Structured Reactive Programming with Céu

Francisco Sant’Anna Roberto Ierusalimschy Noemi Rodriguez
Departamento de Informática — PUC-Rio, Brazil
{fsantanna,roberto,noemi}@inf.puc-rio.br

ABSTRACT
Structured reactive programming (SRP) augments classical structured programming with continuous interaction with the environment. We propose a new SRP abstraction mechanism for the synchronous language Céu: Organisms extend objects with an execution body that composes multiple lines of execution to react to the environment independently. Compositions bring structured reasoning to concurrency and can better describe state machines typical of reactive applications. Organisms are subject to lexical scope and automatic memory management similar to stack-based allocation for local variables. We show that this model does not require garbage collection or a free primitive in the language, eliminating memory leaks by design.

Categories and Subject Descriptors
D.3.3 [Programming Languages]: Language Constructs and Features

General Terms
Design, Languages

Keywords
Concurrency, Determinism, Esterel, Imperative, Structured Programming, Synchronous, Reactivity

1. INTRODUCTION
Reactive applications interact continuously and in real time with external stimuli from the environment [18, 4]. They represent a wide range of software areas and platforms: from games in powerful desktops and “apps” in capable smart phones, to the emerging internet of things in constrained embedded systems.

Research on special-purpose reactive languages dates back to the early 80’s, with the co-development of two complementary styles [5, 24]: The imperative style of Esterel [7] organizes programs with structured control flow primitives, such as sequences, repetitions, and parallelism. The dataflow style of Lustre [17] represents programs as graphs of values, in which a change to a node updates its dependencies automatically.

In recent years, Functional Reactive Programming (FRP) [27] has modernized the dataflow style, inspiring a number of languages and libraries, such as Flapjax [22], Rx (from Microsoft), React (from Facebook), and Elm [11]. In contrast, the imperative style of Esterel is confined to the domain of real-time embedded control systems. As a matter of fact, imperative reactivity is now often associated to the observer pattern, typical in object-oriented systems, because it heavily relies on side effects [20, 25]. However, short-lived callbacks (i.e., the observers) eliminate any vestige of structured programming, such as support for long-lasting loops and automatic variables [3], which are elementary capabilities of imperative languages. In this sense, callbacks actually disrupt imperative reactivity, becoming “our generation’s goto” [13, 15].

We believe that the full range of reactive applications can benefit from the imperative style of Esterel, which we now refer to as Structured Reactive Programming (SRP). SRP extends the classical hierarchical control constructs of Structured Programming (SP) (i.e., concatenation, selection, and repetition [14]) to support continuous interaction with the environment. SRP retains structured and sequential reasoning of concurrent programs which constrains with FRP, bringing the historical dichotomy between functional and imperative languages also to the reactive domain. However, the original rigorous semantics of Esterel, which focuses on static safety guarantees, is not suitable for other reactive application domains, such as GUIs, games, and distributed systems. For instance, the lack of abstractions with dynamic lifetime makes it difficult to deal with virtual resources such as graphical widgets, game units, and network sessions.

In practical terms, SRP provides three extensions to SP: an “await <event>” statement that suspends a line of execution until the referred event occurs, keeping all context alive; parallel constructs to compose multiple lines of execution and make them concurrent; and an orthogonal mechanism to abort parallel compositions. The await statement represents the imperative-reactive nature of SRP, recovering sequential execution lost with the observer pattern. Parallel compositions allow for multiple await statements to coex-

1In this work, the term parallel composition does not imply many-core parallel execution.
ist, which is necessary to handle concurrent events, common in reactive applications. Orthogonal abortion is the ability to abort an activity from outside it, without affecting the overall consistency of the program (e.g., properly releasing global resources).

In this work, we extend the Esterel-based language Cèu [26] with a new abstraction mechanism, the 
organisms, that encapsulate parallel compositions with an object-like interface. In brief, organisms are to SRP like procedures are to SP, i.e., one can abstract a portion of code with a name and manipulate (call) that name from multiple places. Unlike Simula objects [12], organisms react independently to the environment and do not depend on cooperation, i.e., once instantiated they become alive and reactive (hence the name organisms). Furthermore, both static and dynamic allocation of organisms are subject to lexical scope and automatic memory management, not relying on heap allocation at all, and behaving much like local variables in SP.

The rest of the paper is organized as follows: Section 2 presents SRP through Cèu and behaving much like local variables in SP. Section 3 describes the organisms abstraction with static and dynamic instantiation, lexical scope, and automatic memory management. Section 4 discusses related work. Section 5 concludes the paper.

2. SRP WITH CÈU

Cèu is a concurrent language in which the lines of execution, known as trails, react all together continuously and in synchronous steps to external stimuli. The introductory example in Figure 1 starts two trails with the par construct: the first (lines 4-8) increments variable \( v \) on every second and prints its value on screen; the second (lines 10-13) resets \( v \) on every external request to \texttt{RESET}. Programs in Cèu can access C libraries of the underlying platform by prefixing symbols with an underscore (e.g., \texttt{_printf(<...>)}), in line 7).

2.1 Synchronous concurrency

In Cèu, a program reacts completely to an occurring event before handling the next. A reaction represents a logical instant in which all trails awaiting the occurring event awake and execute, one after the other, until they await again or terminate. During a reaction, the environment is invariant and does not interrupt the running trails. If multiple trails react to the same event, the scheduler employs lexical order to preserve determinism, i.e., the trail that appears first in the source code executes first. To avoid infinite execution for reactions, Cèu ensures that all loops contain await statements [26].

As a result of synchronous execution, all consecutive operations to variable \( v \) in Figure 1 are atomic because reactions to events \texttt{1s} and \texttt{RESET} can never interrupt each other. In contrast, in asynchronous models with nondeterministic scheduling, the occurrence of \texttt{RESET} could preempt the first trail during an increment to \( v \) (line 6) and reset it (line 12) before printing it (line 7), characterizing a race condition.

Abortion in asynchronous languages is challenging [6] because the activity to be aborted might be on a inconsistent state (e.g., holding pending messages or locks). This way, the possible (unsatisfactory) semantics for a hypothetical par/or are: either wait for the activity to be consistent before rejoining, making the program unresponsive to incoming events for an arbitrary time; or rejoin immediately and let the activity complete in the background, which may cause race conditions with the subsequent code. In fact, asynchronous languages do not provide effective abortion: \texttt{Java’s Thread.stop} primitive has been deprecated [23]; \texttt{pthread’s pthread_cancel} does not guarantee immediate cancellation [2]; \texttt{Erlang’s exit} either queues a terminating message (which may take time), or unconditionally terminates the process (regardless of its state) [1]; and \texttt{CSP} only supports a composition operator that “terminates when all of the combined processes terminate” [19].

The synchronous model also empowers SRP with an orthogonal abortion construct that simplifies the composition of activities\(^3\). The code that follows shows the \texttt{par/or} construct of Cèu which composes trails and rejoins when either of them terminates, properly aborting the other:

\begin{verbatim}
par/or do
  with <trail-1>
  <trail-2>
  <subsequent-code>
\end{verbatim}

The \texttt{par/or} is regarded as orthogonal because the composed trails do not know when and how they are aborted (i.e., abortion is external to them). This is possible in synchronous languages due to the accurate control of concurrent activities, i.e., in between every reaction, the whole system is idle and consistent [6]. Cèu extends orthogonal abortion to work also with activities that use stateful resources from the environment (such as file and network handlers), as we discuss in Section 2.2.

\(\text{Figure 1: Introductory example in Cèu.}\)

\(\text{The actual implementation enqueues incoming input events to process them in further reactions.}\)

\(\text{We use the term activity to generically refer to a language’s unit of execution (e.g., thread, actor, trail, etc.).}\)
input void START, STOP, RETRANSMIT;
loop do
  await START;
  par/or do
    await STOP;
    with
      loop do
        par/or do
          await RETRANSMIT;
          with
            await <rand> s;
            par/and do
              await 1min;
              with
                <send-beacon-packet>
                end
            end
            <....> // the rest of the protocol
      end
  end
end

The code <....> represents a complex operation with any degree of nested compositions. In the par/and variation, the operation repeats on intervals of at least one second because both sides must terminate before re-entering the loop. In the par/or variation, if the operation does not terminate within 1 second, it is restarted. These SRP archetypes represent, respectively, the sampling and timeout patterns, which are typical of reactive applications.

The example in Figure 2 relies on hierarchical par/or and par/and compositions to describe the state machine of a protocol for sensor networks [16, 26]. The input events START, STOP, and RETRANSMIT (line 1) represent the external interface of the protocol with a client application. The protocol enters the top-level loop and awaits the starting event (line 3). Once the client application makes a start request, the protocol starts three other trails: one monitors the stopping event (line 5); one periodically transmits a status packet (lines 7-18); and one handles the remaining functionality of the protocol (collapsed in line 20). The periodic transmission

Figure 3: Finalization clauses safely release stateful resources.

is another loop that starts two other trails (lines 8-17): one to handle an immediate retransmission request (line 9); and one to await a small random amount of time and transmit the status packet (lines 11-16). The transmission (collapsed in line 15) is enclosed with a par/and that takes at least one minute before looping, to avoid flooding the network with packets. At any time, the client may request a retransmission (line 9), which terminates the par/or (line 8), aborts the ongoing transmission (line 15, if not idle), and restarts the loop (line 7). Also, the client may request to stop the whole protocol at any time (line 5), which terminates the outermost par/or (line 4) and aborts the transmission and all composed trails. In this case, the top-level loop restarts (line 2) and waits for the next request to start the protocol (line 3), ignoring all other requests (as the protocol specifies).

The example shows how parallel compositions can describe complex state machines in a structured way, eliminating the use of global state variables for this purpose [26].

2.2.1 Finalization
The Céu compiler tracks the interaction of par/or compositions with local variables and stateful C functions (e.g., device drivers) in order to preserve safe orthogonal abortion of trails.

Consider the code in the left of Figure 3, which expands the sending trail of Figure 2 (line 15). The buffer packet is a local variable whose address is passed to function _send_enqueue. The call enqueues the pointer in the radio driver, which holds it up to the emission of SEND_ACK acknowledging the packet transmission. In the meantime, the sending trail might be aborted by STOP or RETRANSMIT requests (Figure 2, lines 5 and 9), making the packet buffer go out of scope, and leaving behind a dangling pointer in the radio driver. Céu refuses to compile programs like this and requires finalization clauses to accompany stateful C calls [26]. The code in the right of Figure 3 properly de-queues the packet if the block of buffer goes out of scope, i.e., the finalization clause (after the with) executes automatically on external abortion.

3. ORGANISMS: SRP ABSTRACTIONS
In SP, the typical abstraction mechanism is a procedure, which abstracts a routine with a meaningful name that can be invoked multiple times with different parameters. However, procedures were not devised for continuous input, and cannot retain control across reactions to the environment.

Céu abstracts data and control into the single concept of organisms. A class of organisms describes an interface and
an execution body. The interface exposes public variables, methods, and also internal events (exemplified later). The body can contain any valid code in methods, and also internal events (exemplified later). The interface exposes public variables, methods, and also internal events (exemplified later). The body can contain any valid code in methods, and also internal events (exemplified later). The interface exposes public variables, methods, and also internal events (exemplified later).

The example in Figure 4 introduces static organisms with instantiation can be either static or dynamic. The actual implementation of Céu does not expand the organisms bodies like in CODE-3; instead, a class generates a single code for its body, which is shared by all instances (in the same way as objects share class methods).

The memory model for organisms is similar to stack-living local variables of procedures, employing lexical scope and automatic management. Note that CODE-2 uses a do-end block (lines 11-23) that limits the scope of the organisms for 1 minute (line 22). During that period, the organisms are accessible (through b1 and b2) and reactive to the environment (i.e., blinking continuously). After that period, the organisms go out of scope and not only they become inaccessible but their bodies are automatically aborted, as the expansion of CODE-3 makes clear: The par/or (lines 9-29) aborts the organisms bodies after 1 minute (line 28), just before they go out of scope (line 30). The par/or termination properly triggers all active finalization clauses inside the organism bodies (if any). Lexical scope extends the idea of

Figure 4: Two blinking LEDs using organisms.
orthogonal abortion to organisms, as they are automatically aborted when going out of scope. In this sense, organisms are more than a cosmetic convenience for programmers because they tie together data and associated execution into the same scope.

In addition to properties and methods, organisms also expose internal events which support `await` and `emit` operations. In the example in Figure 5, class `Unit` (lines 1-16) defines the position and destination properties `pos` and `dst` with default values `0` (lines 2-3), and the event `move` to listen requests to move the unit position (line 4). The main program (lines 18-29) creates two units, requests the first to move immediately to `dst=500` (line 20), and the second to move after 1 second to position `500` (line 29). The body of the class enters a continuous loop (lines 6-15) to handle `move` requests (line 8) while performing the ongoing moving operation (lines 10-13) in parallel. The `par/or` (lines 7-14) restarts the loop on every `move` request which updates the `dst` position. The moving operation (collapsed in line 11) can be as complex as needed, for example, using another loop to apply physics over time. The `await FOREVER` (line 13) halts the trail after the move completes to avoid restarting the outer loop. An advantage of event handling over method calls is that they can be composed in the organism body to affect other ongoing operations. In the example, the `await` (line 8) aborts and restarts the moving operation, just like the timeout pattern of Section 2.2.

**3.1 Dynamic organisms**

Static embedded systems typically manipulate hardware with a one-to-one correspondence in software, i.e., a static piece of software deals with a corresponding piece of hardware (e.g., a sensor or actuator). In contrast, more general reactive systems have to deal with resource virtualization that requires dynamic allocation, such as multiplexing protocols in a network, or simulating entire civilizations in a game. Dynamic allocation for organisms extends the power of SRP to handle virtual resources in reactive applications.

Céu supports dynamic instantiation of organisms through the `spawn` primitive. The example that follows spawns a new instance of `Unit` (defined in Figure 5) every second and moves it to a random position:

```plaintext
loop do
  spawn Unit with
    this.pos = _rand() % 500;
    this.dst = _rand() % 500;
  end;
end
```

Dynamic instances also execute in parallel with the rest of the application, but have different lifetime and scoping rules then static ones: A static instance has an identifier and a well-defined scope that holds its memory resources; A dynamic instance is anonymous and outlives the scope that spawns it; in the example, the spawned units outlive the enclosing loop iterations. Due to the lack of an explicit identifier or reference, a dynamic instance can control its own lifetime: once its body terminates, a dynamic organism is automatically freed from memory. This does not apply for a static instance because its memory is statically preallocated and its identifier is still accessible even if its body terminates.

The code that follows redefines the body of the `Unit` class of Figure 5 to terminate after 1 hour, imposing a maximum life span in which a unit can react to `move` requests. After that, the body terminates and the organism is automatically freed (if dynamically spawned):

```plaintext
class Unit with
  <...> // interface
do
  par/or do
    <...> // moving trail
    with
        await 1h;
    end;
end
```

The lack of scopes for dynamic organisms prevents orthogonal abortion, given that there is no way to externally abort the execution of a dynamic instance. To address orthogonal abortion, Céu provides lexically scoped `pools` as containers that hold dynamic instances. The example that follows declares the units pool to hold a maximum of 10 instances (line 3):

```plaintext
input void CLICK;
do
  pool Unit[10] units;
  par/or do
    loop do
      await 1s;
      spawn Unit in units with
        <...> // constructor
    end;
  end
  with
    await CLICK;
end
```
A new unit is spawned in this pool once a second (note the in units, in line 7). Once the application receives a CLICK (line 12), the par/or (line 4) terminates, making the units pool go out of scope and abort/free all units alive.

Pools with bounded dimension (e.g., pool Unit[10] units), have static pre-allocation, resulting in efficient and deterministic organism instantiation. This opens the possibility for dynamic behavior also in constrained embedded systems. If a pool does not specify a dimension (e.g., pool Unit[] units), the instances go to the heap but are still subject to the pool scope. If a spawn does not specify a pool, as in “spawn Unit;”, the instances go to a predefined dimensionless pool in the top of the current class (and are still subject to that pool scope).

Support for lexical scope for both static and dynamic organisms eliminate garbage collection, free primitives, and memory leaks altogether.

3.2 Pointers and References

As organisms react independently to the environment, it is often not necessary to hold pointers to them. Nonetheless, a spawn allocation returns a pointer to the new organism, which can be later dereferenced with the operator ‘->’ (analogous to ‘->’ of C/C++):

```c
var Unit* ptr = spawn Unit;
ptr:pos = 0; // this access is safe
await 2h;
emit ptr:move => 100; // this access is unsafe
```

Pointers can be dangerous because they may last longer than the organisms to which they refer. The code above first acquires a pointer ptr to a Unit. Then, it dereferences the pointer in two occasions: in the same reaction, just after acquiring the reference; and in another reaction, after 2h, when the pointed organism may have already terminated and been freed, leading to unspecified behavior in the program.

As a protection against dangling pointers, CÉU enforces all pointer accesses across reactions to use the watching construct which supervises organism termination, as illustrated in the left of Figure 6. The whole watching construct aborts whenever the referred organism terminates, eliminating possible dangling pointers in the program. The code in the right shows the equivalent expansion of the watching construct into a par/or that awaits the special event __killed (which all classes have internally).

CÉU also refuses to assign the address of an organism to a pointer of greater scope, as illustrated below:

```c
var Unit* ptr;
do
  var Unit u;
  ptr = &u; // illegal attribution
end
ptr:pos = 0; // unsafe access (‘u’ went out of scope)
```

Pools supports iterators to acquire temporary pointers to all alive instances. To preserve safe pointer accesses, iterators cannot await. The example that follows iterates over the units pool to check for collision among units:

```c
var Unit* ptr = spawn Unit;
ptr:pos = 0;
watching ptr do
  await 2h;
  emit ptr:move => 100;
end
with
  await 2h
  emit ptr:move => 100;
end
```

Figure 6: Watching a reference with equivalent expansion.

pool Unit[10] units;
<...>
loop (Unit*)u1 in units do
  loop (Unit*)u2 in units do
    if <check-collision> then
      emit u1:move => _rand() % 500;
      emit u2:move => _rand() % 500;
    end
  end
end
```

4. RELATED WORK

Simula is a simulation language that introduced the concepts of objects and coroutines [12]. The syntactic structure of classes in Simula is very similar to CÉU, exposing an interface that encapsulates an execution body. However, the underlying execution models are fundamentally distinct: CÉU employs a reactive scheduler to resume trails based on external stimuli, while Simula relies on cooperation between processes (i.e., detach and resume calls, at the lowest level). Simula has no notion of compositions, with each process having a single line of execution. In particular, the lack of a par/or precludes orthogonal abortion and many derived CÉU features, such as lexically scoped organisms, finalization, and reference watching. Without scopes, Simula objects and processes have to live on the heap and rely on garbage collection. As far as we know, Simula processes cannot be terminated explicitly from other processes.

Some previous work extend Esterel to provide dynamic synchronous abstractions [9, 8, 10]. In particular, ReactiveML [21] is a functional variant of Esterel with rich dynamic synchronous abstractions through processes. However, these languages rely on heap allocation and/or garbage collection and may not be suitable for constrained embedded systems. They also lack a finalization mechanism that hinders proper orthogonal abortion in the presence of stateful resources.

Finally, the main distinction to existing work is how CÉU incorporates to SRP the fundamental concept in SP of lexically scoped variables. All constructs of CÉU have a clear and unambiguous lifespan that can be inferred statically from the source code. Lexical scope permeates all aspects of the language: Any piece of data or control structure has a well-defined scope that can be abstracted as an organism and safely aborted through finalization. Even dynamic instances of organisms reside in pools with well-defined scopes with the same properties.

Functional Reactive Programming [27] contrasts with SRP as a complementary programming style for reactive applications. We believe that FRP is more suitable for data-intensive applications, while SRP, for control-intensive ap-
plications. For instance, describing a sequence of steps in FRP requires to encode explicit state machines so that functions can switch behavior depending on the current state. In contrast, FRP uses declarative formulas to specify continuous functions over time, such as for physics or data constraints among entities, while SRP requires explicit loops to update data dependencies continuously.

5. CONCLUSION

Céu provides comprehensive support for structured reactive programming, extending classical structure programming with continuous interaction with the environment.

Céu introduces organisms which reconcile data and control state in a single abstraction. In contrast with objects, organisms have an execution body that can react independently to stimuli from the environment. An organism body supports multiple lines of execution that can await events without loosing control context, offering an effective alternative to the infamous “callback hell”. Both static and dynamic instances of organisms are subject to lexical scope with automatic memory management, which eliminates memory leaks and the need for a garbage collector.

Céu is suitable for wide range of reactive applications and platforms. We have been experimenting with it in constrained platforms for sensor networks as well as in full-fledged computers and tablets for games and graphical applications. Céu successfully participated in the Google Summer of Code with a student that had no previous experience with the language. We have also been teaching Céu as an alternative language for sensor networks for the past two years in high-school and undergraduate levels. Our experience shows that students take advantage of the sequential style of Céu which reconciles data and control state in a single abstraction. In contrast with objects, organisms have an execution body that can react independently to stimuli from the environment. An organism body supports multiple lines of execution that can await events without loosing control context, offering an effective alternative to the infamous “callback hell”. Both static and dynamic instances of organisms are subject to lexical scope with automatic memory management, which eliminates memory leaks and the need for a garbage collector.

6. REFERENCES


---

5LabLua GSoC’14: google-opensource.blogspot.com/2014/08/google-summer-of-code-new-organizations.html