論文の要旨

題目：**Study on Privacy-preserving Multi-party Computation of Skyline and its Variants**

（スカイライン問合せ及びその関連問合せに関する個人情報保護に配慮したマルチパーティ計算手法の研究）

氏名 MAHBOOB QAOSAR (ID: **D172517**)

Abstract

Big data is no longer considered merely a large amount of data. Instead, it is regarded as business-driven data with analysis capabilities due to long-term business value. Big data analyses are considered as a revolution in the field of Information Technology. Various organizations all over the world are utilizing advanced analysis techniques to gain new insights from their big data.

Selecting the most influential data objects or samples from a large database is the initial task of big data analyses. The analysis results can be of relatively little value if the samples are not representative of the population from which the results are determined. The skyline query and its variants are functions to find such representative objects.

In principle, the skyline query and its variants select the representative objects from a multi-dimensional database based on the dominance relation. The skyline query returns the non-dominated objects from a multi-dimensional database. Similarly, the *K*-skyband query returns those objects from a database that are not dominated by more than *K* objects from the given database. On the other hand, the top-*k* dominating query returns the *k* data objects from a multi-dimensional database which dominate the highest number of data objects in the given database.

Nowadays, multiple organizations dealing with similar kinds of services want to perform data analysis operations on the union of their databases, referred to as multi-party computation. The multi-party computation of the skyline and its variants can also provide benefits to the participating organization to recognize their most influential data objects. Since the database of individual parties may contain sensitive information relevant to their services or customers or products, any organization does not want to disclose its private database to others. However, it is not possible to compare the dominance relation between multi-party database objects without revealing the database. Realizing this issue, the data-privacy is widely studied in this dissertation for the computation of the multi-party skyline and variants.

Various settings and approaches for the computation of the privacy-preserving multi-party skyline and variants are considered within this dissertation. Several relevant works are reviewed to achieve the goal. The computation and communication complexities of one of proposed frameworks is also analyzed in this study.

This dissertation begins with the discussion and background of the problem in **Chapter 1**. Then, some basic preliminaries and literature surveys on related topics of the dissertation are presented in **Chapter 2**. The rest of this dissertation is split into several parts.

At first, **Chapter 3** discusses a framework for the privacy-preserving multi-party skyline query. The significant advantage of the proposed framework is that it does not require any trusted third-party for the multi-party skyline query. The framework utilizes the Paiillier cryptosystem along with the data anonymization and perturbation techniques for the multi-party computation. It also secures the intermediate computation results during the multi-party skyline query to ensure the highest privacy and security of the participating parties' databases.

After that, this dissertation proposes an efficient approach for the *K*-skyband query in distributed multi-party databases without unveiling the objects' attributes directly. This approach considers that all parties securely transform their objects' attributes without changing their sorting order rank on each dimension of the database objects. Then, a trusted third party computes the multi-party *K*-skyband from the transformed values of the objects' attributes. The detail of this process is explained in **Chapter 4**.

On the other hand, the top-*k* dominating query has drawn massive attention in the database community since it combines the advantages of the top-*k* query and the skyline query. The multi-party top-*k* dominating query can provide more benefits to the participating organizations to identify their most influential products or services. However, in the conventional computation system, it is not possible to compute the multi-party top-*k* dominating objects without revealing the individual parties' databases to others. Therefore, a framework for the cloud-based privacy-preserving multi-party top-*k* dominating query has been proposed in **Chapter 5**.

Finally, a concluding discussion with the future guideline to extend this research work is discussed in **Chapter 6**.