香港浸會大學 HONG KONG BAPTIST UNIVERSITY

DEPARTMENT OF COMPUTER SCIENCE 計算機科學系

DEPARTMENT OF COMPUTER SCIENCE

SEMINAR **2024 SERIES**

Optimal Dynamic Parameterized Subset Sampling

DATE & TIME

5 SEP 2024 (THU) 10:30 – 11:30 AM

VENUE

DLB637, 6/F, David C Lam Building, Shaw Campus

DR. JUNHAO GAN

Senior Lecturer
Computing and Information Systems
The University of Melbourne

ABSTRACT

In this talk, I will present our recent work on the problem of Dynamic Parameterized Subset Sampling (DPSS) in the Word RAM model. In DPSS, the input is a set, S, of n items, where each item, x, has a non-negative integer weight, w(x). Given a pair of query parameters, (\alpha, \beta), each of which is a non-negative rational number, a parameterized subset sampling query on S seeks to return a subset T of S such that each item x from S is selected in T, independently, with probability p_x which is the minimum between w(x) / (\alpha W + \beta) and 1, where W is the sum of the weights of all the items in S. More specifically, the DPSS problem is defined in a dynamic setting, where the item set, S, can be updated with insertions of new items or deletions of existing items. The DPSS problem generalizes several important sampling problems, including the dynamic weighted sampling and the dynamic (non-parameterized) subset sampling problems. Our first main result is an optimal algorithm for solving the DPSS problem, which achieves O(n) pre-processing time, O(1+\mu) expected time for each query parameterized by (\alpha, \beta), given on-the-fly, and O(1) time for each update; here, \mu is the expected size of the query result. At all times, the worst-case space consumption of our algorithm is linear in the current number of items in S. Our second main contribution is a hardness result for the DPSS problem when the item weights are O(1)-word float numbers, rather than integers. Specifically, we reduce Integer Sorting to the DPSS problem with float item weights. Our reduction implies that an optimal algorithm for DPSS with float item weights (achieving all the same bounds as aforementioned) implies an optimal algorithm for Integer Sorting, while the latter still remains an important open problem. Last but not least, some crucial techniques in our first main result are efficient algorithms for the generation of two special types of Bernoulli random variates and the Truncated Geometric random variates in O(1) expected time with O(n) worst-case space in the Word RAM model. Generating these random variates efficiently is of great importance not only to sampling problems but also to encryption in cybersecurity. We believe that these new algorithms may be of independent interests for related research.

SPEAKER'S BIOGRAPHY

REGISTER NOW

Enquiries: 3411-2385 **Email:** comp@comp.hkbu.edu.hk **Website:** https://bit.ly/bucs-events