May 2013

MADALGO seminar by Mikkel Thorup, University of Copenhagen

The Power of Tabulation Hashing

Abstract:

Abstract Randomized algorithms are often enjoyed for their simplicity, but the hash functions used to yield the desired theoretical guarantees are often neither simple nor practical. Here we show that the simplest possible tabulation hashing provides unexpectedly strong guarantees.

The scheme itself dates back to Zobrist [1970]. Keys are viewed as consisting of c characters. We initialize c tablesT1,…,Tc mapping characters to random hash codes. A key x=(x1,…,xc) is hashed toT1[x1]⊕⋯⊕Tc[xc], where ⊕ denotes xor.

While this scheme is not even 4-independent, we show that it provides many of the guarantees that are normally obtained via higher independence, e.g., min-wise hashing for estimating set intersection, and cuckoo hashing.

We shall also discuss a twist to simple tabulation that leads to reliable statistics with Chernoff-type concentration and extremely robust performance for linear probing with small buffers.

Joint work with Mihai Pătraşcu.