

Solution notes

Data Structures and Algorithms 2005 (MR) Paper 4 Question 3, Paper 11 Question 4

Hash tables are a fundamental part of the course.

- (a) Easy bookwork. The hash should be a uniformly distributed pseudo random function of the key yielding a value in the range $0..B-1$. The probe sequence should be a pseudo random permutation of $0..B-1$, or at least have no short loops.
- (b)(i) This is cheap to compute but leads to clustering and so is bad for tables that are getting full. Better use of on-chip caches.
- (ii) The clustering behaviour of this is just as bad as for (i).
- (iii) This remove the primary clustering of (i) and (ii) above, but as the sequence of secondary probes depends only on the initial probe there is a danger of secondary clustering. This is not as bad as primary clustering but could be avoided. B must not be divisible by 13.
- (iv) In this the secondary probe sequence depends on the initial probe *and* the key, and so it is unlikely that probe sequences from different keys will have many probes in common, thus almost eliminating clustering. Note that $\text{Hash2}(\text{key})$ need only be computed once at the start of the secondary probe sequence and, indeed, can be a byproduct of computing the primary probe. Harder to check whether the probe sequences contain loops. OK if B is prime.
- (d) To delete a key-value pair in a closed hash table, find it, then replace it by a tombstone. Probes will step over tombstones. It is hard to remove tombstones since they may lie in the path of several probe sequences. If there is much deletion going on the table gradually fills with tombstones and thus becomes inefficient. When this happens the table should be rebuilt (with no tombstones). If the probe sequence was based on (i) above, a tombstone can be removed if it is followed by a blank entry. This is a slight advantage to (i). The following optimisation is possible. If a tombstone is found during a successful lookup, swap the tombstone with the discovered (k,v) pair. Tombstones will tend to migrate to the end of probe chains reducing the slight inefficiency they cause.