Primitives for secure data analytics on SparkFHE

Milestone 2

Soumyakanti Das

Advisor: Dr. Peizhao Hu
Project Refresh

- Data analytics on private encrypted data
- Apache Spark - Unified data analytics engine for processing large-scale data
- Libraries - Spark core, MLlib, Streaming, GraphX
- Integrate Homomorphic encryption with MLlib library

Source: https://www.cs.rit.edu/~ph/PrivateComputation
Milestone 1 Refresh

- Research papers

- MLlib Deep Dive
  - Vectors & Matrices
  - BLAS (Basic Linear Algebra Subprograms)
Milestone 2 Updates

● BLAS With HE
  ○ fhe_dgemm: matrix - matrix multiplication with HE
  ○ fhe_dgemv: matrix - vector multiplication with HE

● Spark MLlib
  ○ CtxtMatrices & CtxtVectors
  ○ CtxtBLAS

● SparkFHE-Examples
  ○ Examples and test cases for CtxtMatrices & BLAS Algos

sparkfhe-examples.jar
spiritlab.sparkfhe.examples.*
(examples and tests)

sparkfhe-api.jar
spiritlab.sparkfhe.api.*
(swig and utility functions)

libSparkFHE

sparkfhe-plugin.jar
org.apache.spark.*
- MLlib
- GraphX
- SQL
- Streaming

Apache Spark Core
Next Steps - Milestone 3

- Implement a working use case - Logistic Regression
- If possible, implement CtxtSparseMatrix
Thanks

Questions?