Introduction to Reliable Distributed Programming

Figures

Rachid Guerraoui    Luís Rodrigues

Composition model

Component A

Component B
Layering

Layer $n+1$

Layer $n$

Layer $n-1$

(request) → (deliver)

(request) → (receive)

(indication) → (deliver)

(indication) → (receive)
A stack of printing modules

Module Print

Module BoundedPrint

Request: PrintRequest

Confirmation: PrintConfirm

Indication: PrintAlarm

Confirmation: PrintStatus

Request: BoundedPrintRequest
Layers, Sessions, QoS and Channels

- PrintAppLayer
- BoundedPrintLayer
- PrintLayer
- PrintSession
- PrintAppSession
- BoundedPrintSession
- PrintApplSession
- PrintApplLayer
- BoundedPrintLayer
- PrintLayer

CreateChannel
getSession()
Step of a process

Process

- internal computation
  - (modules of the process)
  - (receive)
- (send)
- incoming message
- outgoing message
Failure modes of a process

- Crashes
- Omissions
- Crashes & Recoveries
- Arbitrary
The link abstraction (a)
The link abstraction (b)
The link abstraction (c)
The link abstraction (d)
The *happened-before* relation (a)

\[ p_1 \quad e_1 \quad e_2 \quad p_2 \quad p_3 \]
The \textit{happened-before} relation (b)
The *happened-before* relation (c)

```
p1

\[ e_1 \]

\[ e' \]  \[ e_2 \]

p2

p3
```
Sample execution of basic broadcast

\[ \begin{align*}
  p_1 & \xrightarrow{\text{bebBroadcast}} \\
  p_2 & \xrightarrow{\text{bebDeliver}} \\
  p_3 & \xrightarrow{\text{bebDeliver}} \\
  p_4 & \xrightarrow{\text{bebDeliver}}
\end{align*} \]
Sample executions of reliable broadcast (a)
Sample executions of reliable broadcast (b)
Sample execution of all-ack uniform reliable broadcast

\[ \text{rbBroadcast} \]

\[ p_1 \]

\[ \rightarrow \] \[ \text{rbDeliver} \]

\[ p_2 \]

\[ \rightarrow \] \[ \text{rbDeliver} \]

\[ p_3 \]

\[ \rightarrow \] \[ \text{rbDeliver} \]

\[ p_4 \]
Ack implosion
Ack tree
Epidemic dissemination (a)
Epidemic dissemination (b)
Epidemic dissemination (c)
Causal order of messages (a)
Causal order of messages (b)
Causal order of messages (c)

$p_1 \quad m_1 \quad p_2 \quad m' \quad p_3 \quad m_2$
Sample execution of causal broadcast with complete past

```
rcoBroadcast (m_1)
```

```
rcoBroadcast (m_2)
```

```
rcoDeliver (m_1)
```

```
rcoDeliver (m_2)
```

$p_1$

$p_2$

$p_3$

$p_4$
Sample execution of waiting-causal broadcast

$$rcoBroadcast \ (m_1)$$

$p_1\ [1, 0, 0, 0]$

$p_2$

$p_3\ [1, 1, 0, 0]$

$p_4$

$$rcoBroadcast \ (m_2)$$

$$rcoDeliver \ (m_2)$$

$$rcoDeliver \ (m_1)$$
Non-regular register execution

\[
p_1 \quad \text{write}(5) \quad \text{write}(6) \quad \text{read()} \rightarrow 0 \quad \text{read()} \rightarrow 5 \quad \text{read()} \rightarrow 6
\]

\[
p_2 \quad \text{read()} \rightarrow 0 \quad \text{read()} \rightarrow 5 \quad \text{read()} \rightarrow 6
\]
Regular register execution

\[ p_1 \quad \text{write}(5) \quad \text{write}(6) \quad \text{read()} \rightarrow 5 \quad \text{read()} \rightarrow 5 \quad \text{read()} \rightarrow 6 \quad p_2 \]
Another non-regular register execution

$p_1$

$p_2$

read() $\rightarrow$ 5

write(5)  write(6)
Non-atomic register execution

\begin{itemize}
  \item \texttt{write(5)}
  \item \texttt{write(6)}
  \item \texttt{read()} \rightarrow 5
  \item \texttt{read()} \rightarrow 6
  \item \texttt{read()} \rightarrow 5
\end{itemize}
Atomic register execution

\[ p_1 \quad \text{write}(5) \quad \text{read()} \rightarrow 5 \quad \text{read()} \rightarrow 6 \]

\[ p_2 \quad \text{write}(6) \quad \text{read()} \rightarrow 5 \]
Violation of atomicity in the “Read-One Write-All” regular register algorithm

\[
\begin{align*}
p_1 & \quad \text{write}(5) \quad \text{write}(6) \\
p_2 & \quad \text{read()} \to 5 \quad \text{read()} \to 6 \\
p_3 & \quad \text{read()} \to 5
\end{align*}
\]
Violation of atomicity in the “Majority Voting” regular register algorithm

\[
\begin{align*}
\text{\textbf{p}_1} & \quad \text{write(5)} & \quad \text{write(6)} \\
& \quad \quad \quad \text{sn = 1} & \quad \quad \quad \text{sn = 2} \\
\text{\textbf{p}_2} & \quad \quad \quad \text{sn = 1} \\
\text{\textbf{p}_3} & \quad \quad \quad \text{sn = 1} \\
\text{\textbf{p}_4} & \quad \quad \quad \text{sn = 1} \\
\text{\textbf{p}_5} & \quad \quad \quad \text{sn = 1}
\end{align*}
\]

read() \rightarrow 6
read() \rightarrow 5
A \( (1, N) \) atomic register execution

\[
\begin{align*}
p_1 & \quad \text{write}(5) \quad \text{write}(6) \\
      & \quad \text{sn} = 1 \quad \text{sn} = 2 \\
p_2 & \quad \text{read()} \rightarrow 5 \quad \text{read()} \rightarrow 6 \quad \text{read()} \rightarrow 6 \\
      & \quad \text{sn} = 1 \quad \text{sn} = 2 \quad \text{sn} = 2 \\
p_3 & \quad \text{sn} = 1 \quad \text{sn} = 1
\end{align*}
\]
Sample execution of flooding consensus

round 1

$\text{cPropose } (3)$
$p_1$
$\text{cPropose } (5)$
$p_2$
$\text{cPropose } (8)$
$p_3$
$\text{cPropose } (7)$
$p_4$

round 2

$\text{cDecide } (3 = \min(3, 5, 8, 7))$

$(5, 8, 7)$
$c_{\text{Decide}} (3)$

$(5, 8, 7)$
$c_{\text{Decide}} (3)$
Sample execution of hierarchical consensus

```
round 1  round 2  round 3  round 4

cPropose (3)
p1

cPropose (5)
p2

cPropose (8)
p3

cPropose (7)
p4

(3)  (5)

(5)

(5)

(5)

cDecide (5)

cDecide (5)

cDecide (5)
```
Role of randomization

round 1
phase 1 phase 2

p1
(cPropose (1)

p2
(cPropose (2)

p3
(cPropose (2)
Sample execution of the consensus-based total order broadcast algorithm
Sample execution of consensus-based terminating reliable broadcast

```
trbBroadcast (p₁, m)

p₁

ucPropose (m)

p₂

ucPropose (m)

p₃

ucPropose (F)

p₄

crash (p₁)

uniform consensus

ucDecide (m)

trbDeliver (m)

trbDeliver (m)

trbDeliver (m)
```
Sample execution of the membership algorithm

Introduction to Reliable Distributed Programming

Rachid Guerraoui, Luís Rodrigues

Copyright Guerraoui, Rodrigues, Springer 2006

Introduction
Basic Abstractions
Reliable Broadcast
Shared Memory
Consensus
Consensus Variants

Sample execution of the membership algorithm
Sample execution of the TRB-based view synchronous algorithm

\[
\begin{align*}
\text{membView}(p_2, p_3, p_4) & \quad \text{vsView}(p_1, p_2, p_3, p_4) \\
\text{TRB}(p_1) & \quad \text{TRB}(p_2) \\
\text{TRB}(p_3) & \quad \text{TRB}(p_4) \\
(\text{m}_1, \text{m}_3) & \quad (\text{m}_1, \text{m}_2, \text{m}_3) \\
(\text{m}_2, \text{m}_3) & \quad (\text{m}_1, \text{m}_2, \text{m}_3) \\
\end{align*}
\]