DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE UNIVERSITY OF SOUTHERN DENMARK, ODENSE

COMPUTER SCIENCE COLLOQUIUM

Bloom Filter Alternatives based on Xor-Probing

Stefan Walzer Department of Computer Science University of Cologne

Tuesday, 24 May, 2022 at 14:15

IMADA's Seminar Room

Abstract:

Let \mathcal{U} be some set. An approximate membership data structure (AMQ) for a subset $S \subseteq \mathcal{U}$ tries to decide for a given $x \in \mathcal{U}$ whether or not $x \in S$. A negative answer must always be reliable, but false positives are allowed with a small probability. AMQs can be vastly more space efficient than ordinary set data structures and are widely used in data base technology and networking.

Most AMQs use an array and associate each $x \in \mathcal{U}$ with k positions in the array via hash functions. Whether x is judged to be in S or not only depends on the values v_1, \ldots, v_k found at these positions. In Xor-probing filters, x is judged to be present if $v_1 \oplus \ldots \oplus v_k$ matches a predetermined fingerprint f(x) associated with x. Finding values to put into the array such that the fingerprint matches for each $x \in S$ requires solving a system of linear equations.

This application motivates a strange task for theoreticians: Coming up with a distribution D for linear equations such that a system S of equations drawn i.i.d. from D has certain desirable properties with high probability. The speaker has contributed various ideas to this field in his dissertation and plans to give a high level overview of the subject in this talk.