SOTERIA: Preserving Privacy in Machine Learning

Cláudia Brito, Pedro Ferreira[◆], Bernardo Portela[◆], Rui Oliveira, João Paulo

INESC TEC & University of Minho,
[◆]INESC TEC & Faculty of Sciences, University of Porto

INESCTEC FCT Fundação para a Ciência e a Tecnologia

Privacy and Security in Machine Learning

Motivation

- The exponential growth of data is raising novel challenges for large-scale data analytics.
    - Automation based on ML.
- ML datasets and models are **stored** and **processed** in **plaintext**.

Privacy and Security in Machine Learning

Motivation

- The exponential growth of data is raising novel challenges for large-scale data analytics.
    - Automation based on ML.
- ML datasets and models are stored and processed in plaintext.
- Third-party infrastructures are **untrusted**.

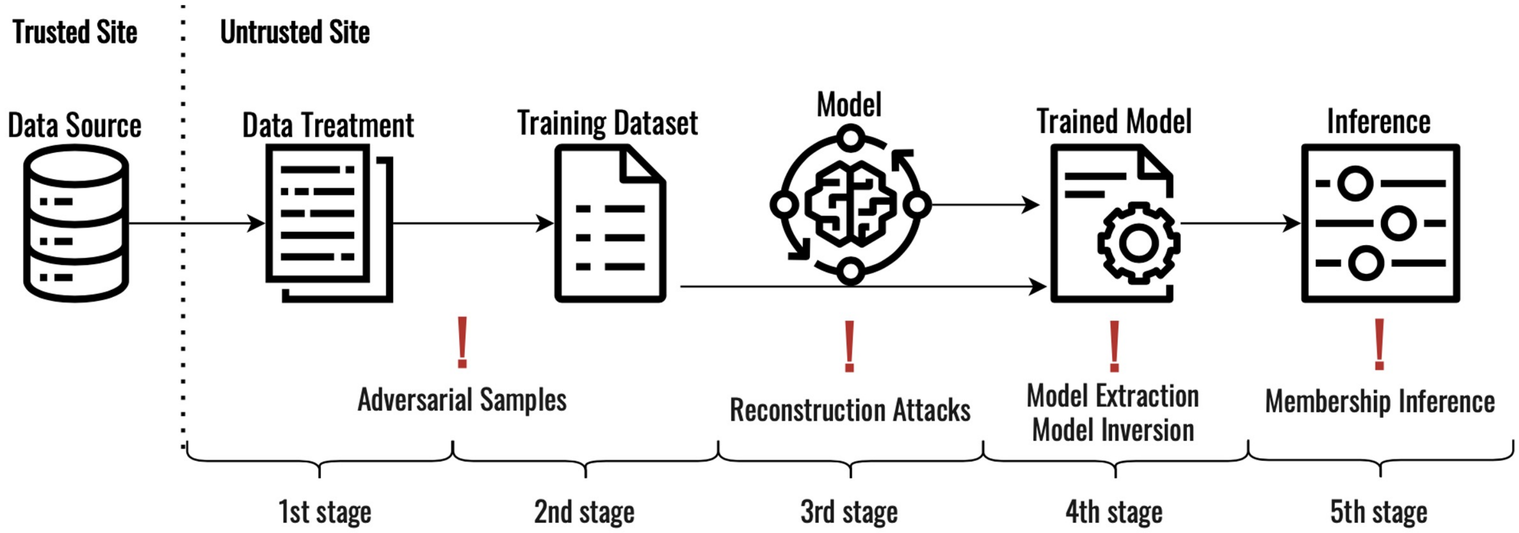
Privacy and Security in Machine Learning

Motivation

- The exponential growth of data is raising novel challenges for large-scale data analytics.
  - Automation based on ML.
- ML datasets and models are stored and processed in plaintext.
- Third-party infrastructures are untrusted.
- Increasing international **legislation** to protect the **privacy** of citizens.

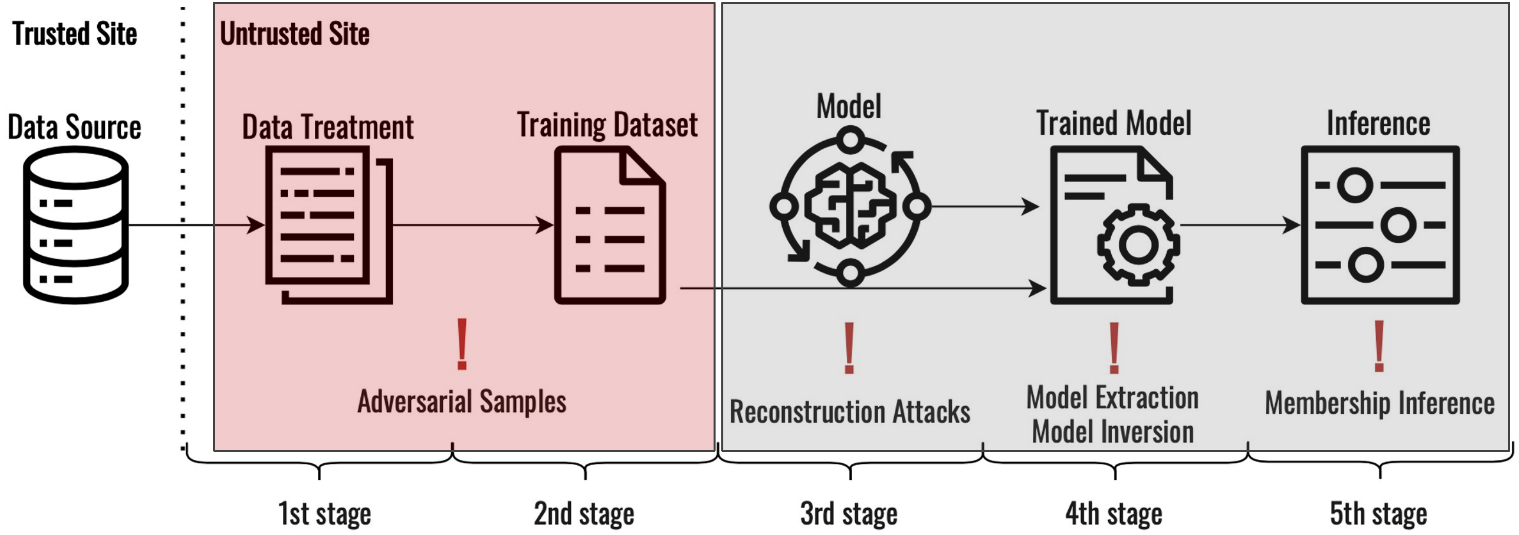
Privacy and Security in Machine Learning

Current ML Pipeline



Privacy and Security in Machine Learning

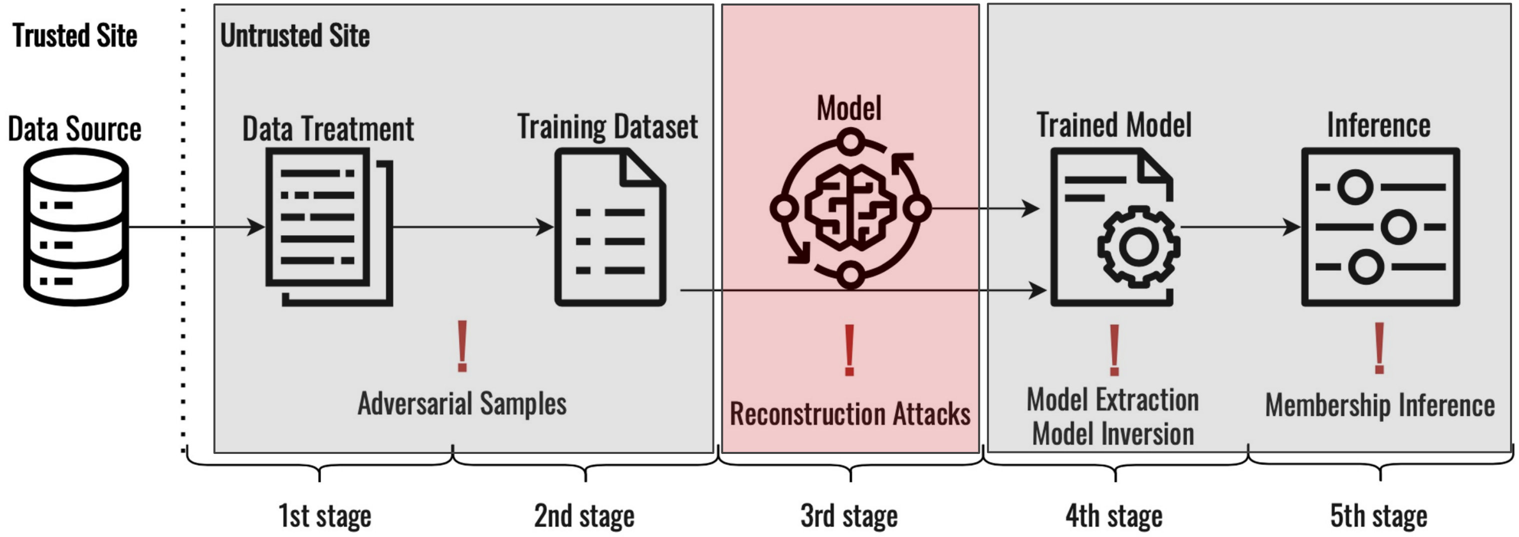
Current ML Pipeline



Adversarial data injection → Direct access to collected data → Wrongly trained models

Privacy and Security in Machine Learning

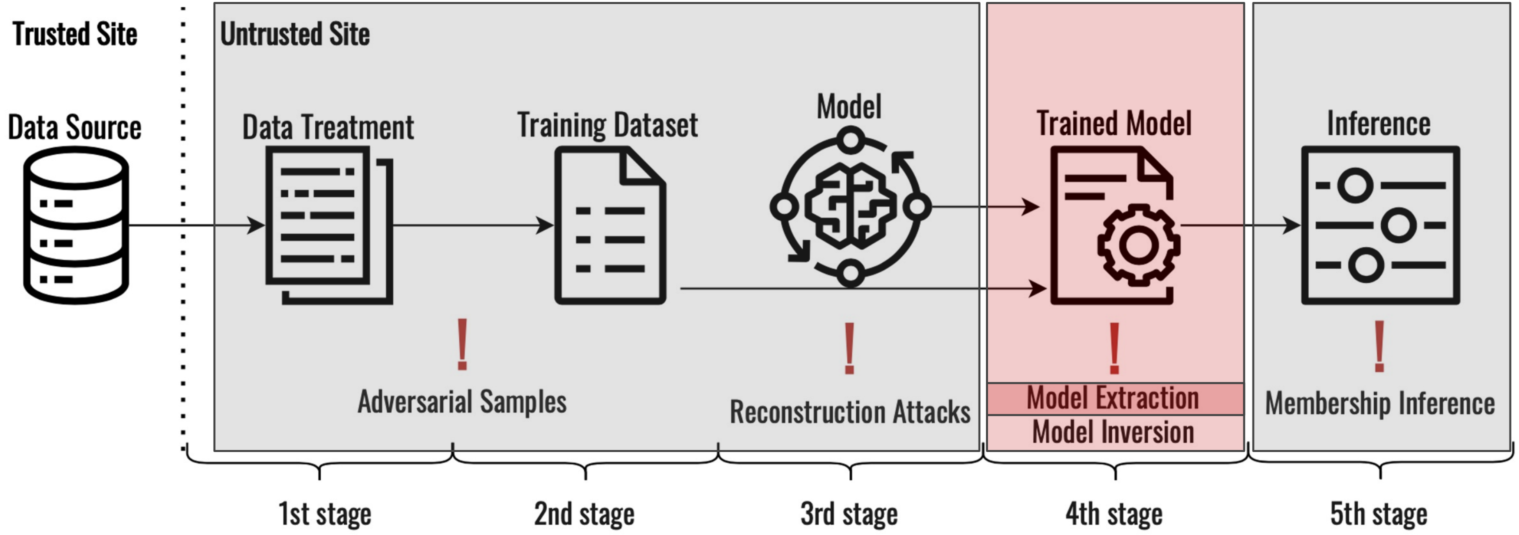
Current ML Pipeline



Feature vectors direct access → Reconstruction of raw data

Privacy and Security in Machine Learning

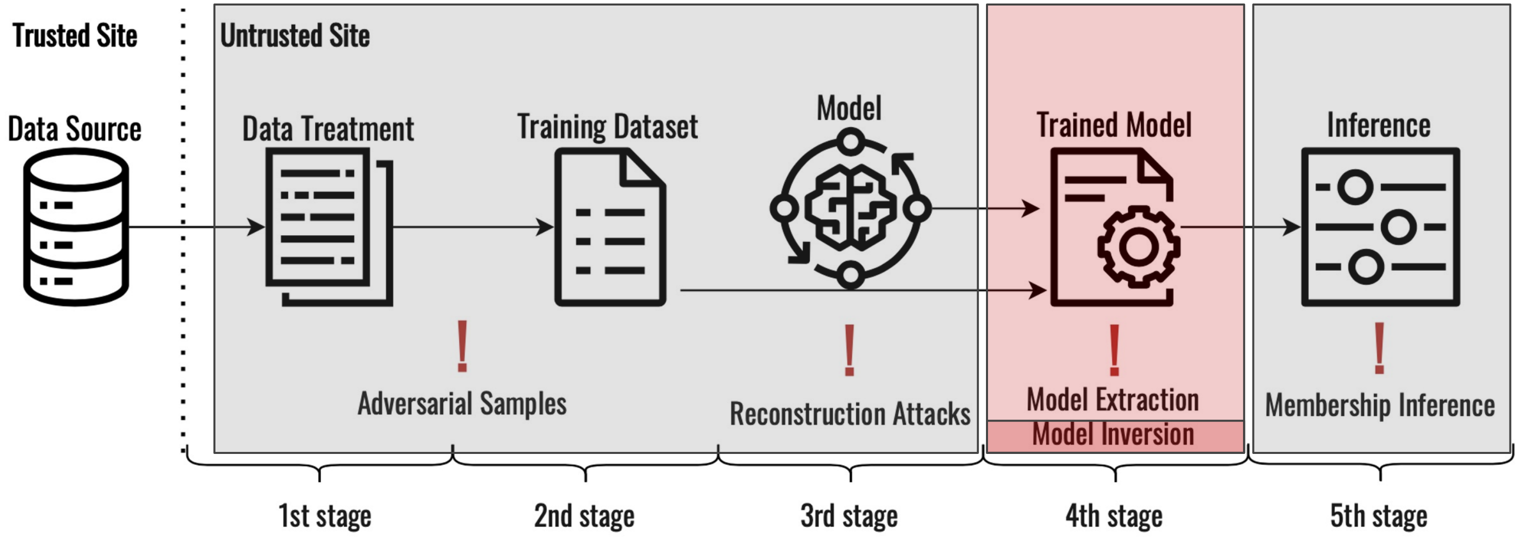
Current ML Pipeline



Confidence results → Direct access to data → Rebuild model parameters

Privacy and Security in Machine Learning

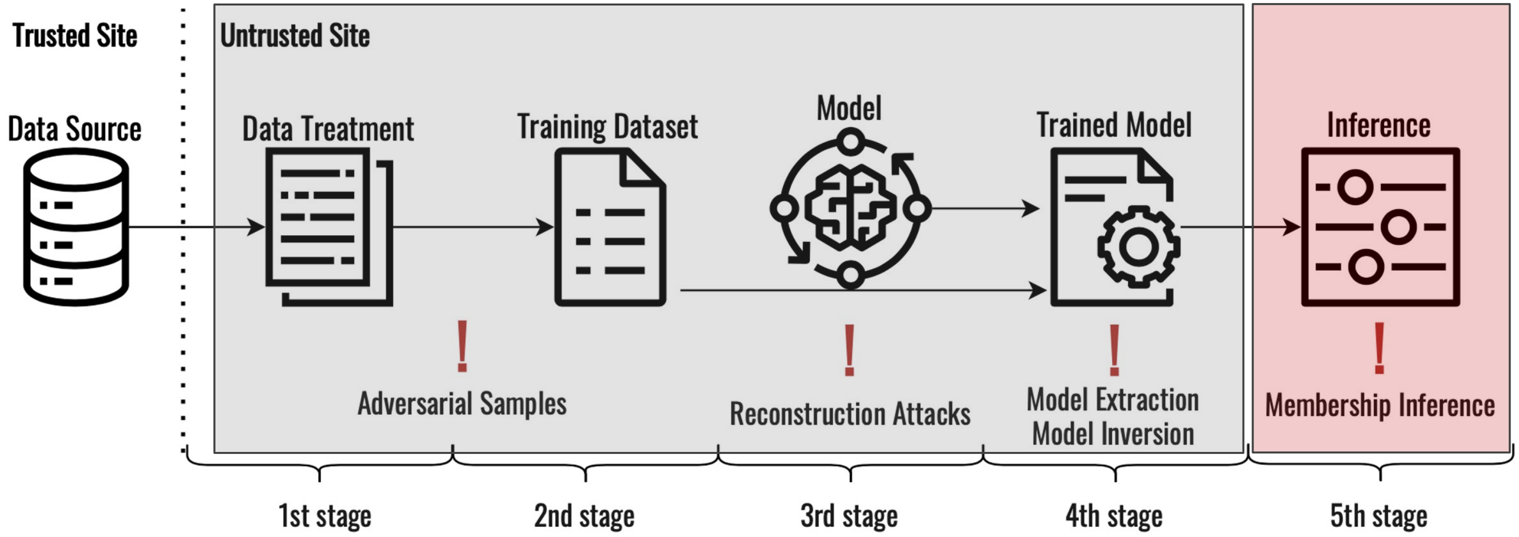
Current ML Pipeline



Direct querying access → Model parameters openly available → Reverse trained data

Privacy and Security in Machine Learning

Current ML Pipeline



Direct querying access → Model parameters openly available → Check for specific data on trained dataset

Privacy and Security in Machine Learning

Limitations

- Common cryptographic schemes impose **impractical overheads.**

Privacy and Security in Machine Learning

Limitations

- Common cryptographic schemes impose **impractical overheads**.
- TEEs’ performance depends on the number of **computations**, **I/O** operations and **trusted computing base** (TCB).
    - **Reducing** the **code base**.
    - **Reducing** the number of **operations**.

SOTERIA

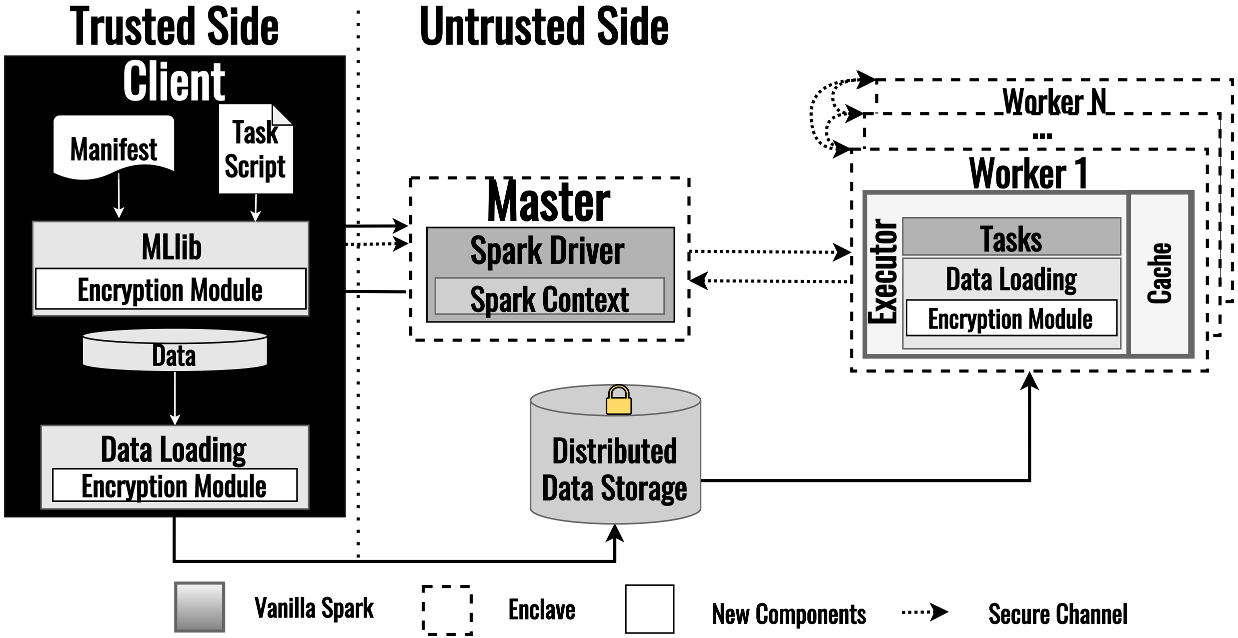
Contributions

- Soteria, a privacy-preserving distributed machine learning solution
  - A **baseline scheme** (SOTERIA-B) for performance and security comparison.
  - A new computation **partitioning scheme** (SOTERIA-P) for running Apache Spark’ MLlib inside SGX.

SOTERIA Architecture

Client Side

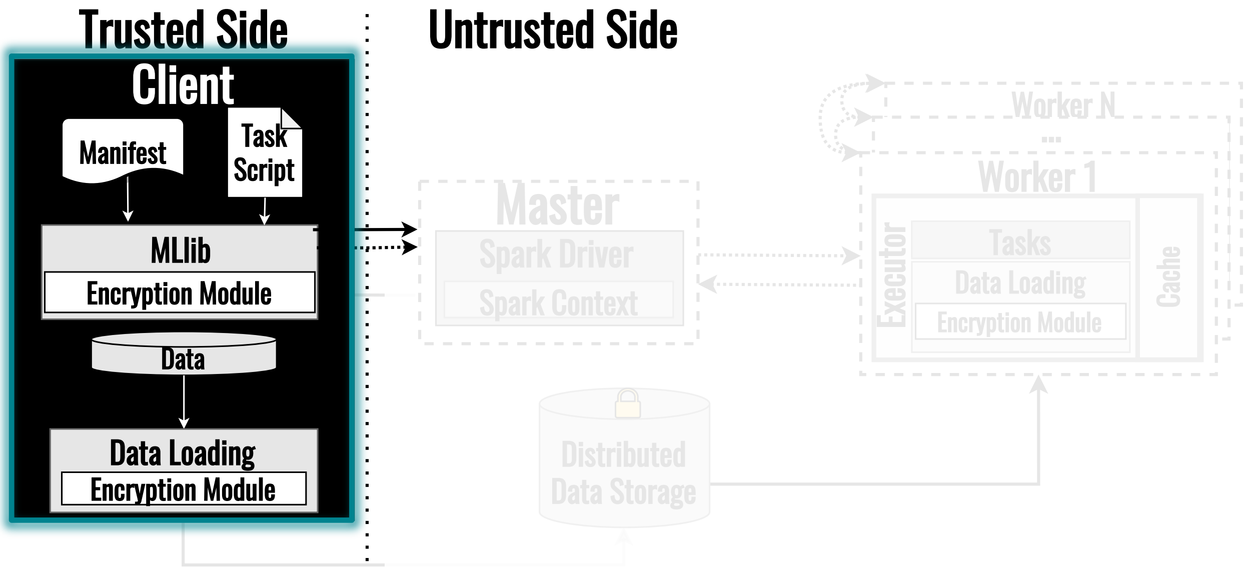
Honest-but-curious



SOTERIA Architecture

Client Side

Clients define their tasks.



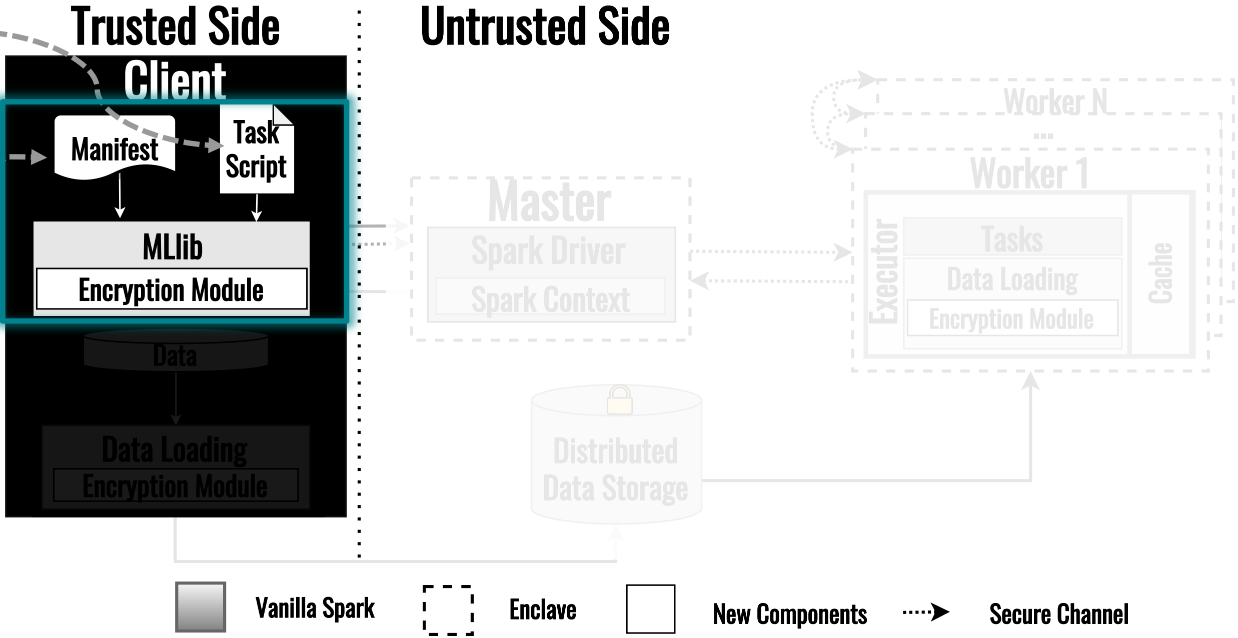
▨ Vanilla Spark ⬚ Enclave □ New Components ┈➤ Secure Channel

SOTERIA Architecture

Task Stage

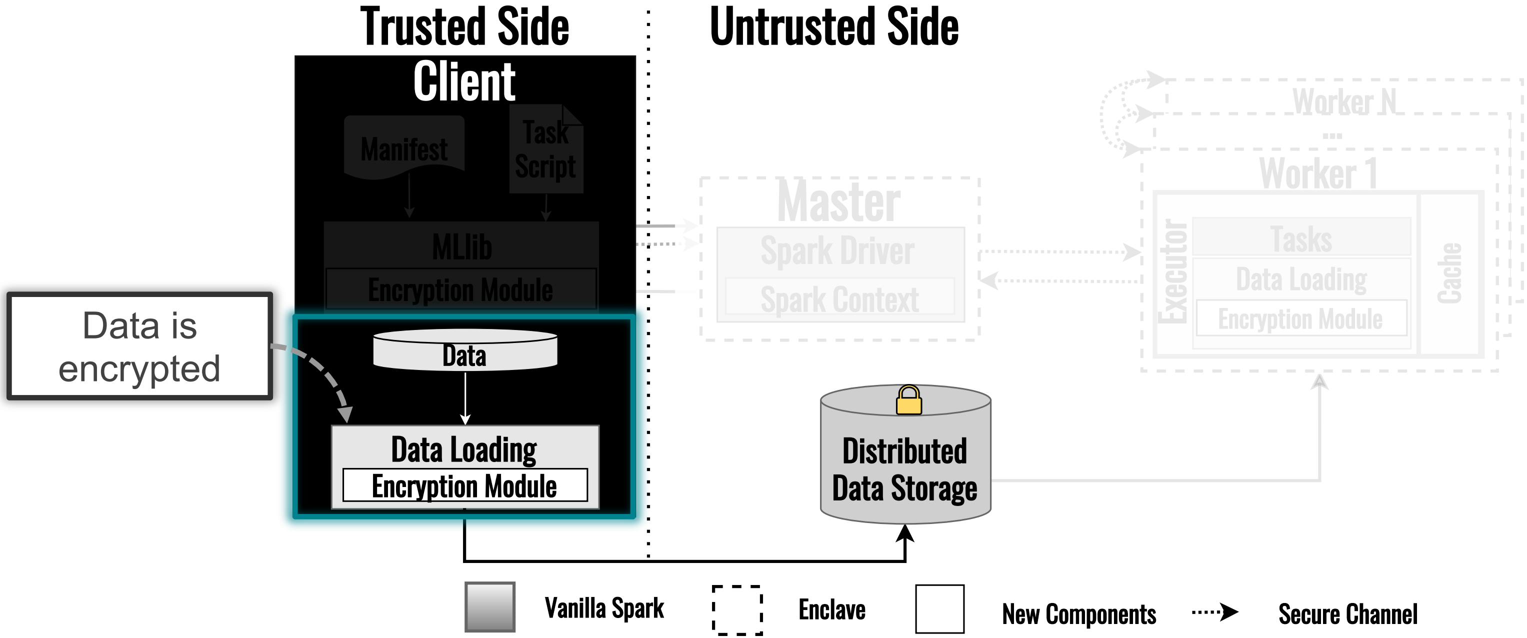
Being training or inference, it needs to be defined previously.

The manifest is predefined.



SOTERIA Architecture

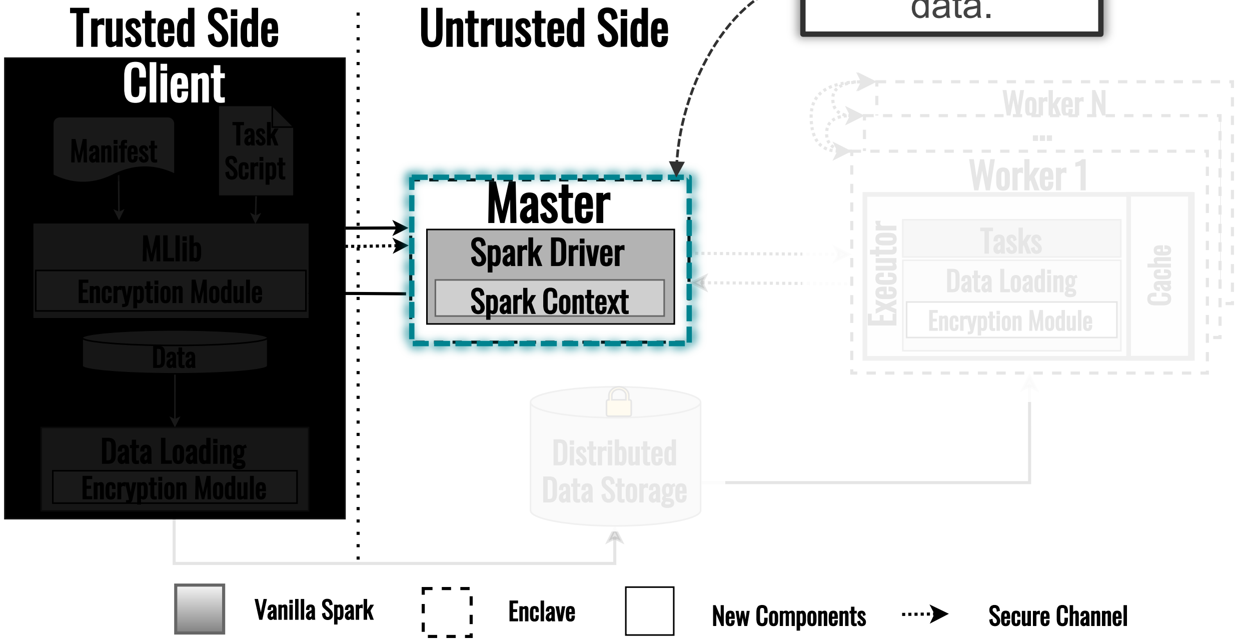
Data Stage



Data is encrypted

SOTERIA Architecture

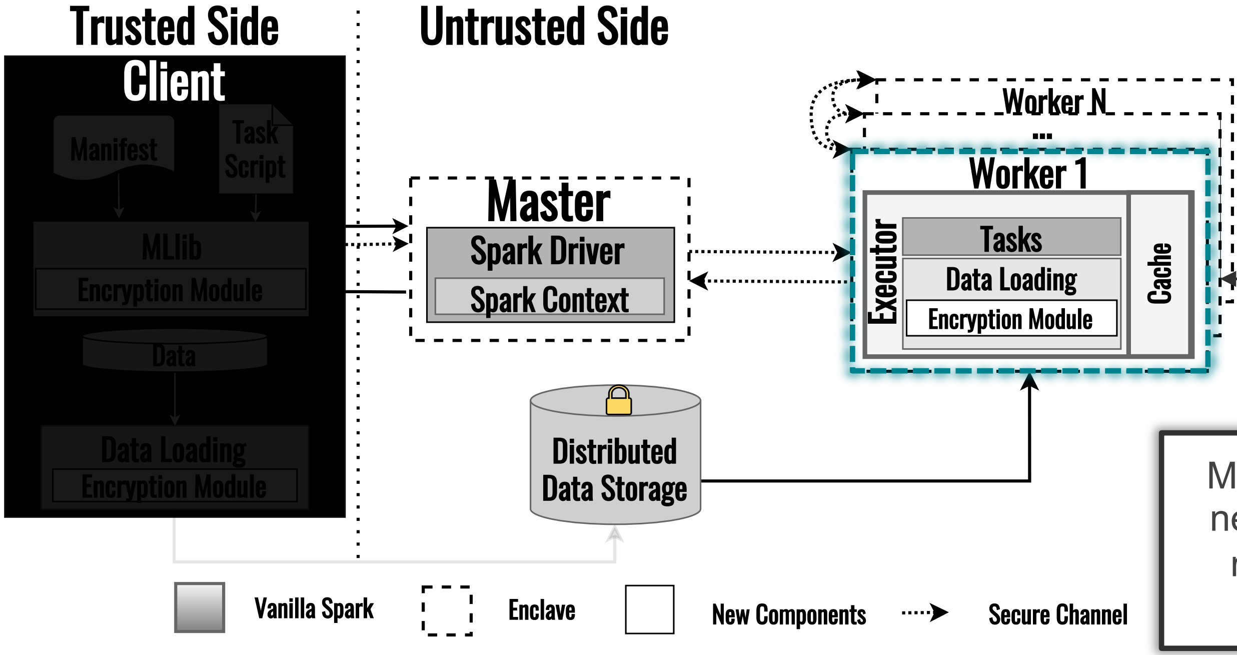
Master Stage



Needs plaintext data.

SOTERIA Architecture

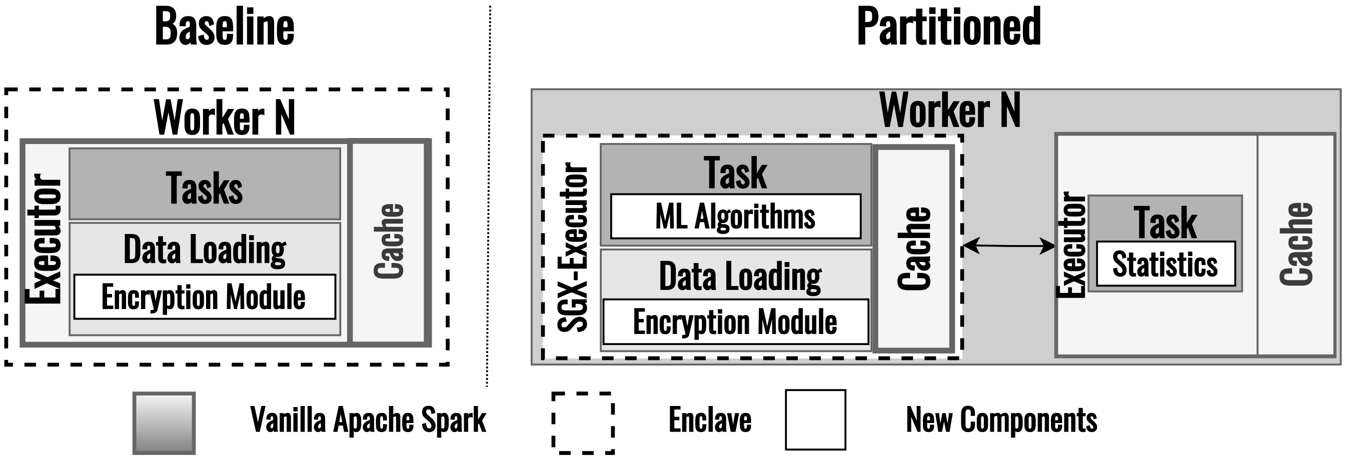
Worker Stage



MLlib and all the needed libraries run inside the enclaves.

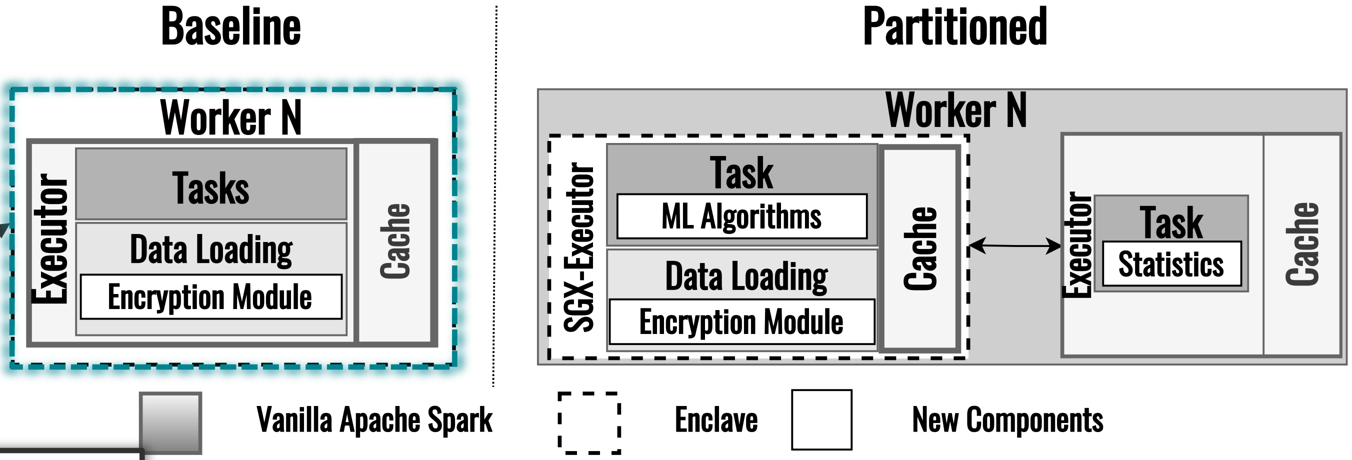
SOTERIA Architecture

Partitioned Design



SOTERIA Architecture

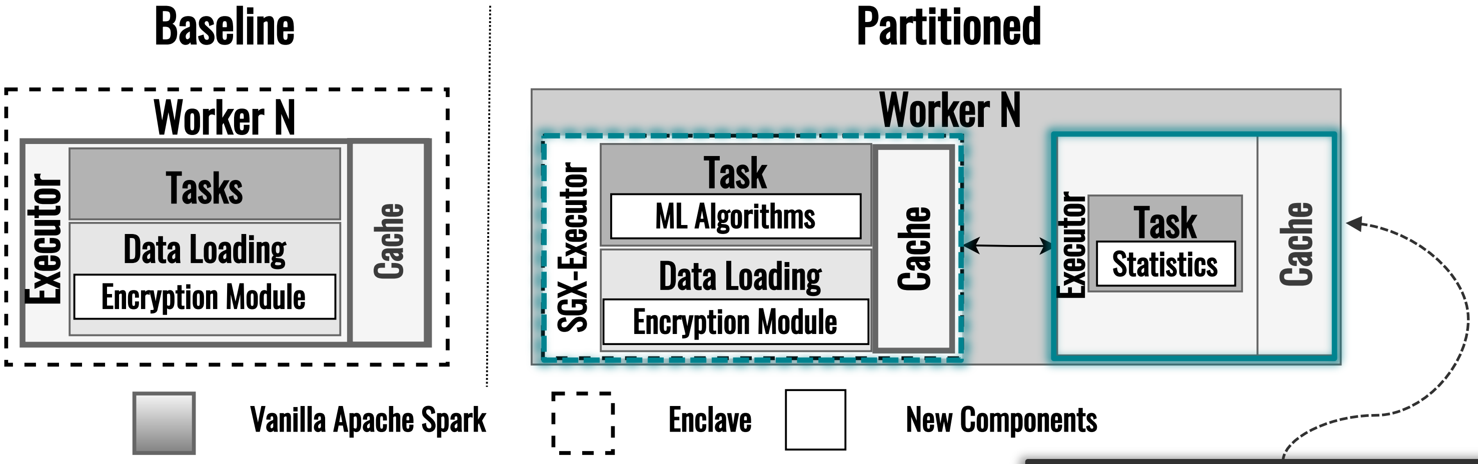
Partitioned Design



Everything runs inside the enclaves.

SOTERIA Architecture

Partitioned Design



Statistics are offloaded and run outside of enclaves.

SOTERIA Architecture

Partitioned Design

Baseline

Partitioned

How does statistical information relate to black-box model access, i.e. does the first imply the second in any way?

□ Vanilla Apache Spark □ Enclave □ New Components

SOTERIA

Security implications of statistical leakage

- Current attacks suggest one is unable to do this in any successful way*.
- **SOTERIA-P** is resilient to any attack that requires black-box access to the model to succeed.

*Varun Chandrasekaran, Kamalika Chaudhuri, Irene Giacomelli, Somesh Jha, et al. Exploring connections between active learning and model extraction. In *29th USENIX Security Symposium*, 2020

SOTERIA

Relation to ML Attacks

- Adversarial Attacks:
  - Authenticated encryption.
- Model Extraction, Model Inversion, Membership Inference and Reconstruction Attacks:
  - Secure channels for communication.
  - Computation on feature vectors is done inside the enclaves.
  - Black-box access to the model.

Evaluation

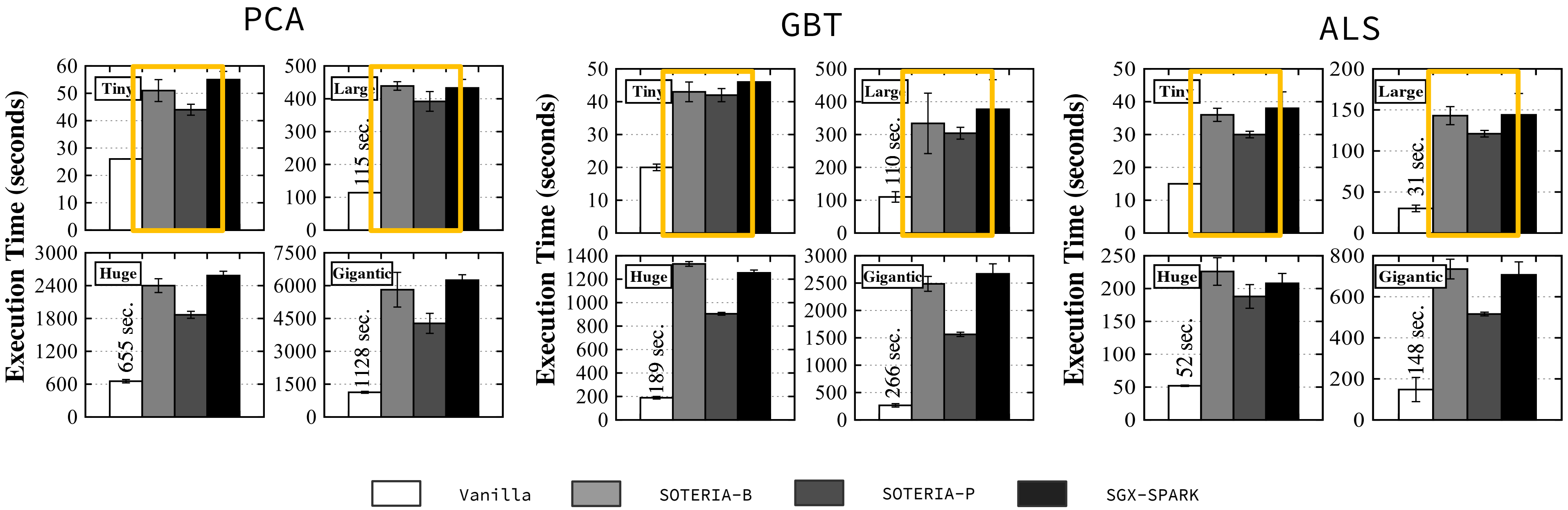
HiBench

- Algorithms:
  - Alternating Least Squares (ALS).
  - Principal Component Analysis (PCA).
  - Gradient Boosted Trees (GBT).
  - Linear Regression (LR).
- Workload sizes ranging from 193KiB to 894GiB.
- Setups:
  - Vanilla, Soteria-B, Soteria-P, SGX-Spark.
- 8 Ubuntu 18.04 servers, Intel Core i5-9500 with 16GiB RAM, 256GiB NVMe.

Evaluation

Dataset Size

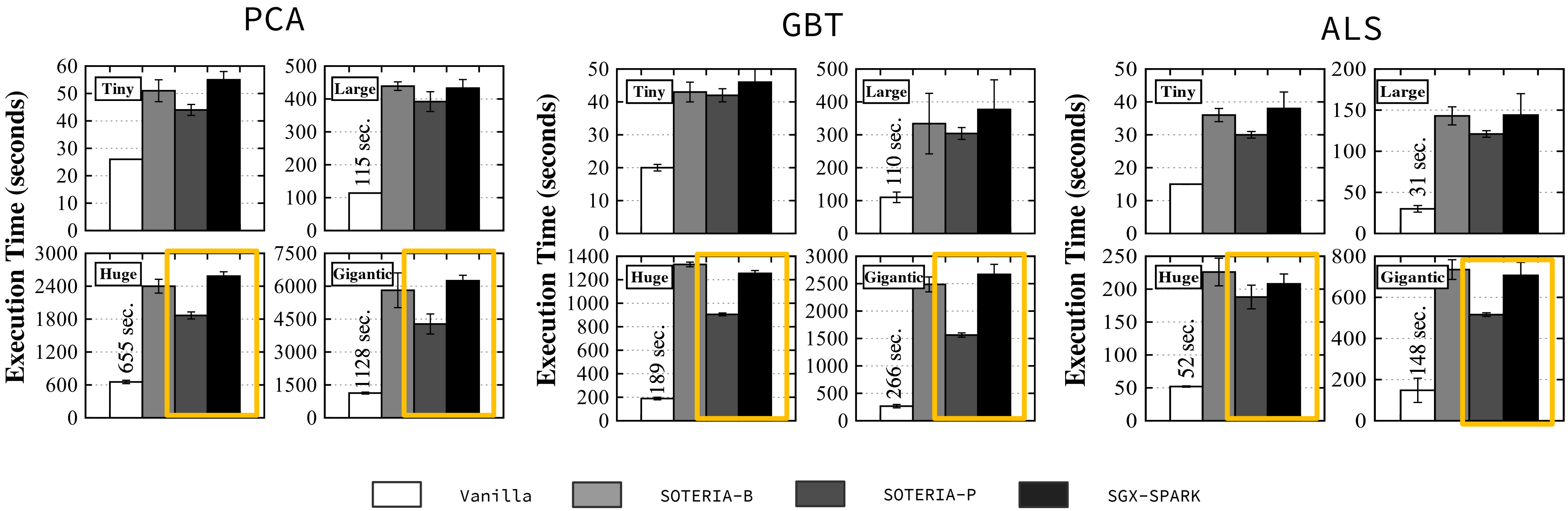
Similar performance with small dataset sizes



Evaluation

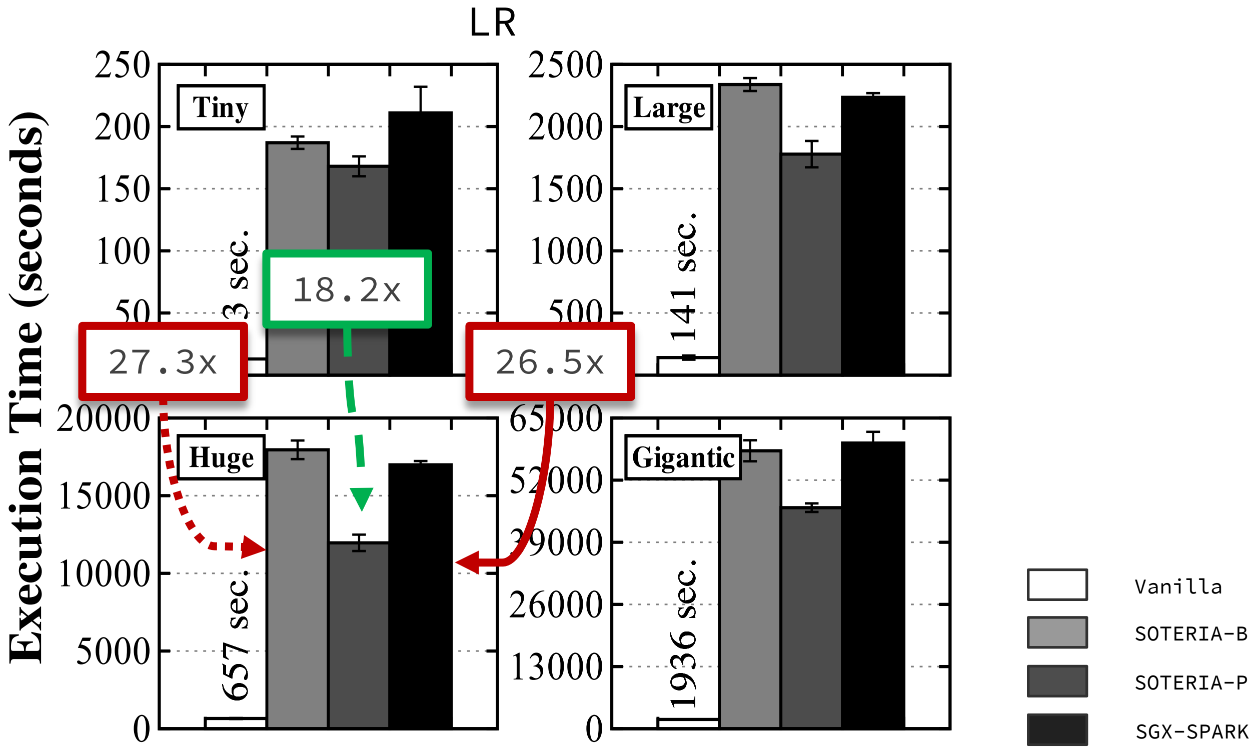
Dataset Size

Up to 41% less execution time



Evaluation

HiBench



27.3x

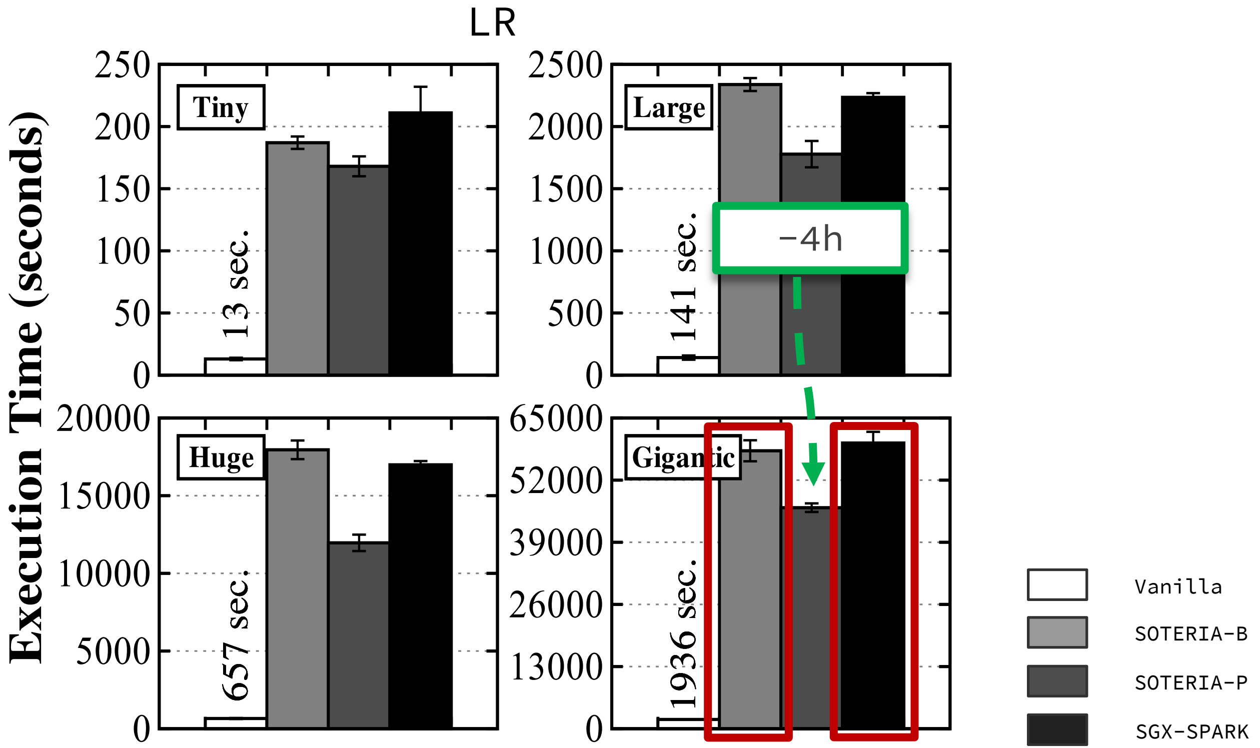
18.2x

26.5x

335GB

Evaluation

HiBench



894GB

Summary

- **SOTERIA**, a system for distributed privacy-preserving ML.
    - A novel **partitioning scheme** (SOTERIA-P) that allows specific ML operations to be deployed outside trusted enclaves.
    - Feasibility of **offloading non-sensitive operations** while still covering a larger spectrum of black-box ML attacks.
    - **Support** of numerous **ML algorithms.**
    - **Non-intrusive** to the **clients** flow.

SOTERIA: Preserving Privacy in Machine Learning

Cláudia Brito, Pedro Ferreira⬩, Bernardo Portela⬩, Rui Oliveira, João Paulo

INESC TEC & University of Minho,

⬩INESC TEC & Faculty of Sciences, University of Porto