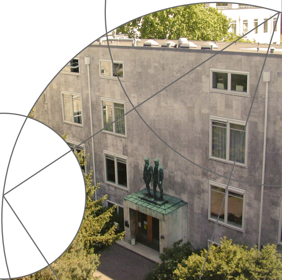
Near-optimal labeling schemes for nearest common ancestors

Stephen Alstrup*
*Esben Bistrup Halvorsen**
Kasper Green Larsen†

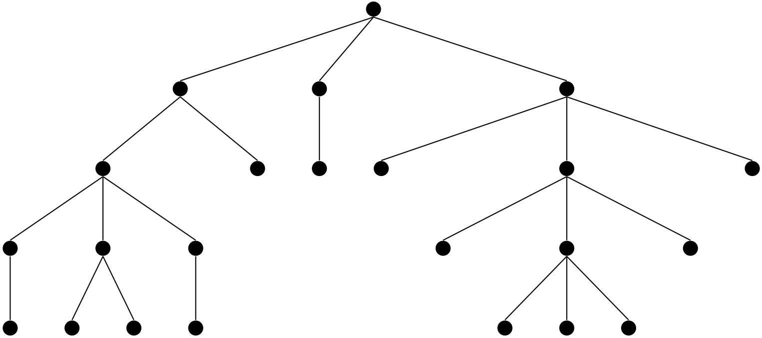
* Department of Computer Science, University of Copenhagen, Denmark

† MADALGO, Department of Computer Science, Aarhus University, Denmark

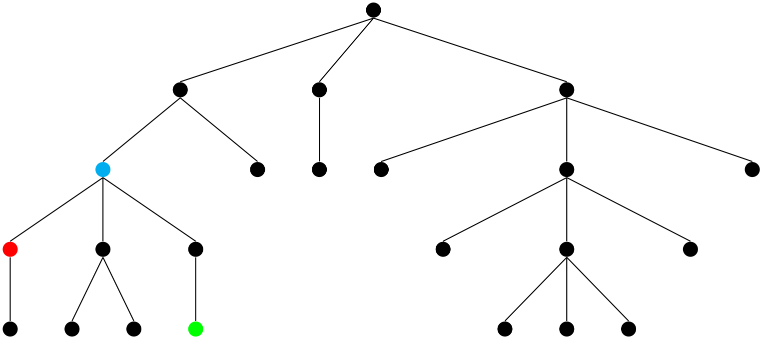
January 6, 2014



Nearest common ancestor (NCA)

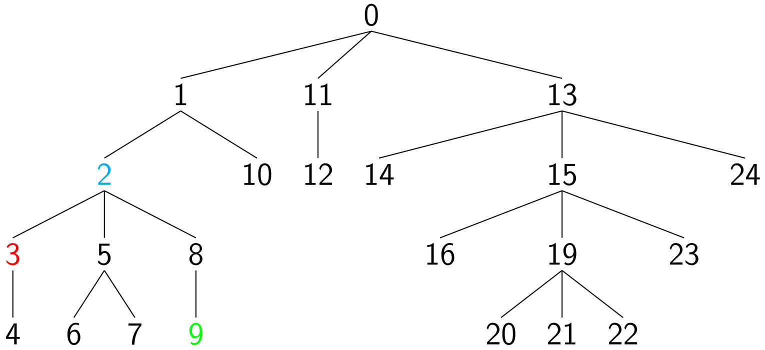


Nearest common ancestor (NCA)

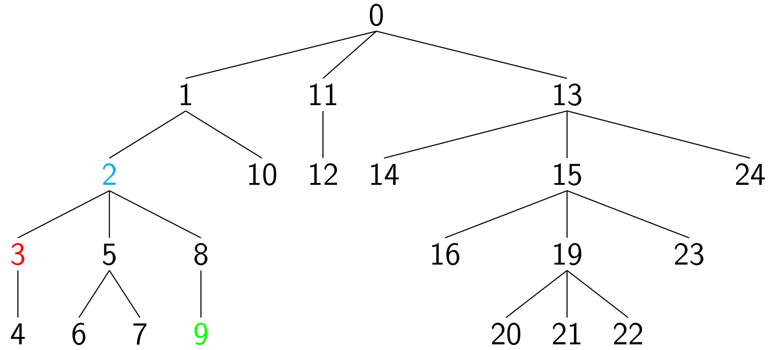


$$\mathrm{nca}(\bullet, \bullet) = \bullet$$

NCA labeling schemes

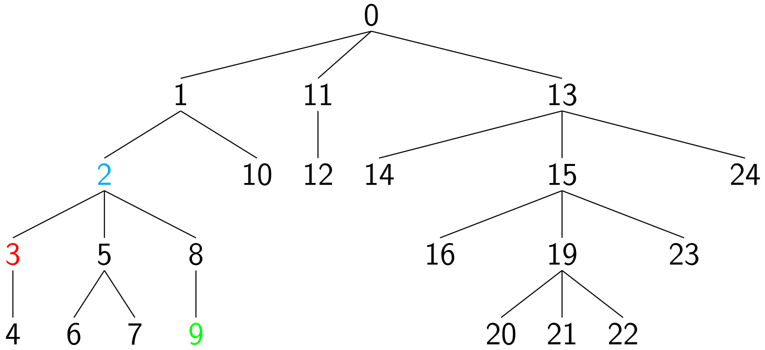


NCA labeling schemes



$l(\bullet) = (0, 1, 2, 3)$ $\quad l(\bullet) = (0, 1, 2, 8, 9)$ $\quad l(\bullet) = (0, 1, 2)$

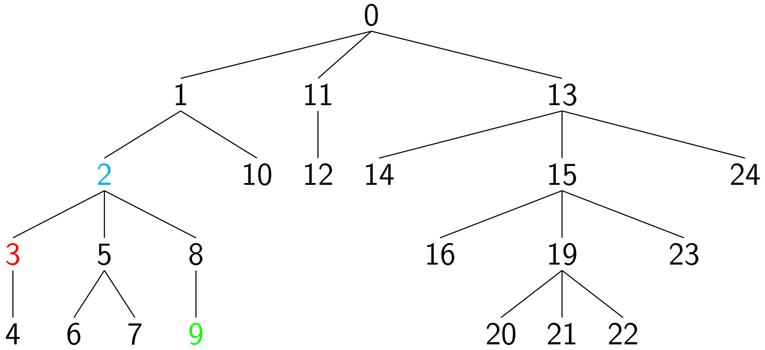
NCA labeling schemes



$$l(\bullet) = (0, 1, 2, 3) \qquad l(\bullet) = (0, 1, 2, 8, 9) \qquad l(\bullet) = (0, 1, 2)$$

$$\max |l(v)| = O(n \log n)$$

NCA labeling schemes



$l(\bullet) = (0, 1, 2, 3)$ $\quad$ $l(\bullet) = (0, 1, 2, 8, 9)$ $\quad$ $l(\bullet) = (0, 1, 2)$

Objective: minimize $\max |l(v)|$

Previous results

Algorithms to compute NCAs

Previous results

Algorithms to compute NCAs

- different settings (static, dynamic,...)

Previous results

Algorithms to compute NCAs

- different settings (static, dynamic,. . .)
- different computational models (RAM-model, pointer model, . . .)

Previous results

Algorithms to compute NCAs

- different settings (static, dynamic,. . .)
- different computational models (RAM-model, pointer model, . . .)
- different applications in mind (computational biology, XML, routing, . . .)

Previous results

Algorithms to compute NCAs*

- different settings (static, dynamic,. . .)
- different computational models (RAM-model, pointer model, . . .)
- different applications in mind (computational biology, XML, routing, . . .)

* see survey by Alstrup, Gavoille, Kaplan, Rauhe (TOCS, 2004)

Previous results

Labeling schemes

Previous results

Labeling schemes

- adjacency, ancestry, distance, routing, NCA, ...

Previous results

Labeling schemes

- adjacency, ancestry, distance, routing, NCA, ...
- static/dynamic

Previous results

Labeling schemes

- adjacency, ancestry, distance, routing, NCA, ...
- static/dynamic
- approximate

Previous results

Labeling schemes

- adjacency, ancestry, distance, routing, NCA, ...
- static/dynamic
- approximate
- probabilistic

Previous results

Labeling schemes*

- adjacency, ancestry, distance, routing, NCA, ...
- static/dynamic
- approximate
- probabilistic

* see survey by Gavoille, Peleg (Dist. Comp., 2003)

Previous results

Harel, Tarjan (Siam J. Computing, 1984):
NCAs in constant time with linear preprocessing time.
(Includes a labeling scheme for complete binary trees.)

Previous results

Kannan, Naor, Rudich (STOC, 1988):
Adjacency and ancestry labeling schemes.

Previous results

Alstrup, Gavoille, Kaplan, Rauhe (TOCS, 2004):
NCA labeling scheme with label size $10 \log n$.

Previous work

Fischer (ESA, 2009):
Experimental results with label size $8 \log n$.

Previous results

Alstrup, Bille, Rauhe (Siam J. Discrete Math., 2005):
NCA labeling schemes have label size $\log n + \Omega(\log\log n)$.

Previous results

Fraigniaud, Korman (STOC, 2010):
Ancestor labeling schemes of $\log n + \Theta(\log \log n)$

New results

Alstrup, Halvorsen, Larsen (SODA, 2014):
Upper bound of $2.772 \log n$ (and $2.585 \log n$ for binary trees).
Efficient upper bound of $3 \log n$.
Lower bound of $1.008 \log n$.

New results

Labels have size $(2 \pm \epsilon) \log n$, $\epsilon < 1$.

New results

Labels have size $(2 \pm \epsilon) \log n$, $\epsilon < 1$.

- Separates NCA from ancestry

New results

Labels have size $(2 \pm \epsilon) \log n$, $\epsilon < 1$.

- Separates NCA from ancestry
- First lower bound of $\log n + \omega(\log \log n)$ for non-distance labeling schemes for trees

New results

Labels have size $(2 \pm \epsilon) \log n$, $\epsilon < 1$.

- Separates NCA from ancestry
- First lower bound of $\log n + \omega(\log \log n)$ for non-distance labeling schemes for trees
- More efficient than previous experimental results

New results

Labels have size $(2 \pm \epsilon) \log n$, $\epsilon < 1$.

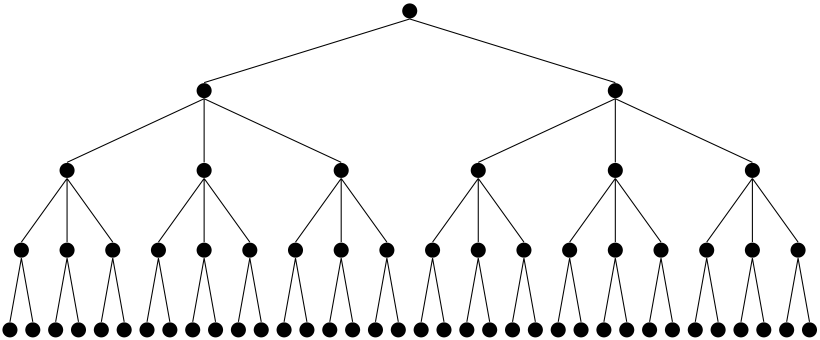
- Separates NCA from ancestry
- First lower bound of $\log n + \omega(\log \log n)$ for non-distance labeling schemes for trees
- More efficient than previous experimental results
- Total space comparable to "normal" NCA algorithms

Lower bound technique

3-2 sequence:

$$(2, 3, 3, 2)$$

3-2 tree:

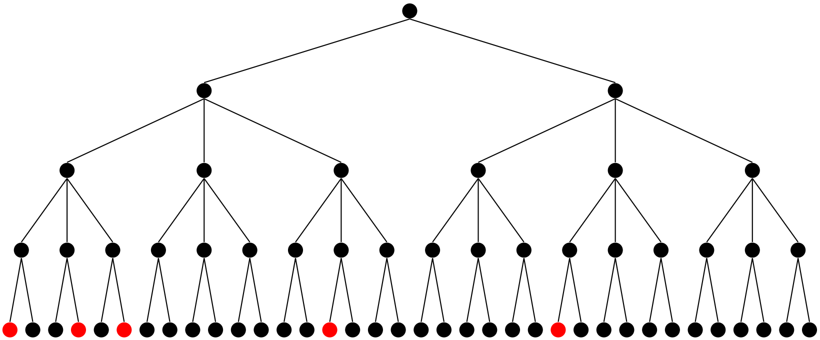


Lower bound technique

3-2 sequence:

$$(2, 3, 3, 2)$$

3-2 tree:

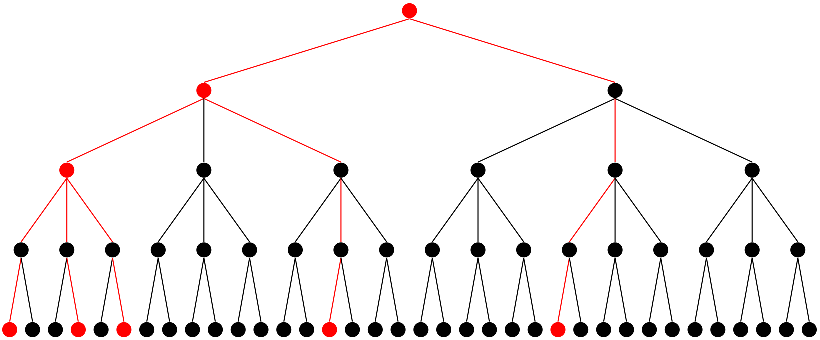


Lower bound technique

3-2 sequence:

$$(2, 3, 3, 2)$$

3-2 tree:

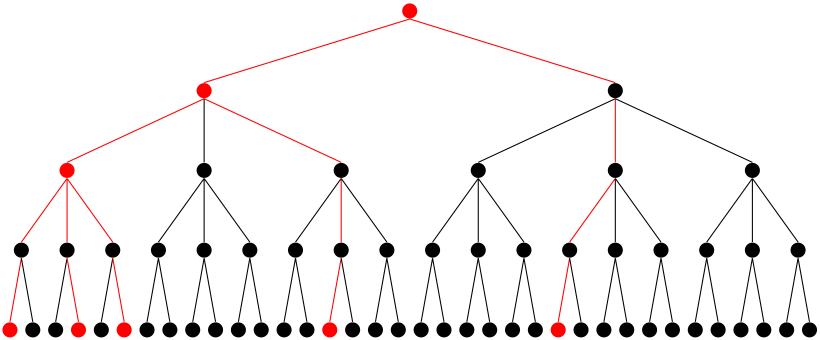


Lower bound technique

3-2 sequence:

$$(2, 3, 3, 2) \qquad (2, 2, 3)$$

3-2 tree:

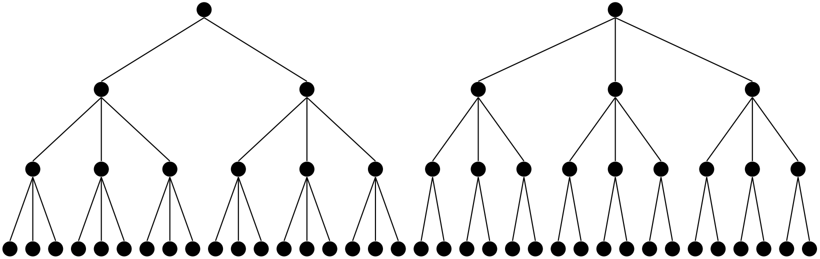


Lower bound technique

3-2 sequences:

$$(2, 3, 3) \qquad\qquad (3, 3, 2)$$

3-2 trees:

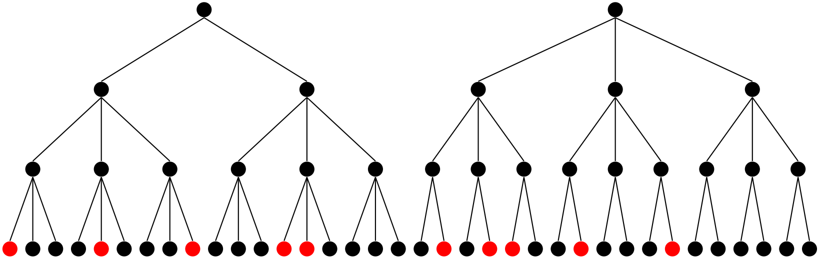


Lower bound technique

3-2 sequences:

$$(2, 3, 3) \qquad\qquad (3, 3, 2)$$

3-2 trees:

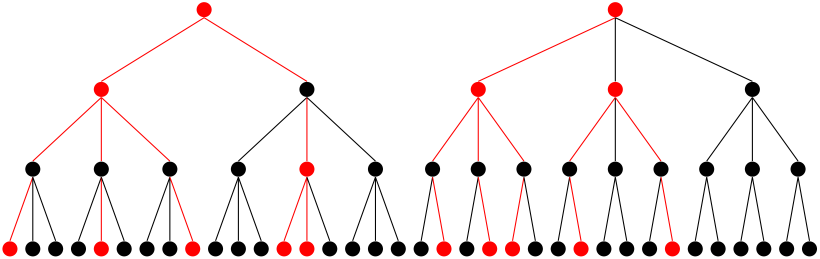


Lower bound technique

3-2 sequences:

$$(2, 3, 3) \qquad\qquad (3, 3, 2)$$

3-2 trees:

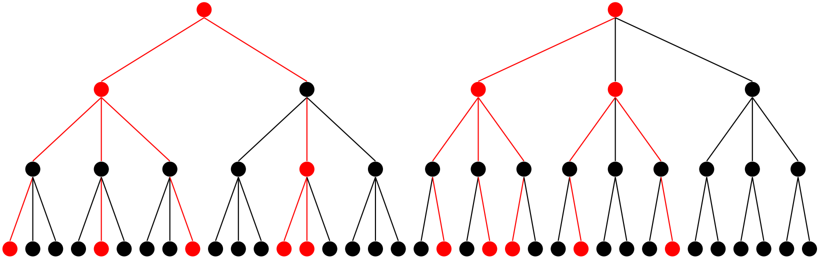


Lower bound technique

3-2 sequences:

$$(2,3,3) \qquad (2,3) \qquad (3,3,2)$$

3-2 trees:



Lower bound technique

$$\{\text{(3-2)-sequences with pairwise distance} > h\}$$

$$\Big\updownarrow \text{1-1}$$

$$\{\text{(3-2)-trees with pairwise fewer than } m \text{ labels in common}\}$$

Lower bound technique

$$\{\text{(3-2)-sequences with pairwise distance} > h\}$$

$$\updownarrow \text{1-1}$$

$$\{\text{(3-2)-trees with pairwise fewer than } m \text{ labels in common}\}$$

$$\Longrightarrow [\ldots, math,\ math,\ math, \ldots]$$

Lower bound technique

$$\{\text{(3-2)-sequences with pairwise distance} > h\}$$

$$\updownarrow \text{1-1}$$

$$\{\text{(3-2)-trees with pairwise fewer than } m \text{ labels in common}\}$$

$$\Longrightarrow [\ldots, math,\ math,\ math, \ldots]$$

$$\Longrightarrow |\{\text{labels}\}| \geq n^{1.008}$$

Lower bound technique

$$\{\text{(3-2)-sequences with pairwise distance} > h\}$$

$$\Big\updownarrow \text{1-1}$$

$$\{\text{(3-2)-trees with pairwise fewer than } m \text{ labels in common}\}$$

$\Longrightarrow [\ldots, math, math, math, \ldots]$

$\Longrightarrow |\{\text{labels}\}| \geq n^{1.008}$

$\Longrightarrow$ labels have worst-case size $1.008 \log n$

Future studies

Future studies

- Tighter bounds than $> 1.008 \log n$ and $< 2.772 \log n$.

Future studies

- Tighter bounds than $> 1.008 \log n$ and $< 2.772 \log n$.
- Time-efficient version

Future studies

- Tighter bounds than $> 1.008 \log n$ and $< 2.772 \log n$.
- Time-efficient version
- Dynamic, probabilistic, ...

Thanks!