Immediate Snapshot

CA 2017

Jingjing Wang, LPD, EPFL
A snapshot has two operations: \textit{update()} and \textit{scan()} and maintains an array \( x \) of size \( n \).

Sequential specification

\textit{scan()}:
- Return \((x)\)

\textit{update}(i, v):
- \( x[i] := v; \)
- Return \((\text{OK})\)
Motivation for immediate snapshot

Snapshot

- Update some state
- Take a “picture” of all states
- Separately

Immediate snapshot

- Immediately take a “picture” of all states after updating a state
Semantics

The memory is accessed via a single $\text{update\_snapshot}$ operation

Semantics: each write operation, in addition to writing, also returns an atomic snapshot

“Weakly atomic” = runs of standard atomic snapshot include runs of immediate snapshot
The power of registers

Can immediate snapshot be implemented by atomic registers?

- Yes. At least for one-shot version

One-shot: Each process invokes at most once that operation
Immediate snapshot

An immediate snapshot has a single operation: `update_snapshot()` and maintains an array `x` of size `n`

Sequential specification

`update_snapshot(vi):`

- `x[i] := vi;`
- `Return {(1, x[1]), (2, x[2]), ..., (n, x[n])}`
Properties

Liveness. An invocation of `update_snapshot()` terminates

Self-inclusion. \((i, \text{vi}) \in \text{viewi}\)

Containment. \(\text{viewi} \subseteq \text{viewj}\) or \(\text{viewj} \subseteq \text{viewi}\)

Immediacy. If \((j, vj) \in \text{viewi}\), then \(\text{viewj} \subseteq \text{viewi}\)
Naive implementation

n processes share an atomic snapshot object x

**update_snapshot(vi):**

- x.update(i, vi);
- a := x.scan();
- Return {{1, a[1]}, {2, a[2]}, ..., {n, a[n]}}
Immediacy?

```
update_snapshot() - {(1, v1), (2, v2)}
```
Immediacy?

update_snapshot() - {((1, v1), (2, v2), (3, v3))}

\[ p_1 \quad \text{update}_1(v_1) \quad \text{snapshot}_1() \]

\[ p_2 \quad \text{update}_2(v_2) \quad \text{snapshot}_2() \]

\[ p_3 \quad \text{update}_3(v_3) \quad \text{snapshot}_3() \]
Snapshot vs. immediate snapshot

An atomic snapshot

An immediate snapshot that satisfies

- Liveness, self-inclusion, containment, immediacy
Possible execution?

\{ (1, v1), (2, v2) \}

update\_snapshot_1(v_1)

\{ (1, v1), (2, v2) \}

update\_snapshot_2(v_2)

\{ (1, v1), (2, v2), (3, v3) \}

update\_snapshot_3(v_3)
Possible execution?

{\( (1, v_1), (2, v_2) \)}

\[ p_1 \]

\text{update\_snapshot}_1(v_1)

\{\( (1, v_1), (2, v_2) \}\}

\text{update\_snapshot}_2(v_2)

\{\( (1, v_1), (2, v_2), (3, v_3) \)\}

\text{update\_snapshot}_3(v_3)

\begin{itemize}
  \item Liveness. ✓
  \item Self-inclusion. ✓
  \item Containment. ✓
  \item Immediacy. ✓
\end{itemize}
A property that follows

(Self-inclusion. \((i, vi) \in \text{viewi}\)

+ Immediacy. If \((j, vj) \in \text{viewi}\), then \(\text{viewj} \subseteq \text{viewi}\))

Property: If \((i, -) \in \text{viewj}\) and \((j, -) \in \text{viewi}\), then \(\text{viewj} = \text{viewi}\)

=> Compared with sequential execution?
Atomicity

Every operation appears to execute at

- Some indivisible point in time (called linearization point) between
- The invocation and reply time events
Atomic execution?

\[
\{(1, v1), (2, v2)\}
\]

\[
\text{update\_snapshot}_1(v_1)
\]

\[
\{(1, v1), (2, v2)\}
\]

\[
\text{update\_snapshot}_2(v_2)
\]

\[
\{(1, v1), (2, v2), (3, v3)\}
\]

\[
\text{update\_snapshot}_3(v_3)
\]
Set linearizability

Linearization replaced by set-linearization:

- These invocations are set-linearized at the same point of the time line

For one-shot immediate snapshot,

- The invocations which are set-linearized at the same point do return the very same view
Key idea for set linearizability

To *update_snapshot()*, a process keeps reading other processes’ updates

For any two processes pi and pj,

- If pi and pj see each other’s update, then pi and pj retry reading until they are going to return the same result
Enforcing set linearizability

The processes share an array of registers $\text{REG}[1]$, $\text{REG}[2]$, $\text{REG}[3]$, …

- $\text{REG}[x]$ is again an array of registers
- $\text{REG}[x]$ contains a view
- $\text{REG}[x][i]$ can only be written by $\pi$

$\pi$ reads $\text{REG}[x]$

- If $\pi$ cannot return $\text{REG}[x]$, then $\pi$ retries, writes and reads the next $\text{REG}$
Enforcing set linearizability

The processes share an array of registers REG[1], REG[2], …, init’ed to ⊥

A recursive implementation:

- **update_snapshot**(vi):
  - my_viewi := rec_update_snapshot(first, vi)
  - Return my_viewi
Enforcing set linearizability

Every process keeps a local array of registers Regi

- **rec_update_snapshot(x, v):**
  - REG[x][i].write(v);
  - For each \( j \in \{1, \ldots, n\} \) do Regi[j] := REG[x][j].read();
  - Viewi := \{ (j, Regi[j]) | Regi[j] \neq \bot \};
  - if( **some condition** ) then resi := viewi;
  - Else resi := rec_update_snapshot(\textbf{next}, v);
  - Return resi
Possible execution?

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>p1</td>
<td>v1</td>
<td>v1</td>
<td>v1</td>
<td></td>
</tr>
<tr>
<td>p2</td>
<td>v2</td>
<td>v2</td>
<td>v2</td>
<td></td>
</tr>
<tr>
<td>p3</td>
<td>v3</td>
<td>v3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Key idea for liveness

If pi and pj see each other’s update, then pi and pj retry

- Pi is waiting for pj’s last-minute view
- So is pj
- Which view is the last one?
Key idea for liveness (cont’d)

Suppose: At most $x$ processes access $\text{REG}[x]$ (invariant)

If $\pi$ sees $\text{REG}[x]$ contains exactly $x$ updates, then

- $\pi$ is one of the **last** processes which access $\text{REG}[x]$
- Or linearized as such

\[
\begin{align*}
\text{p1} & \quad \text{REG}[x][1].\text{write} \\
\text{p2} & \quad \quad \quad \quad \quad \quad \text{REG}[x][2].\text{write} \\
\text{p3} & \quad \quad \quad \quad \quad \quad \quad \text{REG}[x][3].\text{write}
\end{align*}
\]
Key idea for liveness (cont’d)

Suppose: At most x processes access REG[x] (invariant)

If pi sees REG[x] contains exactly x updates, then

- pi is one of the **last** processes which accesses REG[x]
- Or linearized as such

If the invariant is true, then after pi, REG[x] remains the **same**.
Key idea for liveness (cont’d)

Suppose: At most $x$ processes access REG[$x$] (invariant)

If $pi$ sees REG[$x$] contains exactly $x$ updates, then

- $pi$ is one of the last processes which accesses REG[$x$]
- Or linearized as such

If the invariant is true, then after $pi$, REG[$x$] remains the same

- $Pi$ can return REG[$x$]
- As well as other processes who see $pi$’s update
Key idea for set-linearizability & liveness

Recall that we consider one-shot version:

- Each process invokes at most once `update_snapshot()

- This means at most n processes access the `first` REG
Key idea for set-linearizability & liveness

Recall that we consider one-shot version:

- Each process invokes at most once \textit{update\_snapshot()}

- This means at most \( n \) processes access the \textbf{first} \( \text{REG} = \text{REG}[n] \)

If \textbf{some condition} = a process’s view of \( \text{REG}[n] \) contains \( n \) values, then

- Return \( \text{REG}[n] \)
- Otherwise, go to the \textbf{next} \( \text{REG} = \text{REG}[n-1] \)
Key idea for set-linearizability & liveness (cont’d)

The processes share an array of registers $\text{REG}[n]$, $\text{REG}[n-1]$, …, $\text{REG}[1]$

- Each contains a view

Claim:

(a) At most $x$ processes can access $\text{REG}[x]$
(b) At least one process returns $\text{REG}[x]$
Immediate snapshot implementation

- \textit{update\_snapshot}(\textit{vi}):
  - \textit{my\_viewi} := \textit{rec\_update\_snapshot}(\textit{n, vi})
  - Return \textit{my\_viewi}
Immediate snapshot implementation

The processes share an array of registers $\text{REG}[1, \ldots, n]$, init’ed to $\bot$

Every process keeps a local array of registers $\text{Reg}_i$

- $\text{rec_update_snapshot}(x, v)$:
  - $\text{REG}[x][i].\text{write}(v)$;
  - For each $j \in \{1, \ldots, n\}$ do $\text{Reg}_i[j] := \text{REG}[x][j].\text{read}()$;
  - $\text{View}_i := \{ (j, \text{Reg}_i[j]) \mid \text{Reg}_i[j] \neq \bot \}$;
  - if ($|\text{view}_i| = x$) then $\text{res}_i := \text{view}_i$;
  - else $\text{res}_i := \text{rec_update_snapshot}(x-1, v)$;
  - Return $\text{res}_i$
Possible return value?

\[((1, v1), (2, v2))\]

\[update\_snapshot_1(v_1)\]
**Possible execution?**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>v1</td>
<td>v1</td>
<td>...</td>
<td>v1</td>
<td>v1</td>
</tr>
<tr>
<td>v2</td>
<td>v2</td>
<td>...</td>
<td>v2</td>
<td>v2</td>
</tr>
<tr>
<td>v3</td>
<td>v3</td>
<td>...</td>
<td>v3</td>
<td></td>
</tr>
</tbody>
</table>
References
