1 Hash tables

Hash tables with external hash families guarantee an expected runtime of \(O(1)\) for the INSERT, SEARCH, and DELETE operations. What is the meaning of \(\Omega(\log n)\)?

○ It is an average over the choice of the adversary who picks the elements in the table.
○ It is an average over the choice of the algorithm who picks the hash family from the hash family.

In order to conclude an expected runtime of \(O(1)\) for hash table operations, we assumed the following two loopholes in some specific order:

- The adversary selects elements \(u_1, \ldots, u_n\) for the hash table.
- The algorithm picks a hash function from the hash family.

In what order do these loopholes?

○ Algorithm first, and then adversary.
○ Adversary first, and then algorithm.

It does not matter.

Correct

2 Bit lengths

Suppose that there is a toy box with \(M\) toys in it. You have a label printer that can print arbitrary strings of 0s and 1s. If you produce labels for the \(M\) toys in such a way that each toy gets a unique label, what can be said about the longest label’s length?

- \(\Omega(\log n)\)
- \(\Theta(1)\)
- \(\Theta(\log M)\)
- \(\Theta(M)\)
- \(O(\log M)\)

Correct

If our toy box consists of all functions from \(\{0, \ldots, M−1\}\) to \(\{0, \ldots, n−1\}\), what is the minimum label’s length?

- \(\Theta(n)\)
- \(\Theta(1)\)
- \(\Theta(M/n)\)
- \(\Theta(M)\)
- \(O(\log M)\)

Correct

How about the total number of bits used for labeling the \(M\) toys?

- \(\Omega(\log M)\)
- \(\Theta(n)\)
- \(\Theta(1)\)
- \(\Theta(M/n)\)
- \(O(\log M)\)

Correct

3 Modular arithmetic

Suppose that \(M \equiv 1000\pmod{10}\) (the universe size) is a prime number. If we pick \(x \in \{1, \ldots, M−1\}\) and \(n \in \{1, \ldots, M−1\}\), independently and uniformly at random, what is \(\Pr(x \neq \{\overline{0}, \ldots, \overline{n−1}\})\)?

- \(\Theta(n)\)
- \(\Theta(1)\)
- \(\Theta(M/n)\)
- \(\Theta(M)\)
- \(O(\log M)\)

Correct

In fact for any pair of distinct elements \(x, y\) in the universe, \(n = x \times y = b \mod M\) and \(x = a \times y = \overline{b} \mod M\) are uniformly distributed amongst all distinct pairs.

How many elements of \(\{1, \ldots, M−1\}\) are equal to \(0\) modulo \(n\)?

- \(\Theta(n)\)
- \(\Theta(1)\)
- \(\Theta(M/n)\)
- \(\Theta(M)\)

Correct

Infact, for any \(b\), the number of elements of \(\{1, \ldots, M−1\}\) equal to \(b\) modulo \(n\) is \(\lfloor M/n \rfloor\).

Let \(a \in \{0, \ldots, M−2\}\) be uniformly random from \(\{0, \ldots, M−2\}\). What is the maximum \(x\) such that \(x = a \mod n\)?

- \(\Theta(n)\)
- \(\Theta(1)\)
- \(\Theta(M/n)\)
- \(\Theta(M)\)

Correct

You can verify that the answer above is always \(\leq \lfloor M/n \rfloor\). The same answer holds as an upper bound if we changed \(x\) from \(0\) to any other element in \(\{1, \ldots, M−1\}\).

4 Hash family size

Suppose that we have a universe of size \(M\), and our hash table size is \(n\). If \(n \geq M\), what is the minimum size of a universal hash family?

- \(\Theta(1)\)
- \(\Theta(n)\)
- \(\Theta(M/n)\)
- Both of the above

Correct

This means that if \(n\) is a universal, then:

\[
\frac{1}{n} \sum_{i=0}^{n-1} \Pr(x_i = \{\overline{0}, \ldots, \overline{n−1}\}) \geq \frac{1}{n} \frac{n}{M} \geq \frac{1}{n}
\]

or in other words \(M \leq n\). What can be said about the minimum longest \(\Theta(1)\) label length for labeling this hash family?

- \(\Theta(M)\)
- \(\Theta(\log M)\)
- Both of the above.

Correct

For the universal hash family from lecture, how many bits do we need to label the hash functions, if we minimize the longest label’s length?

- \(\Theta(n)\)
- \(\Theta(M)\)
- \(\Theta(\log M)\)

Correct

This shows the hash family from lecture can be labeled by the optimal number of bits \(\Theta(M) = \Theta(\log M)\) when \(M = \omega(n)\).