EXERCISES

(a) Linear Probing

$h(k\ i) = k \% 13 + i$

0	25
1	1
2	14
3	2
4	
5	
6	6
7	
8	8
9	
10	
11	24
12	12

insert these inorder

1
8
14
2
6
24
12
25

load factor $\alpha = 3/4$

how many slots do you need to probe to find...

2	2
5	1
26	5
8	1
18	1
11	7
9	1

expected # of slots = 4

(b) Double Hashing

$$h(k\ i) = (h_1(k) + i h_2(k)) \% m$$
$$= (k\%13) + i(k\%12)$$

0	25
1	1
2	2
3	14
4	
5	
6	6
7	
8	8
9	
10	
11	24
12	12

insert the same elements.

1, 8, 14, 2, 6, 24, 12, 25

load factor $\alpha = 3/4$

expected # of probes = 4 ?

probes to find...

2	1
5	1
26	3
8	1
18	1
11	2
9	1

Recitation 7

Agenda

(11.3.3) ∘ universal hashing
(11.5) ∘ perfect hashing
∘ MD5 & 6 (?)

Reminders

feedback today
Submit early: avoid the rush - soln soon - ret. W
exam conflicts: email (reason, Times R 8am-8pm avail) W 10/15 7:30-9:3

Warmup: linear probing & double hashing problems.

universal hashing

idea: choose hash fcn at random from a family of hash fcns.

- can behave differently on each run.

$\mathcal{H} : U \to \{0 \cdots m-1\}$

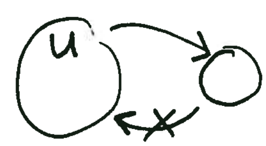
universal: $\forall k_1\, k_2 \in U \quad k_1 \neq k_2$
num ($h \in \mathcal{H}$) for which $h(k_1) = h(k_2)$ (they collide) $< \frac{|\mathcal{H}|}{m}$
so chance of collision is just $1/m$.

now an adversary cannot force worst case run time.

our universal hash: $h_{a,b}(k) = ((ak+b) \bmod p) \bmod m$

$$\boxed{\mathcal{H}_{pm} = \{h_{ab} : a \in \mathbb{Z}_p^*,\ b \in \mathbb{Z}_p\}}$$

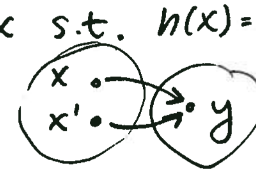
$\mathbb{Z}_p = \{0\ 1 \cdots p-1\}$ integers mod p
$\mathbb{Z}_p^* = \{1\ 2 \cdots p-1\}$
p is prime
and $p > |U|$



more desireable hash properties:

cryptographic

- one way : infeasible given $y \in_R \{0\,1\}^d$ to find any x s.t. $h(x) = y$
- collision resistance : can't find x, x' s.t. $h(x) = h(x')$
- weak " : given x can't find $x' \neq x$ s.t. $h(x') = h(x)$
- psuedorandom : looks random (can't be b/c is also repeatable)
- non-malleability: given $h(x)$ can't produce $h(x')$ for some x' related to x.

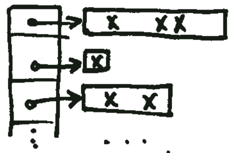


⇒ collisions can be brute forced in $O(2^{d/2})$ (b-day problem)
⇒ inversions " " $O(2^d)$

ex: passwords, file modified

Perfect Hashing

- Static keyset
- 2 level hashing scheme



- universal hash @ each level.
- ~~goal~~ goal: guarantee $\Theta(1)$ for operations. (w/o hugely excess space allocation)

level 1 : same as for hashing w/ chaining. (chosen from $\mathcal{H}_{p,m}$)
level 2 : a secondary hash table for the keys that collide in slot(i)
with its own hash fcn (h_i) chosen from $\mathcal{H}_{p\,m_i}$

NO COLLISIONS

Size $m_i = \left(\begin{matrix}\#\text{ colliding elements}\\ \text{in slot } i\end{matrix}\right)^2$ ← why? b/c the math works out this way.

$\Rightarrow$ this gives probability of any collisions existing $= 1/2$

$\binom{n}{2}$ pairs that could collide w/ pr. $\frac{1}{m_i}$ $m_i = n^2$

so $\binom{n}{2}\frac{1}{n^2} = \frac{n!}{2!(n-2)!}\frac{1}{n^2} = \frac{n(n-1)}{2n^2} < \frac{1}{2}$

— if collide re choose h_i and try again. We expect to have to try 2 h_i's to find a non-colliding one.

overall mem = $O(n)$:)

see book for proof

idea: it is unlikely that the first hash will use only a few of its slots (since it's universal). In order to get bad space performance this would have to happen.

- example ~~excersize~~ exercise.

Cuckoo Hashing

- X can be in $h_1(x)$ or $h_2(x)$ → lookup takes const time (2 slots)
- one elt per slot
- if slots full, evict element there.
  - ∞ loops : set time limit & rehash (can resize too)

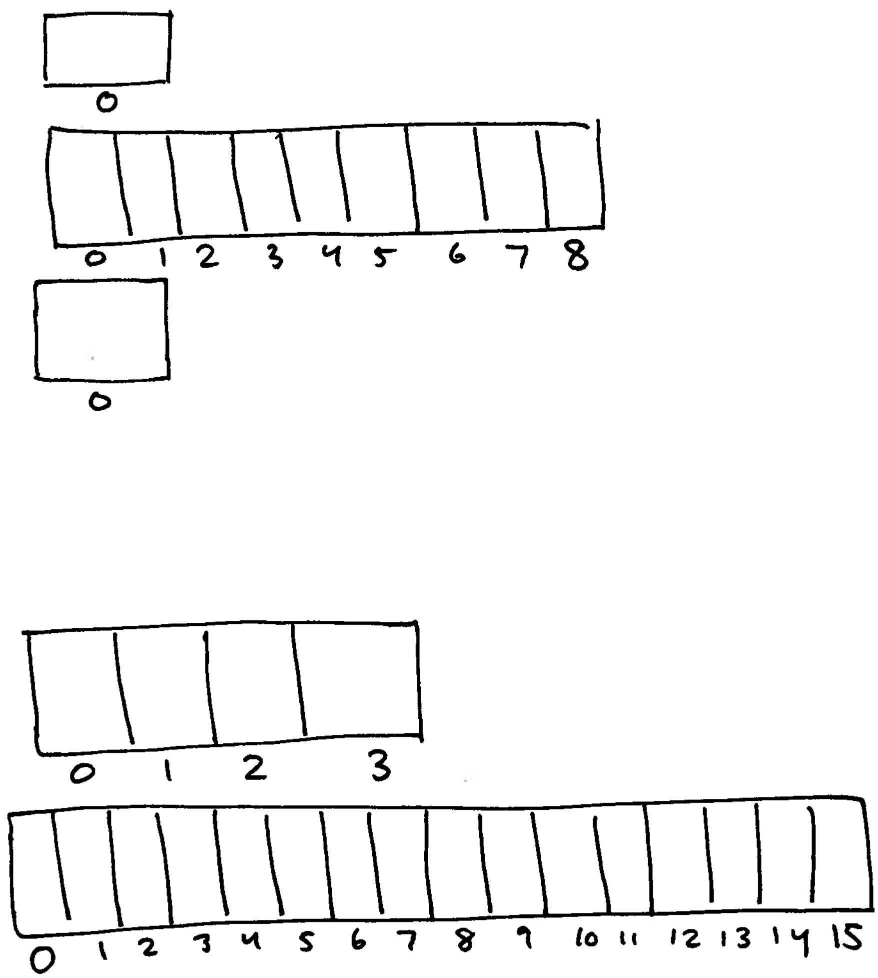
constant amortized cost.

(c) Perfect Hashing

level 1 use hash(k) = k % 7

		a	b
0	/	/	/
1	8	0	0
2	-12 10 23		
3	17	0	0
4	/	/	/
5	39 5		
6	6 -1 41 27		

level 2 for each bucket: choose a, b
$1 \leq a \leq 53$ $0 \leq b < 53$
arbitrarily & see if any collisions occur.

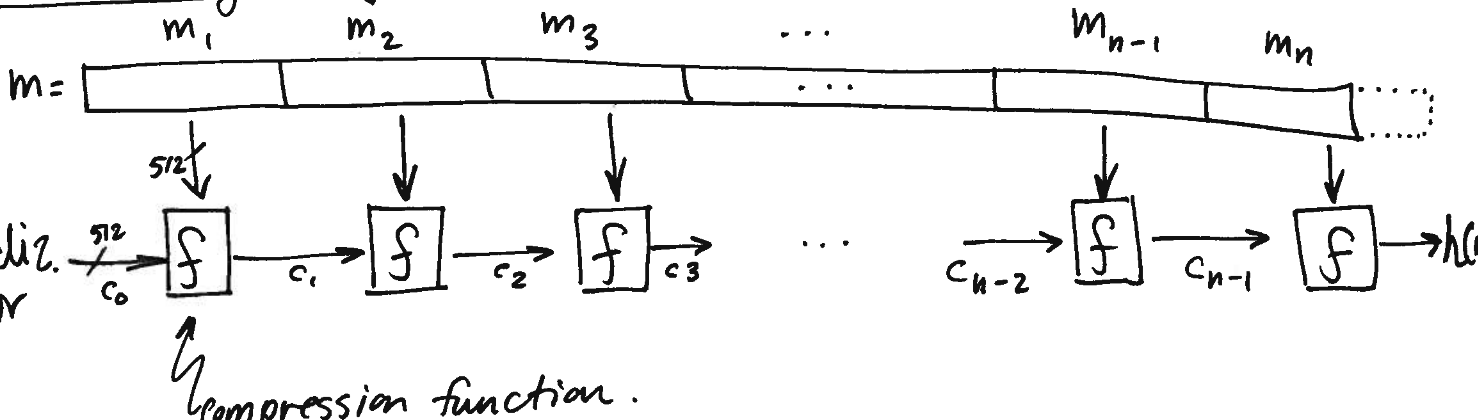


solutions depend on choice of a & b

place the following elts in their level 1 hash bucket

17 8 39 41 -1 10
23 27 -12 5 6

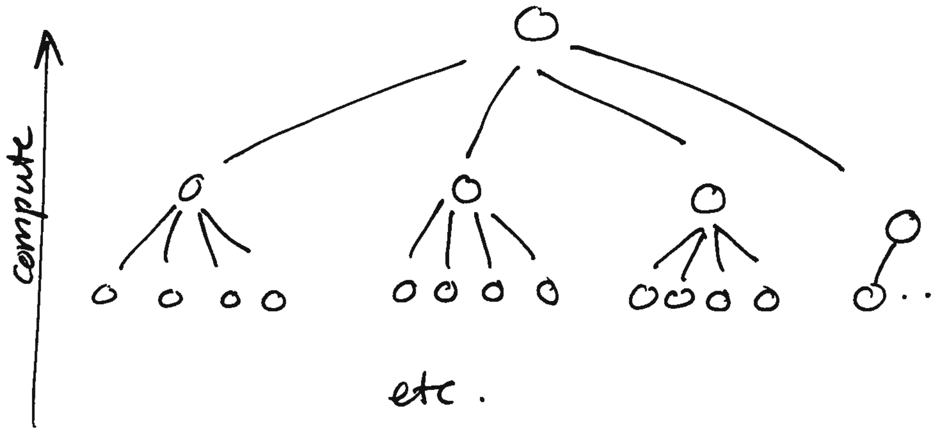
MD5 : message digest 5   1991



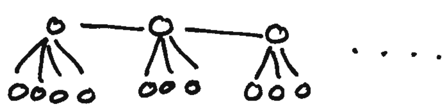
Broken !!

MD6

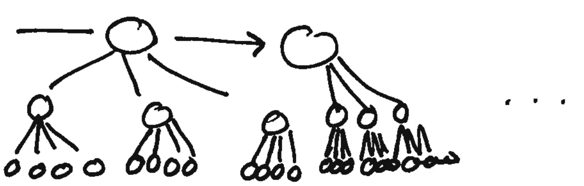
- larger input size
- parallel
- 4:1 compression @ each node



- sequential mode



- in between modes



- every compression fcn is "keyed"

Please comment on

PSETS (difficulty, length, fun?...)

A:

B:

LECTURES (pace, material, technique...)

RECITATIONS (pace, material, technique...)

OTHER

Please comment on

PSETS (difficulty, length, fun?)

A:

B:

LECTURES (pace, material, technique ...)

RECITATIONS (pace, material, technique...)

OTHER