Primitives for secure data analytics on SparkFHE

Milestone 3

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 & 2 Refresh

- Understand HE and Apache Spark
  - Read research papers
  - Deep dive into the Matrix library

- Implementations
  - CtxtMatrix, CtxtDenseMatrix
  - BLAS (Basic Linear Algebra Subprograms) methods for matrix multiplication
  - Examples for Matrix operations and BLAS methods

Diagram:
- Matrix
  - DenseMatrix
  - SparseMatrix
  - Matrices
  - MatrixImplicits
- CtxtMatrix
  - CtxtDenseMatrix
  - CtxtSparseMatrix
  - CtxtMatrices
  - CtxtMatrixImplicits

Java libraries:
- sparkfhe-examples.jar
  - spiritlab.sparkfhe.examples.*
    - (examples and tests)
- sparkfhe-api.jar
  - spiritlab.sparkfhe.api.*
    - (swig and utility functions)
- sparkfhe-plugin.jar
  - org.apache.spark.*
    - MLlib
    - GraphX
    - SQL
    - Streaming
- libSparkFHE
  - fhe.a (HElib)
  - seal.a (SEAL)
  - others.a (others)

Apache Spark Core
Milestone 3 Updates

- BLAS library with HE
  - dgemm, dgemv, drot, dsymv, dsymm, dtrmm
  - Examples and tests

- Logistic Regression
  - Implemented Nesterov’s accelerated Gradient Descent
  - Work in progress...

Challenges

- Cannot compare two ciphertexts
- Cannot divide two ciphertexts
- Ciphertext operations introduce noise
- Missing API in HElib
Conclusions

- Implemented CtxtMatrix - helps future developers
- Added examples and demos
- Implemented various BLAS methods
- Logistic regression - could be a good showcase

Future work

- Expand the MLlib library
- Add more methods in the BLAS library
- Integrate FHE with other libraries of Apache Spark
Thanks

Questions?