Randomized Algorithms 2024-5
Lecture 8b

Cuckoo Hashing*

Moni Naor

We discussed the dictionary problem and various hashing approaches to it. The main 20th-century candidate is linear probing. In general, the type of error probability we have in this area is $1/n^c$ (and not exponentially small). We should mention the important open problem of showing that randomness is essential to achieve efficient ($O(1)$) dictionaries.

The 21st century saw a new invention: **cuckoo hashing** [3]. We went over Cuckoo Hashing and its analysis. We discussed what is needed to detect failure. In particular, we mentioned that a good cycle detection mechanism is useful.

Question: suggest a low memory cycle detection.

The probability of failure of cuckoo hashing after n insertions is $1/n$. We discussed the idea of a stash to reduce the probability of needing to rehash [2]. The point is that using a stash of size c reduces the failure probability to $1/n^{c+1}$ where

Question: We saw the proof that if there are no large components, then the probability of failure (= a component with 2 cycles) is roughly $1/n$. Extend the argument to show that the probability one needs to put c elements into the stash is rough $1/n^{c+1}$.

References

[1] J. L. Carter and M. N. Wegman, Universal classes of hash functions, J. Comput. Syst. Sci. 18 (1979) 143-154.

[2] Kirsch, Mitzenmacher, Wieder, More robust hashing: Cuckoo hashing with a stash, ESA 2008.

[3] Rasmus Pagh and Flemming Friche Rodler: Cuckoo hashing. J. Algorithms 51(2): 122-144 (2004).

[4] R. Panigrahy, Efficient hashing with lookups in two memory accesses, SODA 2005

*These notes summarize the material covered in class, usually skipping proofs, details, examples and so forth, and possibly adding some remarks, or pointers. In the interest of brevity, most references and credits were omitted.