10 van Emde Boas Trees

Dynamic Set Data Structure S:

- $S.\text{insert}(x)$
- $S.\text{delete}(x)$
- $S.\text{search}(x)$
- $S.\text{min}()$
- $S.\text{max}()$
- $S.\text{succ}(x)$
- $S.\text{pred}(x)$

10 van Emde Boas Trees

For this chapter we ignore the problem of storing satellite data:

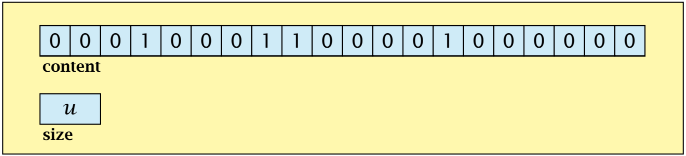
- $S.\mathbf{insert}(x)$: Inserts x into S.
- $S.\mathbf{delete}(x)$: Deletes x from S. Usually assumes that $x \in S$.
- $S.\mathbf{member}(x)$: Returns 1 if $x \in S$ and 0 otw.
- $S.\mathbf{min}()$: Returns the value of the minimum element in S.
- $S.\mathbf{max}()$: Returns the value of the maximum element in S.
- $S.\mathbf{succ}(x)$: Returns successor of x in S. Returns null if x is maximum or larger than any element in S. Note that x needs not to be in S.
- $S.\mathbf{pred}(x)$: Returns the predecessor of x in S. Returns null if x is minimum or smaller than any element in S. Note that x needs not to be in S.

10 van Emde Boas Trees

Can we improve the existing algorithms when the keys are from a restricted set?

In the following we assume that the keys are from $\{0, 1, \ldots, u-1\}$, where u denotes the size of the universe.

Implementation 1: Array



one array of u bits

Use an array that encodes the indicator function of the dynamic set.

Implementation 1: Array

Algorithm 21 array.insert(x)

1: content[x] $\leftarrow$ 1;

Algorithm 22 array.delete(x)

1: content[x] $\leftarrow$ 0;

Algorithm 23 array.member(x)

1: **return** content[x];

- Note that we assume that x is valid, i.e., it falls within the array boundaries.
- Obviously(?) the running time is constant.

Implementation 1: Array

Algorithm 24 array.max()

1: **for** (i = size -1; $i \geq 0$; i−−) **do**
2: **if** content[i] = 1 **then return** i;
3: **return** null;

Algorithm 25 array.min()

1: **for** ($i = 0$; $i <$ size; i++) **do**
2: **if** content[i] = 1 **then return** i;
3: **return** null;

- Running time is $\mathcal{O}(u)$ in the worst case.

Implementation 1: Array

Algorithm 26 array.succ(x)

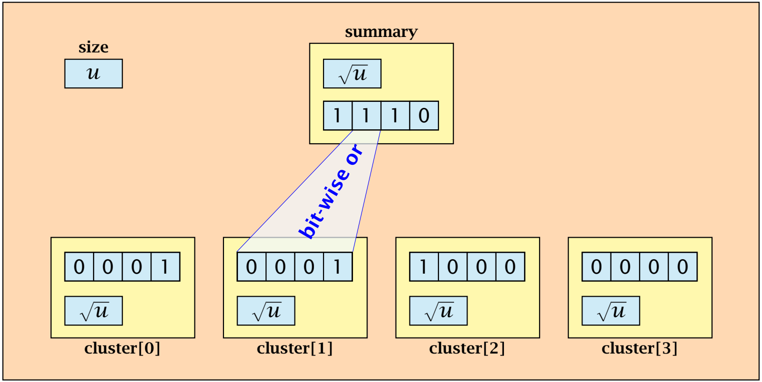
```
1: for (i = x + 1; i < size; i++) do
2:     if content[i] = 1 then return i;
3: return null;
```

Algorithm 27 array.pred(x)

```
1: for (i = x − 1; i ≥ 0; i−−) do
2:     if content[i] = 1 then return i;
3: return null;
```

- Running time is $\mathcal{O}(u)$ in the worst case.

Implementation 2: Summary Array



- $\sqrt{u}$ cluster-arrays of $\sqrt{u}$ bits.
- One summary-array of $\sqrt{u}$ bits. The i-th bit in the summary array stores the bit-wise or of the bits in the i-th cluster.

Implementation 2: Summary Array

The bit for a key x is contained in cluster number $\left\lfloor \frac{x}{\sqrt{u}} \right\rfloor$.

Within the cluster-array the bit is at position $x \bmod \sqrt{u}$.

For simplicity we assume that $u = 2^{2k}$ for some $k \geq 1$. Then we can compute the cluster-number for an entry x as $\mathrm{high}(x)$ (the upper half of the dual representation of x) and the position of x within its cluster as $\mathrm{low}(x)$ (the lower half of the dual representation).

Implementation 2: Summary Array

Algorithm 28 member(x)

---

return cluster[high(x)].member(low(x));

Algorithm 29 insert(x)

---

cluster[high(x)].insert(low(x));
summary.insert(high(x));

- The running times are constant, because the corresponding array-functions have constant running times.

Implementation 2: Summary Array

Algorithm 30 delete(x)

1: cluster[high(x)]. delete(low(x));
2: **if** cluster[high(x)]. min() = null **then**
3: summary . delete(high(x));

- The running time is dominated by the cost of a minimum computation on an array of size $\sqrt{u}$. Hence, $\mathcal{O}(\sqrt{u})$.

Implementation 2: Summary Array

Algorithm 31 max()

1: $\mathit{maxcluster} \leftarrow \text{summary}.\max()$;
2: **if** $\mathit{maxcluster} = \text{null}$ **return** null;
3: $\mathit{offs} \leftarrow \text{cluster}[\mathit{maxcluster}].\max()$
4: **return** $\mathit{maxcluster} \circ \mathit{offs}$;

Algorithm 32 min()

1: $\mathit{mincluster} \leftarrow \text{summary}.\min()$;
2: **if** $\mathit{mincluster} = \text{null}$ **return** null;
3: $\mathit{offs} \leftarrow \text{cluster}[\mathit{mincluster}].\min()$;
4: **return** $\mathit{mincluster} \circ \mathit{offs}$;

The operator $\circ$ stands for the concatenation of two bitstrings. This means if $x = 0111_2$ and $y = 0001_2$ then $x \circ y = 01110001_2$.

- Running time is roughly $2\sqrt{u} = \mathcal{O}(\sqrt{u})$ in the worst case.

Implementation 2: Summary Array

Algorithm 33 succ(x)

```
m ← cluster[high(x)].succ(low(x))
if m ≠ null then return high(x) ∘ m;
succcluster ← summary.succ(high(x));
if succcluster ≠ null then
      offs ← cluster[succcluster].min();
      return succcluster ∘ offs;
return null;
```

- Running time is roughly $3\sqrt{u} = \mathcal{O}(\sqrt{u})$ in the worst case.

Implementation 2: Summary Array

Algorithm 34 pred(x)

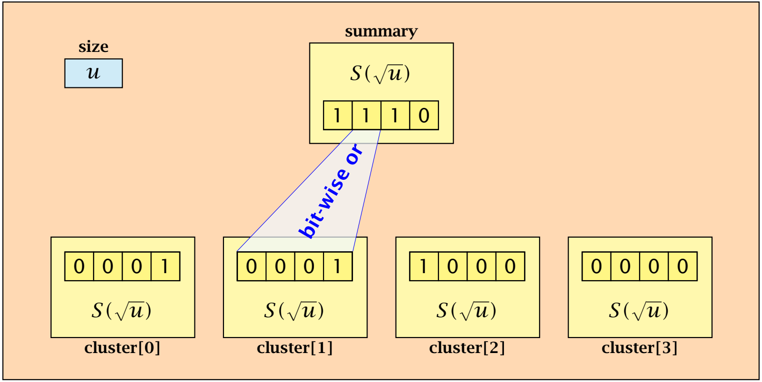
```
m ← cluster[high(x)].pred(low(x))
if m ≠ null then return high(x) ∘ m;
predcluster ← summary.pred(high(x));
if predcluster ≠ null then
    offs ← cluster[predcluster].max();
    return predcluster ∘ offs;
return null;
```

- Running time is roughly $3\sqrt{u} = \mathcal{O}(\sqrt{u})$ in the worst case.

Implementation 3: Recursion

Instead of using sub-arrays, we build a recursive data-structure.

$S(u)$ is a dynamic set data-structure representing u bits:



Implementation 3: Recursion

We assume that $u = 2^{2^k}$ for some k.

The data-structure $S(2)$ is defined as an array of 2-bits (end of the recursion).

Implementation 3: Recursion

The code from Implementation 2 can be used unchanged. We only need to redo the analysis of the running time.

Note that in the code we do not need to specifically address the non-recursive case. This is achieved by the fact that an $S(4)$ will contain $S(2)$'s as sub-datastructures, which are arrays. Hence, a call like $\text{cluster}[1].\text{min}()$ from within the data-structure $S(4)$ is not a recursive call as it will call the function $\text{array}.\text{min}()$.

This means that the non-recursive case is been dealt with while initializing the data-structure.

Implementation 3: Recursion

Algorithm 35 member(x)

1: **return** cluster[high(x)].member(low(x));

- $T_{\text{mem}}(u) = T_{\text{mem}}(\sqrt{u}) + 1$.

Implementation 3: Recursion

Algorithm 36 insert(x)

1: cluster[high(x)].insert(low(x));
2: summary.insert(high(x));

- $T_{\text{ins}}(u) = 2T_{\text{ins}}(\sqrt{u}) + 1$.

Implementation 3: Recursion

Algorithm 37 delete(x)

1: cluster[high(x)]. delete(low(x));
2: **if** cluster[high(x)]. min() = null **then**
3:     summary . delete(high(x));

- $T_{\mathrm{del}}(u) = 2T_{\mathrm{del}}(\sqrt{u}) + T_{\min}(\sqrt{u}) + 1$.

Implementation 3: Recursion

Algorithm 38 min()

1: $\mathit{mincluster} \leftarrow \mathrm{summary.min}();$
2: **if** $\mathit{mincluster} = \mathrm{null}$ **return** null;
3: $\mathit{offs} \leftarrow \mathrm{cluster}[\mathit{mincluster}].\mathrm{min}();$
4: **return** $\mathit{mincluster} \circ \mathit{offs};$

- $T_{\min}(u) = 2T_{\min}(\sqrt{u}) + 1$.

Implementation 3: Recursion

Algorithm 39 succ(x)

```
m ← cluster[high(x)].succ(low(x))
if m ≠ null then return high(x) ∘ m;
succcluster ← summary.succ(high(x));
if succcluster ≠ null then
    offs ← cluster[succcluster].min();
    return succcluster ∘ offs;
return null;
```

- $T_{\mathrm{succ}}(u) = 2T_{\mathrm{succ}}(\sqrt{u}) + T_{\min}(\sqrt{u}) + 1$.

Implementation 3: Recursion

$\boldsymbol{T_{\text{mem}}(u) = T_{\text{mem}}(\sqrt{u}) + 1}$:

Set $\ell := \log u$ and $X(\ell) := T_{\text{mem}}(2^\ell)$. Then

$$\begin{aligned} X(\ell) = T_{\text{mem}}(2^\ell) = T_{\text{mem}}(u) &= T_{\text{mem}}(\sqrt{u}) + 1 \\ &= T_{\text{mem}}(2^{\frac{\ell}{2}}) + 1 = X(\tfrac{\ell}{2}) + 1 \enspace . \end{aligned}$$

Using Master theorem gives $X(\ell) = \mathcal{O}(\log \ell)$, and hence $T_{\text{mem}}(u) = \mathcal{O}(\log\log u)$.

Implementation 3: Recursion

$$\boldsymbol{T_{\text{ins}}(u) = 2T_{\text{ins}}(\sqrt{u}) + 1.}$$

Set $\ell := \log u$ and $X(\ell) := T_{\text{ins}}(2^\ell)$. Then

$$\begin{aligned} X(\ell) = T_{\text{ins}}(2^\ell) = T_{\text{ins}}(u) &= 2T_{\text{ins}}(\sqrt{u}) + 1 \\ &= 2T_{\text{ins}}(2^{\frac{\ell}{2}}) + 1 = 2X(\tfrac{\ell}{2}) + 1 \ . \end{aligned}$$

Using Master theorem gives $X(\ell) = \mathcal{O}(\ell)$, and hence $T_{\text{ins}}(u) = \mathcal{O}(\log u)$.

The same holds for $T_{\max}(u)$ and $T_{\min}(u)$.

Implementation 3: Recursion

$$\boldsymbol{T_{\mathrm{del}}(u) = 2T_{\mathrm{del}}(\sqrt{u}) + T_{\min}(\sqrt{u}) + 1 \le 2T_{\mathrm{del}}(\sqrt{u}) + c\log(u).}$$

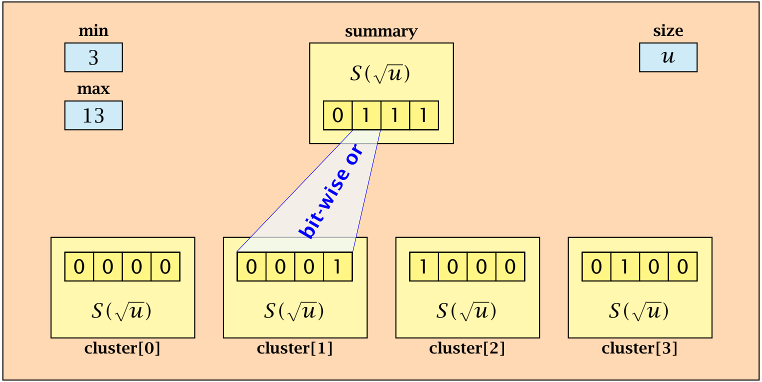
Set $\ell := \log u$ and $X(\ell) := T_{\mathrm{del}}(2^\ell)$. Then

$$\begin{aligned} X(\ell) = T_{\mathrm{del}}(2^\ell) = T_{\mathrm{del}}(u) &= 2T_{\mathrm{del}}(\sqrt{u}) + c\log u \\ &= 2T_{\mathrm{del}}(2^{\frac{\ell}{2}}) + c\ell = 2X(\tfrac{\ell}{2}) + c\ell \enspace . \end{aligned}$$

Using Master theorem gives $X(\ell) = \Theta(\ell \log \ell)$, and hence $T_{\mathrm{del}}(u) = \mathcal{O}(\log u \log\log u)$.

The same holds for $T_{\mathrm{pred}}(u)$ and $T_{\mathrm{succ}}(u)$.

Implementation 4: van Emde Boas Trees



- The bit referenced by min is not set within sub-datastructures.
- The bit referenced by max is set within sub-datastructures (if $\max \neq \min$).

Implementation 4: van Emde Boas Trees

Advantages of having max/min pointers:

- Recursive calls for $\min$ and $\max$ are constant time.
- $\min = \text{null}$ means that the data-structure is empty.
- $\min = \max \neq \text{null}$ means that the data-structure contains exactly one element.
- We can insert into an empty datastructure in constant time by only setting $\min = \max = x$.
- We can delete from a data-structure that just contains one element in constant time by setting $\min = \max = \text{null}$.

Implementation 4: van Emde Boas Trees

Algorithm 40 max()

```
return max;
```

Algorithm 41 min()

```
return min;
```

- Constant time.

Implementation 4: van Emde Boas Trees

Algorithm 42 member(x)

```
1: if x = min then return 1; // TRUE
2: return cluster[high(x)].member(low(x));
```

- $T_{\text{mem}}(u) = T_{\text{mem}}(\sqrt{u}) + 1 \implies T(u) = \mathcal{O}(\log\log u)$.

Implementation 4: van Emde Boas Trees

Algorithm 43 succ(x)

```
if min ≠ null ∧ x < min then return min;
maxincluster ← cluster[high(x)].max();
if maxincluster ≠ null ∧ low(x) < maxincluster then
    offs ← cluster[high(x)].succ(low(x));
    return high(x) ∘ offs;
else
    succcluster ← summary.succ(high(x));
    if succcluster = null then return null;
    offs ← cluster[succcluster].min();
    return succcluster ∘ offs;
```

- $T_{\mathrm{succ}}(u) = T_{\mathrm{succ}}(\sqrt{u}) + 1 \Longrightarrow T_{\mathrm{succ}}(u) = \mathcal{O}(\log\log u)$.

Implementation 4: van Emde Boas Trees

Algorithm 44 insert(x)

if $\text{min} = \text{null}$ **then**
  $\text{min} = x$; $\text{max} = x$;
else
  if $x < \text{min}$ **then** exchange x and min;
  if $\text{cluster}[\text{high}(x)].\text{min} = \text{null}$; **then**
    $\text{summary}.\text{insert}(\text{high}(x))$;
    $\text{cluster}[\text{high}(x)].\text{insert}(\text{low}(x))$;
  else
    $\text{cluster}[\text{high}(x)].\text{insert}(\text{low}(x))$;
  if $x > \text{max}$ **then** $\text{max} = x$;

- $T_{\text{ins}}(u) = T_{\text{ins}}(\sqrt{u}) + 1 \Longrightarrow T_{\text{ins}}(u) = \mathcal{O}(\log\log u)$.

Implementation 4: van Emde Boas Trees

Note that the recusive call in Line 7 takes constant time as the if-condition in Line 5 ensures that we are inserting in an empty sub-tree.

The only non-constant recursive calls are the call in Line 6 and in Line 9. These are mutually exclusive, i.e., only one of these calls will actually occur.

From this we get that $T_{\text{ins}}(u) = T_{\text{ins}}(\sqrt{u}) + 1$.

Implementation 4: van Emde Boas Trees

- **Assumes that x is contained in the structure.**

Algorithm 45 delete(x)

```
if min = max then
        min = null; max = null;
else
        if x = min then                                     find new minimum
                firstcluster ← summary.min();
                offs ← cluster[firstcluster].min();
                x ← firstcluster ∘ offs;
                min ← x;
        cluster[high(x)].delete(low(x));                    delete
```

continued...

Implementation 4: van Emde Boas Trees

Algorithm 45 delete(x)

...continued

fix maximum

```
if cluster[high(x)].min() = null then
    summary.delete(high(x));
    if x = max then
        summax ← summary.max();
        if summax = null then max ← min;
        else
            offs ← cluster[summax].max();
            max ← summax ∘ offs
else
    if x = max then
        offs ← cluster[high(x)].max();
        max ← high(x) ∘ offs;
```

Implementation 4: van Emde Boas Trees

Note that only one of the possible recusive calls in Line 9 and Line 11 in the deletion-algorithm may take non-constant time.

To see this observe that the call in Line 11 only occurs if the cluster where x was deleted is now empty. But this means that the call in Line 9 deleted the last element in $\mathrm{cluster}[\mathrm{high}(x)]$. Such a call only takes constant time.

Hence, we get a recurrence of the form

$$T_{\mathrm{del}}(u) = T_{\mathrm{del}}(\sqrt{u}) + c \ .$$

This gives $T_{\mathrm{del}}(u) = \mathcal{O}(\log\log u)$.

10 van Emde Boas Trees

Space requirements:

- The space requirement fulfills the recurrence

$$S(u) = (\sqrt{u} + 1)S(\sqrt{u}) + \mathcal{O}(\sqrt{u}) \ .$$

- Note that we cannot solve this recurrence by the Master theorem as the branching factor is not constant.
- One can show by induction that the space requirement is $S(u) = \mathcal{O}(u)$. Exercise.

- Let the “real” recurrence relation be

$$S(k^2) = (k+1)S(k) + c_1 \cdot k;\ S(4) = c_2$$

- Replacing $S(k)$ by $R(k) := S(k)/c_2$ gives the recurrence

$$R(k^2) = (k+1)R(k) + ck;\ R(4) = 1$$

  where $c = c_1/c_2 < 1$.
- Now, we show $R(k) \leq k - 2$ for squares $k \geq 4$.
  - Obviously, this holds for $k = 4$.
  - For $k = \ell^2 > 4$ with ℓ integral we have

$$\begin{aligned} R(k) &= (1+\ell)R(\ell) + c\ell \\ &\leq (1+\ell)(\ell-2) + \ell \leq k - 2 \end{aligned}$$

  - This shows that $R(k)$ and, hence, $S(k)$ grows linearly.

van Emde Boas Trees

Bibliography

[CLRS90] Thomas H. Cormen, Charles E. Leiserson, Ron L. Rivest, Clifford Stein:
Introduction to Algorithms (3rd ed.),
MIT Press and McGraw-Hill, 2009

See Chapter 20 of [CLRS90].