COMBINATORICS SEMINAR
The Power of Two Choices: Some Old Results and New Variations

Michael Mitzenmacher
Harvard University

February 9, 2005

Abstract:

In this talk, I present some history and recent results related to the paradigm of the "power of two choices." The initial result is the following: it is well known that when n balls are thrown uniformly at random into n bins, the maximum load (with high probability) grows like (1+o(1))log n / log log n. In 1994, Azar, Broder, Karlin, and Upfal proved the following interesting result: if instead the n balls are thrown sequentially, each ball has two (uniform) choices of where to be placed, and each ball is placed in the least loaded bin at the time of placement, the maximum load is only log log n / log 2. This is useful for several load-balancing applications, including hash tables. We give a combinatorial proof of this result.

In recent applications relating to peer-to-peer and sensor networks, bins may correspond to spaces defined by the underlying geometry that may not be equal in size; locations therefore cannot be modelled as being chosen uniformly at random. This leads us to consider geometrical generalizations of the power of two choices. For search engines, we are concerned with word frequencies; the goal is to keep the number of documents with a given word stored at each server balanced simultaneously for all (sufficiently frequent) words. In this case, balls correspond to documents, but now a ball is really a vector (or word list) instead of a single item. This leads us to consider multi- dimensional generalizations of the power of two choices. In both cases, we find similar improvements as in the standard balls-and-bins scenario