Washington University in St. Louis

Engineering

Analysis of Partition Data Structure

Jon Turner
Computer Science & Engineering
Washington University

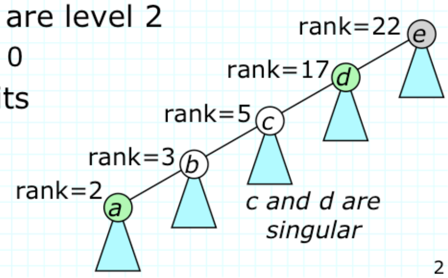
www.arl.wustl.edu/~jst

Towards a More Precise Analysis

- Can improve $O((m+n)\log\log n)$ bound on find steps
  - » improved bound is $O(m\alpha(m,n))$ where $\alpha(m,n)$ is a very slowly growing function
- Similar analysis, but more complicated
  - » will work up to final analysis in several steps
  - » first step gives $O(m+n\log\log n)$ bound
- Define $block(j)=[2^j,2^{j+1})$ for $j\geq 1$, $block(0)=[0,1]$
  - » a node x is on *level 1* if $rank(x)$ and $rank(p(x))$ are both in $block(j)$ for some j; all other nodes are level 2
    - special case: root nodes are on level 0
  - » node x is called *singular* if none of its ancestors is on same level as it is

0 1	2 3	4 .. 7	8 .. 15	16 .. 31	32 .. 63	64 .. 127

block(4)



- At most 2 singular nodes on a find path
  - » so, $O(m)$ find steps involving singular nodes
- To bound # of find steps at non-singular nodes x consider two cases
  - » find steps involving non-singular nodes on level 1
  - » find steps involving non-singular nodes on level 2
- If x is a non-singular node on level 2 before a find op
  - » x has a proper ancestor y on level 2
    - if *rank*(x) in *block*(p) and *rank*(y) in *block*(q), then $q>p$
  - » let z be root of tree containing x and y
    - if *rank*(z) in *block*(r), then $r>q$
  - » after the find $z=p(x)$, so *rank*($p(x)$) moves into a different (and larger) block than it was in previously
  - » there are $O(\log \log n)$ blocks, so the number of find steps at x while x is on level 2 and non-singular is $O(\log \log n)$

- Remains to bound number of find steps at non-singular nodes on level 1
  - » each find involving such a node *x* increases *rank*(*p*(*x*)) by at least 1 (since *x* has a proper ancestor on level 1)
  - » if *rank*(*x*) is in *block*(*j*), there can be at most 2^j find steps at *x* while it is non-singular and on level 1
- Number of nodes with rank *k* is at most $n/2^k$, so number with rank at least *k* is $<2n/2^k$
  - » so, number of nodes with rank in *block*(*j*) is $< 2n/2^{2^j}$
  - » number of find-steps at non-singular nodes *x* on level 1 with *rank*(*x*) in *block*(*j*) is $< 2^j\left(2n/2^{2^j}\right)$
  - » adding up contributions for all values of *j* gives

$$< n + 2n\sum_{j>0} 2^j \Big/ 2^{2^j} \;< n + 2n\sum_{j>0} j \Big/ 2^j = 5n$$

so, # of find steps at non-singular level 1 nodes is $O(n)$

Next Step

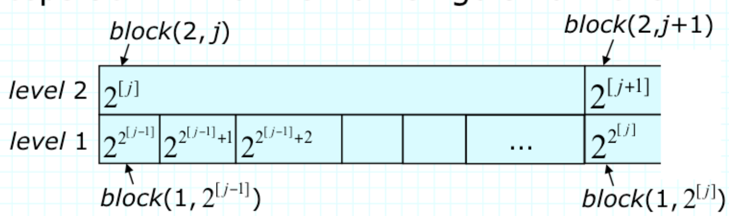
- Taking stock – three cases
  - » singular nodes contribute $O(m)$ find steps
  - » non-singular nodes on level 1 contribute $O(n)$
  - » non-singular nodes on level 2 contribute $O(n \lg \lg n)$
- To take next step, split third case
  - » define $block(1,j)=[2^j,2^{j+1})$ for $j>0$, $block(1,0)=[0,1]$
  - » define $block(2,j)=[2^{[j]},2^{[j+1]})$ for $j>0$, $block(2,0)=[0,3]$ where

$$2^{[1]} = 2^2 \text{ and } 2^{[j]} = 2^{2^{[j-1]}}$$

    so, $2^{[2]}=16$, $2^{[3]}=2^{16}=65{,}536$, $2^{[4]}=2^{65{,}536}$
  - » a non-root node x is on level 1 if there is some $block(1,j)$ that includes both $rank(x)$ and $rank(p(x))$
  - » a non-root node x is on level 2 if it is not on level 1 and there is some $block(2,j)$ that includes both $rank(x)$ and $rank(p(x))$
  - » otherwise, a non-root node x is on level 3

- Define singular nodes as before – a node is singular if none of its proper ancestors is on same level
  - » there are ≤3 singular nodes on a find path, so $O(m)$ find steps
- For non-singular nodes on level 1, previous analysis still valid – so $O(n)$ find steps at these nodes
- If x is a non-singular node on level 3, before a find op
  - » x has a proper ancestor y on level 3
    - if *rank*(x) is in *block*(2,p) and *rank*(y)in *block*(2,q), then $q>p$
  - » let z be root of tree containing x and y
    - if *rank*(z) is in *block*(2,r), then $r>q$
  - » after the find $z=p(x)$, so *rank*($p(x)$) moves into a different (and larger) level 2 block than it was in previously
  - » define $\beta(n)$ =smallest j for which $2^{[j]}>\lg n$; there are $\beta(n)$ level 2 blocks that include values $\leq \lg n$, so number of find steps at x while x is on level 3 and non-singular is at most $\beta(n)$
  - » note $\beta(2^{[k]})=k$, so β grows very slowly

- If x is a non-singular node on level 2, before a find op
  - » x has a proper ancestor y on level 2
    - if *rank*(x) in *block*(1,p) and *rank*(y)in *block*(1,q), then $q>p$
  - » let z be root of tree containing x and y
    - if *rank*(z) in *block*(1,r), then $r>q$
  - » after the find $z=p(x)$, so *rank*($p(x)$) moves into a different (and larger) level 1 block than it was in previously
  - » so, if *rank*(x) is in *block*(2,j), and $b_{2,j}$ is number of level 1 blocks that intersect block(2,j), then the number of find steps at x while x is non-singular on level 2 is $<b_{2,j}<2^{[j]}$



  - » remains to add contributions from all such nodes x

- Number of nodes with rank in *block*(2,*j*) is $< 2n/2^{2^{[j]}}$
- Number of find steps at non-singular nodes *x* on level 2 with rank(*x*) in block(2,*j*)

$$< b_{2,j}\left(2n/2^{2^{[j]}}\right) < 2n\left(2^{[j]}/2^{2^{[j]}}\right)$$

- Adding contributions for all values of *j* gives

$$< n + 2n\sum_{j>0} 2^{[j]} \Big/ 2^{2^{[j]}} < n + 2n\sum_{j>0} j \Big/ 2^{j} = 5n$$

- Summary
  - » singular nodes contribute at most 3*m* find steps
  - » *O*(*n*) find steps at non-singular nodes on level 1
  - » *O*(*n*) find steps at non-singular nodes on level 2
  - » $O(n\beta(n))$ find steps at non-singular nodes on level 3
  - » so overall bound is $O(m+n\beta(n))$

The Final Analysis

- Extend analysis to still more levels ($\alpha(m,n)+1$ levels)

  block$(i,j)=[A(i,j),A(i,j+1))$ for $j>1$ and *block*$(i,0)=[0,A(i,1))$

  where $A(i,j)$ is Ackerman's function

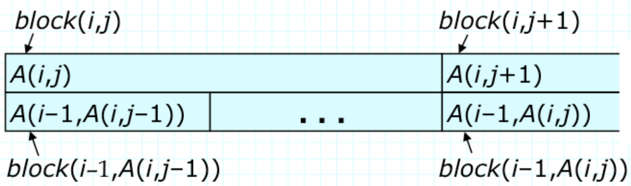
$$A(1,j) = 2^j \quad \text{for } j \geq 1$$

$$A(i,1) = A(i-1,2) \quad \text{for } i \geq 2$$

$$A(i,j) = A(i-1, A(i,j-1)) \quad \text{for } i,j \geq 2$$

$$A(2,j) = A(1, A(2,j-1)) = 2^{A(2,j-1)} = 2^{[j]}$$

  and $\alpha(m,n)=\min\{i\geq 1 \mid A(i,\lfloor m/n \rfloor) > \lg n\}$



so, there are $A(i,j)-A(i,j-1)<A(i,j)$ *blocks at level* $i-1$ *that intersect block*(i,j)

- Levels are defined as before, but for $1 \leq i \leq \alpha(m,n)$
- At most $\alpha(m,n)+1$ singular nodes on a find path, so number of find steps at singular nodes is $O(m\ \alpha(m,n))$
- For non-singular nodes on level 1, earlier analysis gives $O(n)$ find steps
- For a non-singular node x on level $\alpha(m,n)+1$
  - » every find step cause $rank(p(x))$ to move into a larger block on level $\alpha(m,n)$
  - » number of find steps at x while it is in this category is bounded by the number of blocks at level $\alpha(m,n)$ that include values $\leq \lg n$; this is at most m/n
  - » since at most n nodes in this category, at most m find steps occur at non-singular nodes on level $\alpha(m,n)+1$

- For non-singular nodes x on level i in $\{2..\ \alpha(m,n)\}$
  - » assume $rank(x)$ is in $block(i,j)$
  - » each find step at x causes $rank(p(x))$ to move to a larger block on level $i-1$
  - » this can happen to x at most $b_{i,j}$ times, where $b_{i,j}$ is the number of level $i-1$ blocks that intersect $block(i,j)$
    - for $1 \le i \le \alpha(m,n)$, $b_{i0}=2$ and for $j>0$ $b_{i,j}<A(i,j)$
- Let n_{ij}=number of nodes with a final value of $rank(x)$ in $block(i,j)$; $n_{i,0} \le n$ and for $j \ge 1$, $n_{i,j} \le 2n/2^{A(i,j)}$

$$\begin{aligned}\sum_{i=2}^{\alpha(m,n)} \sum_{j\ge 0} b_{i,j} n_{i,j} &= \sum_{i=2}^{\alpha(m,n)} b_{i,0} n_{i,0} + \sum_{i=2}^{\alpha(m,n)} \sum_{j\ge 1} b_{i,j} n_{i,j} \\ &< 2\alpha(m,n)n + 2n \sum_{i=2}^{\alpha(m,n)} \sum_{j\ge 1} A(i,j)/2^{A(i,j)} \\ &\le 2\alpha(m,n)n + 2n\alpha(m,n) \sum_{j\ge 1} j/2^j \\ &= O(\alpha(m,n)n)\end{aligned}$$

Ackerman's Function and $\alpha(m,n)$

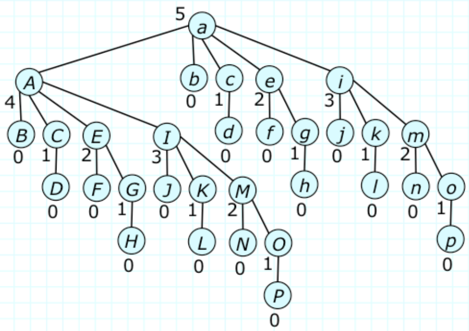
$$A(1,j) = 2^j \quad \text{for } j \geq 1$$
$$A(i,1) = A(i-1,2) \quad \text{for } i \geq 2$$
$$A(i,j) = A(i-1, A(i, j-1)) \quad \text{for } i, j \geq 2$$

i \ *j*	1	2	3	4	5	6
1	2	4	8	16	32	64
2	$4=2^{[1]}$	$16=2^{[2]}$	$2^{16}=2^{[3]}$	$2^{[4]}$	$2^{[5]}$	$2^{[6]}$
3	$2^{[2]}$	$2^{[16]}$				
4	$2^{[16]}$					

- Note, $2^{[4]}=2^{65,536}>10^{19,000}$
- So, while $\alpha(m,n)\rightarrow\infty$, for all practical values of m and n, $\alpha(m,n)\leq 4$

Exercises

1. Consider the tree from the partition data structure that is shown below. Based on the definitions on page 5, which nodes are on level 1, which are on level 2 and which are on level 3? Which nodes are singular?



The following nodes are on level 1: d,h,l,p,D,H,L,P,m,M,A
The following nodes are on level 2: f,g,j,k,n,o,F,G,J,K,N,O
Nodes b,c,e,i,B,C,E,I are on level 3

Nodes d,h,l,m,f,g,j,k,n,o,b,c,e,i,F,G,J,K,N,O,B, C,E,I are singular

Washington University in St. Louis — Engineering

2. Suppose a find operation is done at node *P* in the tree in problem 1. What level do the nodes on the find path now belong to?

   Node A still belongs to level 1, nodes I,M,O and P are all now on level 3

   Does the find operation affect the level of any other node?

   No, since nodes not on the find path have the same parent they had before.

3. Which blocks in level 1 intersect block(2,2)?

   Block(2,2) starts at $2^{[2]}=16$ *and runs up to* $2^{[3]}-1=2^{16}-1$.

   $16=2^4$ *is the start of block(1,4) and* $2^{16}-1$ *is the last value in block(1,15), so block(2,2) intersects 12 blocks in level 1.*

4. Explain why $\alpha(n,n) \geq \alpha(m,n)$ for $m>n$.

   The definition is $\alpha(m,n)=\min\{i\geq 1 \mid A(i,\lfloor m/n \rfloor)>\lg n\}$

   If m=n, this becomes $\alpha(m,n)=\min\{i\geq 1 \mid A(i,1)>\lg n\}$

   Since Ackerman's function increases as either of its arguments increase, if the second argument is restricted to 1, then we need a larger value of i in order for A(i,1) to be greater than lg n than we need if the second argument can be larger than 1. Hence, $\alpha(m,n)$ *is largest when m=n.*