August 2007

ALCOM Seminar by Michael T. Goodrich, University of California, Irvine

Space-Efficient Straggler Identification in Round-Trip Data Streams via Newton's Identities and Invertible Bloom Filters

We introduce the straggler identification problem, in which an algorithm must determine the identities of the remaining members of a set after it has had a large number of insertion and deletion operations performed on it, and now has relatively few remaining members. The goal is to do this in o(n) space, where n is the total number of identities. The straggler identification problem has applications, for example, in determining the set of unacknowledged packets in a high-bandwidth multicast data stream.
We provide a deterministic solution to the straggler identification problem that uses only O(d log n) bits and is based on a novel application of Newton's identities for symmetric polynomials. This solution can identify any subset of d stragglers from a set of n O(log n)-bit identifiers, assuming that there are no false deletions of identities not already in the set. Indeed, we give a lower bound argument that shows that any small-space deterministic solution to the straggler identification problem cannot be guaranteed to handle false deletions. Nevertheless, we show that there is a simple randomized solution using O(d log n log(1/epsilon)) bits that can maintain a multiset and solve the straggler identification problem, tolerating false deletions, where epsilon>0 is a user-defined parameter bounding the probability of an incorrect response. This randomized solution is based on a new type of Bloom filter, which we call the invertible Bloom filter.