1 Hash tables

Hash tables with universal hash families guarantee an expected runtime of $O(1)$ for the INSERT, SEARCH, and DELETE operations. What is the meaning of "expected"?

- This is an average over the choices of the adversary who picks the elements in the table.

- The adversary picks elements $u_1, \ldots, u_N$, for the hash tables.

- The algorithm picks a hash function from the hash family.

- In what order do these happen?

Correct: Algorithm first, and then adversary.

- Advanced first, and then algorithm.

- It does not matter.

2 Bit lengths

Suppose that there is a toy box with $N$ toys in it. You have a label printer that can print arbitrary strings of bits, but in such a way that each toy gets a unique label, what can be said about the longest label's length?

- $O(\log M)$

- $\leq \log(N)$

- $\geq \log(N)$

Correct: $O(\log M)$.

As a remark, for any labeling scheme, the same lower bound of $\Theta(\log N)$ applies even to the average label length, not just the longest label length.

If you produce labels in a way that minimizes the longest label's length, what is this minimum?

- $\Theta(M)$

- $\Theta(M/n)$

- $\Theta(\log M)$

Correct: $\Theta(M/n)$.

If our toy box consists of all functions from $\{0, \ldots, M-1\}$ to $\{0, \ldots, n-1\}$, what is the minimum longest label's length?

- $\Theta(M)$

- $\Theta(M/n)$

- $\Theta(\log M)$

Correct: $\Theta(\log M)$.

How about $\Theta(M^n)$

- $\Theta(M)$

- $\Theta(M^n)$

- $\Theta(M^n)$

Correct: $\Theta(M^n)$.

3 Modular arithmetic

Suppose that $M = 1000$ (the universe size) is a prime number. If we pick $a \in \{1, \ldots, M-1\}$ and $b \in \{0, \ldots, M-1\}$, independently and uniformly at random, what is $P_{a,b}(x) \equiv x^a + b \mod M$?

- $\Theta(M)$

- $\Theta(M/n)$

- $\Theta(\log M)$

Correct: $\Theta(M)$. Let $h \equiv x^a \mod M$. Then $P_{a,b}(x) \equiv x^a + b \mod M$.

In fact for any pair of distinct elements $x, y$ in the universe, $a \equiv x - y \mod M$ and $x \equiv a + y \mod M$ are uniformly distributed amongst all distinct pairs.

How many elements of $\{0, \ldots, M-1\}$ are equal to 0 modulo $a$?

- $[M/a]$

- $[M/a]$

- $[M/a]$

Correct: $[M/a]$.

In fact, for any $x$, the number of elements of $\{0, \ldots, M-1\}$ equal to 0 modulo $x$ is $\lfloor M/x \rfloor$.

Let $a \equiv b \pmod{p}$ uniformly at random from $\{0, \ldots, M-1\}$. What is the chance that $a \equiv b \pmod{p}$?

- $\Theta(1)$

- $\Theta(M/n)$

- $\Theta(1)$

Correct: $\Theta(1)$. You can verify that this answer above is always $\leq 1/n$. The same answer holds as an upper bound if we changed a from 0 to any other element in $\{0, \ldots, M-1\}$.

4 Hash family size

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

- $0$

- $\geq N$

- $\leq N$

Correct: $0$.

Suppose now that $M = n^{10}$ and we have a nonempty hash family $H$. Let $b$ be one of the hash functions of $H$. Since $b$ is uniformly distributed elements $x \in \mathbb{Z}$ in the universe to the same bucket (by the pigeonhole principle). What can be said about

$P_{a,b}(x) \equiv x^a + b \pmod{M}$?

- $b \equiv 0 \pmod{N}$

- $\geq \lceil M/N \rceil$

- $\leq N$

Correct: $0$. This means that if $M$ is even, then

$1/n = \Pr(a \neq b, x \neq b) = \Pr(a \neq b) \Pr(x \neq b) = \Pr(a \neq b) \Pr(x \neq b)$

or in other words $|H| \geq N$. What can be said about the minimum longest $0/1$ label length for labeling this hash family?

- $\geq \lceil M/N \rceil$

- $\geq \lceil M/N \rceil$

- $\geq \lceil M/N \rceil$

Correct: $\geq \lceil M/N \rceil$. Both of the above.

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?

- $\Theta(M)$

- $\Theta(M/n)$

- $\Theta(1)$

Correct: $\Theta(M/n)$. This shows the hash family from lecture can be labeled by the optimal number of bits for $b \geq \lceil M/N \rceil$ when $M = n^{10}$.