**Title:**

**Privacy Preserving Query Processing using Third Parties.**

**What is the research problem?**

Modern day systems, applications and services needs certain data of the user to provide services, challenge lies in providing minimal information to cater to such request and at the same time not exposing information about the end user which can be misused or compromises the privacy of the end user. The paper proposes two protocols for various database operations to answers such queries without revealing any using information about the end-user.

**Related work**

* A scheme to process intersection queries using oblivion transfer and polynomial evaluation based on cryptographic techniques is already recommended but the intersection query takes order of n square where n is the size of the database hence impractical for large databases.
* A third party approach service approach without the need of encryption/decryption is also worked area, however with this design multiple data providers cannot be catered.

**Overview/main points of the proposed approach/architecture**

* Currently Technologies and framework assumes unrestricted access to the end-users data, hence compromising them. Example such as medical record of the user, if needed it should be shared but here the end-user might be interested in sharing only the minimal required information thus maintain their privacy.
* Here the goal of the privacy preserving query to execute the required query, may be on multiple data source, without revealing any extra information thus maintaining the user privacy.
* In case of multiple data sources, operation such as intersection, join, aggregation are fundamental and general purpose queries.
* Currently available techniques include
  + Handing over the data to a trusted third party, which in turn take care of query processing, but again unacceptable in many scenarios
  + Secure multi-party computation, here the given m-parties with input x1, x2….xm, and a function f(x1, x2,….xm), can only learn the output of the given function any nothing more about the original data x.
  + Encryption/decryption is another approach but its expensive operation in terms of computation and it’s an overhead, and doesn’t help with multiple source and aggregation/intersection of data.
* This paper suggests using a hash based P2P system to select third parties that perform computations required for processing the privacy preserving queries using anonymous communication in a P2P system using overlay network.
* Shamir’s secret sharing technique involves distributing secret keys among n peers, such that the knowledge of any k keys (peers) is required to reconstruct the secret. At least k keys are needed and no information about the secret can constructed with k-1 keys.
* The same technique can be applied to set of Databases D1…Dn similar to peers.
* An honest-but-curious environment is one where the parties involved in a query follow the given protocol additionally they may keep any result or information they obtain during the course of the protocol. In the scheme proposed in this paper, queries are answered in a privacy preserving manner using third parties. At the end of query processing, no additional useful information is revealed to any parties involved in query processing but only the query answer is revealed to the query posers.
* The intersection query processing is divided into 3 phases
  + Distribution Phase: Each data source Pi has a list of secret elements Li={e1, e2,…,eLi}, Pi creates n shares one for each of the n third party.
  + Computation at Third party Phase: After receiving their shares from the data sources, each third party calculates bitmaps corresponding to the list share. These bitmaps are used to indicate which element of the list are present in the intersection.
  + Final computation phase: After receiving the bitmaps from all the third parties, each data source determines which elements in its list are present in the result of the intersection query.
* Similarly Equijoin can be done using following two phases
  + Intersection Phase: Peers P1 and P2 computes the intersection set of their values.
  + Join Computation phase: P2 only sends information about the intersection of lists. As a result, nobody gains extra useful information.
* Aggregation operation: such as sum, average, min/max can also be computed with revealing any additional information. One of the way could be computing the aggregate of the values in the union of the lists coming from different peers/data-sources. Such that none of the peers will know the local aggregate of any other peers but still the overall aggregate can be computed.
* In all the above cases the time response will be the sum of the time taken in all the phases, such as for interaction it will be sum of the distribution phase, computation in the third party phase and the final computation phase.

**Conclusions:**

* The paper proposes solution for processing queries across private databases while preserving the data provider’s privacy. The proposed schemes computes the results using third party approach, without violating the privacy policies. The secrets are distributed to the third parties using Shamir’s secret sharing method on every element and enables aggregation queries, intersection, and equijoin. The scheme can be efficiently used to execute queries on large databases