First Critique

Publication:

Title:
Executing SQL over encrypted data in the database-service-provider model.

Summary:
THC-H database benchmark is used to test the model of executing SQL over encrypted data. A 10 MB and a 100 MB databases are tested respectively. In the first experiment, tests with different number of buckets are conducted. Results shows that increase of buckets number used to encrypt data will decrease time consumed to execute query-execution. Three different interactions between client and server are also compared. The proposed communication protocol between a client and a server doesn’t bring in significant overhead comparing with a single server mode. The second experiment takes advantage of TPC-H benchmark to duplicate the first experiment, and get similar results and conclusions.

Pros:
This paper proposed a fascinating model to protect privacy stored in public platform. First, this model prevents others to access the privacy; second, this model even protects personal data from accessing by the service providers. This property is pretty important in cloud age. Nobody wants anyone else can access personal data even if in cloud.

This paper proposed a novel approach to encrypt data in database. Base on the encrypting method with bucket splitting, data are mapped to new values without releasing the original information. Different bucketing approaches are provided in this paper to make it practical. The most significant thing about the encryption is that encrypted data can still be operated by normal SQL sentences, just like the normal database operations. Another innovation is construction of SQL statement on encrypted data. Separating works for server and client is a creative communication approach. This could be a trend in cloud computing.

Cons:
There are three main concerns about this paper:

First, the encryption is just utilized random buckets to mix original data. It doesn’t mention how to deal with the distribution of the data. However, in real world, many data have distributions, such as age. Based on their idea, the encrypted data will
maintain the identical distribution with the original data, which may provide a lot of extra information to crack the encryption.

Second, there are too much for the clients to do. First, it needs to encrypt the data and even keep a large pool for keys. Second, they have to generate the encrypted SQL query for the server. Third, the client might have to complete part of the filtering work on the retrieved data from server.

Third, the experiment doesn’t compare performance of the proposed model and performance of original database without modification.

**Question:**
How to deal with non-numerical data such as text or character string?

**Second Critique**

**Publication:**

**Title:**
Human-powered sorts and joins.

**Summary:**
In this paper, powers and abilities of humans are used to complete sorts and joins operations on the crowd. Four different methods of joining are proposed in this paper, including simple join of object with property, native batching between two objects, smart batching between multiple objects, and some alternative join algorithms. 20 celebrities and matching photos are selected to test the simple join algorithm. Experiments with different pairs per HIT are conducted to test batching algorithms.

Likewise, three different methods are proposed to complete sorting operation, including comparison-based idea to sort objects based on O(N*N) times of mutual comparison, rating-based idea to sort objects based on rating scores given by workers, and hybrid algorithm use rating-based sorting and continuous identifying on subset. 40 synthetically generated squares are used to test comparison batching, comparison batching, and rating granularity in sorting based on size of squares; 25 images of animals with multiple comparison dimensions are selected to compare difference between comparison and rating in sorting. The above two datasets then also utilized to test the hybrid approach in sorting.

An extra end-to-end experiment using movie frames and actor photos is also
conducted to test sorting and joining algorithms simultaneously.

Based on experiments results, the new algorithm can reduce the overall cost from $67 to about 3$, without substantially affecting accuracy or latency. In the end-to-end experiment, the proposed algorithm reduced cost by a factor of 14.5.

**Pros:**
First, this paper regards the crow as a database. It’s pretty interesting to learn about this point, because the idea of using database model could provide good opportunities to make crowd sourcing in a uniform format. Model crowd as a database would make it easier to utilize crowd sourcing.

Second, this paper proposed algorithms to complete sorting and joining operations based on abilities of humans. First, the paper comes up with many ideas of mapping human recognition into computation operations. Second, multiple algorithms are designed for different applications. Third, this paper provides ideas how to assign hits with minimal costs.

**Cons** (the work is so solid and the paper is so clear, so it’s really hard to detect some cons):

First, in the comparison algorithm for sorting, the complexity is $O(N*N)$, which means the sorting could take more time than quick sorting algorithms.

Second, the sample size of experiment is not big enough for practical use.

**Question:**
What is the applicable scope of the new sorting and joining algorithm? Is it only practical to do things about pictures or images?