OSDI 21

ETH zürich

Zeph: Cryptographic Enforcement of End-to-End Data Privacy

Lukas Burkhalter*

Nicolas Küchler*

Alexander Viand

Hossein Shafagh

Anwar Hithnawi

~ **2.5** million terabytes

estimate of data generated
per day in 2020

… much of this data is **personal**

data-driven world ...

… data breaches, data misuse

… privacy laws,
user awareness

Data Privacy Landscape

Compliance

”notice and consent”

Privacy-Enhancing Technologies

”ad-hoc solutions”

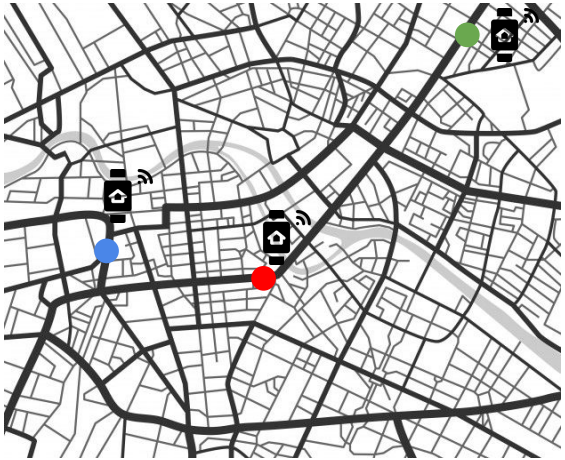
Zeph: Cryptographic Enforcement of End-to-End Data Privacy

User-centric Model for Privacy

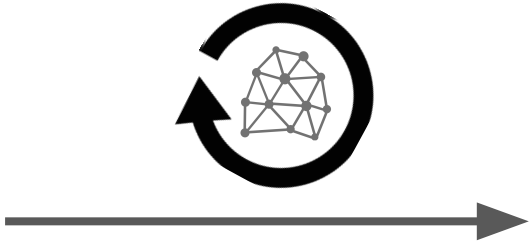
Cryptographically **Enforces** Privacy

One of Many Scenarios

“Raw Location Data”



Privacy Transformation



Policy

- Temporal
- Spatial
- Population

“Daily popular running tracks”

Day 6

Privacy Transformations

Supported in **Zeph**

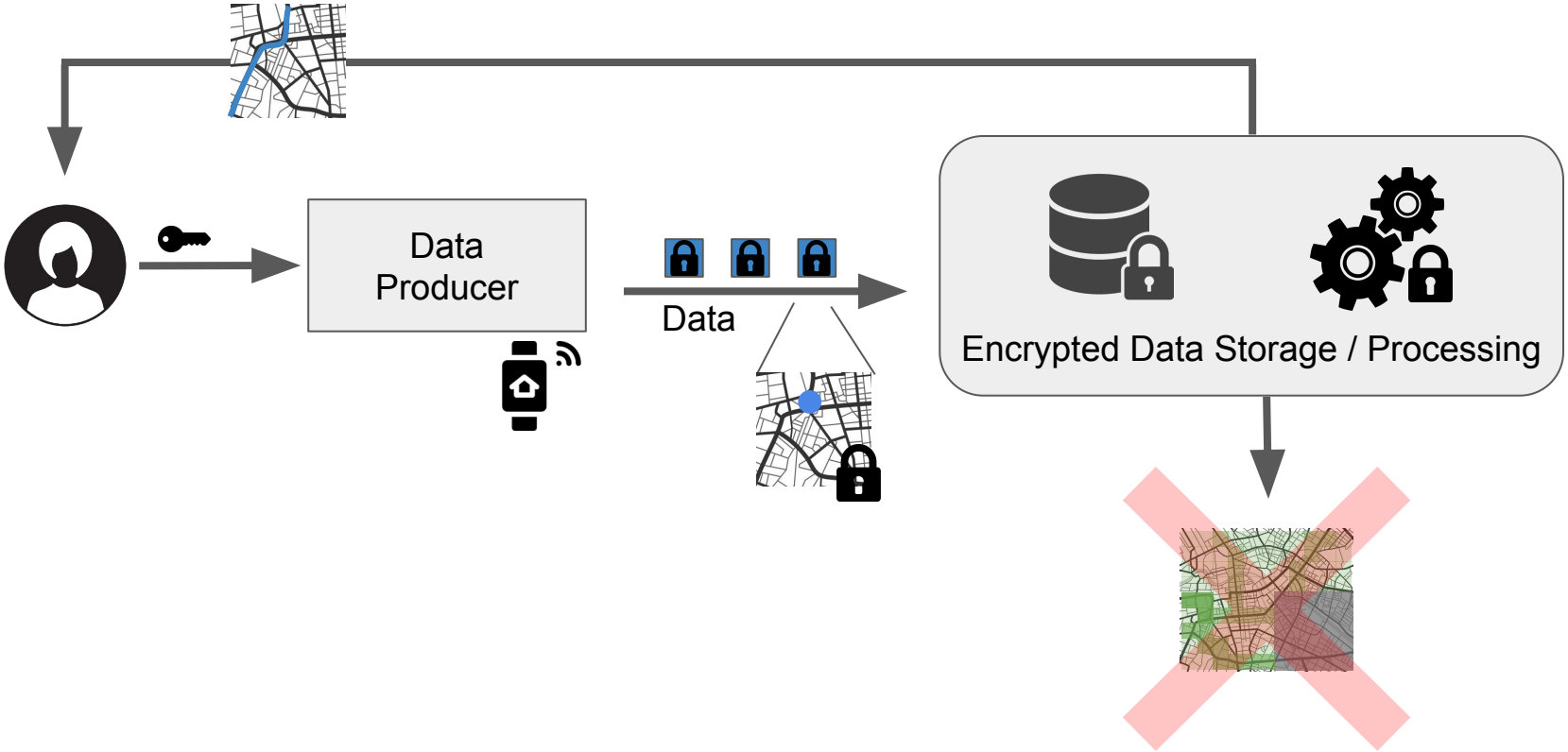
"Practical" Privacy Tools

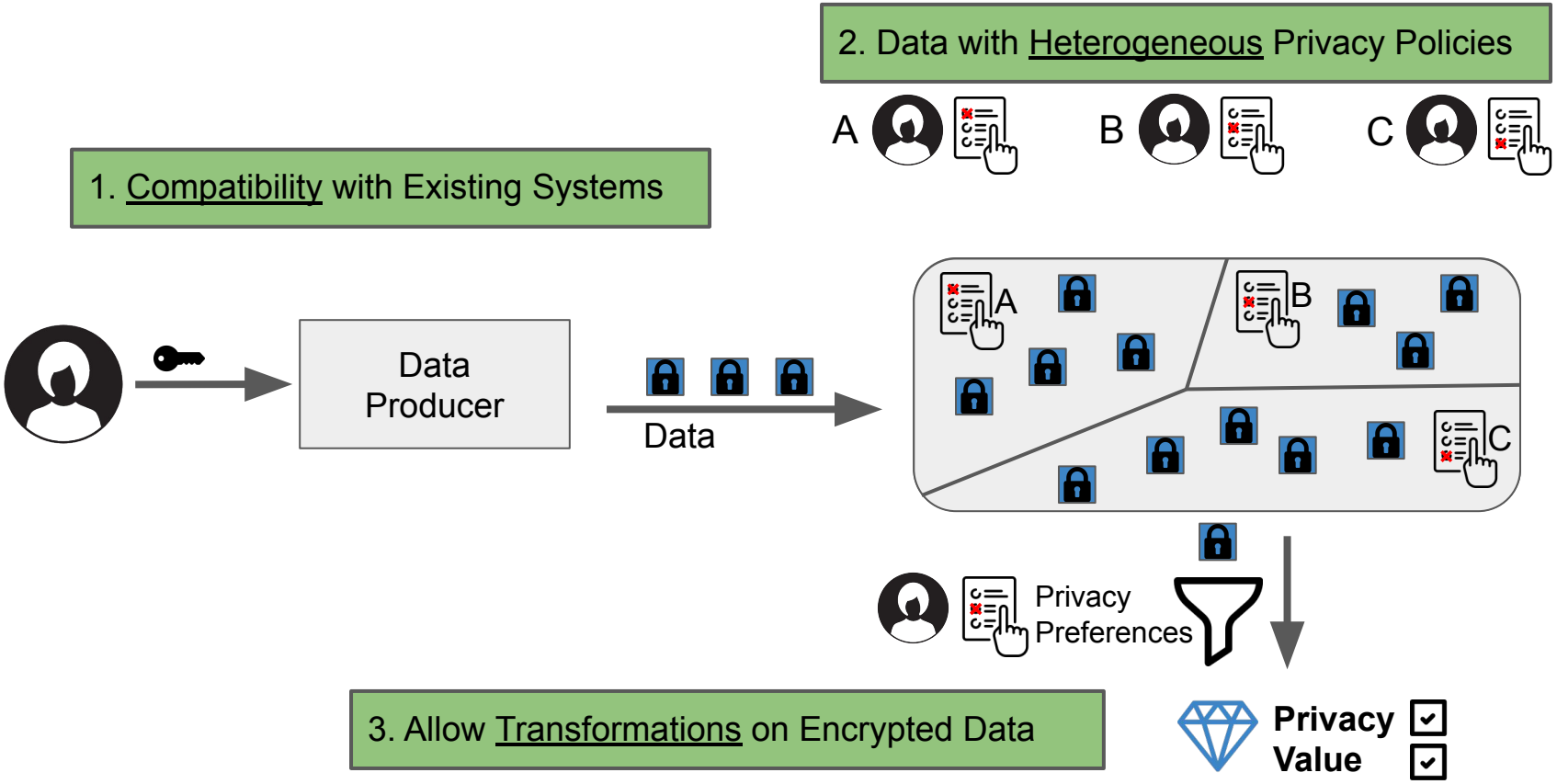
- Data Masking
  - Pseudonymization
  - Tokenization
  - Perturbation ✓
  - Redaction ✓
  - Date Shifting ✓
- Data Generalization
  - Bucketing ✓
  - Time Resolution ✓
  - Population ✓

Formal Privacy Models

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

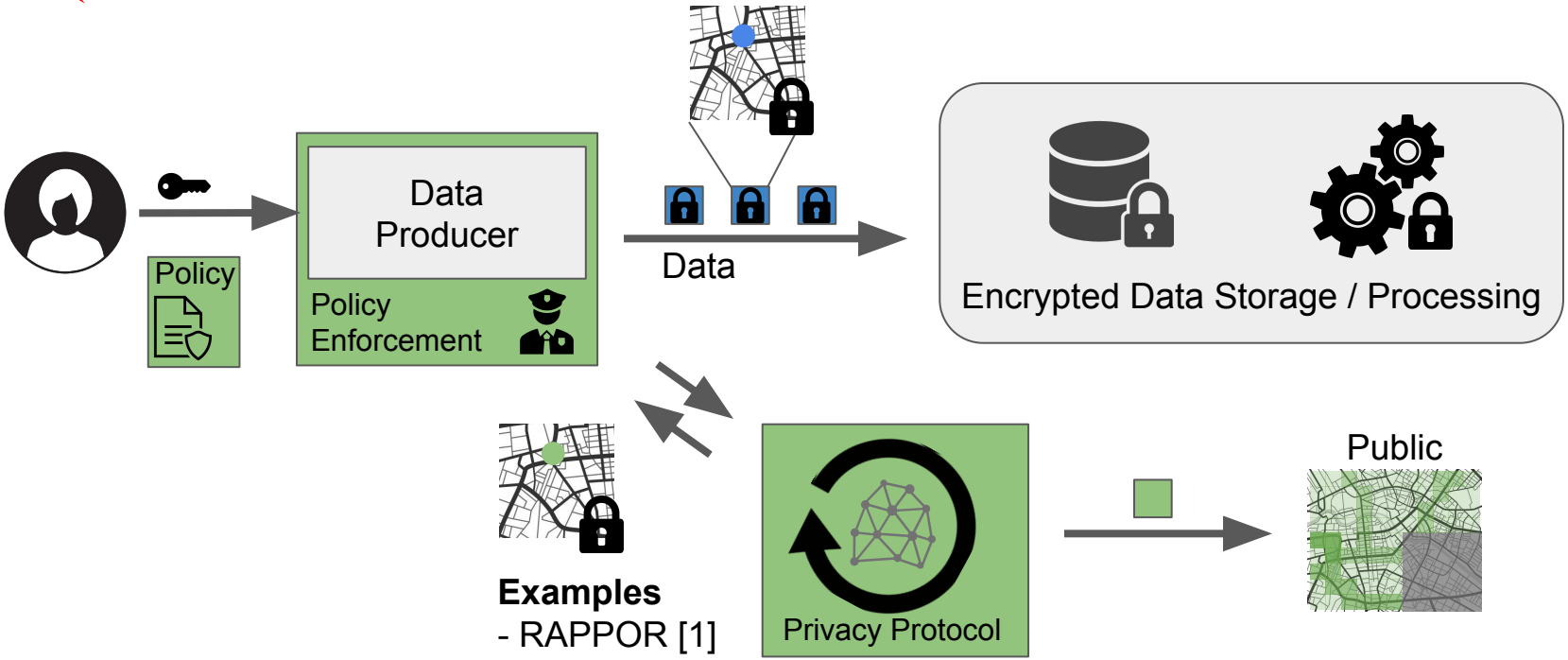
Existing End-to-End Encrypted Streaming Pipeline




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

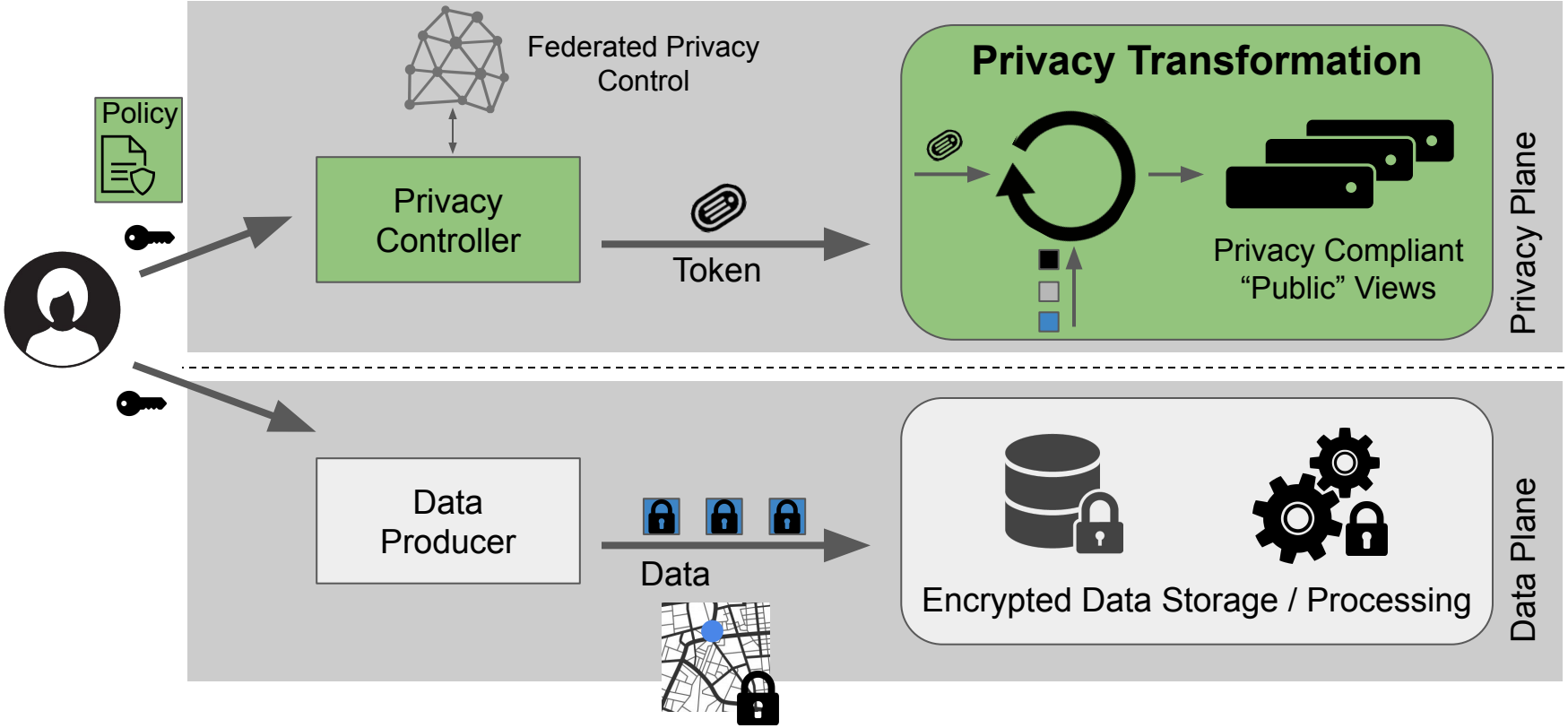
Integrate Privacy Controls into Existing Pipelines

Complex and ad hoc solutions

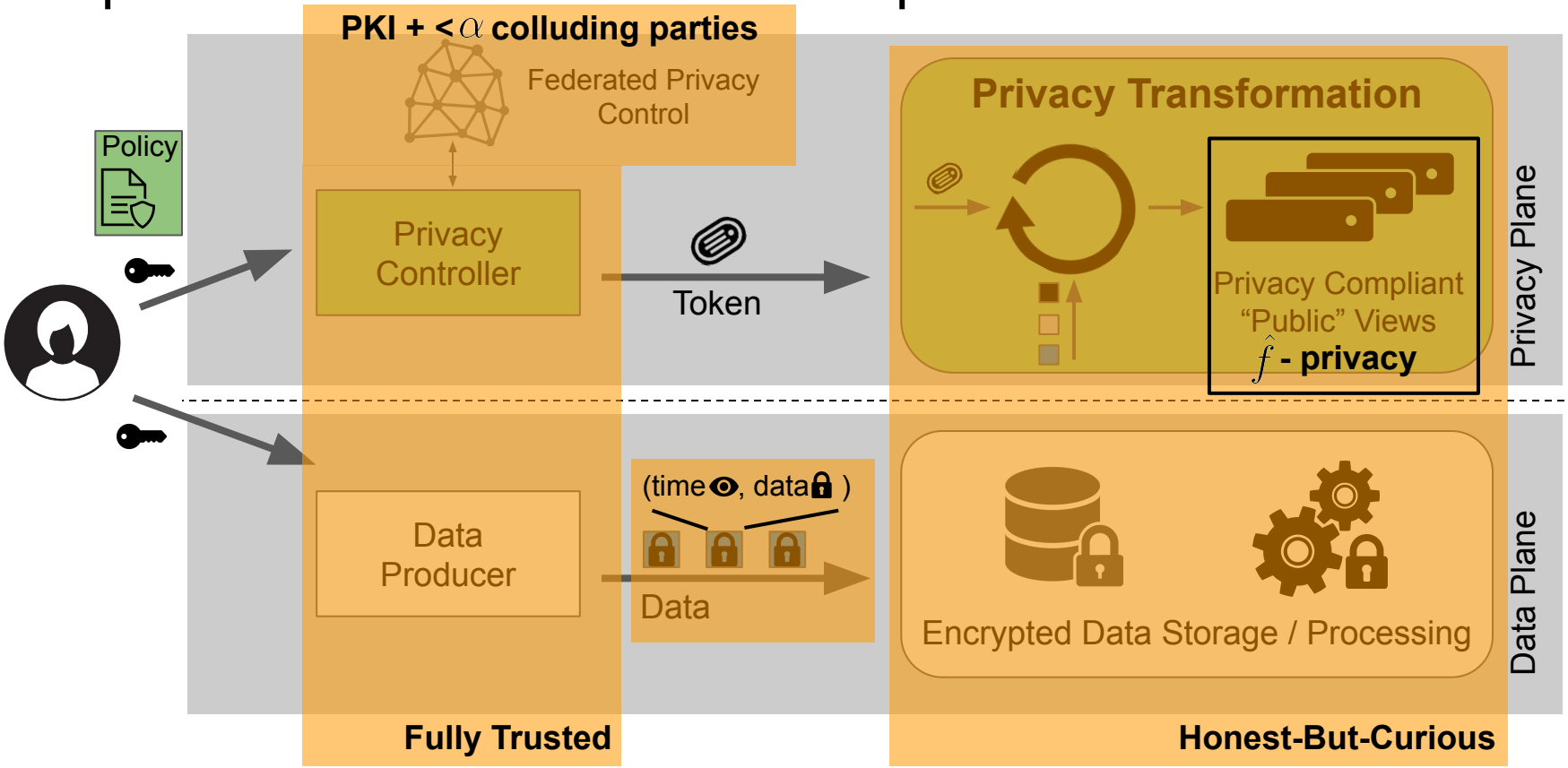


Examples
- RAPPOR [1]
- PRIO [2]
...

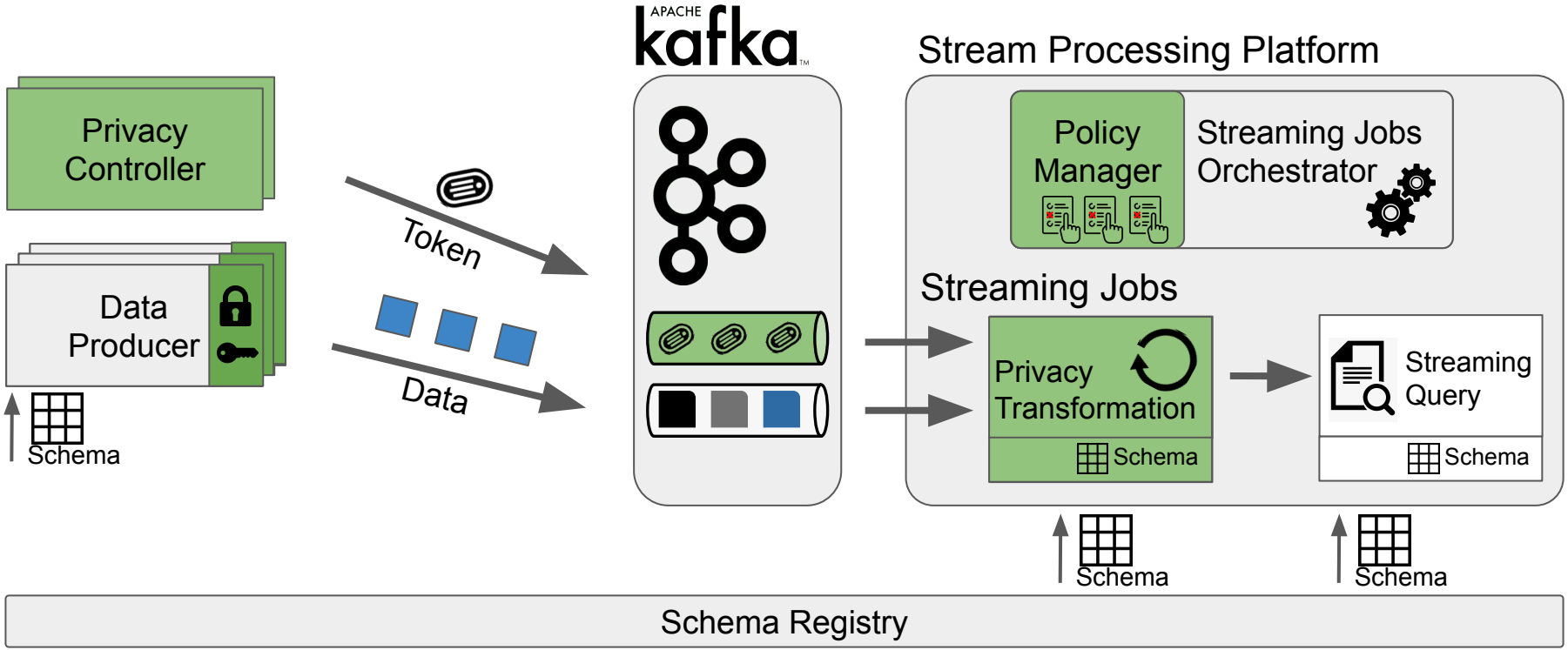
Zeph's End-to-End Approach to Privacy



Zeph's Threat Model and Assumptions



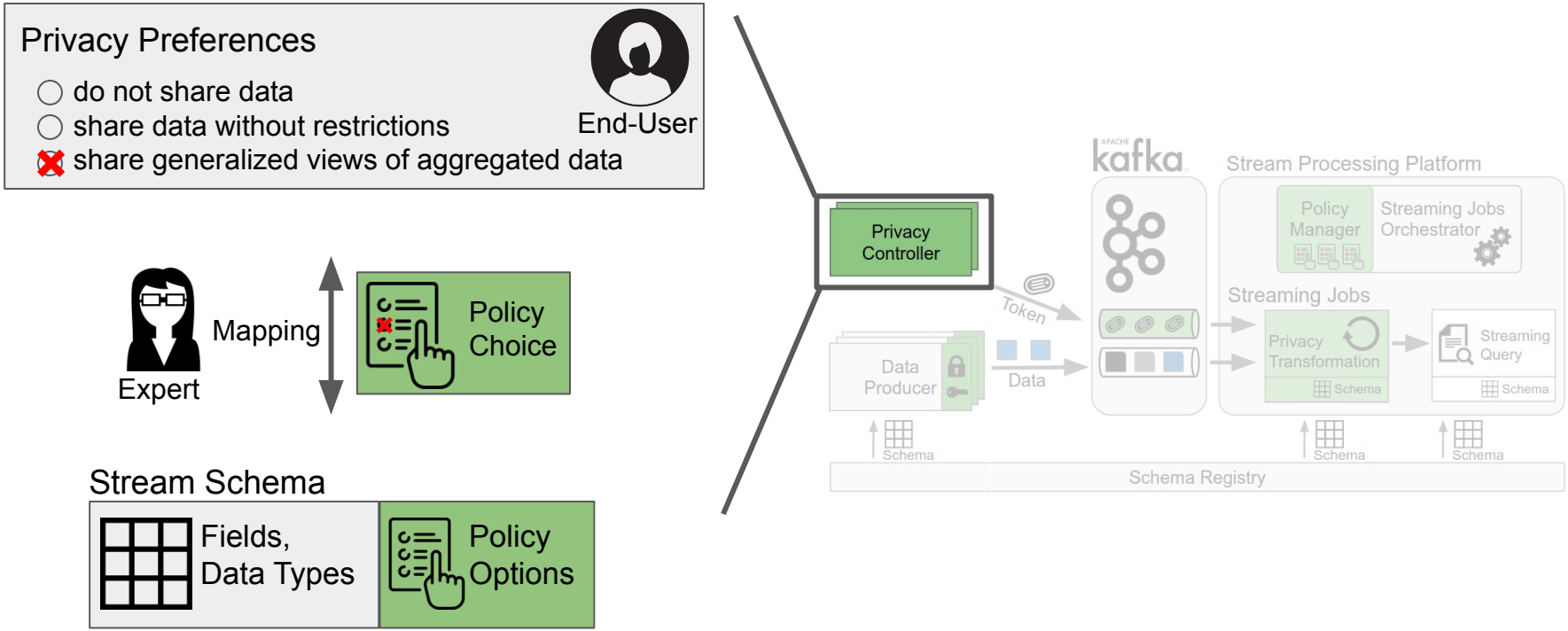
How Zeph augments existing System Designs



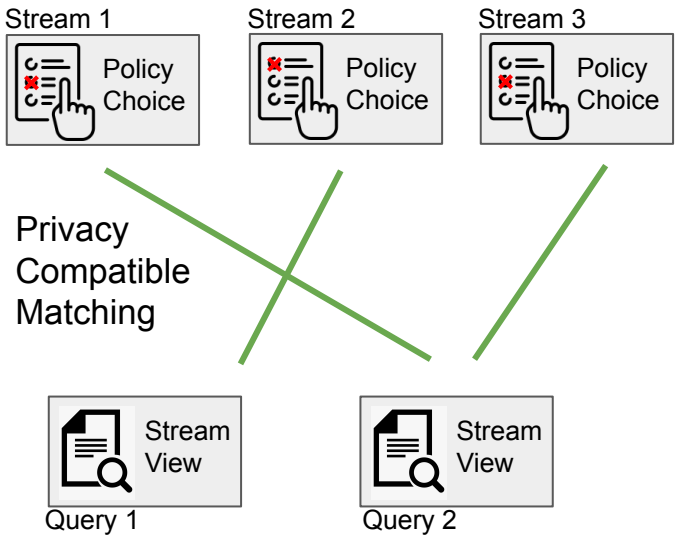
From Privacy Policies to **End-to-End Privacy**

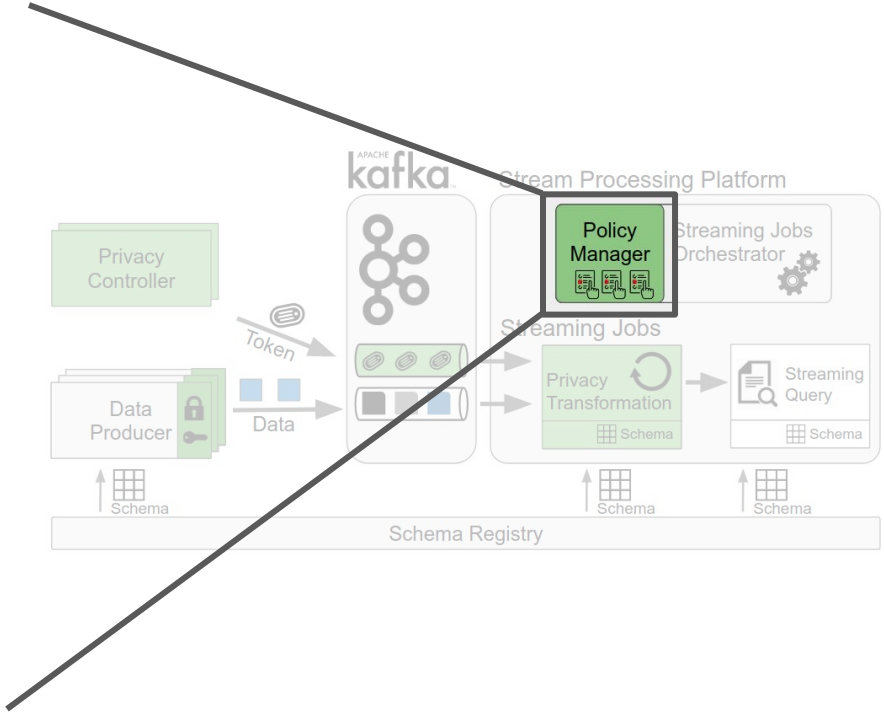
Challenges

Challenge #1: Keep End-User Control Simple



Challenge #2: Organize Privacy Transformations





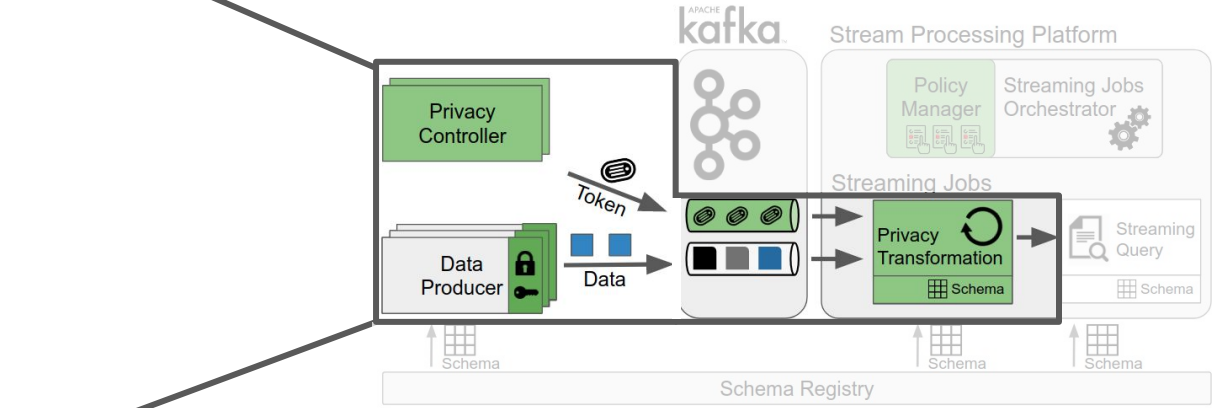
Challenge #3: Meeting Privacy Transformation Requirements

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

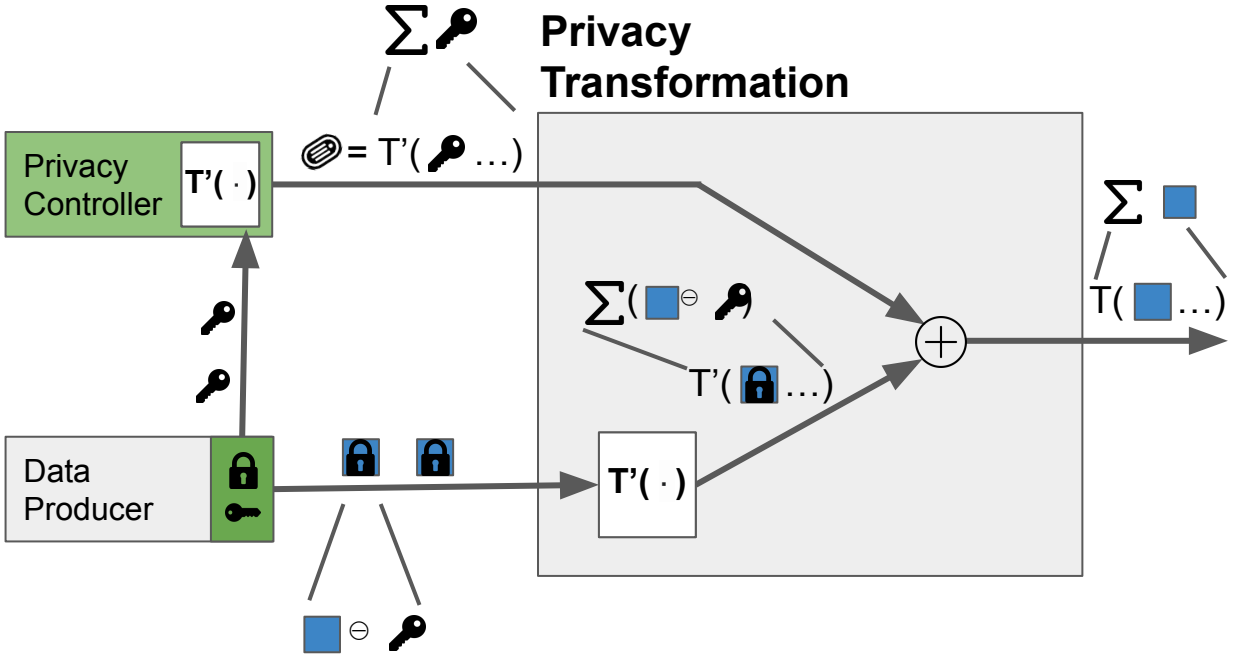
Challenge #3: Meeting Privacy Transformation Requirements

1) **Confidentiality** of data ☑

2) Transformation **Authorization** by Privacy Controller ☑

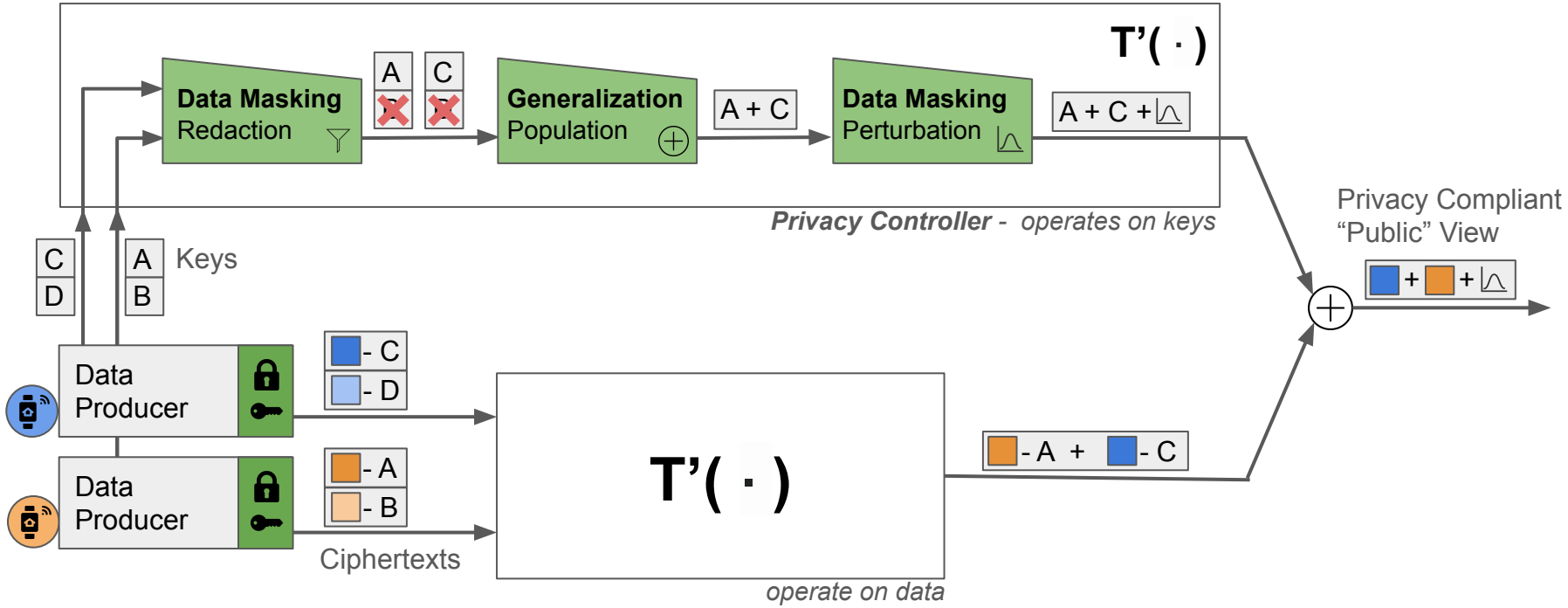
3) **Compute** transformation on confidential data

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



Additive Homomorphic Privacy Transformations

Challenge #3: Meeting Privacy Transformation Requirements



Additive Homomorphic Secret Sharing

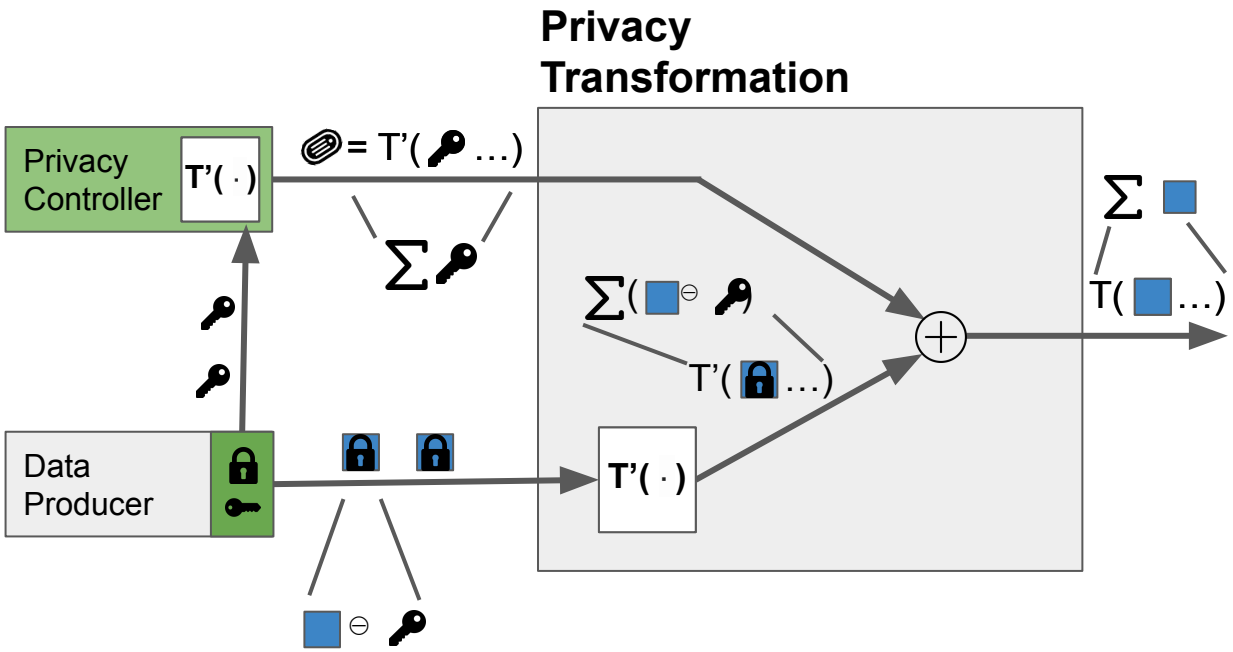
Challenge #3: Meeting Privacy Transformation Requirements

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



Independent and Efficient Privacy Controller

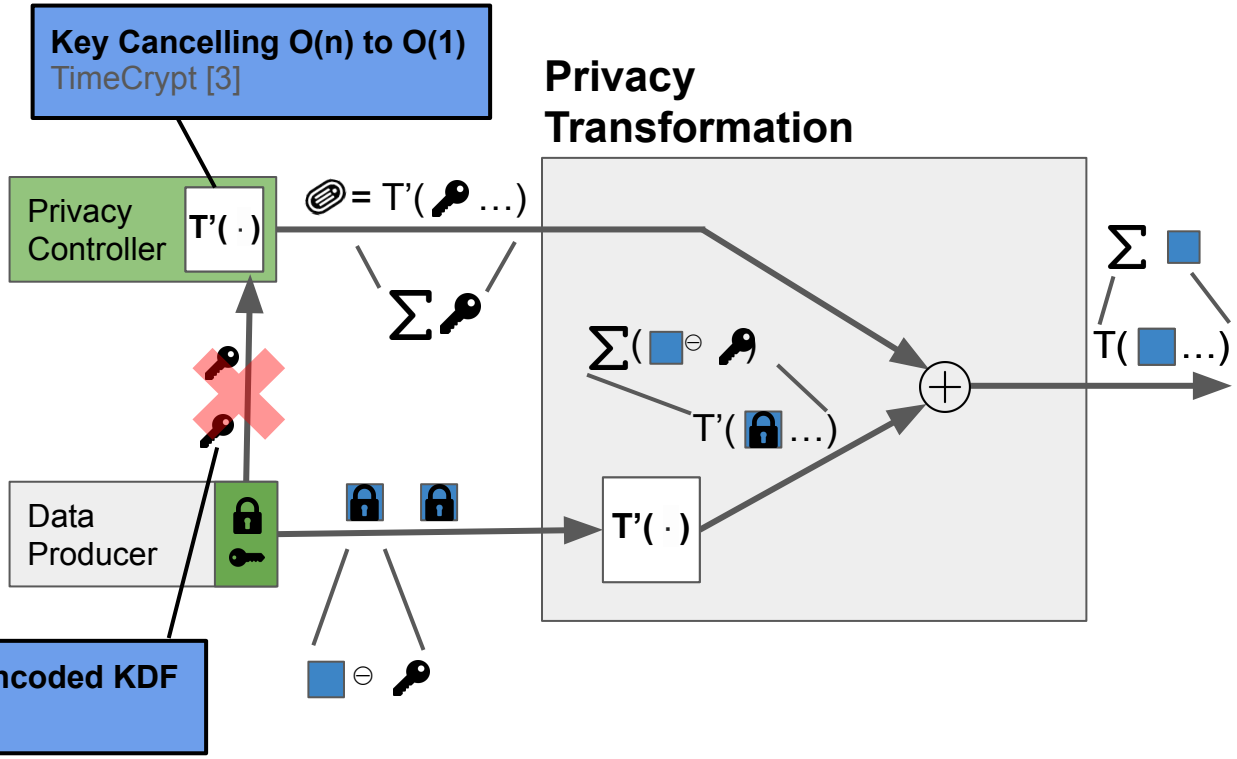
Challenge #3: Meeting Privacy Transformation Requirements

1) **Confidentiality** of data ☑

2) Transformation **Authorization** by Privacy Controller ☑

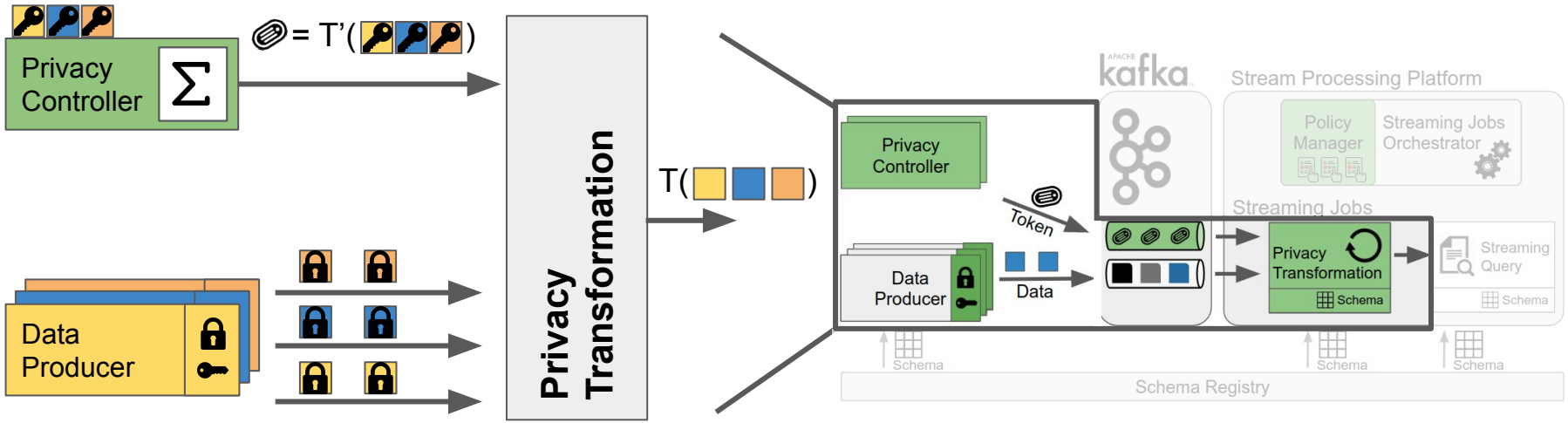
3) **Compute** transformation on confidential data ☑

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



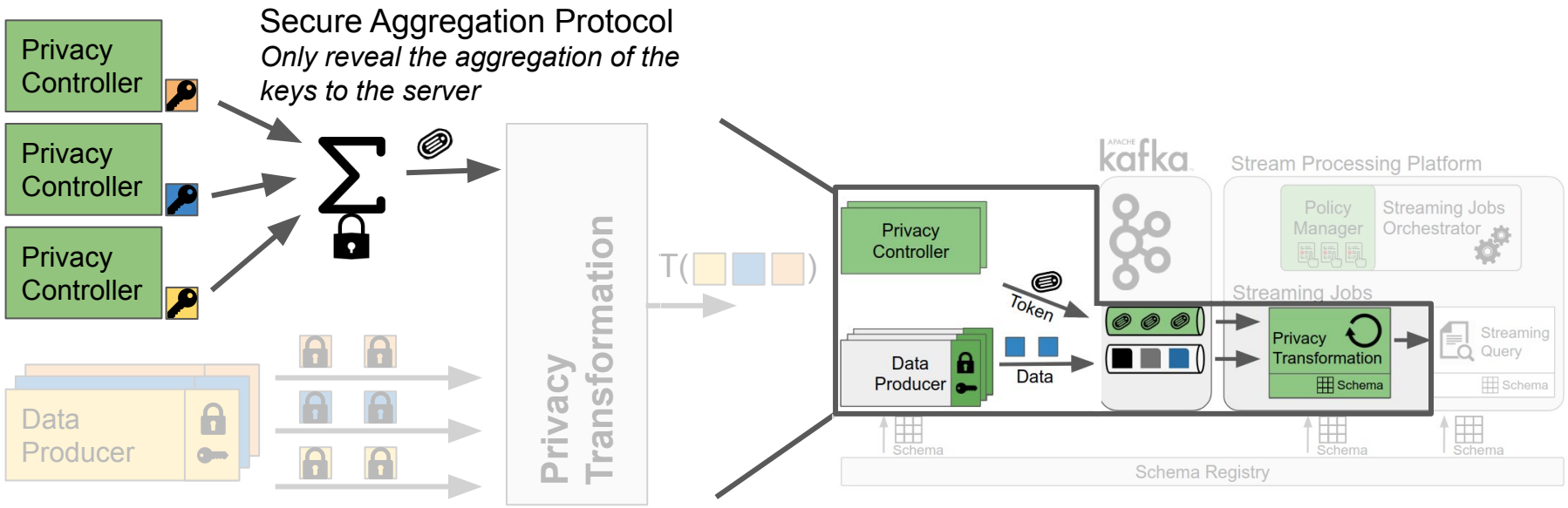
Challenge #4: Enable Federated Privacy Control

"multiple Data Producers - **one** Privacy Controller"

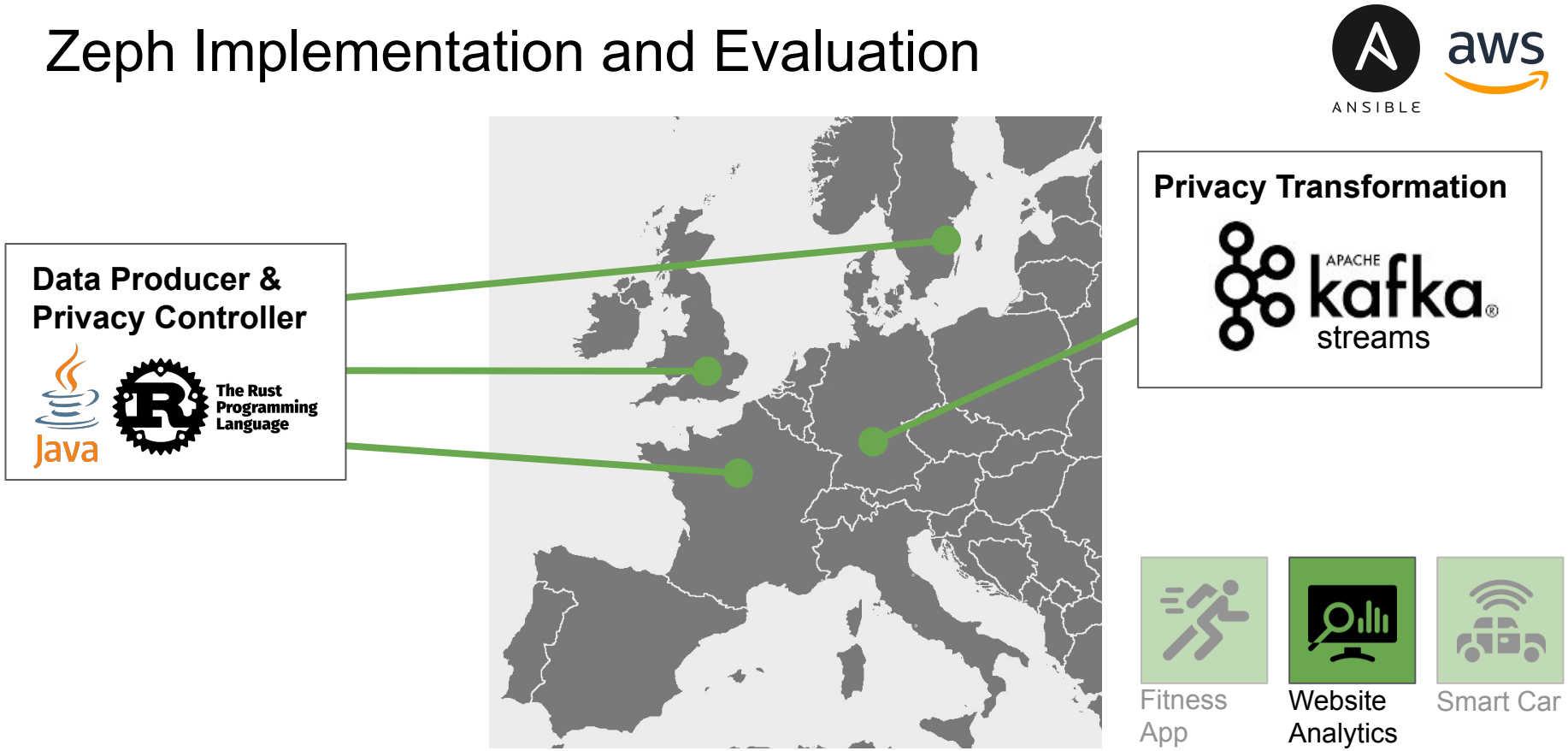


Challenge #4: Enable Federated Privacy Control

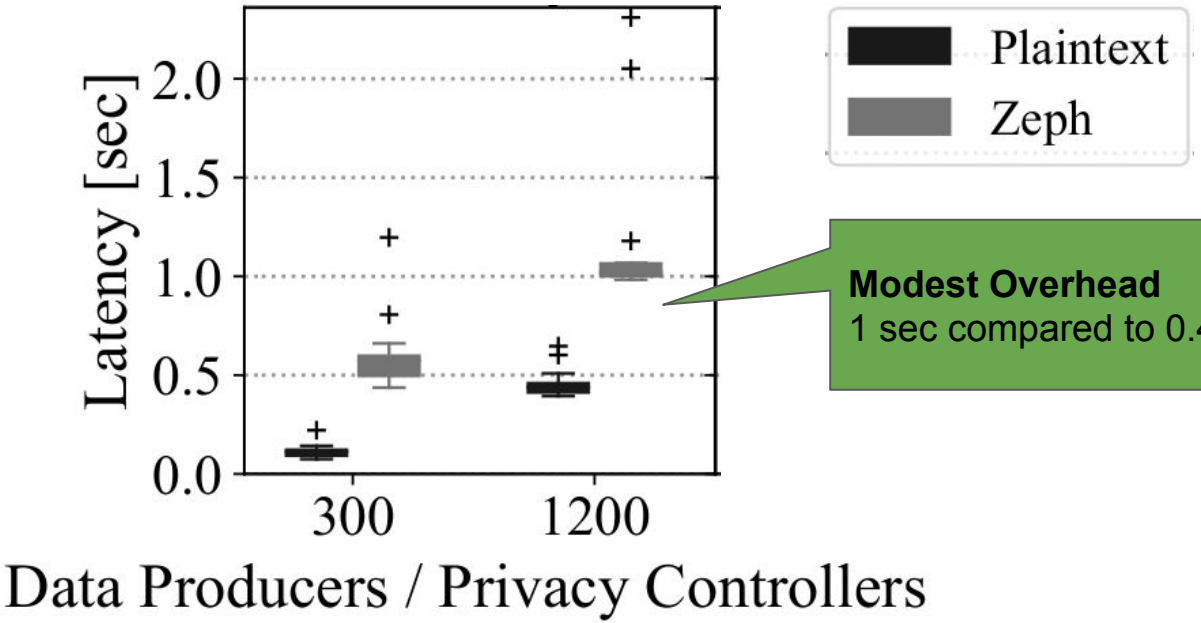
"multiple Data Producers - **multiple** Privacy Controllers"



Zeph Implementation and Evaluation

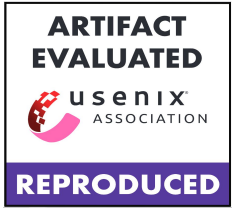


Web Analytics: End-to-End Benchmark



Modest Overhead
1 sec compared to 0.46 sec

pps-lab.com/research/e2e-privacy



github.com/pps-lab/zeph-artifact

Citations

[1] Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. 2014. RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response. In Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security (CCS '14). Association for Computing Machinery, New York, NY, USA, 1054–1067.

[2] Henry Corrigan-Gibbs and Dan Boneh. 2017. Prio: private, robust, and scalable computation of aggregate statistics. In Proceedings of the 14th USENIX Conference on Networked Systems Design and Implementation (NSDI'17). USENIX Association, USA, 259–282.

[3] Lukas Burkhalter and Anwar Hithnawi and Alexander Viand and Hossein Shafagh and Sylvia Ratnasamy. 2020. TimeCrypt: Encrypted Data Stream Processing at Scale with Cryptographic Access Control. 17th {USENIX} Symposium on Networked Systems Design and Implementation (NSDI'20). USENIX Association, Santa Clara, 835-850.