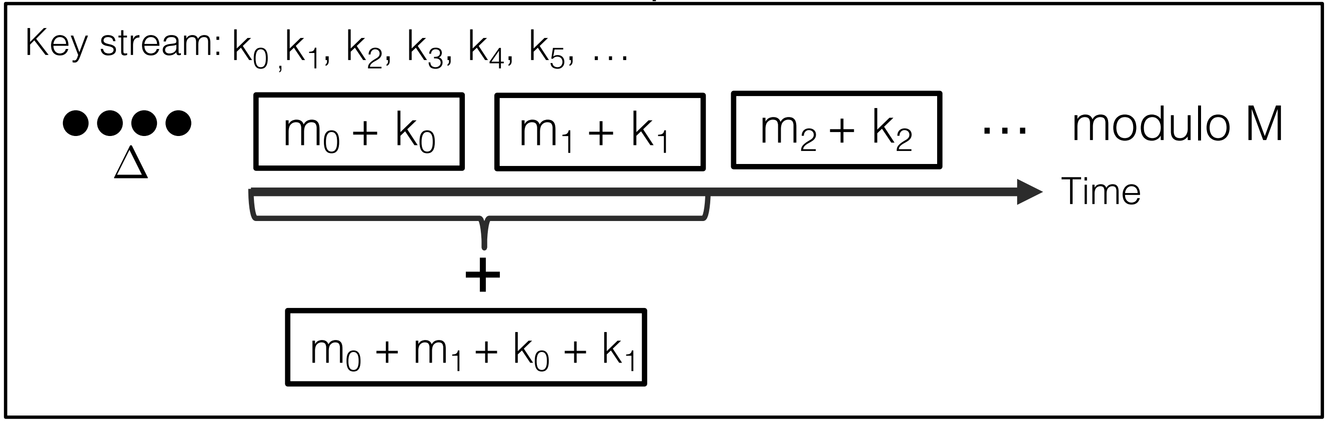
$\boxed{m_0 + m_1 + k_0 + k_1}$

Building Blocks

Efficiency

Additive Symmetric Homomorphic Encryption [Castelluccia]

- Dec. cost $O(n)$ → Key cancelling $O(1)$

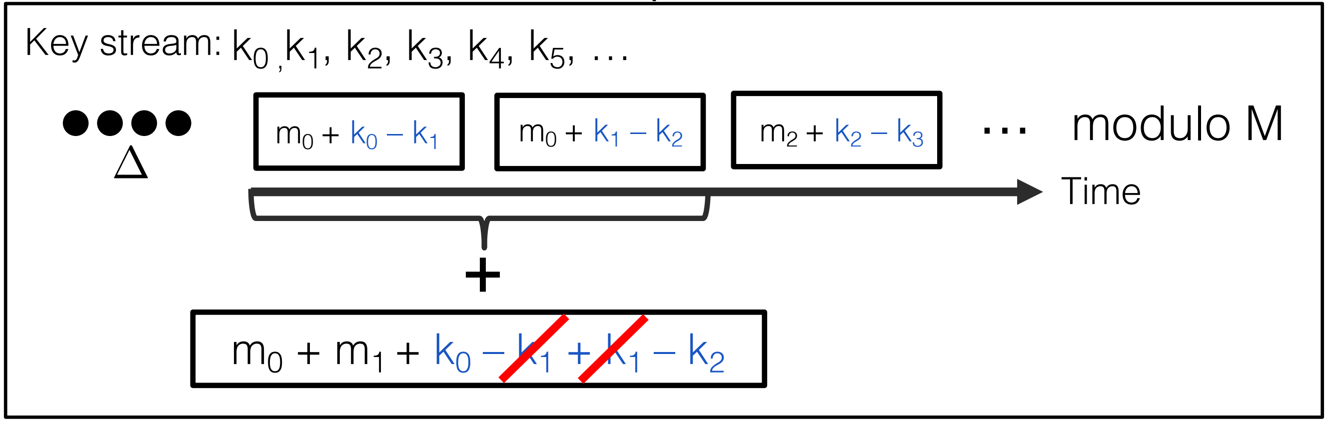


Building Blocks

Efficiency

Additive Symmetric Homomorphic Encryption [Castelluccia]

- Dec. cost O(n) → Key cancelling O(1)



Building Blocks

Efficiency

Additive Symmetric Homomorphic Encryption [Castelluccia]

- Dec. cost O(n) → Key cancelling O(1)
- Key Identifiers → Time-encoded key-streams

Key stream: $k_0, k_1, k_2, k_3, k_4, k_5, \ldots$

$m_0 + k_0 - k_1$ $\quad$ $m_0 + k_1 - k_2$ $\quad$ $m_2 + k_2 - k_3$ $\quad \cdots$ modulo M

Δ

Time

$$\underbrace{(m_0 + k_0 - k_1) \quad (m_0 + k_1 - k_2)}_{+}$$

$$m_0 + m_1 + k_0 - \cancel{k_1} + \cancel{k_1} - k_2$$

Building Blocks

Efficiency	Additive Symmetric Homomorphic Encryption [Castelluccia] ▪ Dec. cost O(n) → Key cancelling O(1) ▪ Key Identifiers → Time-encoded key-streams
Expressiveness	Aggregatable Digests

Building Blocks

Efficiency

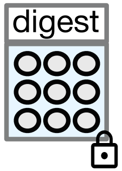
Additive Symmetric Homomorphic Encryption [Castelluccia]

- Dec. cost O(n) → Key cancelling O(1)
- Key Identifiers → Time-encoded key-streams

Expressiveness

Aggregatable Digests

Known encodings: If we can compute sum privately, then we can compute f(·) privately



← vector of encodings of the underling data

Building Blocks

Efficiency

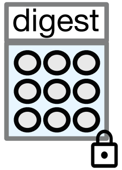
Additive Symmetric Homomorphic Encryption [Castelluccia]

- Dec. cost $O(n)$ → Key cancelling $O(1)$
- Key Identifiers → Time-encoded key-streams

Expressiveness

Aggregatable Digests

Known encodings: If we can compute sum privately, then we can compute $f(\cdot)$ privately



← vector of encodings of the underling data

Statistical queries: average, sum, count, variance, min/max, histograms, least-squares regression, stochastic gradient descent, heavy hitters …

Building Blocks

Efficiency

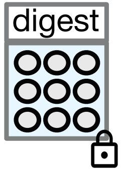
Additive Symmetric Homomorphic Encryption [Castelluccia]

- Dec. cost O(n) → Key cancelling O(1)
- Key Identifiers → Time-encoded key-streams

Expressiveness

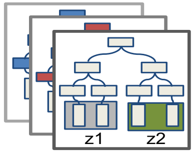
Aggregatable Digests

Known encodings: If we can compute sum privately, then we can compute f(·) privately



← vector of encodings of the underling data

Statistical queries: average, sum, count, variance, min/max, histograms, least-squares regression, stochastic gradient descent, heavy hitters …



Encrypted in-memory statistical index

Building Blocks

Efficiency	Additive Symmetric Homomorphic Encryption [Castelluccia] ▪ Dec. cost O(n) → Key cancelling O(1) ▪ Key Identifiers → Time-encoded key-streams
Expressiveness	Aggregatable Digests Known encodings: If we can compute sum privately, then we can compute f(·) privately
Access Control	New Key Derivation Construction

Building Blocks

Efficiency

Additive Symmetric Homomorphic Encryption [Castelluccia]

- Dec. cost O(n) → Key cancelling O(1)
- Key Identifiers → Time-encoded key-streams

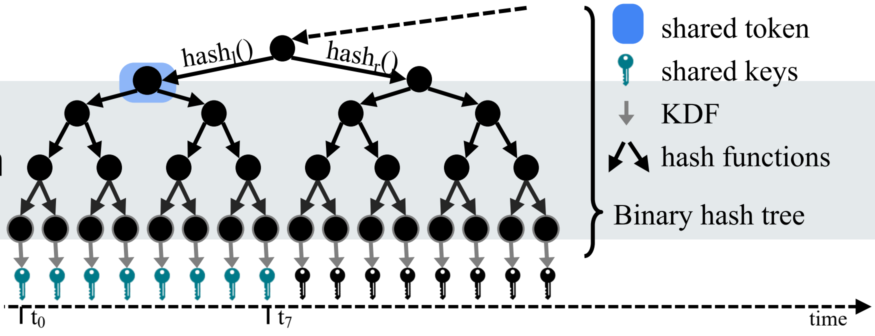
Expressiveness

Aggregatable Digests

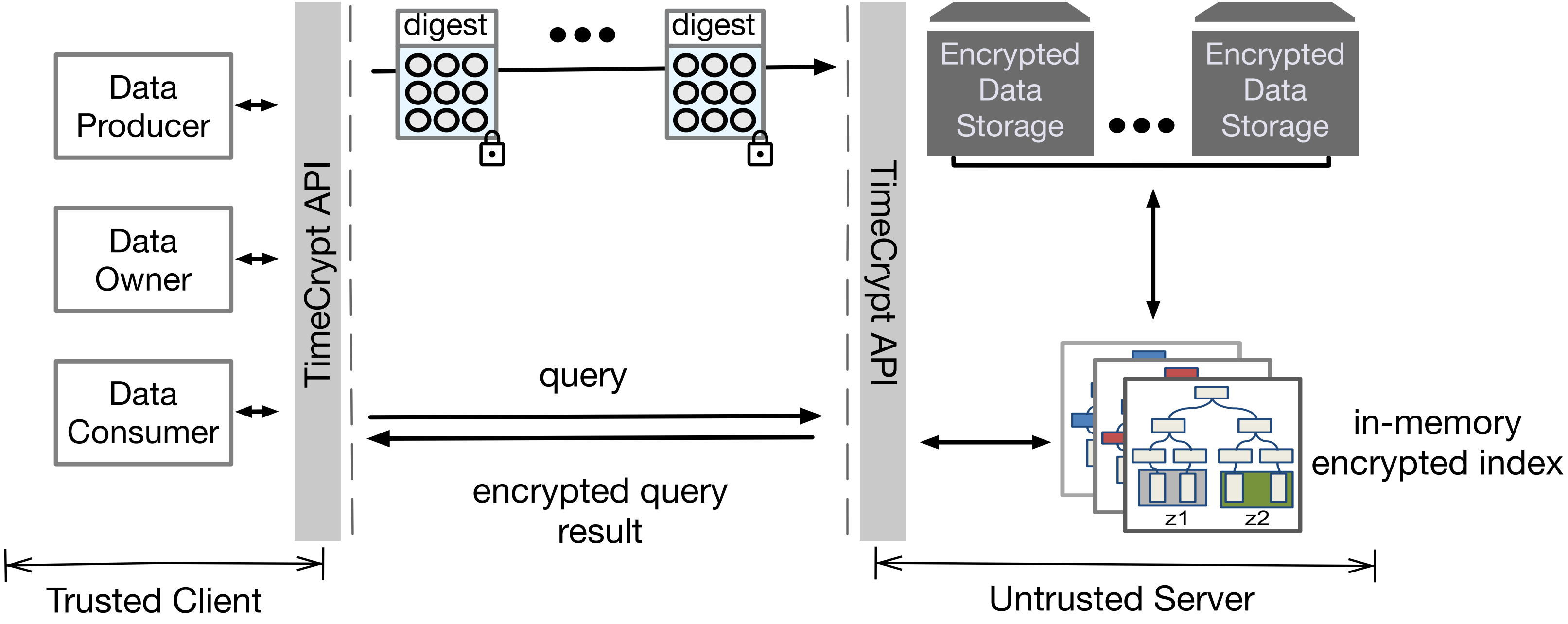
Known encodings: If we can compute sum privately, then we can compute f(·) privately

Access Control

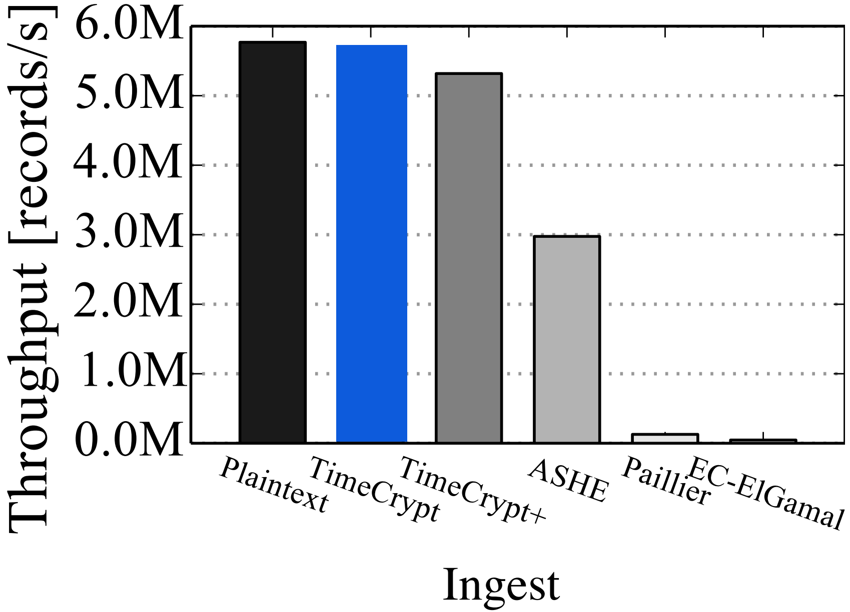
New Key Derivation Construction



TimeCrypt System Architecture



System Performance

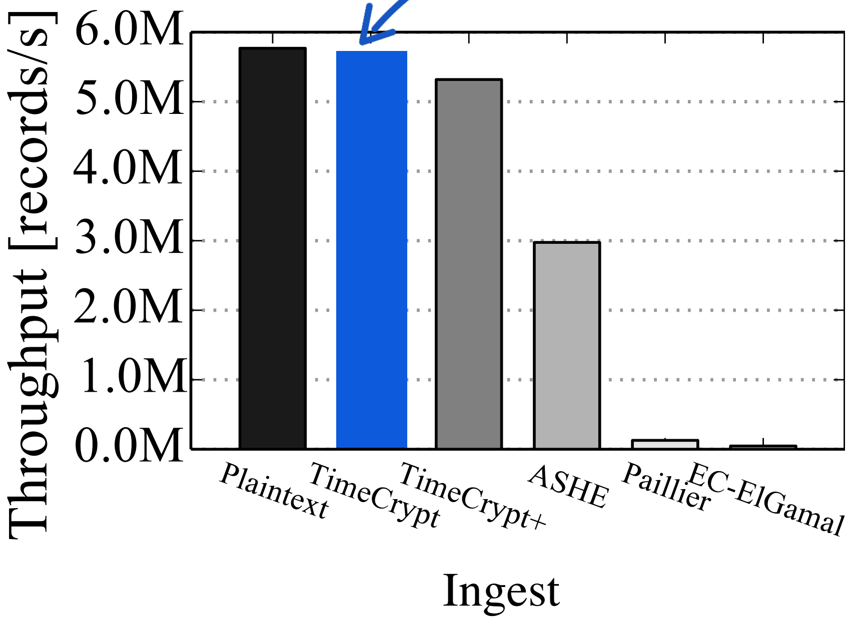


Throughput under load of

4/1 read-write ratio, 49k streams

System Performance

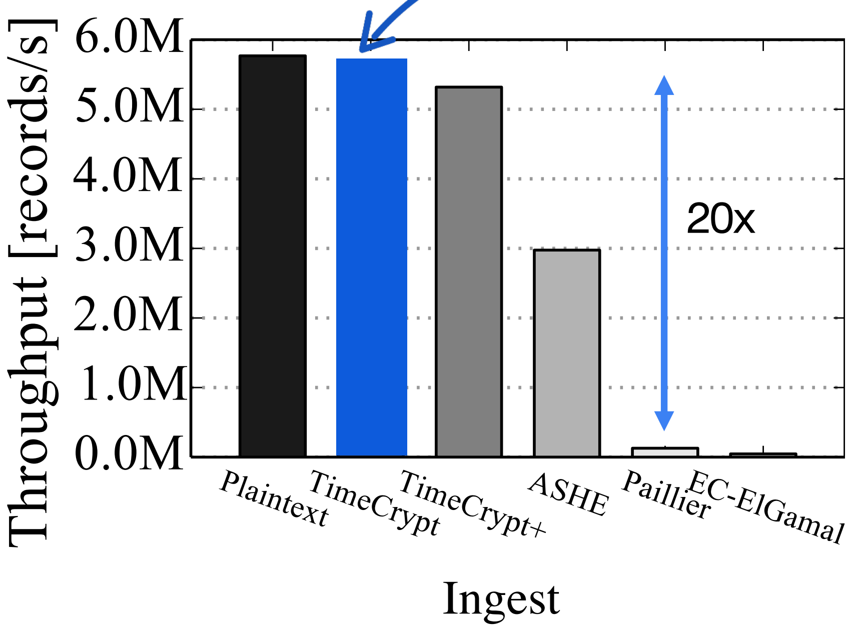
2% slowdown compared to plaintext



Throughput under load of
4/1 read-write ratio, 49k streams

System Performance

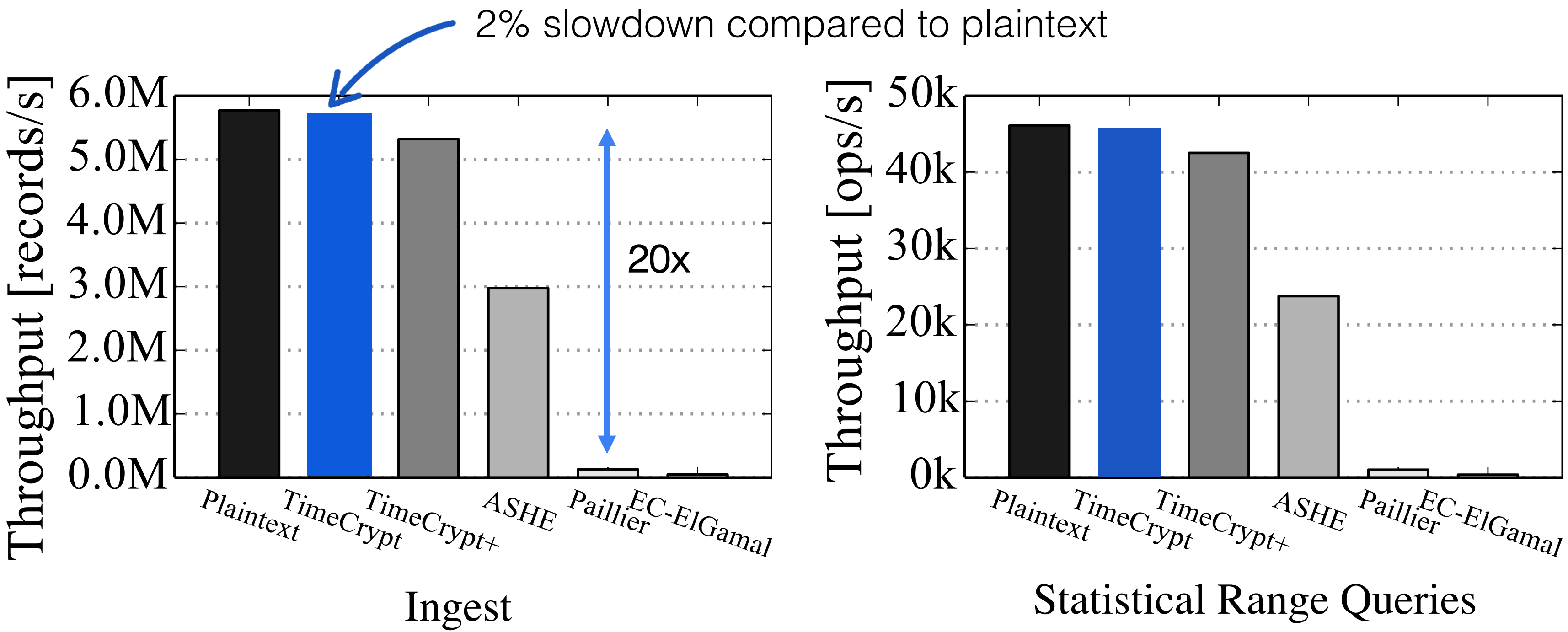
2% slowdown compared to plaintext



Throughput under load of

4/1 read-write ratio, 49k streams

System Performance

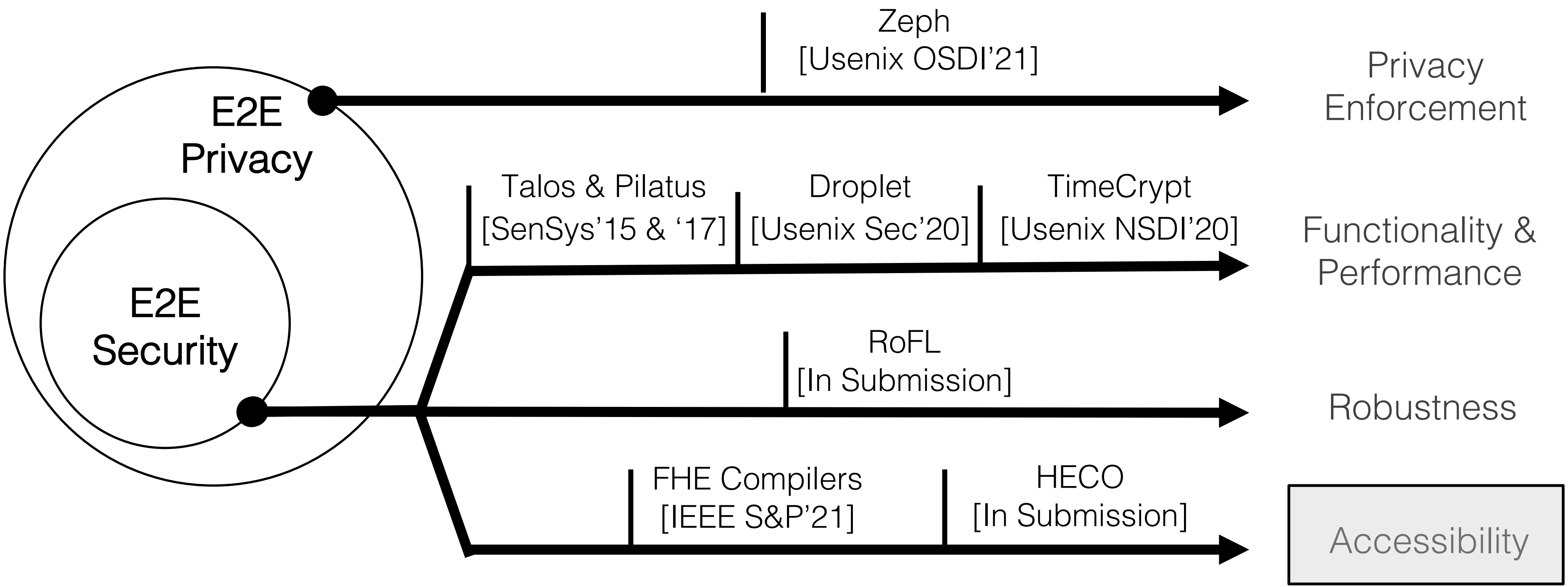


Throughput under load of

4/1 read-write ratio, 49k streams

My Research:

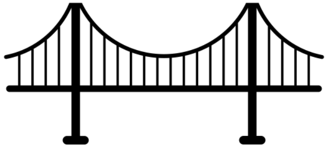
Building practical systems that use cryptography to empower users and preserve their privacy & tools to democratize cryptography



Fully Homomorphic Encryption Accessibility

(IEEE S&P '21)

Advanced Cryptography



Programming Languages

FHE holds huge potential to transforming privacy

Finally "practical" - Real world use of FHE started to emerge

Microsoft Edge
Password Monitor

NuCypher

ZAMA

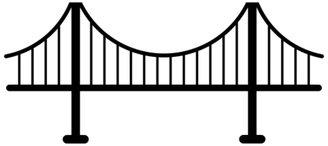
Duality

Developing FHE Applications is

Notoriously Hard

Usable FHE

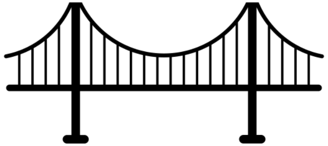
Advanced Cryptography



Programming Languages

Usable FHE

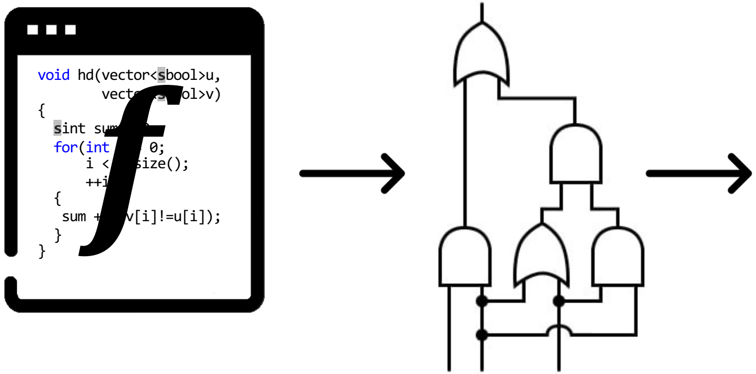
Advanced Cryptography



Programming Languages

1 What makes developing FHE applications hard?

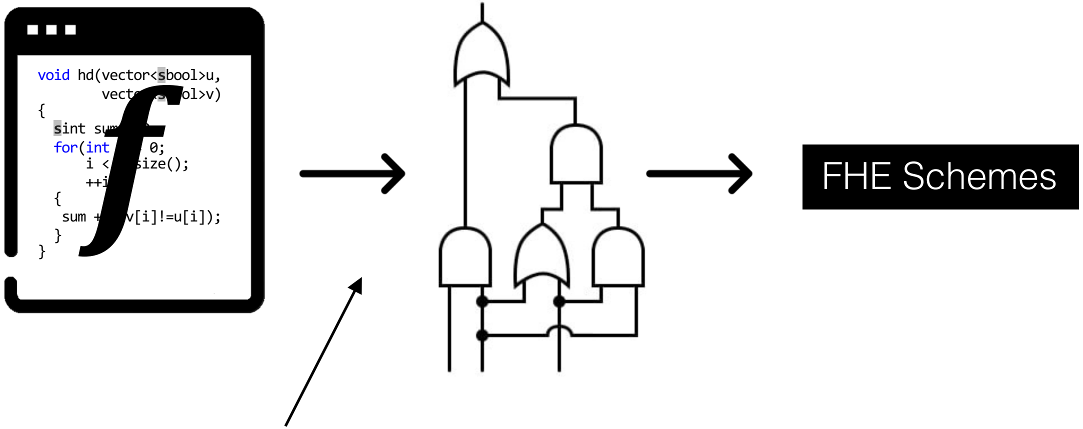
2 How can compilers address these complexities?



FHE Schemes

Functionality and performance depend on $\boldsymbol{f}$'s representation:

- How do we express $\boldsymbol{f}$
- How do we optimize $\boldsymbol{f}$

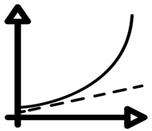


Optimizing this transformation yields better FHE efficiency

Functionality and performance depend on $\boldsymbol{f}$'s representation:

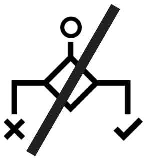
- How do we express $\boldsymbol{f}$
- How do we optimize $\boldsymbol{f}$

FHE Programming Paradigm

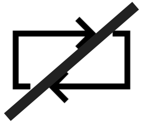


Approximations

Optimizations



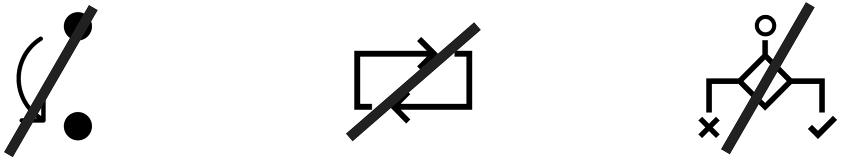
No If/Else



No Loops

SIMD Batching

Data Independence



No Jumps | No Loops | No If/Else

Standard C++

```
int foo(int a, int b) {

   if(a < b) {
      return a * b;
   } else {
      return a + b;
   }


}
```

FHE

```
int foo(int a, int b) {

   int c = a < b;
   int i = a * b;
   int e =  a + b;
   return c*i + (1-c)*e;


}
```

Always worst-case performance

SIMD-like Parallelism

SIMD Batching

Standard C++

```
int foo(int[] x,int[] y){

   int[] r;
   for(i = 0; i < 6; ++i){
      r[i] = x[i] * y[i]
   }
   return r;

}
```

Batched FHE

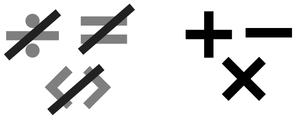
```
int foo(int[] a,int[] b){

   return a * b;

}
```

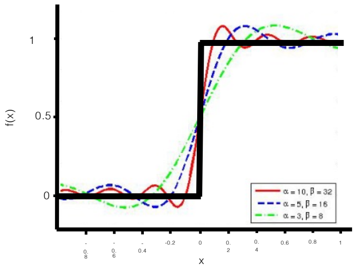
Could get orders-of-magnitude performance difference between different batching schemes.

Complex Design Space

Schemes: BFV, BGV, CKKS

Polynomial Functions

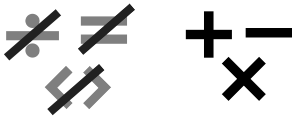


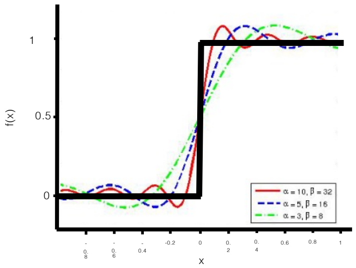


Complex Design Space

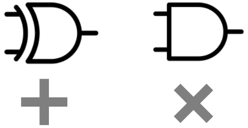
Schemes: BFV, BGV, CKKS

Polynomial Functions

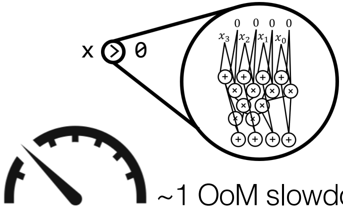




Arbitrary Computation



Schemes: FHEW, TFHE

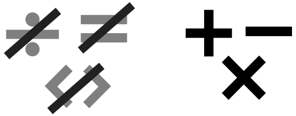


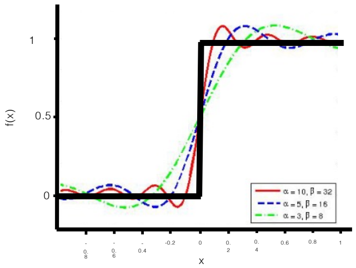
~1 OoM slowdown

Complex Design Space

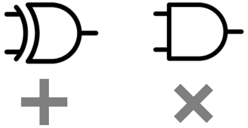
Polynomial Functions

Schemes: BFV, BGV, CKKS

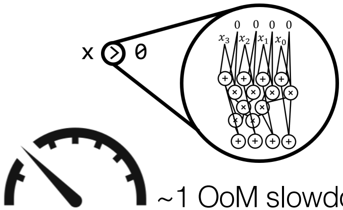




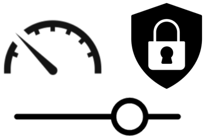
Arbitrary Computation



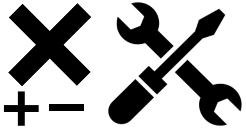
Schemes: FHEW, TFHE



~1 OoM slowdown



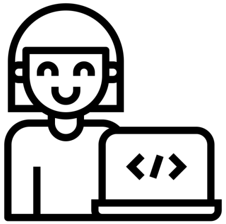
Parameter Selection



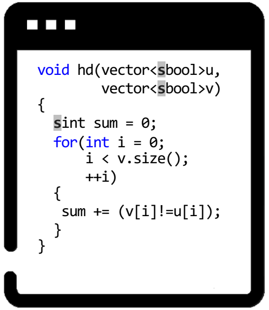
Cost Model

Performance	More complex than overhead of underlying FHE operations
Programming Paradigm	Wide gap between naïve implementations & expert solutions
Compilers	Transform high level programs to efficient FHE circuits

Democratizing Fully Homomorphic Encryption



Developer with no crypto expertise



Automatically generate efficient and secure FHE for any custom workloads?

HECO

FHE Paradigm

Transform high-level programs to efficient FHE solutions

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

HECO

FHE Paradigm

Transform high-level programs to efficient FHE solutions

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

HECO

FHE Paradigm

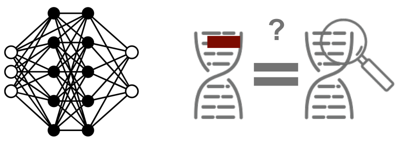
Transform high-level programs to efficient FHE solutions

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

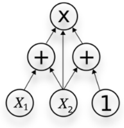
Architecture

End-to-end compilation stack for FHE

Application



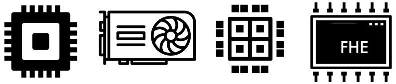
Circuits



Schemes

CKKS, TFHE, BGV, BFV, …

Platforms



HECO

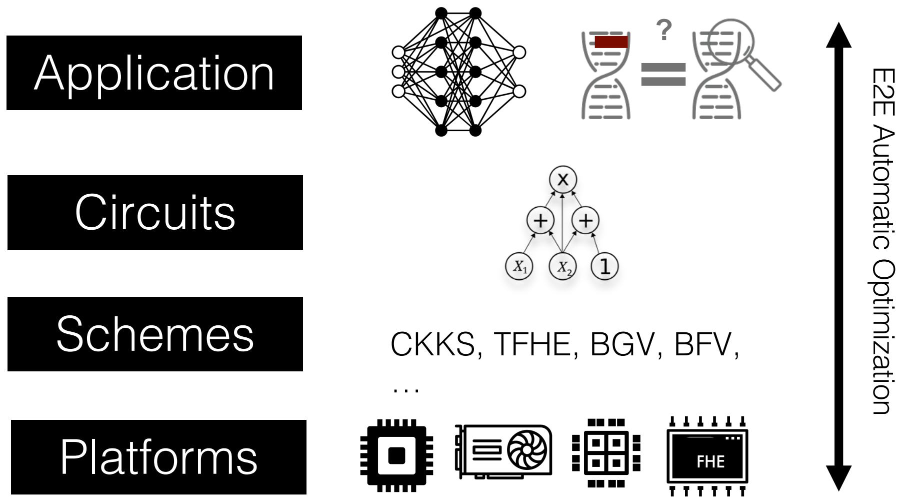
FHE Paradigm

Transform high-level programs to efficient FHE solutions

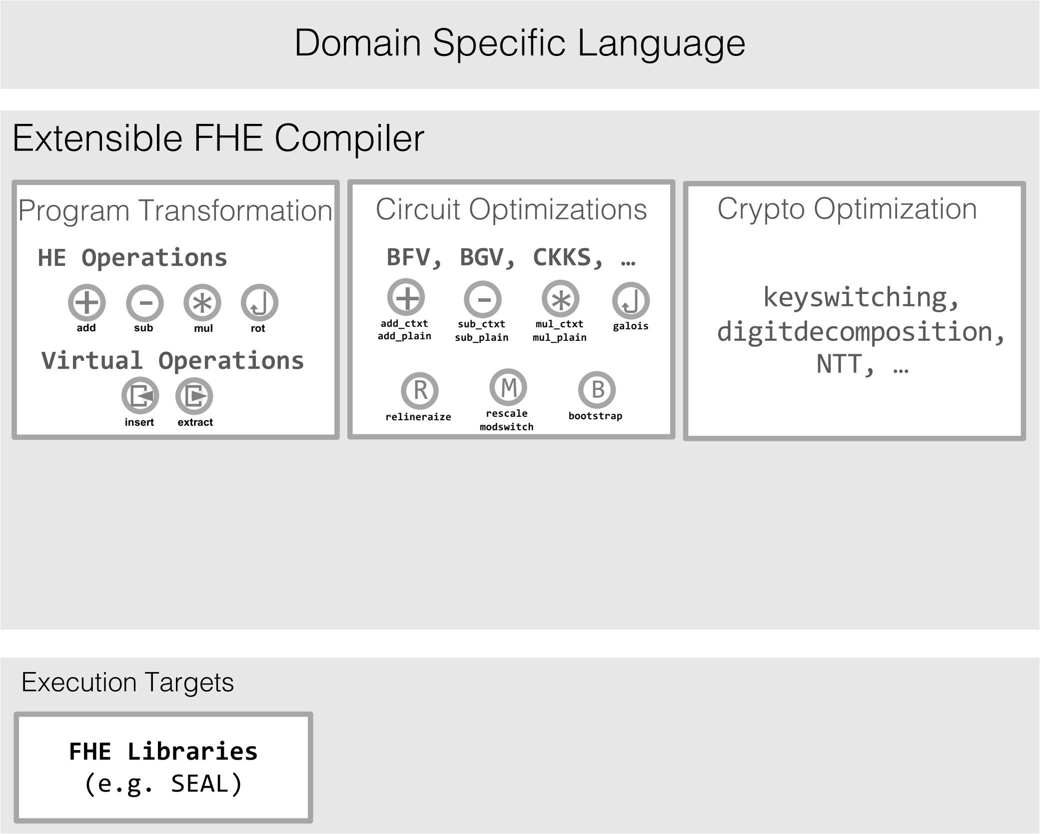
```
void hd(vector<sbool>u,
        vector<sbool>v)
{
  sint sum = 0;
  for(int i = 0;
      i < v.size();
      ++i)
  {
   sum += (v[i]!=u[i]);
  }
}
```

Architecture

End-to-end compilation stack for FHE

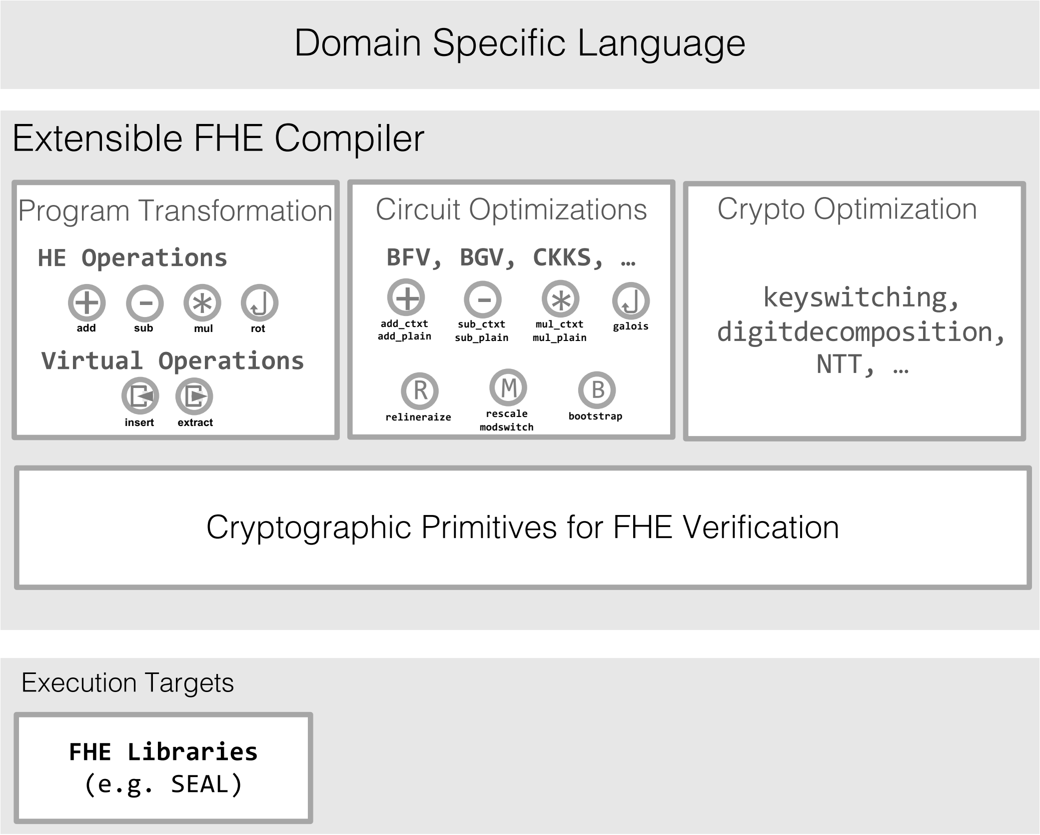


What’s Next?



HECO

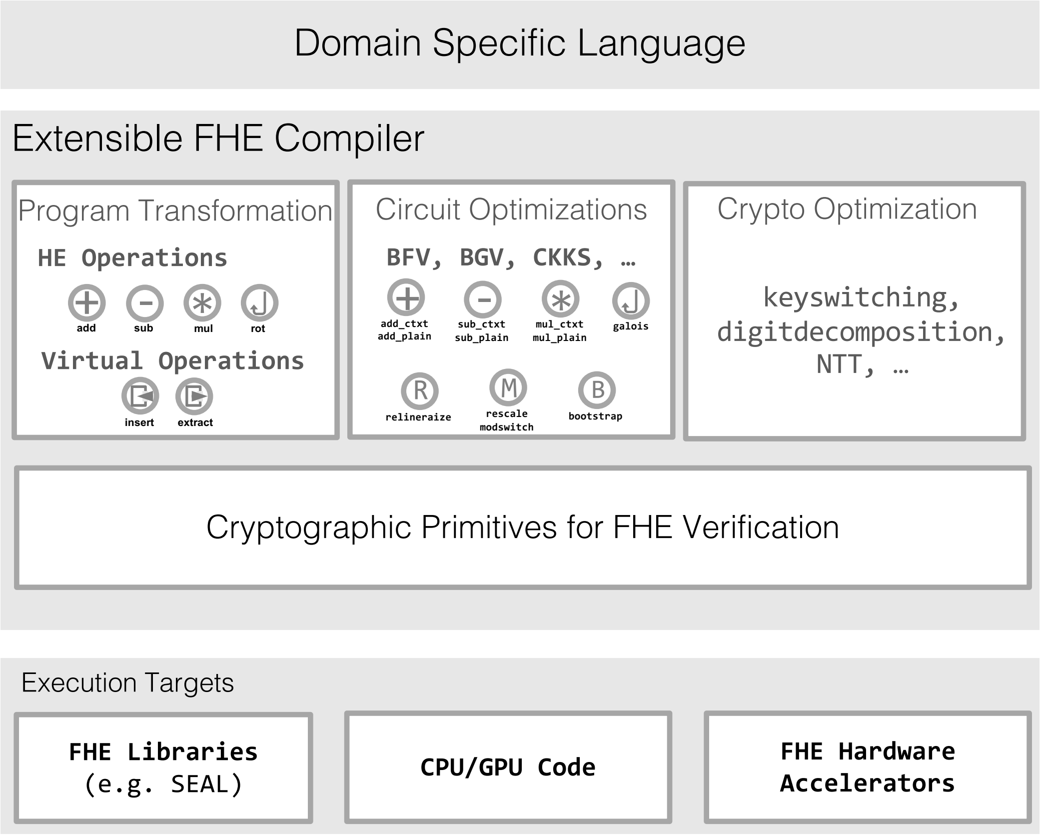
open source, **automated** end-to-end optimization for FHE



HECO

open source, **automated** end-to-end optimization for FHE

Cryptography : Primitives for Verifiable Computation



HECO

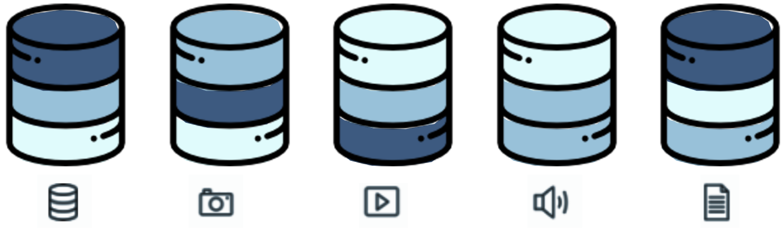
open source, **automated** end-to-end optimization for FHE

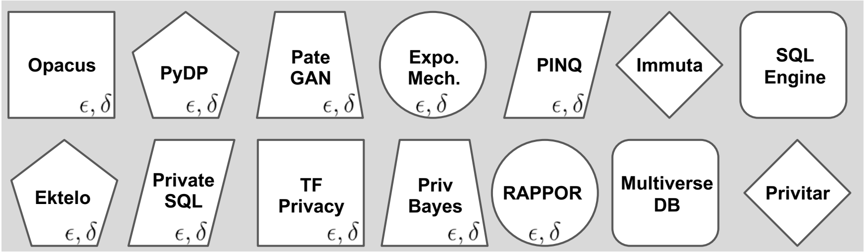
Cryptography : Primitives for Verifiable Computation

Systems: **Target HW directly**, generating code for CPU/GPU, upcoming dedicated FHE accelerators and heterogenous deployments using a mix of these.

End-to-End Privacy

Data with Heterogenous Privacy Restrictions

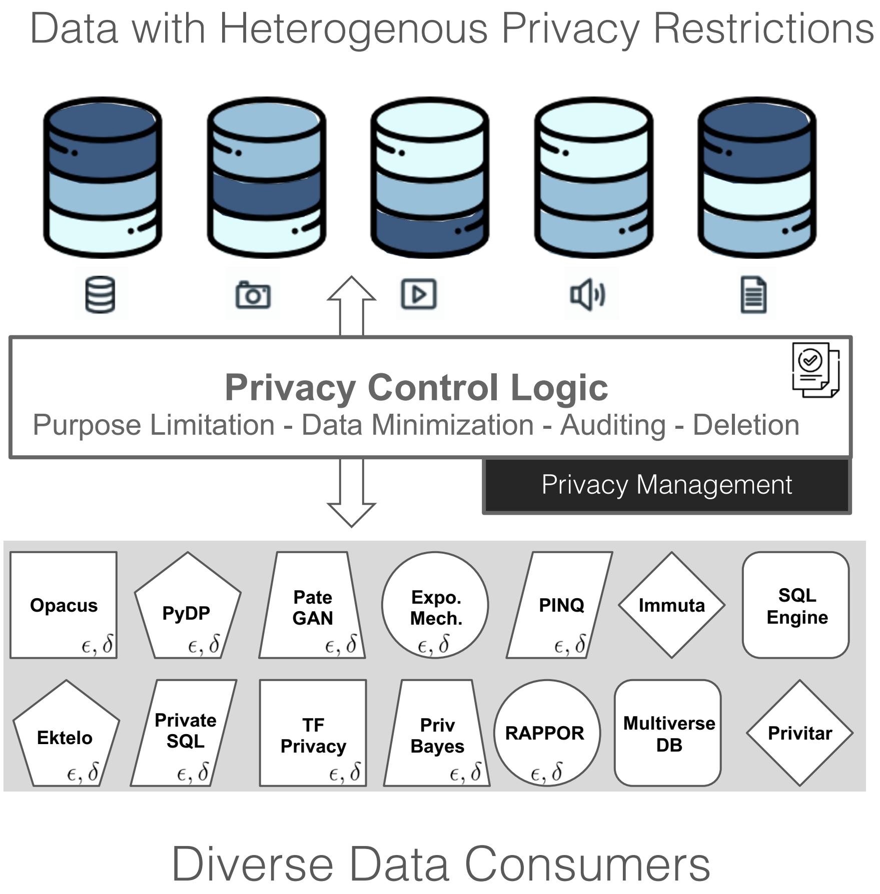




Diverse Data Consumers

End-to-End Privacy

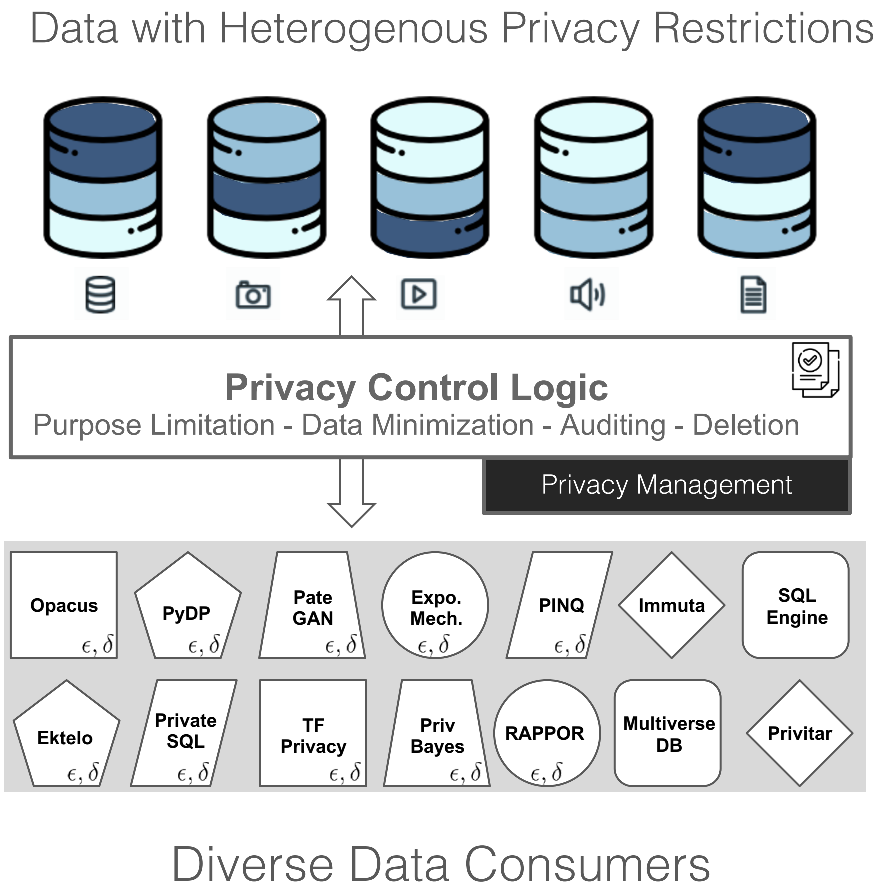
Privacy Management



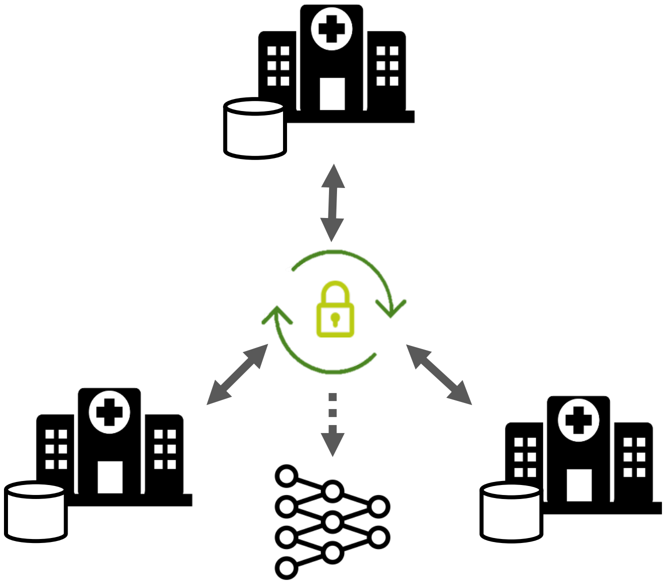
End-to-End Privacy

Privacy Management

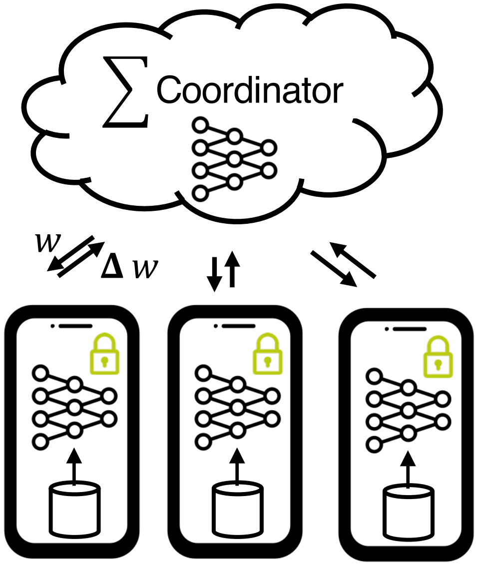
Cryptographically
Enforced Privacy



Secure and Robust Collaborative Learning



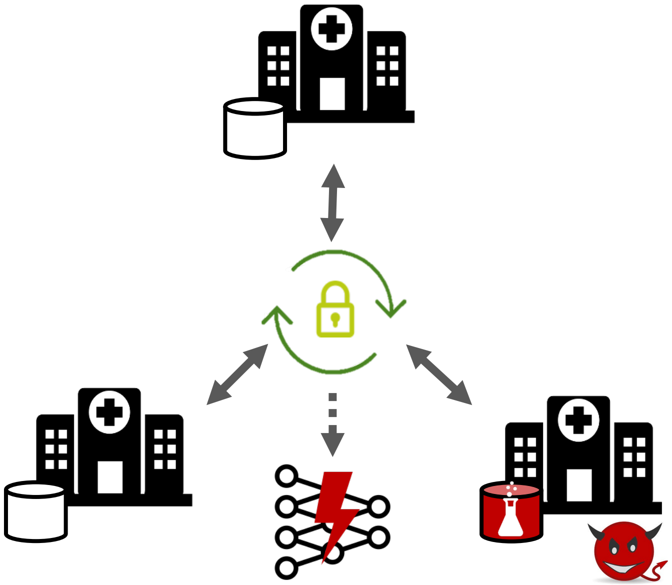
Secure Decentralized Learning



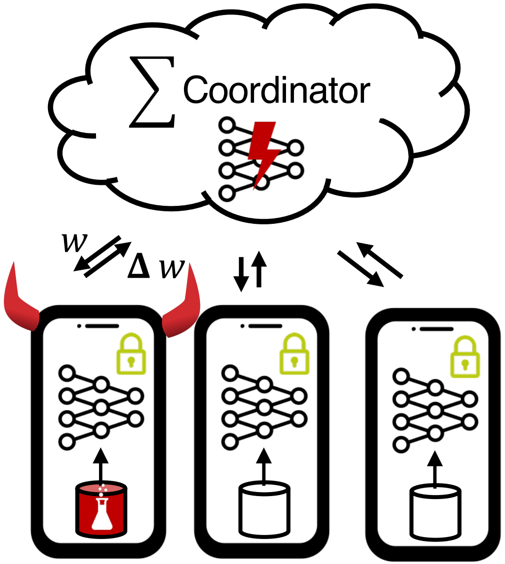
Secure Federated Learning

Secure and Robust Collaborative Learning

Problem: Model integrity

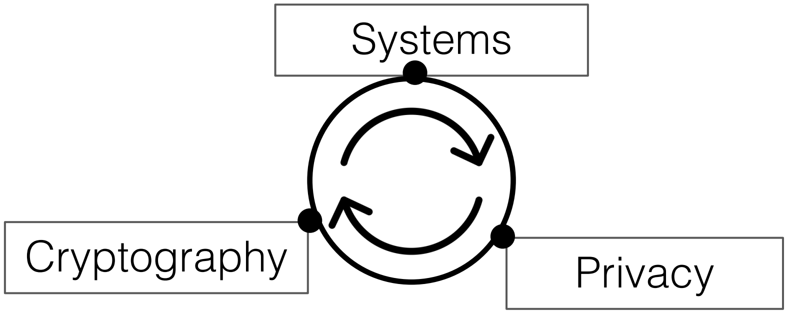


Secure Decentralized Learning



Secure Federated Learning

Retrofit privacy in the fabric of modern systems



Privacy-preserving, functional, and performant systems

End.