Teaching programming methodology using Event B
Dominique Méry

To cite this version:

HAL Id: inria-00287231
https://inria.hal.science/inria-00287231v2
Submitted on 13 Nov 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Teaching programming methodology using Event B

Dominique Méry¹,²

LORIA
Université Henri Poincaré Nancy 1
Vandoeuvre les Nancy, France

Abstract
Event B is supported by the RODIN platform and provides a framework for teaching programming methodology based on the famous pre/post specifications, together with the refinement. We illustrate a methodology based on Event B and the refinement by developing Floyd’s algorithm for computing the shortest distances of a graph, which is based on an algorithm design technique called dynamic programming. The development is based on a paradigm identifying a non-deterministic event with a procedure call and by introducing control states. We discuss points related to our lectures at the university.

Keywords: Event B, refinement, sequential algorithm, pattern, proof, teaching, formal method, recursive procedure, development, correct by construction

1 Foreword
It is a great pleasure to thank Henri Habrias for his lectures on B at the University of Nantes. He understood the rôle of mathematics in the curriculum of computer scientists and the impact of the B methodology in industry and in education. Moreover, he was not only teaching notations but concepts that the young computer scientist, who has attended his lectures, will understand when the maturity will be there. He helps our community to propagate the two mammals of computer science, namely abstraction and refinement and is becoming now the King Henri. He was a solid support to B. It is a very modest exercise to discuss with Jean-Raymond, You and colleagues of our meeting in Nantes. Thanks Henri!

2 Introduction

Overview. Event B is supported by the RODIN platform and provides a framework for teaching programming methodology based on the famous pre/post specifications, together with the refinement. We illustrate a methodology based on Event B and the refinement by

¹ Email: mery@loria.fr
² Work of Dominique Méry is supported by grant No. ANR-06-SETI-015-03 awarded by the Agence Nationale de la Recherche.
developing algorithms for computing the shortest distances of a graph, which is based on an algorithm design technique called dynamic programming. Floyd’s algorithm is redeveloped and we add comments on the complexity of proofs and on the discovery of invariant; it should be considered as an illustration of a technique introduced in a joint paper with D. Cansell[7]. The development is based on a paradigm identifying a non-deterministic event with a procedure call and by introducing control states. We discuss points related to our lectures at different levels of the university. It is also a way to introduce a pattern used for developing sequential structured programs.

Programming methodology. The development of structured programs is carried out either using bottom-up techniques, or top-down techniques; we show how refinement and proof can be used to help in the top-down development of structured imperative programs. When a problem is stated, the incremental proof-based methodology of event B[6] starts by stating a very abstract model and further refinements lead to finer-grain event-based models which are used to derive an algorithm[3]. The main idea is to consider each procedure call as an abstract event of a model corresponding to the development of the procedure; generally, a procedure is specified by a pre/post specification and then the refinement process leads to a set of events, which are finally combined to obtain the body of the procedure. The refinement process can be considered as an unfolding of calls statements under preservation of invariants. It means that the abstraction corresponds to the procedure call and the body is derived using the refinement process. The refinement process may also use recursive procedures and supports the top-down refinement. The procedure call simulates the abstract event and the refinement guarantees the correctness of the resulting algorithm. A preliminary version[7] introduces ideas on a case study and provides an extended abstract of the current paper.

Proof-based Development. Proof-based development methods[5,1,13] integrate formal proof techniques in the development of software systems. The main idea is to start with a very abstract model of the system under development. Details are gradually added to this first model by building a sequence of more concrete events. The relationship between two successive models in this sequence is that of refinement[5,1]. The essence of the refinement relationship is that it preserves already proved system properties including safety properties and termination. A development gives rise to a number of, so-called, proof obligations, which guarantee its correctness. Such proof obligations are discharged by the proof tool using automatic and interactive proof procedures supported by a proof engine[4]. At the most abstract level it is obligatory to describe the static properties of a model’s data by means of an “invariant” predicate. This gives rise to proof obligations relating to the consistency of the model. They are required to ensure that data properties which are claimed to be invariant are preserved by the events of the model. Each refinement step is associated with a further invariant which relates the data of the more concrete model to that of the abstract model and states any additional invariant properties of the (possibly richer) concrete data model. These invariants, so-called gluing invariants are used in the formulation of the refinement proof obligations. The goal of a event B development is to obtain a proved model and to implement the correctness-by-construction[12] paradigm. Since the development process leads to a large number of proof obligations, the mastering of proof complexity is a crucial issue. Even if a proof tool is available, its effective power is limited
by classical results over logical theories and we must distribute the complexity of proofs
over the components of the current development, e.g. by refinement. Refinement has the
potential to decrease the complexity of the proof process whilst allowing for traceability
of requirements. The price to pay is to face possibly complex mathematical theories and
difficult proofs. The re-use of developed models and the structuring mechanisms available
in B help in decreasing the complexity.

Organization of the paper. Section 2 introduces definitions related to the
methodology and details for representing the pattern used for designing the algorithm. Section 3 describes the development of the shortest-path problem and the relationship between
models and programs; it illustrates the methodology for developing structured programs.
Section 4 develops Floyd’s algorithm using the same pattern and discusses issues related to
the implementation in the C programming language. Finally, we conclude our work in the
last section.

3 The modelling framework

We do not recall concepts of the Event B modelling language developed by J.-R. Abrial[2,6];
we sketch the general methodology we are applying. The ingredients for describing the
modelling process based on events and model refinement can be found in [2,6]. We assume
that the goal is to solve a given problem described by a semi-formal mathematical text and
we assume that the problem is defined by a precondition and a postcondition[13]. The
modelling process starts by identifying the domain of the problem and it is expressed using
the concept of CONTEXT. A CONTEXT PB (see Figure 1) states the theoretical notions
required to be able to express the problem statement in a formal way. The CONTEXT PB
declares

• a domain D which is the global set of possible values of the current system.
• a list of constants x, which is specifying the input of the system under development,
  P, which is the set of values for x defining the precondition, and Q, which is a binary
  relation over D defining the postcondition of the problem.
• a list of axioms assigns types to constants and adds knowledges to the RODIN environ-
  ment; for instance, the axiom 5 states that there is always a solution y, when the input
  value x satisfies the precondition P.

A CONTEXT may include a clause THEOREMS containing properties derivable in the
theory defined by sets, contants and axioms; theorems are discharged using the proof as-
assistant of the tool RODIN. The underlying language is a set-theoretical language partially
given in Table 1. When an expression E is given, a well-definedness condition is generated
by the tool; this point llows us to check that some side conditions are true. For instance, the
expression \( f(x) \) generates a condition as \( x \in \text{dom}(f) \).

The first model provides the declaration of the procedure call. Variables y are call-by-
reference parameters, constants x are call-by-value parameters and carrier sets s are used
to type informations and also for defining a generic procedure:
MÉRY

**CONTEXT PB**

**SETS**

\[ D \]

**CONSTANTS**

\( x, P, Q \)

**AXIOMS**

\[ 
axm_1 : x \in D \quad / \quad x \text{ belongs to a general set of the problem domain} \\
axm_2 : P \subseteq D \quad / \quad P \text{ is a set defining the precondition} \\
axm_3 : Q \subseteq D \times D \quad / \quad Q \text{ is a binary relation over } S \text{ defining the postcondition} \\
axm_4 : x \in P \quad / \quad x \text{ is supposed to satisfy the precondition } P \\
axm_5 : \forall a . a \in P \Rightarrow ( \exists b . a \mapsto b \in Q ) \quad / \quad \text{there is at least one solution for each data } x \text{ satisfying the precondition } P \\
\]

**END**

---

**Fig. 1. Context for modelling the problem PB**

**Table 1**

<table>
<thead>
<tr>
<th>Name</th>
<th>Syntax</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binary relation</td>
<td>( s \leftrightarrow t )</td>
<td>( \mathcal{P}(s \times t) )</td>
</tr>
<tr>
<td>Composition of relations</td>
<td>( r_1 ; r_2 )</td>
<td>( { x, y</td>
</tr>
<tr>
<td>Inverse relation</td>
<td>( r^{-1} )</td>
<td>( { x, y</td>
</tr>
<tr>
<td>Domain</td>
<td>( \text{dom}(r) )</td>
<td>( { a</td>
</tr>
<tr>
<td>Range</td>
<td>( \text{ran}(r) )</td>
<td>( \text{dom}(r^{-1}) )</td>
</tr>
<tr>
<td>Identity</td>
<td>( \text{id}(s) )</td>
<td>( { x, y</td>
</tr>
<tr>
<td>Restriction</td>
<td>( s \triangleleft r )</td>
<td>( \text{id}(s) ; r )</td>
</tr>
<tr>
<td>Co-restriction</td>
<td>( s \triangleright s )</td>
<td>( r ; \text{id}(s) )</td>
</tr>
<tr>
<td>Anti-restriction</td>
<td>( s \triangleleftleft r )</td>
<td>( ( \text{dom}(r) - s ) \triangleleft r )</td>
</tr>
<tr>
<td>Anti-co-restriction</td>
<td>( r \trianglerightleft s )</td>
<td>( r \triangleright ( \text{ran}(r) - s ) )</td>
</tr>
<tr>
<td>Image</td>
<td>( r[w] )</td>
<td>( \text{ran}(w \triangleleft r) )</td>
</tr>
<tr>
<td>Overriding</td>
<td>( q \triangleleftleft r )</td>
<td>( ( \text{dom}(r) \triangleleftleft q ) \cup r )</td>
</tr>
<tr>
<td>Partial Function</td>
<td>( s \mapsto t )</td>
<td>( { r</td>
</tr>
</tbody>
</table>

---

**procedure call**(\( x \); var \( y \))

**precondition** \( y = y_0 \land Init(y_0, x, D) \land \sim P(x) \)

**postcondition** \( \sim Q(x, y) \)
Figure 2 describes the complete model for the problem PB; it is expressed by a generic procedure stating the pre/post-specification. The term *procedure* can be substituted by the term *method*. The current status of the development can be represented as follow:

\[
\text{call}(x, y) \xrightarrow{\text{call-as-event}} \text{PREPOST} \xrightarrow{\text{SEES}} \text{PB}
\]

The statement of a given problem in the Event B modelling language is relatively direct, as long as we are able to express the mathematical underlying theory using the mechanism of contexts. The existence of a solution \( y \) for each value \( x \) is assumed to be an axiom; however, it would be better to derive the property as a theorem and it means that we should develop a way to validate axioms to ensure the consistency of the underlying theory.

The next section illustrates the technique used for developing new algorithms. We think that it is a good way to teach the design of algorithms. HOARE logic[10] provides a very interesting framework for dealing with specifications an development and our work shows how the ingredients of HOARE logic can be used to provide a general framework for developing sequential programs correct by construction. Event B and the RODIN platform can be used to teach basic notions like pre and postconditions, invariants, verification and finally design-by-contract.

**Teacher's note:** The challenge of the teacher is to relate the Event B notations to the notations of the programming language. We have used the Event B notations in lectures on fixed-point theory and on the explanation of sequential algorithms. It is then clear that we should provide more systematic rules for deriving algorithms. The management of definitions using a tool, like RODIN, helps students to understand why a function call like \( f(x) \) generates conditions like \( x \in \text{dom}(f) \). Nobody can cheat with the tool. Moreover, when a tool is available for a free download, it is really a teatchermate.
4 The Shortest Path Problem

4.1 Summary of the problem

Floyd’s algorithm computes the shortest distances of a graph and is based on an algorithmic design technique called dynamic programming: simpler subproblems are first solved before the full problem is solved. It computes a distance matrix from a cost matrix: the costs of the shortest path between each pair of vertices are in $O(|V|^3)$ time.

**Teacher’s note:** In the case of Floyd’s algorithm, there is a mathematical definition of the matrix we have to compute from a starting state defining the initial basic link between nodes with cost. The function is called $d$ and should be first defined in a context of the problem.

The set of nodes $N$ is $1..n$, where $n$ is a constant value and the graph is simply represented by the distance function $d$ ($d \in N \times N \times N \rightarrow \mathbb{N}$) and when the function is not defined, it means that there is no vertex between the two nodes. The relation of the graph is defined as the domain of the function $d$. $n$ is clearly greater than 1 and it means that the set of nodes is not empty.

The distance function $d$ is defined inductively from bottom to top according to the dynamic programming principle and the next axioms define this function:

- **axm1**: $d \in N \times N \times N \rightarrow \mathbb{N}$

- **axm5**: $\forall i \cdot i \in N \Rightarrow 0 \mapsto i \mapsto i \in \text{dom}(d) \land d(0 \mapsto i \mapsto i) = 0$

- **axm6**: $\forall i, j, k$.

  $\left( k \mapsto i \mapsto j \in \text{dom}(d) \\
  \land (k \mapsto i \mapsto k \notin \text{dom}(d) \lor k \mapsto j \notin \text{dom}(d)) \right)$

  $\Rightarrow$

  $\left( k \mapsto i \mapsto j \in \text{dom}(d) \land d(k \mapsto i \mapsto j) = d(k \mapsto i \mapsto j) \right)$

- **axm7**: $\forall i, j, k$.

  $\left( k \mapsto i \mapsto j \in \text{dom}(d) \\
  \land k \mapsto i \mapsto k \in \text{dom}(d) \\
  \land k \mapsto k \mapsto j \in \text{dom}(d) \right)$

  $\land d(k \mapsto i \mapsto j) \leq d(k \mapsto i \mapsto k) + d(k \mapsto i \mapsto j)$

  $\Rightarrow$

  $\left( k \mapsto i \mapsto j \in \text{dom}(d) \\
  \land d(k \mapsto i \mapsto j) = d(k \mapsto i \mapsto j) \right)$
The optimality property is derived from the definition of \( d \) itself, since it starts by defining bottom elements and applies an optimal principle summarized as follows: 

\[
D_{i+1}(a, b) = Min(D_i(a, b), D_i(a, i+1) + D_i(i+1, b))
\]

and means that the distances in \( D_i \) represent paths with intermediate vertices smaller than \( i \); \( D_{i+1} \) is defined by comparing new paths including \( i+1 \). \( D_i \) is defined by a partial function over \( N \times N \times N \). The partiality of \( d \) leads to some possible problems for computing the minimum and when at least one term is not defined, we should define a specific definition for the resulting term. Floyd’s algorithm provides an algorithmic process for obtaining a matrix of all shortest possible paths with respect to a given initial matrix representing links between nodes together with their distance. Our first attempt was based on the computation of a shortest path between to given nodes \( a \) and \( b \). The resulting matrix is called \( R \) and a boolean variable \( FD \) tells us if the shortest path exists. By the way, this first attempt is not the strict Floyd’s algorithm but it will use the same principle of computation for the resulting matrix \( R \).

The first step defines the context of the problem and the context is validated by the RODIN platform[14]. We decide to design an algorithm which is computing the value of the shortest path between two given nodes but using the same principle than Floyd’s algorithm.
Teacher's note: The validation of the context SHORTESTPATH0 helps us to define carefully the function \( d \). The translation of mathematical properties is made easier by the notion of partial function. The expression \( D_{i+1}(a, b) = \min(D_i(a, b), D_i(a, i+1) + D_i(i+1, b)) \) hides possible underfinedness and generally the non-existence of an edge between two nodes is defined by an extra value like \( \infty \). We have to compute the following value \( \lambda_{i, j} \in \mathbb{N} \cdot d(l \mapsto i \mapsto j) \) but the \( \lambda \) notation is not directly usable in the B notations. However, we are computing in fact the value of \( d \) for the triple \( l \mapsto i \mapsto j \) because it seems to be simpler to state.

4.2 Writing the function call

The first model provides the declaration of the procedure \( \text{shortestpath} \). Variables \( D \) and \( FD \) are call-by-reference parameters, constants \( l, a, b, D \) are call-by-value parameters:

\[
\begin{align*}
\text{procedure} & \quad \text{shortestpath}(l, a, b, G; \text{var} \quad D, FD) \\
\text{precondition} & \quad G = d_0 \land FD = \text{FALSE} \land l > 0 \land a \in \mathbb{N} \land b \in \mathbb{N} \\
\text{postcondition} & \quad (FD = \text{true} \Rightarrow D = d(l, a, b))
\end{align*}
\]

We apply the \textit{Call as Event principle} and we have to define a new model called SHORTESTPATH1, which is defining an event corresponding to the action of calling the procedure.

Teacher's note: The event is considered as a function call; we can explain at this time that the event is triggered because the guard is true. It is not a precondition.

The new model SHORTESTPATH1 is using definitions of the context SHORTESTPATH0. The event FLOYDKO models the fact that the call of floyd is returning a value FALSE for FD: there is no path between \( a \) and \( b \). The event FLOYDOK returns the value TRUE for FD and the value of the minimal path from \( a \) to \( b \). The two events are also interpreted by a procedure which is called with respect to the existence of a path.
MÉRY

MACHINE SHORTESTPATH1 SEES SHORTESTPATH0

VARIABLES
\[ D \hspace{1cm} FD \]

INVARIANTS
inv1 : \( D \in \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N} \)
inv2 : \( FD \in \text{BOOL} \)

EVENTS

INITIALISATION
BEGIN
act1 : \( D : \{ D' \in \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N} \land (\forall i, j : 0 \rightarrow i \rightarrow j \in \text{dom}(d) \Rightarrow i \rightarrow j \in \text{dom}(D') \land D'(i \rightarrow j) = d(0 \rightarrow i \rightarrow j))\} \)
act2 : \( FD := \text{FALSE} \)
END

EVENT shortestpathOK
WHEN
\( \text{grd}1 : l \rightarrow a \rightarrow b \in \text{dom}(d) \)
THEN
act1 : \( D(a \rightarrow b) := d(l \rightarrow a \rightarrow b) \)
act2 : \( FD := \text{TRUE} \)
END

EVENT shortestpathKO
WHEN
\( \text{grd}1 : l \rightarrow a \rightarrow b \notin \text{dom}(d) \)
THEN
act1 : \( FD := \text{FALSE} \)
END

END

Now, we have two events really non-deterministic, since they are defined using the constant \( d \) which should be computed in fact!. The solution is to refine the model SHORTESTPATH1 into a new model SHORTESTPATH2 which reduces non-determinism.

Teacher’s note: It is very important to explain the difference between a flexible [11] variable and a rigid variable. Rigid variable like \( d \) denotes values which are defined as mathematical static objects and flexible variables denotes a name which is assigned to a value depending on the current state.

4.3 Refining the procedure call

The main idea is to unfold the calls or to refine the events to get a model which is closer to an algorithm. We introduce several new variables:

- \( D \) and \( FD \) are both variables of the models SHORTESTPATH1 and SHORTESTPATH2.
- \( c (\text{inv}1 : c \in C) \) expresses the control flow and the possible values of \( c \) are in the set \( C (\text{axm}15 : C = \{ \text{start, end, step1, step2, step3, finalstep} \}) \).
- \( D1, D2 \) and \( D3 \) are three variables storing the values required for computing the next value of \( D \) at a given step; the values may be undefined and the undefinedness is controlled by the three variables \( FD1, FD2 \) and \( FD3 \). Variables are typed according to the following part of the invariant:
  \[ \text{inv2} : D1 \in \mathbb{Z} \]
  \[ \text{inv3} : D2 \in \mathbb{Z} \]
MÉRY

· \(\text{inv} 4 : D3 \in \mathbb{Z}\)

· \(\text{inv} 5 : F D1 \in \text{BOOL}\)

· \(\text{inv} 6 : F D2 \in \text{BOOL}\)

· \(\text{inv} 7 : F D3 \in \text{BOOL}\)

We do not give more details for the invariant and we will give later the details of the invariant of the current model. First we give the different events of the model \text{SHORT-ESTPATH2}.

The event \text{INITIALISATION}
is simply setting the variables as follows: \(\text{act} 1 : D := D0, \text{act} 2 : F D := \text{FALSE}, \text{act} 3 : F D1 := \text{FALSE}, \text{act} 4 : F D2 := \text{FALSE}, \text{act} 5 : F D3 := \text{FALSE}, \text{act} 6 : D1 \in \mathbb{Z}, \text{act} 7 : D2 \in \mathbb{Z}, \text{act} 8 : D3 \in \mathbb{Z}, \text{act} 10 : c := \text{start}\).

Since \(d(0 \leftrightarrow i \leftrightarrow j)\) models the existence of an elementary path from \(i\) to \(j\), \(D0\) is defined by the following axioms:

- \(\text{axm} 12 : D0 \in \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}\)

- \(\text{axm} 13 : \text{dom}(D0) = \{i \leftarrow j | 0 \leftarrow i \leftarrow j \in \text{dom}(d)\}\)

- \(\text{axm} 14 : \forall i, j \cdot i \leftarrow j \in \text{dom}(D0) \Rightarrow D0(i \leftarrow j) = d(0 \leftarrow i \leftarrow j)\)

Now, we can introduce refinement of existing events of \text{SHORTESTPATH1} and new events which are not in the abstraction.

4.3.1 Refining events of \text{SHORTESTPATH1}

First, we give elements for competing the invariant; the typing informations can be completed as follows and they correspond to an analysis of the definition of \(d\). We introduce a new variable \(c\) which is expressing the control state and whose possible values are given by the set \(C\): \(C = \{\text{start, end, step1, step2, step3, finalstep}\}\). We summarize the different steps for computing \(D\).

\[\text{Teacher's note: Using a graphical notation helps to communicate the meaning of control assertions. The steps of the algorithm appear. Moreover, steps provide a guide for defining the invariant which is based on the construction of } d.\]
The invariant is based on the decomposition into steps and each step analyses the definition of values required for computing the minimum of $D_1$ and $D_2+D_3$. The invariant should take into account the definedness of these values and the tool helps us to complete the invariant.

The analysis step provides a decision depending on the values of $D_1$, $D_2$ and $D_3$, if they are defined. The boolean expression $FD_1 \land (FD_2 \lor FD_3)$ is the key for updating $D(a \mapsto b)$ and it is triggered, when the control is finalstep.

When the control is at start:

- when $l$ is initially equal to 0, $D$ and $d$ are equal too; $D$ is defined when $d$ is defined and reciprocally:

  - **inv20**: $c = \text{start} \land a \mapsto b \notin \text{dom}(D) \land l = 0 \Rightarrow 0 \mapsto a \mapsto b \notin \text{dom}(d)$

  - **inv8**:

    \[
    \begin{align*}
    & c = \text{start} \\
    & \land a \mapsto b \in \text{dom}(D) \\
    & \land l = 0 \\
    \Rightarrow \\
    & \left(0 \mapsto a \mapsto b \in \text{dom}(d) \\
    & \land D(a \mapsto b) = d(0 \mapsto a \mapsto b) \right)
    \end{align*}
    \]

- when $l$ is not equal to 0 and there is no path from $a$ to $b$ with intermediates whose numbers is smaller than $l - 1$, $a \mapsto b$ is not in $D$. 

Teacher’s note: The expression $D_{i+1}(a, b) = \text{Min}(D_i(a, b), D_i(a, i+1) + D_i(i+1, b))$ should be carefully analysed and it allows us to derive specific conditions for structuring the algorithm.
When the control is at end:
If the control is at end, the invariant enumerates the different cases for the resulting computation. The variable $D$ should contain the values correspondin to $l$.

\begin{itemize}
  \item $inv_{12}: c = end \land FD = TRUE \Rightarrow \left( a \mapsto b \in dom(D) \land l \mapsto a \mapsto b \in dom(d) \land D(a \mapsto b) = d(l \mapsto a \mapsto b) \right)$
  \item $inv_{14}: c = end \land FD = FALSE \Rightarrow a \mapsto b \notin dom(D) \land l \mapsto a \mapsto b \notin dom(d)$
  \item $inv_{18}: c = end \land l \mapsto a \mapsto b \notin dom(d) \Rightarrow FD = FALSE$
  \item $inv_{19}: c = end \land a \mapsto b \notin dom(D) \Rightarrow FD = FALSE$
  \item $inv_{21}: c = end \land a \mapsto b \in dom(D) \Rightarrow FD = TRUE$
\end{itemize}

When the control is at finalstep:
The invariant states that the variables \( FD_1, FD_2 \) and \( FD_3 \) are related to the definition of the expression \( \min(D(a, b), D(a, l) + D(l, b)) \). \( \min(D(a, b), D(a, l) + D(l, b)) \) is defined, if, and only, if \( FD_1 \land (FD_2 \lor FD_3) \). The invariant explores the different cases for the definition of \( D \) for the given pairs. Moreover, the values are stored in the variables \( D_1, D_2 \) and \( D_3 \) when defined.

\[
c = \text{finalstep} \land FD_3 = \text{TRUE}
\]

- \text{inv11}:
  \[
  l - 1 \mapsto l \mapsto b \in \text{dom}(d) \land D_3 = d(l - 1 \mapsto l \mapsto b)
  \]

\[
c = \text{finalstep} \land FD_1 = \text{TRUE}
\]

- \text{inv15}:
  \[
  l - 1 \mapsto a \mapsto b \in \text{dom}(d) \land D_1 = d(l - 1 \mapsto a \mapsto b)
  \]

\[
c = \text{finalstep} \land FD_2 = \text{TRUE}
\]

- \text{inv16}:
  \[
  l - 1 \mapsto a \mapsto l \in \text{dom}(d) \land D_2 = d(l - 1 \mapsto a \mapsto l)
  \]

\[
\left(\begin{array}{c}
c = \text{finalstep} \\
\land FD_1 = \text{FALSE} \\
\land (FD_2 = \text{FALSE} \lor FD_3 = \text{FALSE})
\end{array}\right) \\
\Rightarrow \\
\left(\begin{array}{c}
l \mapsto a \mapsto b \notin \text{dom}(d) \\
\land a \mapsto b \notin \text{dom}(D)
\end{array}\right)
\]

- \text{inv13}:

\[
c = \text{finalstep} \land FD_3 = \text{FALSE} \Rightarrow l - 1 \mapsto l \mapsto b \notin \text{dom}(d)
\]

- \text{inv24}:

\[
c = \text{finalstep} \land FD_2 = \text{FALSE} \Rightarrow l - 1 \mapsto a \mapsto l \notin \text{dom}(d)
\]

- \text{inv27}:

\[
c = \text{finalstep} \land FD_1 = \text{FALSE} \Rightarrow l - 1 \mapsto a \mapsto b \notin \text{dom}(d)
\]

- \text{inv29}:
The diagram shows that shortestpath is made up of three steps.

When the control is in \{step1, step2, step3\}:

- When the control is in \{step1, step2, step3\}, since
  \[ \text{inv28} : c \neq \text{start} \land c \neq \text{end} \Rightarrow l \neq 0 \]
  \( l \) is not equal to 0.

- When the control is at step1, \( l \) is not equal to 0. There are two conditions for the unde-
  finedness of \( D \) in relationship to \( d \).
  \[ \text{inv33} : \Rightarrow \quad a \mapsto b \notin \text{dom}(D) \]

- When the control is in step2, either the evaluation of \( D1 \) is successful or not.
  \[ \text{inv9} : c = \text{step2} \land FD1 = \text{TRUE} \Rightarrow l - 1 \mapsto a \mapsto b \in \text{dom}(d) \land D1 = d(l - 1 \mapsto a \mapsto b) \]
\[ \text{inv22} : c = step2 \land FD1 = FALSE \Rightarrow l - 1 \mapsto a \mapsto b \notin \text{dom}(d) \]
\[ c = step2 \land FD1 = FALSE \land l - 1 \mapsto l \mapsto b \notin \text{dom}(d) \]

\[ \text{inv32} : \Rightarrow \]
\[ a \mapsto b \notin \text{dom}(D) \]

\[ c = step2 \land l - 1 \mapsto a \mapsto l \notin \text{dom}(d) \land FD1 = FALSE \]

\[ \text{inv35} : \Rightarrow \]
\[ a \mapsto b \notin \text{dom}(D) \]

- When the control is in \textit{step3}, either the evaluation of \(D2\) is successful or not.

\[ c = step3 \land FD2 = TRUE \]

\[ \text{inv10} : \Rightarrow \]
\[ l - 1 \mapsto a \mapsto l \in \text{dom}(d) \land D2 = d(l - 1 \mapsto a \mapsto l) \]

\[ c = step3 \land FD1 = TRUE \]

\[ \text{inv17} : \Rightarrow \]
\[ l - 1 \mapsto a \mapsto b \in \text{dom}(d) \land D1 = d(l - 1 \mapsto a \mapsto b) \]

\[ \text{inv23} : c = step3 \land FD2 = FALSE \Rightarrow l - 1 \mapsto a \mapsto l \notin \text{dom}(d) \]

\[ \text{inv25} : c = step3 \land FD1 = FALSE \Rightarrow l - 1 \mapsto a \mapsto b \notin \text{dom}(d) \]

\[ \text{inv26} : c \neq \text{finalstep} \land c \neq \text{end} \land 0 \mapsto a \mapsto b \notin \text{dom}(d) \Rightarrow a \mapsto b \notin \text{dom}(D) \]

\[ \text{inv30} : c = step3 \land FD1 = FALSE \land FD2 = FALSE \Rightarrow a \mapsto b \notin \text{dom}(D) \]

\[ c = step3 \land FD1 = FALSE \land l - 1 \mapsto l \mapsto b \notin \text{dom}(d) \]

\[ \text{inv31} : \Rightarrow \]
\[ a \mapsto b \notin \text{dom}(D) \]
Refining shortestpathOK

Now, we define each transition between the different steps according to the invariant. We consider several possible cases depending on \( l \) and other conditions. When the value of \( l \) is 0 and when \( D \) is defined for the pair \( a \mapsto b \), it means that there is a path between \( a \) and \( b \) without any intermediate node. It is the basic case and one returns the value TRUE for \( F D \). The control is set to \textit{end}, since the procedure is completed:

\[
\begin{align*}
\text{EVENT shortestpathOK} & \\
\text{REFINES} & \text{shortestpathOK} \\
\text{WHEN} & \\
\quad \text{grd2} : l = 0 \\
\quad \text{grd1} : a \mapsto b \in \text{dom}(D) \\
\quad \text{grd3} : c = \text{start} \\
\text{THEN} & \\
\quad \text{act2} : F D := \text{TRUE} \\
\quad \text{act3} : c := \text{end} \\
\text{END}
\end{align*}
\]

When the control expresses the accessibility of the last control point (\( c = \text{finalstep} \)) and when the three values \( D_1, D_2 \) and \( D_3 \) are defined and satisfy the condition \( D_1 \leq D_2 + D_3 \), we can update \( D \) in \( a \mapsto b \) by \( D_1 \). In fact, the value is not modified. The control is set to the final control point called \textit{end}. There is a path and \( F D \) is set to TRUE.

\[
\begin{align*}
\text{EVENT shortestpathcallOKmin} & \\
\text{REFINES} & \text{shortestpathOK}
\end{align*}
\]
WHEN
grd1 : FD1 = TRUE ∧ FD2 = TRUE ∧ FD3 = TRUE
grd2 : D1 ≤ D2 + D3
grd3 : c = finalstep
THEN
act1 : D(a ↦→ b) := D1
act2 : FD := TRUE
act3 : c := end
END

The next case is stating that there is a new path from \(a\) to \(b\), which is shortest than the current one (\(grd3 : D1 > D2 + D3\)) and we should update \(D\) by the new value \(D2 + D3\).

---

EVENT shortestpathcallOKmax
REFINES shortestpathOK

WHEN
grd1 : FD1 = TRUE ∧ FD2 = TRUE ∧ FD3 = TRUE
grd2 : c = finalstep
grd3 : D1 > D2 + D3
THEN
act1 : D(a ↦→ b) := D2 + D3
act2 : c := end
act3 : FD := TRUE
END

The next possible case is that the value \(D1\) is not defined; it means that there is not yet a path from \(a\) to \(b\) and we have discovered that there is a node which can be reached from \(a\) and which can reach \(b\). Hence, the variable \(D\) is defined in \(a ↦→ b\) by the value \(D2 + D2\).

---

EVENT shortestpathFD2FD3
REFINES shortestpathOK

WHEN
grd1 : c = finalstep
grd2 : FD1 = FALSE ∧ FD2 = TRUE ∧ FD3 = TRUE
THEN
act1 : D(a ↦→ b) := D2 + D3
act2 : FD := TRUE
act3 : c := end
END
Finally, when either $D_2$ or $D_3$ is not defined, the value of $D$ is not modified and remains equal to $D_1$.

**EVENT** shortestpathFD1  
**REFINES** shortestpathOK  
**WHEN**  
\[ \text{grd}_1 : c = \text{finalstep} \]  
\[ \text{grd}_2 : FD_1 = \text{TRUE} \land (FD_1 = \text{FALSE} \lor FD_2 = \text{FALSE}) \]  
**THEN**  
\[ \text{act}_1 : D(a \rightarrow b) := D_1 \]  
\[ \text{act}_2 : c := \text{end} \]  
\[ \text{act}_3 : FD := \text{TRUE} \]  
**END**

The refinement of abstract events should be completed by events which compute the values $D_1$, $D_2$ and $D_3$.

**Refining shortestpathKO**  
We consider several possible cases depending on $l$ and other conditions.

When the value of $l$ is 0 and when $D$ is not defined for the pair $a \rightarrow b$, it means that there is no elementary path between $a$ and $b$. It is the basic case and one returns the value FALSE for $FD$. The control is set to end, since the procedure is completed:
EVENT shortestpathKO
REFINES shortestpathKO
WHEN
  grd2 : l = 0
  grd1 : a ↦ b ∉ dom(D)
  grd3 : c = start
THEN
  act1 : FD := FALSE
  act2 : c := end
END

When the value of \( l \) is not 0 and when \( D1 \) is not defined and either \( D2 \) is not defined, or \( D3 \) is not defined, for the pair \( a ↦ b \), it means that there is no path between \( a \) and \( b \). One returns the value FALSE for \( FD \). The control is set to end, since the procedure is completed:

EVENT shortestpathKOelse
REFINES shortestpathKO
WHEN
  grd1 : c = finalstep
  grd2 : FD1 = FALSE ∧ (FD2 = FALSE ∨ FD2 = FALSE)
THEN
  act1 : c := end
  act2 : FD := FALSE
END

4.3.2 Introducing new events in SHORTESTPATH2
The first new event models the calling step of the procedure floyd and it transfers the control to the control point step1.

EVENT shortestpathcallone
WHEN
  grd1 : l > 0
  grd2 : c = start
THEN
  act1 : c := step1
END

Now, we consider the three steps for computing \( D1, D2 \) and \( D3 \).
Calling the procedure \texttt{floyd} for evaluating $D_1$ and $F D_1$

The event \texttt{shortestpathcalltwook} simulates the procedure for computing $D_1$, which is $d(l - 1 \mapsto a \mapsto b)$ and which is successfully computed, since $F D_1$ is TRUE. The event \texttt{shortestpathcalltwoko} simulates the procedure for computing $D_1$, which is $d(l - 1 \mapsto a \mapsto b)$ and which is unsuccessfully computed, since $F D_1$ is FALSE.

\begin{verbatim}
EVENT floydcalltwook
WHEN
  grd1 : c = step1
  grd2 : l - 1 \mapsto a \mapsto b \in \text{dom}(d)
THEN
  act1 : D1 := d(l - 1 \mapsto a \mapsto b)
  act2 : FD1 := TRUE
  act3 : c := step2
END

EVENT shorttestpathcalltwoko
WHEN
  grd1 : l - 1 \mapsto a \mapsto b \notin \text{dom}(d)
  grd2 : c = step1
THEN
  act1 : FD1 := FALSE
  act2 : c := step2
END
\end{verbatim}

Calling the procedure \texttt{floyd} for evaluating $D_2$ and $F D_2$

The event \texttt{shortestpathcallthreeok} simulates the procedure for computing $D_2$, which is $d(l - 1 \mapsto a \mapsto l)$ and which is successfully computed, since $F D_2$ is TRUE. The event \texttt{shortestpathcallthreeko} simulates the procedure for computing $D_2$, which is $d(l - 1 \mapsto a \mapsto l)$ and which is unsuccessfully computed, since $F D_2$ is FALSE.

\begin{verbatim}
EVENT shortestpathcallthreeok
WHEN
  grd1 : c = step2
  grd2 : l - 1 \mapsto a \mapsto l \in \text{dom}(d)
THEN
  act1 : D2 := d(l - 1 \mapsto a \mapsto l)
  act2 : FD2 := TRUE
  act3 : c := step3
END

EVENT shortestpathcallthreeko
WHEN
  grd1 : c = step2
  grd2 : l - 1 \mapsto a \mapsto l \notin \text{dom}(d)
THEN
  act1 : c := step3
  act2 : FD2 := FALSE
END
\end{verbatim}

Calling the procedure \texttt{floyd} for evaluating $D_3$ and $F D_3$

The event \texttt{shortestpathcallfourok} simulates the procedure for computing $D_3$, which is $d(l - 1 \mapsto l \mapsto b)$ and which is successfully computed, since $F D_3$ is TRUE. The event \texttt{shortestpathcallfourko} simulates the procedure for computing $D_3$, which is $d(l - 1 \mapsto l \mapsto b)$ and which is unsuccessfully computed, since $F D_3$ is FALSE.
4.4 Producing the shortestpath procedure

The shortestpath procedure can be derived from the list of events of the model SHORT-ESTPATH2 and we structure events into conventional programming structures like \texttt{while} or \texttt{if} statements. J.-R. Abrial\cite{Abrial} has proposed several rules for producing algorithmic statements. The next diagram gives the complete description of the process we have followed:

The procedure header is \texttt{shortestpath(l, a, b, G, D, FD)} and the text of the procedure is given by the algorithms 1 and 2.
**Algorithm 1:** Algorithm Version 1

- **precondition:** \( l \in 1..n \)
- **postcondition:** \( D, F_D \)
- **local variables:** \( F_{D1}, F_{D2}, F_{D3} \in BOOL \)

\[
\begin{align*}
FD & := FALSE; \\
F_{D1} & := FALSE; \\
F_{D2} & := FALSE; \\
F_{D3} & := FALSE;
\end{align*}
\]

if \( l = 0 \) then

- if \( (a, b) \in \text{dom}(D) \) then
  
  \[
  \begin{align*}
  FD & := TRUE; \\
  R & := D[a, b];
  \end{align*}
  \]

- else
  
  \[
  FD := FALSE;
  \]

else

\[
\begin{align*}
& floyd(l - 1, a, b, D, F_{D1}); floyd(l - 1, a, l, D, F_{D2}); floyd(l - 1, l, b, D, F_{D3}); \\
& \text{case } F_{D1} \land F_{D2} \land F_{D3}
  
  & \quad \text{if } D_1 < D_2 + D_3 \text{ then}
  
  & \quad \quad \quad R := D_1; \\
  & \quad \text{else}
  
  & \quad \quad \quad R := D_2 + D_3; \\
  & \quad \quad \quad \quad \quad FD := TRUE;
  \]

\[
\begin{align*}
& \text{case } F_{D1} \land (\neg F_{D2} \lor \neg F_{D3}) \\
& \quad R := D_1; \\
& \quad FD := TRUE;
  \]

\[
\begin{align*}
& \text{case } \neg F_{D1} \land (F_{D2} \land F_{D3}) \\
& \quad R := D_2 + D_3; \\
& \quad FD := TRUE;
  \]

\[
\begin{align*}
& \text{case } \neg F_{D1} \land (\neg F_{D2} \lor \neg F_{D3}) \\
& \quad FD := FALSE;
  \]

\]

The two next frames are containing C codes produced for the two algorithms 4.4 and 4.4; we have produced the C codes by hand and we have forgotten that C arrays starts by 0 and it means that our initial calls were wrongly written. It is clear that we need a way to produce codes in a mechanized way. Moreover, there are some conditions to check and
some interactions to manage with the user to help in choices.

**Teacher’s note:** It is the time to recall that we are planning to use a real programming language and that we should represent abstract objects by concrete objects. It would be better to add informations on the integers of computer scientists and it is easy to add the constraint.

```c
/* N = 1..n-1 */
void shortestpath (int l, int a, int b, int g[][n], int *D, int *FD)
{
    int D1, D2, D3, FD1, FD2, FD3;
    *FD = 0; FD1=0;FD2=0;FD3=0;
    if (l==0)
    {
        if (g[a][b] != NONE) {
            *FD = 1; *D = g[a][b];
        }
    }
    else
    {
        shortestpath(1..a-1..b..g..&D1..&FD1);
        shortestpath(1..a..b-1..g..&D2..&FD2);
        shortestpath(1..a..b-1..g..&D3..&FD3);
        if ((FD1 == 1 && (FD2==1 && FD3==1))
        {
            if (D1 < D2+D3)
            {
                *D= D1;
            }
            else
            {
                *D=D2+D3;
            }
            *FD = 1;
        }
        else if (FD1==1 && ( FD2==0 || FD3==0))
        {
            *D= D1;
        }
        else if (FD1==0 && ( FD2 == 1 & FD3==1)) (*D=D2+D3; *FD=1;)
        else /* (FD1==0 & ( FD2==0 || FD3==0)) */ (*FD = 0;)
    }
}
```

The complete development has a cost related to proof obligations. The refinement generates 493 proof obligations and 328 proof obligations were automatically discharged. 165 proof obligations were manually discharged with minor interactions.
Algorithm 2: Algorithm Version 2

precondition : \( l \in 1..n \land a, b \in \mathbb{N} \land G \in \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N} \)

postcondition : \( D, FD \)

local variables: \( FD1, FD2, FD3 \in \text{BOOL} \)

\[
FD := FALSE; \\
FD1 := FALSE; \\
FD2 := FALSE; \\
FD3 := FALSE;
\]

if \( l = 0 \) then

if \( (a, b) \in \text{dom}(D) \) then

\[
FD := TRUE; \\
R := D[a, b];
\]

else

\[
FD := FALSE;
\]

else

\[
\text{shortestpath}(l - 1, a, b, D_1, FD_1);
\]

if \( FD_1 \) then

\[
\text{shortestpath}(l - 1, a, l, D_2, FD_2);
\]

if \( FD_2 \) then

\[
\text{shortestpath}(l - 1, l, b, D_3, FD_3);
\]

if \( FD_3 \) then

\[
\text{if } D_1 < D_2 + D_3 \text{ then} \\
R := D_1; \\
\text{else} \\
R := D_2 + D_3;
\]

\[
FD := TRUE;
\]

else

\[
R := D_1; \\
FD := TRUE;
\]

else

\[
R := D_1; \\
FD := TRUE;
\]

else

\[
\text{if } FD_2 \land FD_3 \text{ then} \\
R := D_2 + D_3; \\
FD := TRUE;
\]

else

\[
FD := FALSE;
\]

\[
; 
\]
### MÉRY

<table>
<thead>
<tr>
<th>model</th>
<th>Total</th>
<th>Auto</th>
<th>Manual</th>
<th>Reviewed</th>
<th>Undischarged</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHORTESTPATH0</td>
<td>8</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SHORTESTPATH1</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SHORTESTPATH2</td>
<td>493</td>
<td>328</td>
<td>165</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Global</td>
<td>506</td>
<td>340</td>
<td>166</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Teacher’s note:** Proof obligations are not very difficult to discharge; there were based on the properties of $d$ and it was boring to click the tool for discharging mechanically them. Efforts were made on the definition of $d$.

Now, it turns that our goal was to get Floyd’s algorithm and we have an algorithm for computing the existence or non-existence of a shortest path between two nodes. The next section address the question.

## 5 Floyd’s algorithm

We can use the developed algorithm to produce a result equivalent to Floyd’s execution. In fact, we apply our algorithm on each pair of possible nodes and we store it in a matrix. The algorithm 5 describes the real algorithm which can be found in any lecture notes.

**Algorithm 3:** Floyd’s Algorithm Wikipedia

- **precondition:** $l \in 1..n \land \text{matrix} \in \mathbb{N} \times \mathbb{N} \Rightarrow \mathbb{N}$
- **postcondition:** $\text{matrix} \in \mathbb{N} \times \mathbb{N} \Rightarrow \mathbb{N}$
- **local variables:** $FD1, FD2, FD3 \in \text{BOOL}$

```plaintext
foreach $k = 1; k <= n; k + +$ do
  foreach $i = 1; i <= n; i + +$ do
    foreach $j = 1; j <= n; j + +$ do
      if $\text{matrix}[i][j] > (\text{matrix}[i][k] + \text{matrix}[k][j])$ then
        $\text{matrix}[i, j] = \text{matrix}[i][k] + \text{matrix}[k][j]$  
```

Now, we are considering the problem of derivation of this solution. In fact, the development starts from the same context. Two new constants are defined namely $DF$ and $Daf$. $Df$ is the final value of the matrix $D$ corresponding to $d$ for the value $l$. $C$ is simpler and is defined as follows:

\[
\text{axm15} : C = \{\text{start}, \text{end}, \text{call}, \text{finalstep}\}
\]

New axioms define new constants:

- **axm39:** $DF \in \mathbb{N} \times \mathbb{N} \Rightarrow \mathbb{N}$
- **axm40:** $\text{dom}(DF) = \{u \mapsto v|l \mapsto u \mapsto v \in \text{dom}(d)\}$
**axm1 :** $\forall u, v \cdot u \mapsto v \in \text{dom}(Df) \Rightarrow Df(u \mapsto v) = d(l \mapsto u \mapsto v)$

**axm2 :** $Df \in N \times N \rightarrow N$

**axm3 :** $l \neq 0 \Rightarrow \text{dom}(Daf) = \{u \mapsto v|l - 1 \mapsto u \mapsto v \in \text{dom}(d)\}$

**axm4 :** $l \neq 0 \Rightarrow (\forall u, v \cdot u \mapsto v \in \text{dom}(Daf) \Rightarrow Daf(u \mapsto v) = d(l - 1 \mapsto u \mapsto v))$

**axm5 :** $l = 0 \Rightarrow Df = D0$

**axm6 :** $l \neq 0 \Rightarrow D0 \subseteq Daf$

**axm7 :** $l \neq 0 \Rightarrow Daf \subseteq Df$

**axm8 :** $l \neq 0 \Rightarrow (\forall u, v \cdot u \mapsto v \in \text{dom}(Daf) \Rightarrow Daf(u \mapsto v) = d(l - 1 \mapsto u \mapsto v))$

**axm9 :** $\forall u, v \cdot u \mapsto v \in \text{dom}(Df) \Rightarrow Df(u \mapsto v) = d(l \mapsto u \mapsto v)$

**axm10 :** $\forall u, v \cdot u \mapsto v \in \text{dom}(Df) \Rightarrow D0(u \mapsto v) = d(0 \mapsto u \mapsto v)$

The problem is to refine the model FLOYD1 to get a list of events which lead to an algorithm. The two constants $Df$ and $Daf$ are used to state the final step and the intermediate step:
• $Daf$ is the result of the call of the under construction algorithm for $l - 1$
• $Df$ is the final value which is computed from $Daf$.

We obtain the following diagram for expressing events corresponding to Floyd’s algorithm:

The new model has three variables: $c$, $D$, $TD$.

• $inv6 : TD \in N \times N \rightarrow N$
• $inv1 : c \in C$
• $inv2 : c = start \Rightarrow TD = D0 \land D = D0$
• $inv3 : c = end \Rightarrow D = Df$

• $inv4 : c = call \Rightarrow TD = D0 \land l \neq 0$

• $inv5 : c = finalstep \Rightarrow TD = Da \land l \neq 0$

Initial conditions over variables are defined by $act1 : D := D0$, $act2 : c := start$, $act3 : TD := D0$. Events are very simple to write from the diagram:

EVENT floyd-ok
REFINES floyd
WHEN
    $grd1 : c = finalstep$
THEN
\[
\left( D' \in N \times N \mapsto N \land c' = \text{end} \right)
\land \forall w, v. \left( \begin{array}{l}
w \in N \land v \in N \\
\land w \mapsto v \in \text{dom}(TD) \\
\land w \mapsto l \in \text{dom}(TD) \\
\land l \mapsto v \in \text{dom}(TD) \\
\land TD(w \mapsto v) > TD(w \mapsto l) + TD(l \mapsto v)
\end{array} \right)
\Rightarrow
\left( \begin{array}{l}
w \mapsto v \in \text{dom}(D') \\
\land D'(w \mapsto v) = TD(w \mapsto v)
\end{array} \right)
\land \forall u, v. \left( \begin{array}{l}
\land (u \mapsto l \notin \text{dom}(TD) \lor l \mapsto v \notin \text{dom}(TD))
\land u \mapsto v \notin \text{dom}(TD)
\land (u \mapsto l \in \text{dom}(TD))
\land l \mapsto v \in \text{dom}(TD)
\end{array} \right)
\Rightarrow
\left( \begin{array}{l}
\land D'(u \mapsto v) = TD(u \mapsto v)
\land u \mapsto v \notin \text{dom}(TD)
\land (u \mapsto l \in \text{dom}(TD))
\land l \mapsto v \in \text{dom}(TD)
\end{array} \right)
\land \forall u, v. \left( \begin{array}{l}
\land (u \mapsto l \notin \text{dom}(TD) \lor l \mapsto v \notin \text{dom}(TD))
\land u \mapsto v \notin \text{dom}(TD)
\land (u \mapsto l \notin \text{dom}(TD))
\land l \mapsto v \notin \text{dom}(TD)
\end{array} \right)
\Rightarrow
\left( \begin{array}{l}
\land D'(u \mapsto v) = TD(u \mapsto l) + TD(l \mapsto v)
\land u \mapsto v \notin \text{dom}(TD)
\land (u \mapsto l \notin \text{dom}(TD) \lor l \mapsto v \notin \text{dom}(TD))
\land u \mapsto v \notin \text{dom}(D')
\end{array} \right)
\]
The event

**EVENT floyd-ok**

uses a structure of Event B, which is assigning a value to variables and values are in a set. The set can be either empty, a singleton or a general set. In our case, the statement defines only one possible singleton and then the statement is clearly deterministic. However, we can substitute the event by a call of a new procedure and we should starts a new development in another development using the same principle. We get the nested loops.

The two algorithms 5 and 5 are produced from the set of events. The recursive version is simply derived using the control points. The second algorithm is the iterative version which is produced by applying the classical transformations over recursive algorithms. The function \textit{nld} is derived from an independant development by applying the same pattern.

```
EVENT floyd0
REFINES floyd
    WHEN
        grd1 : c = start ∧ l = 0
    THEN
        act2 : c := end
END
```

```
EVENT tofloydcall
    WHEN
        grd1 : c = start ∧ l > 0
    THEN
        act1 : c := call
END
```

```
EVENT floydcall
    WHEN
        grd1 : c = call
    THEN
        act1 : c := finalstep
        act2 : TD := Daf
END
```
Algorithm 4: Recursive algorithm floyd

precondition : \( l \in 1..n \land G \)
postcondition : \( D \)
local variables: \( TD \)

\[
TD := D0; \quad \text{if} \ l \neq 0 \quad \text{then} \\
\quad floyd(l - 1, G, TD); \quad D := nld(TD); \\
\quad \text{else} \\
\quad \quad D := TD; \\
\]

Algorithm 5: Non-recursive algorithm floyd

precondition : \( l \in 1..n \land G \)
postcondition : \( D \)
local variables: \( TD \)

\[
TD := D0; \quad l := 0; \quad \text{while} \ l \neq 0 \quad \text{do} \\
\quad TD := d(TD); \\
\quad D := TD; \\
\]

6 Concluding Remarks and Perspectives

The main objective of the paper is to show how we can develop a sequential structured algorithm using a one-step refinement strategy. We have illustrated the technique introduced by Cansell and Méry in [7] and we have made more precise details left unspecified in the paper. The paper has tried to give hints and advocates to the teacher who wants to use the technique for teaching correct-by-construction algorithmics using a tool which is clearly a very good mate for controlling the development. You may have questions on the treatment of arithmetics. The technique of development is a top/down approach, which is clearly well known in earlier works of Dijkstra[8,13], and to use the refinement for controlling the correctness of the resulting algorithm. It relies on a more fundamental question related to the notion of problem to solve. It is also an illustration of the application of the call-as-event pattern.

What we have learnt from the case study is summarized as follows:

(i) Developing a first abstract one-shot model using pre/post-condition. It provides the declarations part of the procedure (method) related to the one-shot model. The basic structure to express is the function \( d \) which the key of the problem. Constants of the model are defined as call-by-value parameters and variable of the model are call-by-reference parameters, The context SHORTESTPATH0 is clearly reusable and we
have reused it for the effective algorithm of Floyd.

(ii) Refining the abstract model to obtain the body of procedure. New variables are defined as local variables. The refinement introduces control states which provide a way to structure the body of the procedure. We have clearly the first control point namely \textit{start} and the last control point namely \textit{end}. The diagram helps to decompose the procedure into steps of the call and a special control point called \textit{call} is introduced. The main question is to obtain a deterministic transition system in the new refinement model.

(iii) If there are still remaining non-deterministic events, we can eliminate the non-deterministic events by developing each non-deterministic event in a specific B development starting by the statement of a new problem expressed by the non-deterministic event itself. In fact, it is what is done with the last version of Floyd’s algorithm and the event computing $D'$ from $TD$ is clearly refined to get two nested loops.

(iv) Proof obligations are relatively easy to check because the invariant is written by a list of properties of $d$ according to $d$. Even if the number of manual proof obligations is high, it was very easy to discharge them using the prover and to reuse former interactive ones.

(v) The translation of Event B model into a C program was carried out by hand and we did a mistake. We forgot that C arrays are starting the index by 0 and it leads to a bad call. We should mechanize this step to avoid this mistake.

Now, if we have to teach concepts, it is easier to teach how to write concepts and definitions using notations provided by Event B (see for instance the table1). You will get a way to check definitions and the type checker is sometimes cruel. We can discuss on many questions using this methodology: coding of numbers, preconditions, postconditions, invariant, assertions, proofs, ids and questions can lead to replies which are pertinent because of the proof tool.

Future works will provide more case studies and tools for supporting the link between models and codes. We aim to enrich the RODIN tools[14] by specific plug-ins managing libraries of models and implementing new proof obligations.

References


