CERIAS

the center for education and research in information assurance and security

Private Searching for Nearest Neighbors

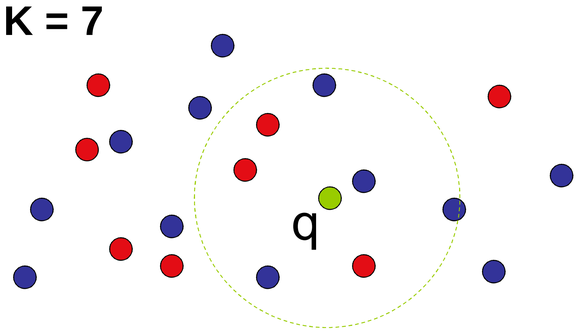
Yinian Qi, Mikhail Atallah

Privacy Requirements

- Assuming both being “honest but curious”, neither Alice nor Bob can learn anything about the other party’s inputs during the K-NN set computation.
- Both the intermediate results and the final answer are in additively split form to prevent information leakage
  - Additively split: v = v' + v" where v' is with Alice, v" is with Bob, and each of v' and v" look random

K-NN Query: K = 7

- Alice
- Bob



Building Blocks

- Blind and Permute Protocol (BP)
- Secure Scalar Product Protocol (SPP)
- Secure Comparison Protocol (SCP)
- Secure Selection Protocol (SELECT)

Single-Step K-NN Protocol

Scenario: Low-dimensional data, easy to compute the distance between objects (e.g. Euclidean distance). The query point q is from Alice.

Idea: Alice locally computes her K-NN to q, then coordinates with Bob in computing the global K-NN from her local K-NN and Bob’s inputs.

Protocol Steps:

- Alice and Bob generate a private and public key pair respectively in a homomorphic cryptosystem and exchange public keys.
- Alice then locally computes her K-NN list.
- For each item in Bob’s input, they jointly compute the distance to q in split form.
- They engage in a BP protocol to blind and permute all distance values they have so far.
- Alice and Bob run the SELECT protocol to select the k smallest distances.

Multi-Step K-NN Protocol

Scenario: High-dimensional data, expensive to compute the accurate distance between two objects.

Idea: Use a cheaper distance function df (feature distance) to facilitate pruning, satisfying df ≤ do (the actual distance). Refer to multi-step K-NN algorithm (T.Seidl and H.P.Kriegel).

Protocol Sketch: Alice and Bob first run BP on their inputs (result split), then securely compute df between query q and all inputs, which is organized into a split priority queue PQ. The list L of the current k-NN is maintained also in split form. Both PQ and L are updated with the help of SCP. The final k surviving items in L are returned as K-NN.

Example: sequence data where do is the edit distance.

Privacy preserved
Efficient protocols
Wide applicability

Summary:

- All privacy requirements are satisfied. Both protocols are provably secure.
- The single-step protocol is efficient, with linear computation and communication complexity while the multi-step protocol uses df for efficient pruning.
- Since the K-NN result is split, our protocols can safely be used as building blocks in privacy-preserving data mining tasks, such as classification and outlier detection.

PURDUE UNIVERSITY

CERIAS

Discovery Park e-Enterprise Center