Verifiable Outsourcing of Data Mining Computations

Wendy Hui Wang
Stevens Institute of Technology, Hoboken, NJ

Objective

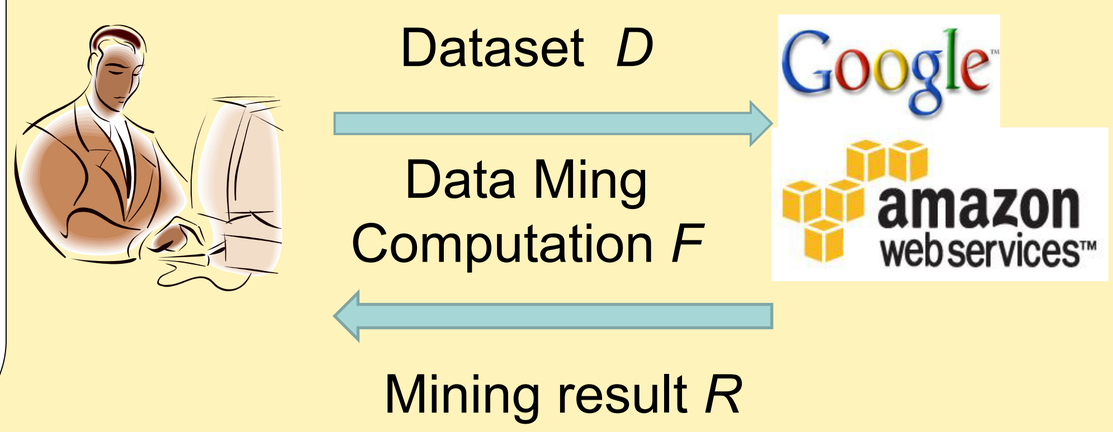
- *Data-mining-as-a-service (DMaS) model*
  - ***Service provider (server)*** - provides storage and computational power as the service.
  - ***Data owner (client)*** - outsources data and data mining computations to the server.

•Security issue: the server may return incorrect mining results.

•Problem: how to verify $R = F(D)$ in an efficient way?

Project goal

Design efficient and practical verification mechanisms that enable the client with weak computational power to verify the correctness of the returned results of outsourced data mining computations.



Approach

- Verification of non privacy-preserving data mining computations;
- Verification of outsourced privacy-preserving data mining computations;
- Analyzing and strategic modeling of the service provider's misbehaviors.

Progress

- Verification of non privacy-preserving data mining computations
  - Frequent itemset mining: deterministic approach (ICDM'13) and probabilistic approach (DBSec'13) ;
  - Outlier mining (PKDD'12);
  - K-means clustering (SDM'13);
  - Bayesian network structure learning (SDM'14);
- Verification of privacy-preserving data mining computations
  - Frequent itemset mining (SDM'15);
- Verification of MapReduce data mining computations (VLDB'12).
- Verification of record matching (IRI'16).

Interested in meeting the PIs? Attach post-it note below!

NSF Secure and Trustworthy Computing Principal Investigator Meeting
Jan. 9 -11 2017
Arlington, VA