COL 351: Analysis and Design of Algorithms

Lecture 18

Hash Function

Definition: A function that can be used to map data of arbitrary size to fixed-size output.

Examples:

$$x \mapsto x \ (\mathbf{mod}\ n)$$

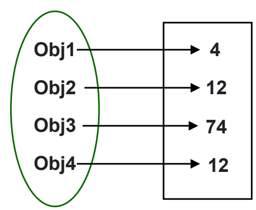
$$x \mapsto (2x^2 + 4x - 5) \ (\mathbf{mod}\ n)$$

Applications:

Message Digest

Password Verification

Data-structures



Block-chains

Two applications of Hash Functions

1. Set Membership

2. Pattern Matching

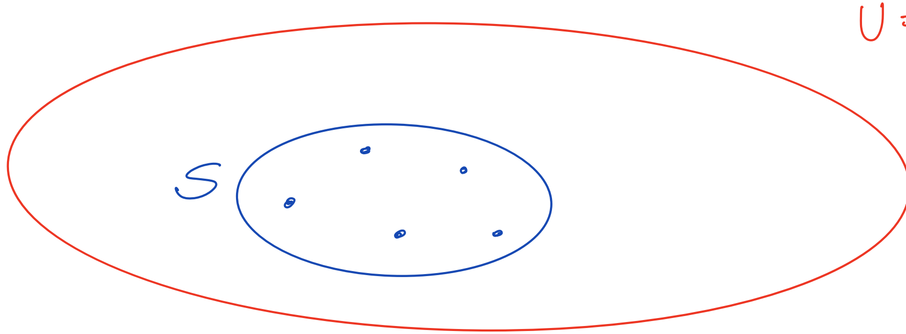
Set Membership

Given: A universe $U = [1, 2, \ldots, M]$, and a set $S \subsetneq [1, M]$ of size n.

Goal: Find a data-structure of $O(n = |S|)$ size that answers for any $x \in [1, M]$ query of form:

"Does $x \in S$?"

$U = [1, 2, 3, \ldots]$



Set Membership

Given: A universe $U = [1, 2, \ldots, M]$, and a set $S \subsetneq [1,M]$ of size n.

Goal: Find a data-structure of $O(n = |S|)$ size that answers for any $x \in [1,M]$ query of form:

"Does $x \in S$?"

	Search-Time	Space
Boolean Array (Yes/No)	$O(1)$	$O(M)$
Link List storing S	$O(n)$	$O(n)$
AVL Tree storing S	$O(\log n)$	$O(n)$

Hash Table

Eg: $H(3) = 3 \bmod n$

Given: Hash Function $H : U \to [0, n-1]$.

Table T of size n:

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

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"

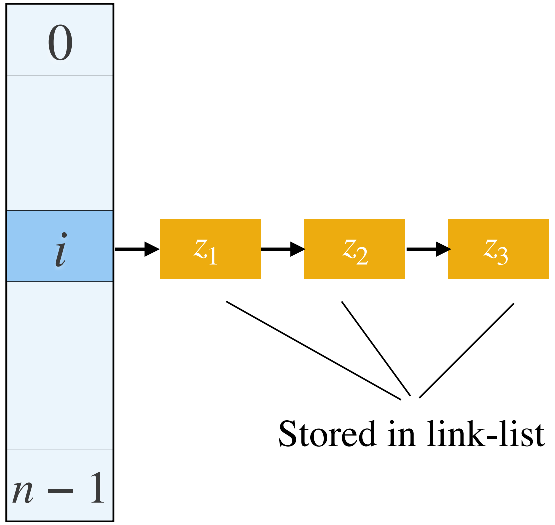


Table T

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

Simplest Hash Function

$$H(z) = z \mod n$$

- Bad for sets like $S = \{n, 2n, 3n, \ldots, n^2\}$
- Good for a **random** S

Reason:

We will have

$|T[0]| = n$

$|T[i]| = 0$, for $i > 0$

Simplest Hash Function

$$H(z) \;=\; z \mod n$$

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 collision probability $\big($probability that $H(x) = H(y)\big)$?

Solution:

Suppose $i = H(x)$.

$$\text{Prob}\big(H(y) = i\big) \;=\; \frac{|\{i,\, n+i,\, 2n+i,\, \ldots\}|}{M} \;\approx\; \frac{1}{n}$$

Eg. $M = 1000$
$|S| = 10$
$x = 17$

Now, for random $y \in [1, 1000]$

$\text{Prob}(y \bmod 10 = 7) = \frac{1}{10}$, i.e. $\frac{1}{n}$

Simplest Hash Function

$$H(z) = z \mod n$$

Suppose $S = \{s_1, s_2, \ldots, s_n\}$ where every s_i is a uniformly random integer in $U = [1, M]$.

Question: For a given $x \in [1, M]$, what is expected time to verify if $x \in S$?

Solution:

Suppose $i = H(x)$.

$$\mathrm{Exp}\big(|T[i]|\big) = 1 + \underbrace{\sum_{y \in S \setminus \{x\}} \mathrm{Prob}\big(H(y) = i\big)}_{} = 1 + (n-1)\left(\frac{1}{n}\right) = O(1)$$

$$|T[i]| = 1 + \sum_{y \in S \setminus \{x\}} \mathbb{1}_{H(y)=i}$$

Each $y \in S \setminus \{x\}$ contributes one unit to $T[i]$ iff $H(y) = i$.

Thus, time to search x is sum of

(i) Time to compute $i = H(x)$, and

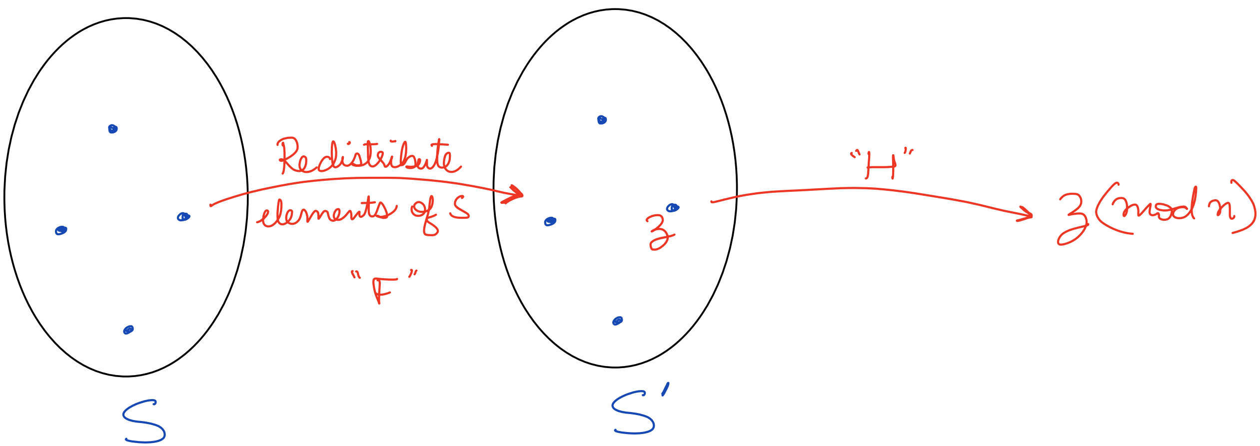
(ii) $|T[i]|$ which is $O(1)$ on expectation.

Simplest Hash Function

$$H(z) \;=\; z \mod n$$

- Works well for a random S

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



Modular Arithmetics

$$F(z) = (r \cdot z) \mod p \qquad \text{(Here, } p \text{ is a prime).}$$

Eg. of a different number system

$$A = \{0, 1, 2, 3, 4, 5, 6\}$$

Must be prime

New addition "$\oplus$" : $a \oplus b = a + b \mod 7$

New product "$\otimes$" : $a \otimes b = a \otimes b \mod 7$

$2 \oplus ? = 0$ Ans $= 5$	$5 \oplus ? = 0$ Ans $= 2$	$5 \otimes ? = 1$ Ans $= 3$	$4 \otimes ? = 1$ Ans $= 2$

Additive inverse

Multiplicative inverse

Modular Arithmetics

Remark

$^pC_i = 0 \pmod p$

for $i \in [1, p-1]$

$F(z) \;=\; (r \cdot z) \mod p$ (Here, p is a prime).

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

Proof:

Suppose claim holds for $r \leq p-1$.

$$(r+1)^p = \underbrace{r^p}_{= r} + \underbrace{\sum_{i=1}^{p-1} {}^pC_i}_{= 0} + 1 \pmod p = r + 1 \pmod p$$

So, $\dfrac{(r+1)^p - (r+1)}{p} = \dfrac{(r+1)\left((r+1)^{p-1} - 1\right)}{p}$ is integer

As p doesn't divide $r+1$, we prove the claim.

Modular Arithmetics

$F(z) \;=\; (r \cdot z) \mod p$ (Here, p is a prime).

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

Proof:

$$z \xrightarrow[F]{\cdot r} \underset{\text{mod } p}{rz} \xrightarrow[F^{-1}]{\cdot r^{p-2}} \underbrace{r^{p-2} \cdot r}_{=1} \cdot z \pmod p$$

Modular Arithmetics

$$F(z) \;=\; (r \cdot z) \;\bmod p \qquad \text{(Here, } p \text{ is a prime).}$$

Claim 3: If $r \in [1, p-1]$ was random, then for any $z, i \in [1, p-1]$, we have

$$\text{Prob}\big(F(z) = i\big) = \frac{1}{p-1}.$$

Proof:

Note: z, i are fixed, but r is random.

$F(z) = i \iff rz \pmod p = i \iff r = z^{p-2}\, i \pmod p$

So, $\text{Prob}_r\big(F(z) = i\big) = \underbrace{\text{Prob}_r\left(r = z^{p-2}\, i \pmod p\right)}_{\text{This is } \left(\frac{1}{p-1}\right) \text{ as } r \text{ has } p-1 \text{ possiblities}}$

Modular Arithmetics

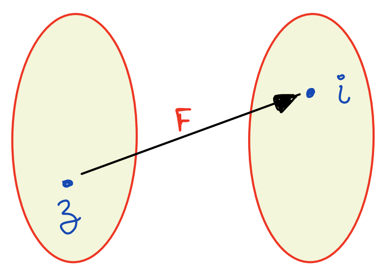
$F(z) \;=\; (r \cdot z) \mod p$ (Here, p is a prime).

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

Claim 2: $F(z)$ 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

$$\text{Prob}\big(F(z) = i\big) = \frac{1}{p-1}.$$



Both input/output sets have size $p-1$