論文の要旨

題目：**Study on Privacy-preserving Data Manipulation and Secure Computation of Skyline Objects on MapReduce**

（プライバシー保護データ操作および

スカイラインオブジェクトの安全なMapReduce計算法に関する研究）

**氏名 SALEH AHMED**

## Dissertation Summary

Database systems commonly control the access in the database to regulate authorization of confidential data. This process preserves the privacy of sensitive information, and the provided data is obtained by utilizing the required database operation interfaces. However, access control for vital and private data is often insufficient. Attacks against computer systems have confirmed that the security of data can be compromised, if an unlawful user can get access to the data files generated by the database management system, avoiding the access constraint mechanism of a database completely. For example, the Toronto Star issued an article. The article reports an occurrence where certain bank sold old disk in eBay without deleting the potential private data of the hundreds of clients. The worldwide privacy law does not support a disk holding the records of several hundred clients was being sold on eBay. So, the privacy of client data is a vital issue. As a result, organizations must encrypt the data before storing it in the drive.

People are concern about the privacy of sensitive data, such as salary, merit positions, tender evaluation, and so on, stored in a database. Several privacy preserving methods have been proposed. On the contrast, such encryption degrades the performance of the database operations. This degradation occurs due to decryption of values in the first place, then execution of the operations. Order-preserving encryption system (OPES) may solve these issues and improve the performance degradation issues. However, the order of numeric values itself is sensitive information. In such a case, it is necessary to hide the order information as well as solve the performance degradation issues.

The author proposed three methods which run on the top of OPES able to hide the order information in the values as well as solve the performance degradation issues.

On the contrary, every day, different computing organizations generate a huge amount of information. Such a massive amount of data is accountable for information overwhelm issues. Various related works to retrieve valuable information from large data have been studied as a solution to the matter. The essential fundamental operation of information selection is to gather a little number of objects that represent the whole data from a big database. Skyline Query in one of the popular tools to select representative objects from large scale databases.

Skyline query that is also appreciated as a successful information retrieval method has been successfully utilizing to filter out dominated objects. Skyline query usually utilized to retrieve objects that are good for all organizations whose evaluation mechanism are not alike. However, it may generate a large number or a very few numbers of objects. Moreover, whenever the organizations want to get the desired result of the skyline query, it is essential to reveal the attribute values of the objects in the datasets. In various situation, to calculate the skyline query, it needs to reveal valuable, sensitive information. The author introduced a skyline object picking tool in this dissertation that accumulates the favorite skyline objects for all the organizations; meantime, it also guarantees the privacy of sensitive attribute value throughout the process of skyline calculation.

Recently, people frequently have to retrieve important objects using mobile devices like a smartphone or phablets or tablets. In such a situation, it is difficult to explain complex query requirements like top-k query evaluation function. Clients are willing to get desired objects by defining only keywords. The proposed system must be serviceable for such circumstances. To achieve the query as mentioned above, the author used the skyline query function. To handle potential “big data", The author proposed a distributed algorithm in MapReduce framework to calculate the skyline query. For big data processing, MapReduce is a very popular, distributed open-source computing framework. The proposed method utilizes the MapReduce framework to manage the large-scale database.

For calculating a skyline query in conventional distributed algorithms, the attribute values of the individual object of a local database must be revealed to other organizations. Nowadays, people are concern about the privacy of their data; as a result, such revelations of private data in traditional distributed schemes are intolerable. In the proposed scheme, the security and privacy of the distributed algorithm are improved by the author in such a way that the confidentiality of the data during the processing of the skyline query remained intact. A novel and efficient strategy is introduced by the author to calculate the skyline from data of multiple organization in distributed computing condition without revealing the local private attribute values of objects to other organizations.

The author investigates the background of the problem and provides the introduction of the problem in Chapter 1. Then, the literature reviews on related work of the dissertation are provided in Chapter 2. After that, the author divides this dissertation into different parts. In the following section of this dissertation, the author discusses the Semi-order preserving encryption. Conventional encryption techniques encrypt the database values, but query execution on encrypted values is a severe performance degradation issue. Order preserving encryption is introduced to solve this problem. But in several cases, the order in the data is also a security concern. Therefore, semi-order preserving encryption effectively hides the order information and solve the performance degradation issues. The author discusses semi-order preserving encryption in detail in Chapter 3. In Chapter 4 of this dissertation, the author discusses a novel scheme to compute the skyline query in a secured manner in a distributed way on MapReduce. The author has considered the circumstances where the owner of the dataset is multi-party rather than a private entity. They desire to compute the skyline query result but never willing to reveal the attribute values during computation. The individuals do not desire to disclose attribute value as the values may be considered as sensitive information. The author introduced an efficient solution to settle such circumstances and compute the skyline query without disclosing any attribute values of the organizations. The suggested algorithm has used MapReduce programming infrastructure to guarantee its capacity to handle big data in a distributed manner.

Finally, a concluding study with a future guideline for enhancing the work has been given in Chapter 5.