PrivApprox

Privacy-Preserving Stream Analytics

https://privapprox.github.io

[USENIX ATC'17]

Do Le Quoc, Martin Beck,
Pramod Bhatotia, Ruichuan Chen, Christof Fetzer, Thorsten Strufe

Motivation

Clients   **Analysts**

Private data
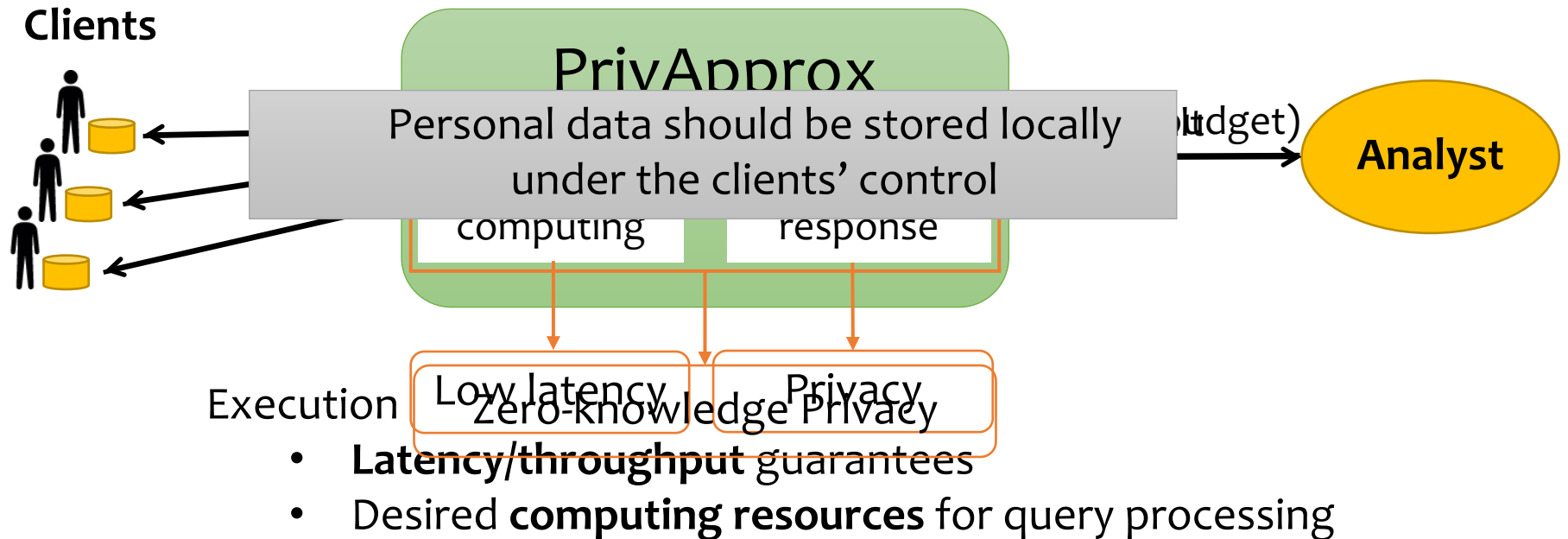
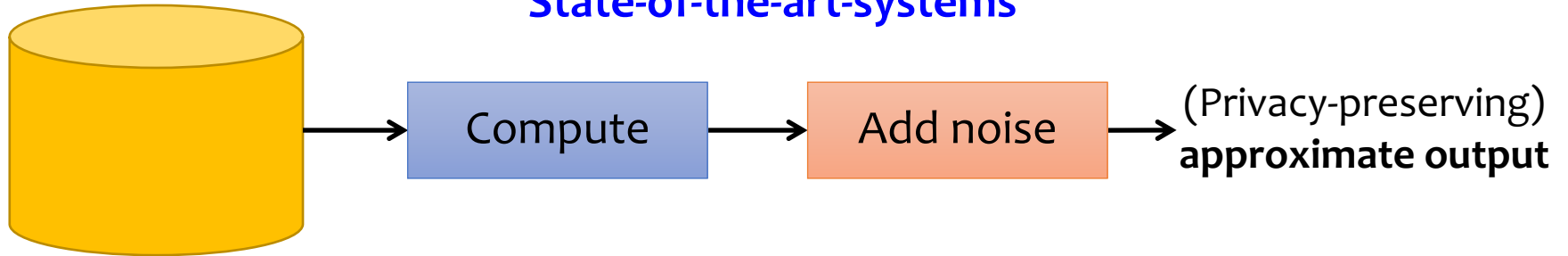Recommendation, Ads

Strong privacy guarantees

High utility analytics in real-time
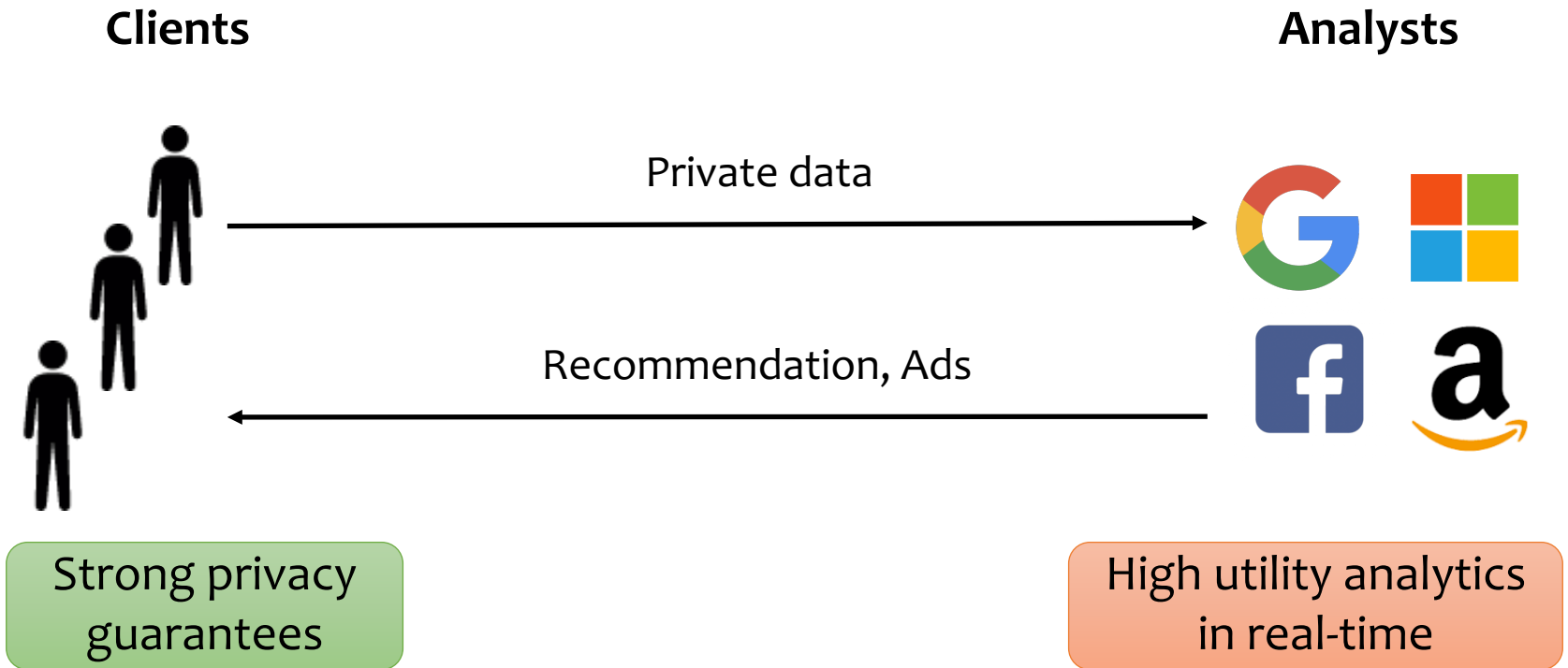
How to preserve users' privacy while supporting high-utility data analytics for low-latency stream processing?

PrivApprox

Clients

PrivApprox

Personal data should be stored locally
under the clients' control

computing   response

(budget)

Analyst

Execution   Low latency   Privacy
Zero-knowledge Privacy

- **Latency/throughput** guarantees
- Desired **computing resources** for query processing

#1: Approximate computing

State-of-the-art-systems

Compute → Add noise → (Privacy-preserving) **approximate output**

Idea: To achieve low latency, compute over a sub-set of data items instead of the entire data-set

Approximate computing

Take a sample → Compute → **Approximate output** ± Error bound
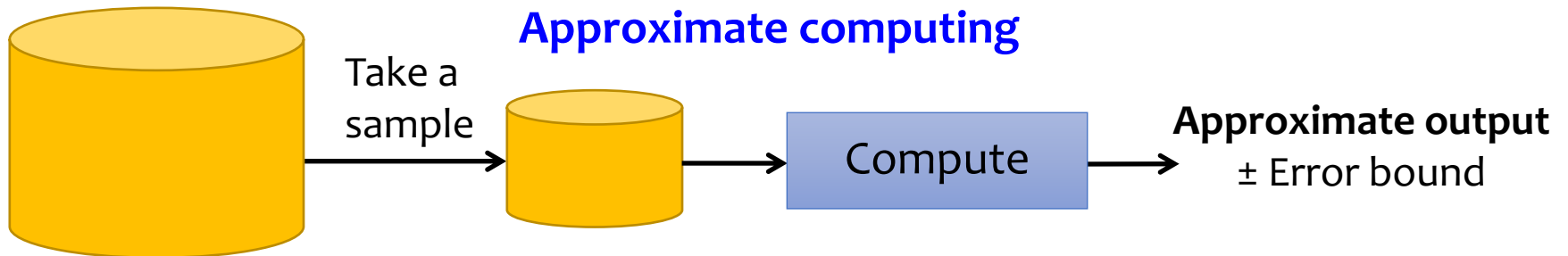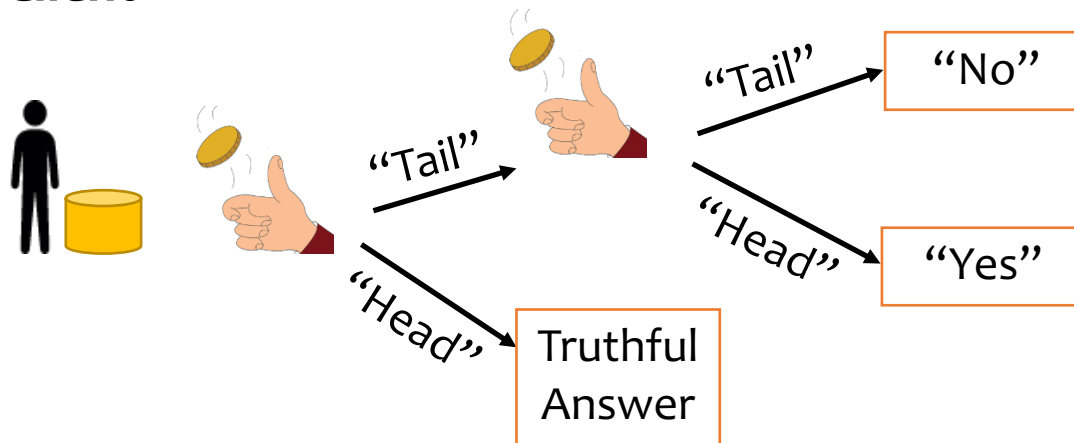
#2: Randomized response

Idea: To preserve privacy, clients may not provide **truthful answers** all the time

Client

"Tail" → "No"

"Head" → "Yes"

"Head" → Truthful Answer

Provides **plausible deniability** for clients responding to sensitive queries; achieves **differential privacy** (RAPPOR [CCS'14])

Summary

PrivApprox: a privacy-preserving stream analytics system over distributed datasets

Privacy	Zero-knowledge privacy

Practical	Adaptive execution based on query budget

Efficient	Randomized response & sampling techniques

Thank you!
https://privapprox.github.io