Component-based Modeling in Umple

Mahmoud Husseini Orabi, Ahmed Husseini Orabi and Timothy C. Lethbridge
School of Electrical Engineering and Computer Science, University of Ottawa,
800 King Edward Avenue, Ottawa, Canada

Keywords: Umple, Active Object, Composite Structure, UML.

Abstract: Modelling tools provide weak or no support for the rich semantics of composite structure, such as enforcing connection constraints and maintaining referential integrity. Tools that generate code from composite structure typically depend on excessive and complex internal class representations such as Actor or BasePort. In this paper, we present easy-to-comprehend syntax describing composite structure in Umple. We describe a novel protocol-free approach that dynamically extracts communication protocols as a way to ease component-based modelling, and lead to concise and optimized code generation. We outline Umple composite structure features, and the related code generation patterns that resolve difficulties around connections and the integrity of multiplicity constraints.

1 INTRODUCTION

Composite structure development refers to the implementation of concurrent components that interact and communicate via ports and connectors (Orabi, Orabi, & Lethbridge, 2016).

In UML, interactions among components are handled as messages and signals. There are two message-passing actions, one-way call (asynchronous call) and calls that block waiting for a response (synchronous call). An asynchronous call is referred to as a one-way message passing, since it does not support a scheduling mechanism to receive results. Typically, events are triggered when receiving messages by invoking a corresponding method. Event handling is often done in a component’s state machine.

Handling message flow among ports and components usually depends on protocols. Typically, defining a protocol involves repetitive steps with redundant information to define how in and out events are managed. Hence, complexity is added to the development process.

Open-source tools such as eTrice and ArgoUml do not support all the major features of composite structure (Orabi et al., 2016). On the other hand, commercial tools that provide strong support to composite structure typically restrict users to certain libraries, such as Connexis, which is tightly integrated with RSARTE (Lakkimsetti, 2014).

Motivated by the above, we show how we extend Umple, an open-source tool, to support composite structure and overcomes such limitations.

Umple is a text- and model-oriented programming language that allows for generative programming using many target languages such as C++, Java, and PHP (Badreddin, Forward, & Lethbridge, 2014; Badreddin, Lethbridge, & Forward, 2014; Lethbridge, Abdelzad, Husseini Orabi, Husseini Orabi, & Adesina, 2016; Orabi et al., 2016). Umple implements the core features of UML, such as state machines, associations, and attributes. In addition, Umple provides other usable features to ease development such as traits, mixins, and aspect-orientation. The selected target language in this paper is C++. Umple features can be fully tried and tested using the UmpleOnline website (try.umple.org).

Our key contributions can be summarized as follows:

- We support major composite structure features using compact keywords and shortened syntax.
- We provide generic extensible communication and transport definitions to support distributed examples. We implement support for the TCP/UDP communication protocol and JSON as a message interchange format to be the default medium for communication among distributed components.
We incorporate the active object pattern (Lavender & Schmidt, 1996) to extend and enable the request-scheduling mechanism.

We implement a protocol-free approach, in which protocols are inferred from ports, connectors, and components.

Our implementation works for distributed applications, and does not restrict users to certain network paradigms or any particular transport data format.

Composite structure features are introduced to Umple as a part of this research, as a major step towards the development of connected embedded devices.

2 COMPONENT-BASED MODELLING AND STANDARDS

Many of the existing component-based modelling tools (Orabi et al., 2016) adapt specifications such as UML (OMG, 2011), SysML (Mallet, Peraldi-Frati, & André, 2009), Specification and Description Language (SDL), and Real-time Object-Oriented Modelling (ROOM) (Selic & ObjectTime, 1996).

In our work, we are more inspired by UML, since, 1) OMG uses UML as the de facto modeling notation for object-oriented systems (Grady Booch, James Rumbaugh, 2005), and 2) UML is well-known for design specification for real-time systems, especially embedded devices. As well, in terms of real-time modelling, we follow many of the ROOM specifications. However, we have an extended the active object pattern to handle concurrency (Orabi, 2017).

SDL’s main purpose is to provide unambiguous software system specifications (Olsen, Faergemand, Møller-Pedersen, Smith, & Reed, 1994). SDL has mainly been used in the modeling of real-time communication systems. It provides similar functionality as UML but with different terminology and notation (Table 1).

ROOM was developed by ObjectTime and introduced in the ObjectTime Developer Tool (ODT) (Selic & ObjectTime, 1996). ROOM incorporates a variant of Harel’s statecharts. In 1996, RoomLanguage represented the actor as the primary element that communicates with other elements using port interfaces. A port is an instance of a protocol class that defines the message communication between actors. The concept of Actor was later referred to as a Capsule in UML-RT. ROOM supports hierarchical modeling and incremental refinement of complex behaviour.

<table>
<thead>
<tr>
<th>UML</th>
<th>SDL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class</td>
<td>Type</td>
</tr>
<tr>
<td>Interface</td>
<td>Interface</td>
</tr>
<tr>
<td>Associations</td>
<td>Channels</td>
</tr>
<tr>
<td>Operations</td>
<td>Signal List</td>
</tr>
<tr>
<td>Variable</td>
<td>Attribute</td>
</tr>
<tr>
<td>Sub</td>
<td>Block</td>
</tr>
<tr>
<td>Abstract</td>
<td>Abstract</td>
</tr>
<tr>
<td>Implementation</td>
<td>Process</td>
</tr>
<tr>
<td>Type</td>
<td>Gate</td>
</tr>
<tr>
<td>Inheritance</td>
<td>Inheritance</td>
</tr>
</tbody>
</table>

Table 1: UML to SDL term mapping.
3 COMPOSITE STRUCTURE IN UMLE

In Umple, components, ports, and connectors are used to describe structural implementation of an object, while state machines are used to define its behaviour. A component is viewed as an active object entity with a unique identifier (UID) that handles its own thread of execution and encapsulates its well-defined behaviour.

The implementation of composite structure in Umple uses an extended version the active object pattern (Orabi, 2017). In Umple, an active object class is referred to as a component. Communication among components is established via ports and connectors, with necessary protocols being internally generated and hidden from the developer.

Each component has a public interface, UID, internal router, connection type, and transport data format. An internal router is used to manage components in a distributed system.

We use template meta-programming (TMP) to avoid stub generation, reduce the volume of code the user has to write, and provide an extensible communication stack (Orabi, 2017; Smaragdakis & Batory, 2000).

The public interface refers to the methods exposed for communication, which can be synchronous, asynchronous, or future asynchronous. They are internally represented as a generic template proxy that enables a publish-subscribe mechanism, and uses the internal router of their owning class. Standard public methods are synchronous, while port public interface methods are asynchronous. A special case is future asynchronous, which represents an active method with a return type. This means that in the distributed mode, a component that initiates future asynchronous calls is going to receive a scheduled response asynchronously from the other component.

A component handles both inter-process communication (IPC) and remote method invocation (RMI). The internal routing structure provides the essential blocks to ease the building of components that can communicate. The internal routing table is used to handle send-reply and request-scheduling mechanisms between component’s internal methods and respondents. A component can be either in local or distributable state. Being in distributable state means that either the component is listening (acting as a server) or initiating communication (acting as a client). There is no restriction to a specific network paradigm or multi-party communication.

A communication stack provides abstract definitions for connection mechanisms and data interchange transport formatters. By default, we support TCP/IP as a connection mechanism, and JSON as a data interchange format. The code can be easily extended to support different connection mechanisms such as UDP and Bluetooth, and other formats such as XML.

3.1 Components and Parts

A class becomes a component if it has at least one active method, port, or connector. An active method is defined as a regular method proceeded by the active keyword (Snippet 1 - Lines 2 and 9). When invoking an active method, it executes asynchronously, since it has its own thread.

```
class A { // A component
    active method1 {
        cout << "Method without parameters" << endl;
    }
}

class B {
    active method2 (int someParam) {
        cout << "Parameter value" << someParam << endl;
    }
}

class C {
    A a;
    B b;
}
```

Snippet 1: An example of a component definition.

A part, or subcomponent, is an instance of a component, and it is owned by the component structure of some component. The instance type of a part, or subcomponent, can be the same as its owning component; this is similar to the programming patterns, in which an instance of a class is created in that class’s definition in places such as constructors, attributes, or methods. When a component owns multiple parts, it is referred to as a composite component. A subcomponent can possibly be composite. A component has a composition relationship to each class typed by its owned parts.

```
C
A B
```

Figure 3: Multiple instances of different components.
In Figure 3, the parts "a" and "b" are instances of A and B respectively, and they are owned by a component "c" of the type C. The Umple code is in (Snippet 1; see definitions of classes A, B, and C). This means that C has composition relationships to A and B.

3.2 Ports

A port in Umple is defined as an attribute, and additionally has a direction, which can be in, out, or both (specified using the keyword port and meaning bi-directional, and also referred to as dual). A port attribute is lazy. In Umple, lazy attributes are not initialized through their owning class’s constructor.

The keywords in, out, and port are used to set a port direction (Snippet 2 - Lines 2-4).

```
class Component{
  in Integer inPort;
  out Integer outPort;
  port Integer dualPort;
  internal in Integer privatePort;
  in SomeClass someComplexPort;
  port CompoundPort compoundPort;
  CompoundPort
    active compoundPort_ActiveMethods()
    { someInPort }
    active void handleDefaultDirection()
    { handleConjugatedDirection() }
    // CompoundPort inversion is as below
    //out Integer someInPort;
    //in Integer someOutPort;
  }
  void active someMethod()
  { pIn1Statemachine
    receive
    { stateEvent(pIn1 + 1);
    } done
  }
}
```

Snippet 2: Port examples.

A port can be simple, complex, or compound. A port is simple if its attribute type is simple such as string, integer, or double (Snippet 2- Lines 2-5); otherwise, it is considered complex (Line 6). A compound port consists of a number of subports, as a way to encompass a number of events for transmission (Lines 7, 8, and 30-33). A port attribute type has no restrictions. For instance, a port attribute can be typed by a component.

Within a component-based application, communication is established among a number of components instances. The boundary of instances created is managed using class associations, similarly to any normal application (Snippet 3 - Line 7). Messages propagated will be received by all instances associated (Line 23).

3.3 Connectors

A connector associates between two ports in order to establish a communication channel for data transmission. A class is considered a component if it has a connector defined, even if this class does not own active methods or ports.

```
class Client{
  in String cp;
  }

class Server{
  out String sp;
  * -- * Client; //Many to many association
  }

class Sys{
  Client c;
  Server s;
  s.sp->c.cp;
  public static void main(int argc, char *argv[]){
    Server* server= new Server();
    Client* c1= new Client();
    Client* c2= new Client();
    Client* c3= new Client();
    server->addClient(c1);
    server->addClient(c2);
    server->addClient(c3);
    s->sp("Broadcast a message to all instances");
  }
}
```

Snippet 3: Basic active objects in Umple.

The operator ":-" is used to define a connector, such that the port on the left is the source, and the port on the right is the target. A connector can associate between ports in the same component (Snippet 4 - Line 8) or different components (Lines 24 and 25).
The composite structure of "D" defined in Snippet 4 is visualized in Figure 4 (generated by UmpleOnline).

A connector can only connect between ports if they have opposite directions; i.e. an in port versus out port, dual port versus in port, dual port versus out port, and dual port versus dual port.

The notion of "->" that we use to define associations and connectors can be confusing to C/C++ developers, since it is exactly the same as using pointers (Snippet 4 - Line 21). For future research, we will look into trying to reduce the need to having to use pointers in target-language code such as methods that the users embed in the Umple code.

![Figure 4: Composite structure of connected components.](image)

<p>| 1 | class A { |
| 2 |   out Integer outPort; |</p>
<table>
<thead>
<tr>
<th>3</th>
<th>}</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>class B {</td>
</tr>
<tr>
<td>5</td>
<td>in Integer inPort;</td>
</tr>
<tr>
<td>6</td>
<td>out Integer outPort;</td>
</tr>
<tr>
<td>7</td>
<td>inPort -&gt; outPort;</td>
</tr>
<tr>
<td>8</td>
<td>}</td>
</tr>
<tr>
<td>9</td>
<td>class C {</td>
</tr>
<tr>
<td>10</td>
<td>in Integer inPort;</td>
</tr>
<tr>
<td>11</td>
<td>}</td>
</tr>
<tr>
<td>12</td>
<td>class D {</td>
</tr>
<tr>
<td>13</td>
<td>A a;</td>
</tr>
<tr>
<td>14</td>
<td>B b;</td>
</tr>
<tr>
<td>15</td>
<td>C c;</td>
</tr>
<tr>
<td>16</td>
<td>void active someMethod() {</td>
</tr>
<tr>
<td>17</td>
<td>a-&gt;outPort(120); // Send a signal of 120</td>
</tr>
<tr>
<td>18</td>
<td>a.outPort-&gt;b.inPort;</td>
</tr>
<tr>
<td>19</td>
<td>b.outPort-&gt;c.inPort;</td>
</tr>
<tr>
<td>20</td>
<td>}</td>
</tr>
</tbody>
</table>

Snippet 4: Connector examples.

### 3.4 Protocols and Our Protocol-Free Approach

Typically, an active method uses port *attaches* to listen to port events (Snippet 5 – Line 13, 21, and 38). In our protocol-free approach, a port attachment holds the information about incoming and outgoing ports, which makes it possible to generate protocols based on such information.

In terms of code generation, a protocol class is generated to handle communication for each component via its ports. When a port type is complex, we apply an appropriate serialization/deserialization technique. A port value is serialized into an intermediary object transmitted in the form of messages. When messages transmitted are received, they are deserialized back to the original object form.

Data is sent through connectors as signals. A port is expected to be able to receive signals simultaneously from different connectors. This means that there must be a queue mechanism to handle signals appropriately based on their priority and/or receiving order.

In our implementation, we have a priority FIFO queue, in which requests are ordered based on their priorities (i.e. Snippet 4 - Line 21), and then based on their receiving order.

At the level of the generated code, queuing depends on an internally generated helper class, MessageService. The utility class MessageService is also generated when processing any Umple model that uses composite structure features.

Communicating components can exist in different applications or locations. Hence, it was important to support buffered message transmission. This is handled using a generic API MessageDescriptor we implemented, which is also generated as needed. MessageDescriptor follows a publisher-subscriber pattern. MessageDescriptor and MessageService APIs are available in UmpleOnline when generating an Umple model that has ports or connectors defined. By default, the maximum size of a transfer request is 512 Kilobytes. If a message to be transmitted exceeds this maximum size, it will be divided into a number of smaller chunks, such that each chunk size will not exceed that size. When all chunks are received, they will be assembled into a message, which will be deserialized into the form of the object data originally sent. We selected a small maximum size in order to make sure that it will not exceed the maximum transmission unit (MTU) of a network, such that it will take less memory and process fast. For future work, we will experiment with other values to see which can be better, and investigate whether we can adjust the value at the model level.

MessageService works closely with the publisher-subscriber API existing in MessageDescriptor. Upon receiving incoming events, new messages will be created based on a subscriber list, which contains the active methods subscribed. The created messages will be added to the message queue using the MessageService API.
4 PORT TYPES

A port has a type that can be either conjugated or base. By default, a port is base (Snippet 2 - Lines 2-6). A conjugated port (Lines 11 and 12) can be alternatively referred to as an invert port. For instance, a conjugated in port also acts as an out port, and a conjugated out port also acts as an in port. Conjugation is only used with compound ports. When a port is compound, its conjugated version will have all of the subports inverted, as commented in (Snippet 2 - Lines 13-15).

An in or dual port can additionally be a relay or end port (Selic, 1998). Relay and end ports are called external ports. Any out port is a relay port (Snippet 4-Lines 2 and 7). Ports that propagate signals to other ports are considered relay ports (Line 6). Signal propagation stops at end ports (Lines 12).

The process of signal propagation changes whether a port is a service or nonservice port, and whether it is a behaviour or nonbehaviour port.

A service port expects to receive inputs from, or send outputs to, its environment. Service ports are drawn on the boundary of its owning part. In Umple, public ports are considered service ports (i.e. Snippet 2 - Lines 2-4).

On the other hand, nonservice ports are only visible within its part, and thus they are drawn within the internal region of its part. In Umple, private ports are considered nonservice (Snippet 2 - Line 5).

A nonservice port can still receive or send signals to or from other components via another relay port, which will act as an intermediary port.

When a port triggers a state machine event, it is considered a behaviour port (i.e. Snippet 2 - Lines 19-28).

A port cannot be nonservice and nonbehaviour at the same time. On the other hand, a service port can possibly be a behaviour or nonbehaviour port.

Associations are typically used to manage the number of port instances in a class. A replicated port means that this port can have multiple instances. When a port is connected to other ports, it is referred to as a wired port. An unwired port can still connect dynamically to other ports during runtime.

Table 2 summarises the different types of ports. Figure 5 shows ports of different types visualized using UmpleOnline. We follow the notations in specifications such as UML (OMG, 2011) and AUTOSAR (AUTOSAR, 2014).

When a port is a service port, it is drawn on the boundary of the composite structure; otherwise, it is drawn within the composite. Hence, all ports in Figure 5 are service ports, since they are all drawn on the boundary of their owning component.

![Figure 5: Visualization of different port types using Umple.](image-url)

We recognize wired ports if they are connected to other ports via connectors; Figure 4 for instance.

We distinguish between in and out ports using the crescent symbol, such that the open end of the crescent refers to the out port, and the closed end refers to the in port (Orabi et al., 2016).

At the moment, we do not visually distinguish between behaviour and nonbehaviour ports, since this will require parsing users’ code to check whether there are invocations to state events (Snippet 2 - Line 20). For future work, we will assess the necessity of supporting this feature.

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behavior</td>
<td>Triggers state machine events</td>
</tr>
<tr>
<td>Nonbehaviour</td>
<td>Does not propagates signals via state machines</td>
</tr>
<tr>
<td>Complex</td>
<td>Encompasses a number of attributes rather than a single attribute as in simple ports</td>
</tr>
<tr>
<td>Base</td>
<td>A port designed to send out signals</td>
</tr>
<tr>
<td>Conjugated</td>
<td>A port that also defines an inverse port that operates in the reverse manner</td>
</tr>
<tr>
<td>Service</td>
<td>Used to communicate between ports in its environment and ports in other environments; i.e. public or external</td>
</tr>
<tr>
<td>Nonservice</td>
<td>Only visible within its part; i.e. private</td>
</tr>
<tr>
<td>Replicated</td>
<td>Can have multiple instances.</td>
</tr>
<tr>
<td>In</td>
<td>Provides a service for other ports. It can be conjugated, replicated, or service</td>
</tr>
<tr>
<td>Out</td>
<td>Requires a service from other ports. It can be conjugated, replicated, or service</td>
</tr>
<tr>
<td>Wired</td>
<td>Means that a port is connected to other ports at the model level</td>
</tr>
<tr>
<td>Unwired</td>
<td>Means that a port is not connected to other ports at the model level, but possibly can still connect at runtime</td>
</tr>
</tbody>
</table>

---

252
5 CASE STUDY

In this section, we demonstrate a case study (Snippet 5) that follows component-based and event driven programming. In this case study (Figure 6), we report on a component-interface version of the one in (Orabi et al., 2016), such that the port interface is the

interaction point communicating with other components (Bauer, Hennicker, & Legay, 2013).

The example encompasses many features reported in this paper, such as compound ports, constraints, redefinition, and peer-to-peer communication.

![Figure 6: Structure diagram of the ping-pong example.](image)

Snippet 5: Case study.

```
interface IPinger{
    void ping(int pIn);
}

interface IPonger{
    void pong(int pOut);
}

class PingPongPort{
   [pongPort]
   active void pong(int num) {
       pongPort(num + 1);
       cout << "CMP 1 : Pong Out data = " << pOut1 << endl;
   }
}

class Pinger {
    isA IPinger;
    port PingPongPort pingPort;
}

class Ponger {
    isA IPonger;
    port PingPongPort pongPort;
    [pongPort, num < 10]
    active void pong(int num) {
       pongPort(num + 1);
       // Initiates communication in the constructor
       cmp1->ping(startValue);
    }
    cmp1.pongPort -> cmp2.pongPort;
}
```

6 EVALUATION

We evaluate Umple models that utilize different features explained in this paper. We use McCabe cyclomatic complexity (Kan, 2003) and lines of code (LOC) metrics.

Cyclomatic complexity counts logical conditions and is a proxy measure for the difficulty of maintaining and testing different parts of code logic, which requires cognitive efforts from developers. On the other hand, LOC is the standard measure of code size, and is useful because more code to read means more time is required for understanding it.

We calculate cyclomatic complexity based on Boolean satisfaction constraints, each of which is weighted as two. Hence, if there are two constraints, they will valued as four. The complexity ratio is calculated as $100 - (\text{Umple McCabe/McCabe}) \times 100$. We use LocMetrics tool (http://www.locmetrics.com/) to calculate the cyclomatic complexity and LOC of the generated code.

The generated code of an Umple model provides a built-in lightweight library that supports many features such as distributed communication and
multi-threading. We exclude the generated code of this library in order to avoid evaluation bias.

There is a high statistical significant (\(p < 0.0001\) and \( t = 7.4558 \)) in terms of lines of code reduction between Umple models and the generated C++ code (Figure 7). The average of reduction is 849 LOC and 96.4%, which is roughly constant meaning that it is independent from the model size.

The reduction in percentage for six Umple test models is shown in Figure 8. More details about the models used in our evaluation are in (Orabi, 2017). The reduction of cyclomatic complexity averages about 95.05%. Figure 9 is a doughnut chart showing the cyclomatic differences between C++ generated code and Umple in the test models. We can see the differences between Umple models and the generated C++ code. Note that for three of the models the cyclomatic complexity of the Umple code is zero.

A threat to validity of our evaluation is that the C++ user code might be different from the code generated by Umple. Some developers may argue that they may be able to come up with C++ code that is more compact. However, compact code might in fact be more obfuscated meaning that it could lead to yet more complexity.

7 CONCLUSIONS

Umple provides the major features required for component-based development. Our focus was on showing how development can be simplified using Umple. Specifically, we showed our protocol-free workflow, in which protocols are inferred from the definitions of ports, components, connectors, interfaces, and state events.

We discussed how we extended Umple to overcome many of the limitations in existing modeling tools that lack the support for many composite structure features or tend to have complicated workflows (Orabi et al., 2016) due to, for instance, needing to additionally define protocols.

We showed a component-interface-based case study, which used a number of composite structure features. The lines of C++ code generated from this use case is more than 2700, as compared to the Umple model that consists of only 54 lines.

In our evaluation, we used a number of Umple models that utilized the composite structure features discussed in this paper. We used both McCabe cyclomatic complexity and lines of code (LOC) metrics to assess the extent to which cognitive effort can be saved when using Umple as opposed to C++.

For future work, we will conduct an empirical evaluation, in which developers of different levels of expertise will use Umple. Based on which, we will be able to assess the usability of the composite structure features Umple.
REFERENCES


Component-based Modeling in Umple

255