Lecture 27

Wednesday, 20 October 2021 6:14 PM

COL 351: Analysis and Design of Algorithms

Lecture 27

Two Algorithms involving

Modulo arithmetics

Universal Hashing **String Matching**

Problem

Given:

- $U = [1, 2, \ldots, M]$, universe with M elements.
- $S \subseteq U$ of size n
- $n << M$

Find:

For the given S, find a data-structure of $O(n = |S|)$ size to answer in $O(1)$ time queries of form:

"Does $z \in S$?" $1 \leq z \leq M$

	Search-Time	Space
Array	$O(1)$	$O(M)$
Link List	$O(n)$	$O(n)$
AVL Tree	$O(\log n)$	$O(n)$

Hashing

$$\sum_{i=0}^{n-1} |T[i]| = O(n)$$

- Given: Hash Function $H : U \to [0, n-1]$.
- Using H compute hash-table " T " of size n.

Property of Table T :

$T[i]$ — List storing $\{z \in S \mid H(z) = i\}$

($T[i]$ can also be empty)

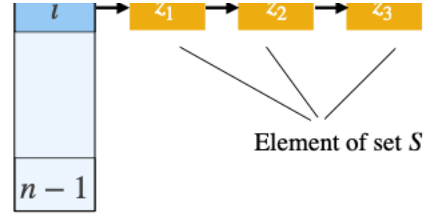

Table T

Search-Query(z)

1. Compute $i = H(z)$
2. Scan the link-list stored at $T[i]$
3. If $z \in T[i]$ return "Found", else return "Not-found"

Simplest Hash Function

$$\boxed{H(z) := \underbrace{z \bmod n}_{\{0,1,..,n-1\}}}$$

$H : [1, M] \to [0, n-1]$

- Bad for sets like $S = \{n, 2n, 3n, \ldots, n^2\}$
- Reason: $|T[0]| = n$

- Good for a random S

"COLLISION-PROB"

If $0 \le i \le n-1$ and $x \in [1, M]$ is random, then

$Pr[x \bmod n] = i$ is $\dfrac{\{i, n+i, 2n+i, .., \lfloor \frac{M}{n} \rfloor (n-1)+i\}}{M} \approx \dfrac{1}{n}$

So, for random x, y, Prob($x \bmod n = y \bmod n$) is

$$\sum_{i=0}^{n-1} \Pr[x \bmod n = y \bmod n = i] \approx \sum_{i=0}^{n-1} \frac{1}{n} \cdot \frac{1}{n} = \frac{1}{n}$$

Simplest Hash Function

$$H(z) := z \bmod n$$

"COLLISION"

• Suppose $S = \{s_1, s_2, \ldots, s_n\}$ where every s_i is a uniformly random integer in $U = [1, M]$.	• ***Question:*** For random $x, y \in U$, what is the probability that $H(x) = H(y)$? • Prob($H(x) = H(y)$) $= \sum_{i=0}^{n-1} \frac{\lvert\{i,\ n+i,\ 2n+i,\ \ldots\}\rvert^2}{M^2}$ $= \frac{1}{n}$	• ***Question:*** For a given $x \in [1, M]$, what is expected time to check if $x \in S$? • Let $i = H(x)$. $\mathrm{Exp}(\lvert T[i]\rvert) = 1 + \sum_{y \in S \setminus \{x\}} \mathrm{Prob}(H(x) = H(y)) = O(1)$ • Therefore, time to search x is (i) Time to compute $i = H(x)$, and (ii) $\lvert T[i]\rvert$ which is O(1) on expectation.

Simplest Hash Function

$$H(z) := z \bmod n$$

Works well for a random S

What if S is <u>not</u> random?

How to achieve O(1) search time for ALL possible sets S?

Lemma: No single hash function can work for all possible sets S

Reason :

For any Hash Fn "H"

$$\max_{0 \le i \le n-1} \begin{pmatrix} \text{No of elements} \\ x \text{ in } U \text{ s.t.} \\ H(x) = i \end{pmatrix} \geq \frac{M}{n}$$

Goal: Finding a good hash function for a given set S

c-Universal Hash Family:

C- Universal Hash Family

HASH FUNCTION :

HASH Family :

$H(z) := z \mod n$	$Hash - Family\ \mathscr{H}$ { Collection of Hash Functions
KEY PROPERTY:	KEY PROPERTY:
• For random $x, y \in U$, the probability that $H(x) = H(y)$ is at-most $1/n$.	• For $x, y \in U$, and random hash-function $H \in \mathscr{H}$, the probability that $H(x) = H(y)$ is at-most c/n.
• Works well with random sets.	• To show: Works well with any choice for set S

Hash Family

Redistributing elements of S

$$\boxed{H_r(z) := (r\,z \mod p) \mod n}$$

- Here $p \in [M+1, 2M]$ is a prime number, and
- r is a integer in range $[1, p-1]$ which is independent of set S

Hash Family: $\{H_r(z) \mid r \in [1, p-1]\}$

Claims from Lec 26

$$\boxed{F(z) := (r \cdot z \mod p)}$$

$$\boxed{\begin{aligned} H_r(z) &= F(z) \bmod n \\ &= rz \bmod p \bmod n \end{aligned}}$$

Claim 1: For any $r \in [1, p-1]$, we have $r^{p-1} = 1 \mod p$

Claim 2: The function $F(z)$ is invertible, and its inverse is given by $F^{-1}(y) := (r^{p-2}\, y) \mod p$

Claim 3: If $r \in [1, p-1]$ was random, then for any $z, i \in [1, p-1]$, we have $\text{Prob}\big(F(z) = i\big) = \dfrac{1}{p-1}$.

If $z \in [1, M]$ was random then $\text{Prob}[z = i] = \cancel{1/m}\ 1/M$	If $z \in [1, M]$ is deterministic, but r is random then $Pr(F(z) = i) = \frac{1}{p-1}$

Hash Family

$$\boxed{H_r(z) \;:=\; (r\,z \mod p) \mod n}$$

- Here $p \in [M+1, 2M]$ is a prime number, and
- r is a random integer in range $[1, p-1]$ which is independent of set S

Question: For distinct $x, y \in [1, M]$, and random $r \in [1, p-1]$, what is probability $H_r(x) = H_r(y)$?

Hash Family

$$\boxed{H_r(z) \;:=\; (r\,z \mod p) \mod n}$$

Question: For distinct $x, y \in [1, M]$, and random $r \in [1, p-1]$, what is probability $H_r(x) = H_r(y)$?

Solution:

$$\text{Prob}\big(H_r(x) = H_r(y)\big) = Prob\big((rx \mod p \mod n) = (ry \mod p \mod n)\big)$$

$rx \bmod p - ry \bmod p = \pm in \ (\bmod p)$

$$= Prob\big((rx - ry \mod p) \in \{0, n, -n, 2n, -2n, \ldots\} \mod p\big)$$

By Claim 3

$$\le \boxed{\frac{1}{p-1}} \cdot |\{0, n, -n, 2n, -2n, \ldots\} \mod p|$$

$$\approx \frac{1}{p-1} \cdot 2 \cdot \frac{p-1}{n} = \frac{2}{n}$$

THEOREM: $\mathcal{H} = \{ H_r \mid 1 \le r \le p-1 \}$ is a 2-universal hash family

Hash Family

$$H_r(z) := (r\,z \mod p) \mod n$$

Contribution of y in $T[i]$

$$= \begin{cases} 1 & \text{if } H_r(x) = H_r(y) \\ 0 & \text{o/w} \end{cases}$$

• Suppose $S = \{s_1, s_2, \ldots, s_n\}$ is a subset of $U = [1, M]$.	• ***Question:*** For distinct $x, y \in [1, M]$, and random $r \in [1, p-1]$, what is probability $H_r(x) = H_r(y)$?? • $\text{Prob}(H_r(x) = H_r(y))$ $\le \frac{2}{n}$	• ***Question:*** For a given $x \in [1, M]$, what is expected time to check if $x \in S$? • Let $i = H(x)$. $\text{Exp}(\ \lvert T[i]\rvert\ ) = 1 + \sum_{y \in S \setminus \{x\}} \text{Prob}(H_r(x) = H_r(y)) = O(1)$ • Therefore, time to search x is (i) Time to compute $i = H_r(x)$, and (ii) $\lvert T[i] \rvert$ which is O(1) on expectation.

Hash Family

$$H_r(z) := (r\,z \mod p) \mod n$$

Question: What is expected value of:

$$\max_{x \in S} \text{(Time to check if } x \in S \text{) ?}$$

Solution: Coming lecture / tutorial

Example

Expected versus Max-of-Expected

Suppose X_1 and X_2 are values obtained of independent dice throws

Then $E(X_1) = E(X_2) = \dfrac{1+2+3+4+5+6}{6} = 3.5$

However, $E(\max(X_1, X_2))$

$$= 1\left(\frac{1}{36}\right) + 2\left(\frac{3}{36}\right) + 3\left(\frac{5}{36}\right) + 4\left(\frac{7}{36}\right) + 5\left(\frac{9}{36}\right) + 6\left(\frac{11}{36}\right)$$

$$= 4.472 > 3.5$$