Concepts of Concurrent Computation

Bertrand Meyer
Sebastian Nanz

Lecture 5: an overview of SCOOP
The issue that SCOOP addresses

Can we bring concurrent programming to the same level of abstraction and convenience as sequential programming?
Then and now

**Sequential programming:**
Used to be messy
Still hard but key improvements:
- Structured programming
- Data abstraction & object technology
- Design by Contract
- Genericity, multiple inheritance
- Architectural techniques

**Concurrent programming:**
Used to be messy
**Still messy**
Example: threading models in most popular approaches
Development level: sixties/seventies
Only understandable through operational reasoning
### Previous advances in programming

<table>
<thead>
<tr>
<th>Feature</th>
<th>&quot;Structured programming&quot;</th>
<th>&quot;Object technology&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use higher-level abstractions</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Helps avoid bugs</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Transfers tasks to implementation</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Lets you do stuff you couldn't before</td>
<td>NO</td>
<td>✓</td>
</tr>
<tr>
<td>Removes restrictions</td>
<td>NO</td>
<td>✓</td>
</tr>
<tr>
<td>Adds restrictions</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Has well-understood math basis</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Doesn’t require understanding that basis</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Permits less operational reasoning</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
The chasm

Theoretical models, process calculi (see forthcoming lectures)

Elegant theoretical basis, but

- Remote from the ordinary practice of programming
- Handle concurrency aspects only

Practice of concurrent & multithreaded programming

- Low-level, e.g. threads, semaphores
- Poorly connected with rest of programming model (O-O structure of modern programs)
Simple Concurrent Object-Oriented Programming

First version described in CACM article (1993) and chapter 32 of Object-Oriented Software Construction, 2nd edition, 1997

Prototype implementation at ETH (2005-2008)
Recent production implementation at Eiffel Software, part of EiffelStudio

Recent descriptions: Piotr Nienaltowski's 2007 ETH PhD dissertation; Morandi, Nanz, Meyer (2011)
The design of SCOOP (and this presentation)

To achieve the preceding goals, SCOOP makes a number of restrictions on the concurrent programming model.

This presentation explains and justifies these restrictions one after the other.

The goal is not to limit programmers but to enable them to reason about the programs.
The design of SCOOP

SCOOP intends to make concurrent programming as predictable as sequential programming.

A key criterion is “reasonability” (not a real word!): the programmer’s ability to reason about the execution of programs based only on their text.

- As in sequential O-O programming, with contracts etc.

SCOOP is not a complete rework of basic programming schemes, but an incremental addition to the basic O-O scheme: one new keyword

- “Concurrency Made Easy”
Handling concurrency simply

SCOOP narrows down the distinction between sequential and concurrent programming to five key properties, studied next:

- **(A)** Single vs multiple “processors”
- **(B)** Synchronous vs asynchronous calls
- **(C)** Semantics of argument passing
- **(D)** Semantics of resynchronization (lazy wait)
- **(E)** Semantics of preconditions
The starting point (A): processors

To perform a computation is
- To apply certain actions
- To certain objects
- Using certain processors

Sequential: one processor
Concurrent: any number of processors
What makes an application concurrent?

**Processor:**

*Thread of control* supporting sequential execution of instructions on one or more objects

Can be implemented as:

- *Computer CPU*
- *Process*
- *Thread*
- *AppDomain (.NET)* …

The SCOOP model is abstract and does not specify the mapping to such actual computational resources
Reasoning about objects: sequential

Only $n$ proofs if $n$ exported routines!

$\{\text{INV and } \text{Pre}_r \} \quad \text{body}_r \quad \{\text{INV and } \text{Post}_r \}$

$\{\text{Pre}_r \} \quad x.r(a) \quad \{\text{Post}_r \}$

Priming represents actual-formal argument substitution

The concurrent version of this rule will come later!
In a concurrent context

Only $n$ proofs if $n$ exported routines?

Client 1  Client 2  Client 3

$r1$  $r2$  $r3$

No overlapping!

$\{\text{INV and } \text{Pre}_r\}$  body$_r$  $\{\text{INV and } \text{Post}_r\}$

$\{\text{Pre}_r',\} \ x.r(a) \ {\text{Post}_r'}$
SCOOP restriction: one handler per object

- **One processor per object:** "handler"

- At most one feature (operation) active on an object at any time
The notion of handler implies a partitioning of the set of objects:

- The set of objects handled by a given processor is called a region.
- Handler rule implies one-to-one correspondence between processors and regions.
(B) The sequential view: O-O feature calls

\[ x.r(a) \]

Client

\textit{previous}

\textit{x.r(a)}

\textit{next}

Supplier

\[ r(x : A) \]

\textit{do}

\textit{...}

\textit{end}

Processor
(B) The concurrent form of call: asynchronous

Client

previous

\(x.r(a)\)

next

Client’s handler

Supplier

\(r(x : A)\)

do

... 

de

Supplier’s handler
The two forms of O-O call

To wait or not to wait:

- If same processor, synchronous
- If different processor, asynchronous

Difference must be captured by syntax:

- $x: T$
- $x: \text{separate}\ T\ --\ \text{Potentially different processor}$

Fundamental semantic rule: a call $x.r(a)$

- Waits (i.e. is synchronous) for non-separate $x$
- Does not wait (is asynchronous) for separate $x$
Why potentially separate?

*separate* declaration does not specify processor: only states that the object *might* be handled by a different processor

- In class A: \( x: \text{separate B} \)
- In class B: \( y: \text{separate A} \)

In some execution the value of \( x.y \) might be a reference to an object in the current region (including *Current* itself)
Call vs application

With asynchrony we must distinguish between feature call and feature application.

The execution

\[ \mathbf{x \cdot r (...) \ldots} \]

is the call, and (with \( \mathbf{x} \) separate) will not wait (the client just logs the call).

The execution of \( \mathbf{r} \) happens later and is called the feature application.
Consistency rules: avoiding traitors

\[
\text{nonsep}: T
\]

\[
\text{sep}: \text{separate } T
\]

\[
\text{nonsep} \equiv \text{sep}
\]

\[
\text{nonsep}.p(a)
\]

Traitor!

More traitor protection through the type system! (next lectures)
Tusting what you read ("reasonability")

remote_stack: separate STACK[T]

...

remote_stack.put(a)

... Instructions not affecting the buffer...

y := remote_stack.item
SCOOP requires the target of a separate call to be a formal argument of enclosing routine:

```plaintext
put (b: separate QUEUE [T]; value: T)
  -- Add value, FIFO-style, to b.
do
  b.put (value)
end
```
The target of a separate call must be a formal argument of enclosing routine:

```plaintext
put (buffer: separate QUEUE [T]; value : T)
   -- Store value into buffer.
   do
      buffer.put (value)
   end
```

To use separate object:

```plaintext
my_buffer: separate QUEUE [INTEGER]
create my_buffer
put (my_buffer, 10)
```
(C) Separate argument rule

The target of a separate call must be an argument of the enclosing routine

Separate call: $x.f(...) \text{ where } x \text{ is separate}$
A routine call guarantees exclusive access to the handlers (the processors) of all separate arguments.

\[ a_{\text{routine}}(\text{nonsep}_a, \text{nonsep}_b, \text{sep}_c, \text{sep}_d, \text{sep}_e) \]

Exclusive access to \[ \text{sep}_c, \text{sep}_d, \text{sep}_e \] within \[ a_{\text{routine}} \]

Background for this rule: “reasonability” again
An example: from sequential to concurrent

transfer (source, target: separate ACCOUNT; amount: INTEGER)
   -- Transfer amount from source to target.
   require
      source.balance >= amount
   do
      source.withdraw (amount)
      target.deposit (amount)
   ensure
      source.balance = old source.balance - amount
      target.balance = old target.balance + amount
end
Dining philosophers in SCOOP

class PHILOSOPHER inherit PROCESS
    rename setup as getup
    redefine step end

feature {BUTLER}
    step
        do
            think; eat (left, right)
        end

    eat (l, r: separate FORK)
        -- Eat, having grabbed l and r.
        do ... end
end
Typical traditional code

Listing 4.33: Variables for Tanenbaum’s solution

1. state = ['thinking'] * 5
2. sem = [Semaphore(0) for i in range(5)]
3. mutex = Semaphore(1)

The initial value of state is a list of 5 copies of 'thinking'. sem is a list of 5 semaphores with the initial value 0. Here is the code:

Listing 4.34: Tanenbaum’s solution

1. def get_fork(i):
2.     mutex.wait()
3.     state[i] = 'hungry'
4.     test(i)
5.     mutex.signal()
6.     sem[i].wait()
7.
8. def put_fork(i):
9.     mutex.wait()
10.    state[i] = 'thinking'
11.    test(right(i))
12.    test(left(i))
13.    mutex.signal()
14.
15. def test(i):
16.    if state[i] == 'hungry' and
17.        state (left (i)) != 'eating' and
18.        state (right (i)) != 'eating':
19.            state[i] = 'eating'
20.       sem[i].signal()
A **PROCESS** library class

SCOOP integrates inheritance and other O-O techniques with concurrency, seamlessly and without conflicts ("inheritance anomaly")

No need for built-in notion of **active object**: it is **programmed** through a library class such as **PROCESS**:

```plaintext
class process feature
  setup do end
  step do end
  over : BOOLEAN
  tear_down do end
  live
    do
    from setup until over loop step end
    tear_down
  end
  end
end
```
What the wait rule means

Beat enemy number one in concurrent world: atomicity violations

- Data races
- Illegal interleaving of calls

Data races cannot occur in SCOOP

- Why? See computational model ...
Older SCOOP literature (OOSC, Nienaltowski, Morandi...) says that feature application “waits” until all the separate arguments’ handlers are available
This is not necessary! What matters is **exclusive access**: implementation does not have to wait unless semantically necessary
The current implementation performs these optimizations

```plaintext
f (a, b, c: separate T)
  do
    something_else
    a.r
    b.s
  end
```

No need to wait for `a` and `b` until here

No need to wait for `c`!
Resynchronization: lazy wait

How do we resynchronize after asynchronous (separate) call?
No explicit mechanism!
The client will wait only when it needs to:

\[
\begin{align*}
  &x.f \\
  &x.g(a) \\
  &y.f \\
  &... \\
  &value := x.some\_query
\end{align*}
\]

Lazy wait (also known as wait by necessity)
(D) Synchrony vs asynchrony revisited

For a separate target $x$:

- $x \cdot \text{command(...)}$ is asynchronous
- $v := x \cdot \text{query(...)}$ is synchronous
What becomes of contracts, in particular preconditions, in a concurrent context?
put\( (b: \text{BUFFER}[G]; v: G) \)
-- Store v into b.

require
not \(b.\text{is_full}\)

do
...
ensure
not \(b.\text{is_empty}\)
end

my_queue: \text{BUFFER}[T]

... if not my_queue.is_full then

put (my_queue, t)
end
(E) Contracts

\[
\text{put}(\text{buf} : \text{BUFFER [INTEGER]}; \ v : \text{INTEGER})
\]

-- Store \( v \) into buffer.

\[
\text{require}
\]

\[
\text{not} \ \text{buf}\text{.is_full}
\]

\[
v > 0
\]

\[
\text{do}
\]

\[
\text{buf}\text{.put}(v)
\]

\[
\text{ensure}
\]

\[
\text{not} \ \text{buf}\text{.is_empty}
\]

\[
\text{end}
\]

...

\[
\text{put}(\text{my_buffer}, 10)
\]
**Contracts**

\[ \text{put} \left( \text{buf} : \text{BUFFER [INTEGER]} ; v : \text{INTEGER} \right) \]

-- Store \( v \) into buffer.

\[ \text{require} \]

\[ \text{not} \ \text{buf.is_full} \]
\[ v > 0 \]

\[ \text{do} \]

\[ \text{buf.put} \left( v \right) \]

\[ \text{ensure} \]

\[ \text{not} \ \text{buf.is_empty} \]

\[ \text{end} \]

...

\[ \text{put} \left( \text{my_buffer}, 10 \right) \]
(E) Full synchronization rule

A call with separate arguments waits until:
- The corresponding objects are all available
- Preconditions hold

\[ x.f(a) \] -- where \( a \) is separate
Which semantics applies?

```plaintext
put (buf : separate BUFFER [INTEGER]; i : INTEGER)
  require
    not buf.is_full
    i > 0
  do
    buf.put (i)
  end

my_buffer : separate BUFFER [INTEGER]
put (my_buffer, 10)
```

**Wait condition**: 
- `not buf.is_full`
- `i > 0`

**Correctness condition**: 
- `buf.put (i)`
Generalized semantics of preconditions

The different semantics is surprising at first:
- Separate: wait condition
- Non-separate: correctness condition

At a high abstraction level, however, we may consider that:
- Wait semantics always applies in principle
- Sequentiality is a special case of concurrency
- Wait semantics boils down to correctness semantics for non-separate preconditions.
  - Smart compiler can detect some cases
  - Other cases detected at run time
What about postconditions?

\[ \text{spawn\_two\_activities} (\text{loc1, loc2: separate LOCATION}) \]

\[
\begin{align*}
do & \quad \text{loc1.do\_job} \\
& \quad \text{loc2.do\_job} \\
\text{ensure} & \quad \text{loc1.is\_ready} \\
& \quad \text{loc2.is\_ready} \\
\end{align*}
\]

\[ \text{spawn\_two\_activities} (\text{zurich, lausanne}) \]
\[ \text{do\_local\_stuff} \]
\[ \text{get\_result} (\text{zurich}) \]

Should we wait for \text{zurich.is\_ready}?
Reasoning about objects: sequential

\[
\{\text{INV and } \text{Pre}_r\} \quad \text{body}_r \quad \{\text{INV and } \text{Post}_r\}
\]

\[
\{\text{Pre}'_r\} \quad \text{x} \cdot r(a) \quad \{\text{Post}'_r\}
\]

Only \( n \) proofs if \( n \) exported routines!
Refined proof rule (partial correctness)

\[
\{INV \land Pre_r(x)\} \text{ body}_r \{INV \land Post_r(x)\}
\]

\[
\{Pre_r(a^{\text{cont}})\} \ e.r(a) \ \{Post_r(a^{\text{cont}})\}
\]

Hoare-style sequential reasoning

Controlled expressions (known statically as part of the type system) are:

- Attached (statically known to be non-void)
- Handled by processor locked in current context
SCOOP highlights

- Close connection to O-O modeling
- Natural use of O-O mechanisms such as inheritance
- Built-in guarantee of no data races
- Built-in fairness
- Removes many concerns from programmer
- Supports many different forms of concurrency
- Retains accepted patterns of reasoning about programs
- Simple to learn and use