論 文 の 要 旨

題目：**Study on Secure MapReduce Computation**

of Representative Object Selection from Big Data

（ビッグデータからの代表的オブジェクト選別の安全なMapReduce計算法に関する研究）

氏名　　**ASIF ZAMAN**

Recent computing infrastructure makes a large amount of information available to consumers. Such large volume of data are responsible for information overload problems. A lot of works related to useful information retrieval have been considered to solve the issue. One of the most important primitive function of information retrieval is to select small number of representative objects from a large scale database.

Top-k queries have been extensively used to make a choice of preferable objects from large dataset. A scoring function and a number k, for number of objects to be selected, are specified by users. Then, top-k query returns k objects based on the user defined scoring function. Specifying a scoring function is sufficient and users do not have to define any complex query conditions for retrieving k objects. Therefore, top-k query is more preferable query interface for hand-held devices like smart-phones, tablets and tablets. On the other hand, it is quite possible that scoring functions of every user may not be similar for selecting top-k objects, which indicates that the top-k query results are valuable for those users who share an identical scoring function.

Skyline query, which is also known as a popular information retrieval tool, has been used to eradicate dominated objects. Skyline query can be used to select objects that are preferable for all users whose scoring functions are not identical. However, it may retrieve too many or too few objects.

Moreover, in order to retrieve the result of top-k or skyline query, it is necessary to disclose the values of each data object. In some cases, to compute top-k or skyline query, we have to disclose sensitive information.

In this dissertation, the author proposed a k-object selection mechanism that chooses various k objects which are preferable for all users who may have non-identical scoring function;

meanwhile, it also ensures the privacy of attribute value during the process of computation. Now a days, we often have to retrieve necessary objects using hand-held devices like a smart phone or tablets or tablets. In such environment, it is tough to define complex query conditions like scoring function. Users want to retrieve objects by specifying only keywords and the number of objects k. The proposed method must be useful for such situation.

To achieve above mentioned query, the author uses skyline query function. In order to handle so called "big data", the author has considered a distributed algorithm for computing a skyline query. MapReduce is a popular distributed computing framework for big data applications. To handle large-scale database, proposed algorithm has been developed on MapReduce framework. In conventional distributed algorithms for computing a skyline query, the values of each object of a local database have to be disclosed to another.

Recently, we have to be aware of privacy in a database, in which such disclosures of privacy information in conventional distributed algorithms are not allowed. In this work, the author has enhanced the security of the distributed algorithm so that the privacy of the data during processing kept intact. In other words, the author proposes a novel approach to compute the skyline in a multi-parties computing environment without disclosing individual values of objects to another party.

The author starts this dissertation from discussion and background of the problem in **Chapter 1**. Then, literature surveys on related topics of the dissertation is presented in **Chapter 2**.

After that, the author has split this dissertation in several parts. The first part of this dissertation considers a novel way of secure skyline computation on MapReduce. The author has considered the situation where owner of dataset are multiparty rather than a single entity. They want to find skyline query result but do not want to disclose any domain values. Even if parties are not willing to share domain value ranking or order information - as order itself may be considered as sensitive information. The author proposed a novel way to resolve the situation and find the result of skyline query without knowing any domain value. The proposed algorithm has used MapReduce programming model to ensure the its capability to process big data efficiently. Details of this process in expressed in **Chapter 3**.

In the second part of this dissertation, the author studies the problem of securely k-object selection. When dataset belong to multiparty and privacy of data has become a vital issue, conventional k-object selection algorithms are useless. Proposed model have addressed the issue and it is capable to find k objects from dataset whose attribute values are not discloseable. The author has discussed the issue in **Chapter 4**.

The third part of this dissertation, we have considered one of the most important application of the secure k objects selection function, which is the issue of finding key person from social media. Conventional social media mining techniques use graph mining algorithm to mine social media. But, the author has proposed a parallel model to mine social media using skyline query. The author discussed details about this model in **Chapter 5**.

Finally, a concluding discussion with future guideline for extending the work have been discussed in **Chapter 6**.