【Grant-in-Aid for Scientific Research (S)】

Broad Section J

Title of Project : Big Data Processing with Compressed Secure Computation

SADAKANE Kunihiko
(The University of Tokyo, Graduate School of Information Science and Technology, Professor)

Research Project Number : 21H05052 Researcher Number : 20323090
Term of Project : FY2021-2025 Budget Allocation : 122,300 Thousand Yen
Keyword : secure computation, succinct data structure

【Purpose and Background of the Research】

In Society 5.0, we can share various information and knowledge. However, there are following problems.

•Computation cost for analyzing and learning from huge amount of data.

•Privacy problem for sharing personal information

To do computation with preserving privacy, we need to anonymize and encrypt data, and we need to analyze and learn from the data without decryption. Concerning the cost of computation for learning and analyses, we need to reduce the computation time and computational resources such as the amount of computer memory.

As for a technology of doing computation with preserving privacy, secure computation is known. This is a technology to perform computation on encrypted data. For an application of the secure computation, we can construct databases of DNA sequences. By collecting and analyzing DNA sequences of many patients, we can heal genetic disorders. However, DNA sequence information is the “ultimate personal information”, and it is necessary to encrypt them to protect privacy. If a standard encryption scheme is used, we need to decrypt data once, and then re-encrypt after computation. This means that secret information is leaked to the computation side. On the other hand, if secure computation is used, we can perform computation without decrypting the data, and therefore the party to perform computation does not obtain the information of individual data and obtain only the result of the computation. That is, we can analyze and learn from data while preserving privacy of personal.

Concerning the computation cost, the problem cannot be resolved by simply using super computers. The problem is due to the amount of data. Though we can reduce data size by compression, if standard compression algorithms are used, we need to decompress data before computation. Therefore, we need computers with huge amount of memory. To solve this problem, we can use succinct data structures. This is a technology to perform computation without decompressing data. This enables us to process big data quickly using computers with small memory. An application of succinct data structure is DNA sequence assembly.

【Research Methods】

As shown above, secure computation and succinct data structures are basic technologies for information sharing. It is however not obvious if we can apply these techniques simultaneously. In this project, we aim to develop technologies to process various data while encrypting and compressing them. Especially we consider the following themes.

1. Oblivious Search

For DNA databases, multiple uses will perform data search queries to a server. We want to fulfill: (a) to keep the query secret, (b) fast search (polylogarithmic time in the number n of data in the server), (c) to minimize the memory space in the server.

2. Secure Learning and Analyses

We advance the concept of computation on compressed data, and work on learning from compressed data. In general, to increase the accuracy of learning, we need to make learning models more complex, that is, increase the number of parameters in learning models. However, we need more data to learn complex models. By compressing data, we can reduce the amount of data. This means that we can learn from data with fewer number of parameters.

In particular, we develop computational models for securely analyzing unstructured data such as natural languages. Though various secure computation methods based on the homomorphic encryption have been proposed, research on secure computation for unstructured data is not enough. In this project, we propose a mechanism such that both data providers and data analysts obtain merits by developing technologies for analyzing data without anonymization and decryption.

【Expected Research Achievements and Scientific Significance】

Compressed Secure Computation proposed in this project is a technology to store data in compressed and encrypted form, and to analyze and learn from data without decompression and decryption. Furthermore, we can reduce time for analyses using economical computers.

As just described, compressed secure computation will be an indispensable basic technology in Society 5.0.

【Publications Relevant to the Project】

· K. Shimizu, K. Nuida, G. Rätsch. Efficient privacy-preserving string search and an application in genomics. Bioinformatics, 32(11):1652-1661, 2016.
· Kunihiko Sadakane. Succinct Data Structures. Algorithm Science Series 8, Kyoritsu Shuppan Co., Ltd., 2018.

【Homepage Address and Other Contact Information】

https://researchmap.jp/sada/?lang=en