VDM Specification of an Algorithm for Graph Decomposition

Abdul Huq
Middle East College of Information Technology, Sultanate of Oman
huq@mecit.edu.om

Narayanan T. Ramachandran
Middle East College of Information Technology, Sultanate of Oman
narayanan@mecit.edu.om

ABSTRACT: Complex software systems need a precise specification of their intended behaviour. While graph theory plays an important role in several aspects of building such software systems, formal methods provide a rigorous mathematical framework within which a system can be described. In this paper, an algorithm for decomposing a graph into disjoint paths is presented. Such decompositions are useful, particularly in implementing software testing. VDM representation of the proposed algorithm facilitates automation of testing process without any ambiguity.

Categories and Subject Descriptors
D.2.6 [Programming Environments]: Graphical Environments: D.2.5 [Testing and debugging]; I.3.6 [Methodology and Techniques]; Graphics data structures and data types

General Terms
Graph Theory, Software Testing, Graph Decomposition

Keywords: Vienna development method, Graphs decomposition algorithm, VDM

1. Introduction

There are several options to describe software properties. Natural languages are expressive but imprecise. Natural language descriptions carry considerable noise, ambiguities and contradictions [16], necessitating alternative techniques. Formal methods based on mathematical specifications have been found quite useful in this context. Extensive studies have been conducted to analyse their potential influence in complex and critical applications [9,10,21].

Discussing the need for and the application of formal methods to complex systems, Alagar et al. [1] present a case study on the specification of a robot based assembly system. Larsen et al. [14] assess the effectiveness of employing formal methods in different activities of the development of a security critical system. Emphasising the need for practical guidelines to make the best use of formal specifications, Palshikar provides tips to be used in formal specification in industrial applications[19].

Choice of appropriate formal method is influenced by several factors. These are discussed by Bacherini et al. [2] for an application in the development process of a railway signalling manufacturer. Hui et al. [11] discuss details of two kinds of specification languages viz., TLA and Larch in software development. Industrial applications of the specification language TRYIO are illustrated by Ciapessoni et al. [7].

One popular formal method is the Vienna Development Method (VDM). A rigorous approach to this model-oriented specification method is found in [12]. VDM has been explored by several authors for different applications. Pedersen and Klein use it to formalise a communication protocol [20]. Comparing specifications in software cost reduction method and in VDM for a safety-critical system, Williams identifies several key assessment issues [24].

In this paper, an application of VDM is considered for the purpose of implementing an algorithm developed for decomposing a graph. Such a decomposition can be used in problems relating to software testing. More specifically, a decomposition algorithm that produces several disjoint paths, the union of which results in the original graph has been developed and VDM specification of the process is described. The advantage of specifying the algorithm in VDM is to introduce rigour and to eliminate inconsistencies and ambiguity, facilitating code development for the purpose of automated implementation of the proposed algorithm.

The remainder of the paper is organised as follows: The algorithm is presented in Section 2. In Section 3, an application of the algorithm is described. The next Section discusses the VDM specification. It gives details of the terminology and notations used as well as the data structures and functions.

2. Graph Decomposition Algorithm

Graphs are very useful discrete structures in Computer science. Describing how graph properties are valuable for understanding the characteristics of the underlying software systems, usefulness of graph theory in object oriented systems has been explored in [6].

In the following paragraphs we describe a decomposition algorithm that produces disjoint paths, whose union is the original graph. In other words, given a graph G with edge set E and node set V, the algorithm yields paths P1 (with edge set E1 and node set V1), path P2 (with edge set E2 and node set V2) ...., path Pn (with edge set En and node set Vn) such that each member of E is in exactly one of E1, E2, ..., En i.e., E1∪E2∪...∪En = E and E1∩Ej = ø, for i≠j.

Different paths are obtained by successively collecting adjacent nodes, starting with an arbitrary initial node. The search for adjacent nodes to be included in the path continues until a node is repeated or a final node is reached. However, our search is on for the next path, if any.

Whenever an edge (v, v') is included in a path, it is deleted from E. Also whenever there is an isolated node we shall delete the same from V. The following algorithm describes more rigorously the operations involved in obtaining different paths which constitute the components of the decomposition.
Algorithm

Input: Graph G = (V, E)
Output: Set of components, whose disjoint union is G

1. Set G ← Graph considered for decomposition
   V ← Set of nodes of G
   m ← 0 //path index
2. While (V ≠ ∅) do
3.   Set E ← Set of edges in G
   m ← m + 1
4.   Choose an initial node j in V
   Set P_m ← {j}
   t ← j
   repeated ← false // repeated is a boolean variable that becomes true when any node occurs for the second time
5.   While (there is a k such that (t, k)
       is in E and repeated = false) do
6.     If k is in P_m then
   Set repeated ← true
   Else
   Add k to P_m
   Remove (t, k) from E
   Set t ← k
7. EndIf
8. EndWhile
9. V ← V - {isolated nodes}
10. EndWhile
11. Set Components ← {P_1, P_2, ..., P_m}
12. Return Components

3. An Application of the Algorithm

Testing is an important phase in any software development life cycle. It involves a systematic approach to imagine, explore and locate the weaknesses of a complex system and demonstrate how it needs to be corrected [3,13]. For this purpose software structures have to be abstracted. Graphs have been frequently used for such abstractions. Several authors have attempted a graph theoretic approach to testing issues [4,15,18,23].

In large real time systems, the following scenario is common: one or more modules call another module, or control is passed from one module to another. Considering the number of modules and the transfers of control from one module to another with associated parameters, there is a need to organise a systematic testing strategy. Such a strategy must enumerate and represent all transfers of control as well as find the most appropriate way to test them.

All the modules in a system along with their transfers of control can be represented as a graph with nodes representing modules and edges indicating transfers of control. For the purpose of testing, it is more convenient to study all the sub graphs of a graph than the graph itself. Hence we aim at decomposing a graph into several sub graphs which are collectively represent the original graph and no two sub graphs have an edge in common.

There are several advantages in such a decomposition viz.

a. it provides a step-by-step process to identify and represent all the transfers of control that exist in the system
b. it ensures that no transfer of control is repeated; hence none of the transfers of control is tested more than once c. it provides a disjoint list of transfers of control thus facilitating simultaneous testing of different transfers of control d. it provides systematically the order in which the tests need to be carried out.

4. VDM Specification for the Decomposition Algorithm

Benefits of using VDM in different areas of software engineering are well documented. Studies on the suitability and benefits of applying VDM for specific activities have been carried out, illustrating the use of VDM in different stages of software development [5,22]. Using VDM specifications, Nadeem and Ur-Rehman [17] propose an approach to automated testing that generates the required C code as well as test data. In another study [8], Fenkam et al. report a technique for constructing a CORBA-supported VDM oracle for automated testing starting from a VDM specification.

In this section, we develop VDM specification for the decomposition algorithm presented in Section 2. It demonstrates how a VDM specification represents an algorithm rigorously without ambiguity or inconsistency.

4.1 Terminology and Notations

We begin with a discussion of the data structures required for the specification of the algorithm.

Length of the path is the number of edges in the path, denoted by len path. The first element of a path is called its head and the rest of the path is called its tail. In VDM, head and tail of a path are indicated hd path and tl path respectively, whereas the element in the ith position of the path is denoted by path(i).

If A is a set, then A-set denotes the power set of A and A' denotes the set of all sequences of the elements of A (including the empty sequence), where repetitions are allowed. More precisely, A' = θ ∪ A ∪ A² ∪ A³ ∪ ... = ∪_{i=0}^∞ A^i, where A' = A × A × A × ... × A, the product taken i times with A^θ = θ.

In data structures, a variable is also referred to as an object. Sometimes an object may contain in itself other objects. Such an object is called a composite object which is similar to the records of Pascal or the structures of C. More precisely, a composite object has a number of fields, each such field is a variable.

A make function, when applied to appropriate values for the fields, yields a value of the composite type. A make function is specific to a type; its name is formed by prefixing mk- to the name of the type. Thus if D is a composite object, with fields f₁, f₂, ..., fₙ, then mk-D: f₁ × f₂ × ... × fₙ → D.

Often we may have to impose some restrictions on the objects. In formal specifications, such restrictions are known as invariants. They are written as a part of the type definition with key word (inv) followed by the character _.

A precise statement of all external characteristics of a system used is called functional specification or implicit definition. A direct definition of a function provides a rule for computing the result of applying the function to its arguments. The pre-condition of a function is a truth-valued function, which defines the elements of the domain of a partial function (operation) for which the existence of a result is guaranteed. The pre-condition of a function or operation defines the state/inputs to which it
can be applied. The post-condition is a truth-valued function, which defines the required relation between the input and output.

There are several ways of decomposing a composite object. One method uses the name of the make function in a context which makes it possible to associate names with the variables. Alternatively, one can use local variables which are introduced with the let command.

## 4.2 Two Composite Objects

Formal specification of the decomposition algorithm begins with the definition of two composite objects viz. Directed graph and Restricted graph.

The idea of a graph as a pair comprising a set of nodes and a set of edges is brought out in the following definition of the composite object Graph:

**Graph::**

- **nodes**: $\mathbb{Z}$-set
- **edges**: $(\mathbb{Z} \times \mathbb{Z})$-set

**inv_Graph**(mk-Graph(n,e)) $\Delta e \subseteq n \times n$

The labels of the nodes have been assumed to be integers – negative as well as positive. The invariant stipulates that the nodes connected by the edges of the graph are only those from the given set of nodes. For example, $\{(a,b,c), \{<a,b>, <b,c>, <c,d>\}\}$ will not be accepted as a graph since $<c,d> \in n \times n$, where $n = \{a, b, c\}$. We note that the graph, as defined here, is actually a directed graph.

A restricted graph is a directed graph with the additional restrictions that there is exactly one initial node as well as exactly one final node. Hence the following definition of the composite object RGraph:

**RGraph::**

- **nodes**: $\mathbb{Z}$-set
- **edges**: $(\mathbb{Z} \times \mathbb{Z})$-set

**inv_Rgraph**(mk-Rgraph(n,e)) $\Delta$

$$\begin{align*}
es &\subseteq n \times n \\
is - unique - initial(g) &\land 
is - unique - final(g)
\end{align*}$$

The boolean functions testing the uniqueness of the initial and final nodes are developed in the next section.

## 4.3 Functions: Uniqueness of Initial and Final Nodes

The process of checking whether the initial node is unique is carried out in two stages. We first collect all the initial nodes, count them, and make sure that the count is one. However, the collection of initials assumes that there is an initial node. For this reason we have the following specification:

**exists-initial**(g: RGraph) $t : \mathbb{B}$

**pre** $g \neq \emptyset$

**post** let mk-RGraph(n,e)=g in

$$\begin{align*}
t &\equiv \exists k \in n \exists j \in n (k,j) \in e \land 
\exists r \in n (r,k) \in e
\end{align*}$$

The function exists-initial returns the value true if there is at least one initial node. Otherwise it returns false. The post-condition is a predicate expression for the mathematical definition of an initial node, stated as follows: a node $k$ is an initial node if and only if there is an edge going out of it and there is no edge coming into it. This boolean valued function provides the pre-condition for the function that collects all the initial nodes given below:

**initials**(g: RGraph) $s : \mathbb{Z}$-set

**pre** exists-initial(g)

**post** $\forall k \in s (\exists j \in n (k,j) \in e \land 
\neg \exists r \in n (r,k) \in e)$

The definition of an initial node is used again to verify that every element of the output set is an initial node. We are now in a position to specify the function that examines the uniqueness of the initial node:

**is-unique-initial**(g: RGraph) $r : \mathbb{B}$

**pre** exists-initial(g)

**post** $r = (\text{card initials}(g) = 1)$

The post-condition is made simpler by the use of the pre-defined operator card that stands for the cardinality of the set.

The function is-unique-final and the associated functions can be specified in a similar manner as follows:

**exists-final**(g: RGraph) $t : \mathbb{B}$

**pre** $g \neq \emptyset$

**post** let mk-RGraph(n,e)=g in

$$\begin{align*}
t &\equiv \exists k \in n \exists j \in n (j,k) \in e \land 
\neg \exists r \in n (k,r) \in e
\end{align*}$$

**finals**(g: RGraph) $s : \mathbb{Z}$-set

**pre** exists-final(g)

**post** let mk-RGraph(n,e)=g in

$$\begin{align*}
\forall k \in s (\exists j \in n (j,k) \in e \land 
\neg \exists r \in n (k,r) \in e)
\end{align*}$$

**is-unique-final**(g: RGraph) $r : \mathbb{B}$

**pre** exists-final(g)

**post** $r = (\text{card finals}(g) = 1)$

The details are similar as in the case of testing for the uniqueness of initial node. For the identification of the final node, the post-condition in the exists-final function stipulates that there exists one node $k$ such that there is an edge coming into it and there is no edge going out of it.

## 4.4 Functions: Path Extraction

As seen in Section 2, extraction of paths plays a key role in the decomposition algorithm. VDM provides ways of manipulating such paths. We can specify a function that extracts a path. Once a path is extracted, there is a need to delete the edges included in the extracted path. As a result of such a deletion some nodes may become isolated and the structure of the restricted graph may also be violated. To convert the resulting graph into a restricted graph, dummy nodes and dummy edges are introduced.
Suppose, for example, there occur two initial nodes, say \( m \) and \( n \). A dummy node \( k \) is introduced at this stage and two dummy edges \( (k, m) \) and \( (k, n) \) are created. To distinguish the dummy nodes from the real nodes, negative integers are used to label the former. Multiple final nodes are handled in the same manner.

Once the path is extracted, the dummy nodes, if any, are removed from the path. We use the path number to label the dummy node consistently. If the graph used for the extraction of \( k \)th path has multiple initials, a dummy initial node with label \(-2k – 1\) is added; in case the path has multiple finals, the dummy final node added will have the label \(-2k\). The process then continues with the extraction of the next path.

We shall specify the function that extracts a path, starting from the initial node of the restricted graph until either the final node of the graph or a node already included is encountered. The function also needs to delete the associated edges included in the extracted path.

Central to the process of decomposition is the function \( \text{create}(k, g) \). With the path number and the restricted graph as the input parameters, it specifies the conditions to be satisfied by the \( k \)th path. We shall now give the specification for \( \text{create}(k, g) \).

\[
\text{create}(k : N_1, g : \text{RGraph}) \quad \text{path} : Z^*
\]

**pre** \( g \neq \emptyset \)

**post** (\( \text{hd } \text{path} \in \text{initials}(g) \))

\[
\text{hd } \text{path} = -(2^k – 1) \quad \land
\]

\[
\text{let } m = \text{len } \text{path} \text{ in}
\]

\[
\forall i \in \{1, 2, \ldots, m\}.
\]

\[
\langle \text{path}(i), \text{path}(i+1) \rangle \in \text{edges}(g) \quad \land
\]

\[
\langle \text{path}(m+1) \rangle \in \text{finals}(g)
\]

\[
\text{path}(j) \text{ path}(m+1) = 2^k
\]

It is worth noting that \( g \) represents the graph after the extraction of \( k-1 \) paths.

The process is to be terminated when \( g \) becomes empty. Hence the pre-condition. The post condition ensures the following: \( a. \) first node of the path is either the initial node or a dummy node \( b. \) consecutive nodes of the path correspond to some edge of the graph and \( c. \) the last node of the path is the final node or a node already included in the path or a dummy final node.

Clearly, extraction of paths requires dummy nodes, which, in turn, depend on the path number. Thus the process of decomposition, defined by the function \( \text{decompose}(g) \), passes the path number 1 along with the graph to the function \( \text{decompose}(m, g) \), which is executed recursively.

\[
\text{decompose}: \text{RGraph} \rightarrow N_1^* - \text{set}
\]

\[
\text{decompose}(g) \quad \Delta
\]

\[
\text{decompose}(1, g)
\]

\[
\text{decompose}(N_1 \times \text{Graph}) \rightarrow N_1^* - \text{set}
\]

\[
\text{decompose}(m, g) \quad \Delta
\]

\[
\text{if } (g = \emptyset) \quad \text{then } \emptyset
\]

\[
\text{else}
\]

The function \( \text{decompose} \) is executed, as long as the graph is non-empty. It performs five sub-tasks: \( a. \) extracts the \( m \)th path \( b. \) removes dummy initials and/or dummy finals, if any \( c. \) updates the graph by removing the edges of \( p \) as well as the resulting isolated nodes from \( g \) \( d. \) merges the initial nodes, in the event of multiple initials \( e. \) merges the final nodes, in the event of multiple finals.

Functions have been defined to carry out each of these subtasks. Before we discuss them, it is necessary to note the data type of the input \( g \) coming into the function \( \text{decompose} \). While \( g \) is a restricted graph initially, introduction of dummy nodes produces a graph. Thus \( g \) is taken to be a graph. Also, since the dummy nodes are removed before the path is added to the decomposition, the output set is still of type \(-\text{set}\).

- remove-dummy: \( Z \rightarrow N_1^* \)

**remove-dummy(p) \( \Delta \)** let

\[
\text{let } p_1 = \text{remove-dummy-initial}(p) \quad \land
\]

\[
\text{let } p_2 = \text{remove-dummy-final}(p) \text{ in}
\]

\[
P_2
\]

**remove-dummy-initial: \( Z \rightarrow Z \)**

**remove-dummy-final: \( Z \rightarrow N_1^* \)**

**remove-dummy-final(p) \( \Delta \)** let

\[
\text{let } p = \text{remove-dummy-final}(p) \quad \land
\]

\[
\text{if } (\text{hd } p > 0) \quad \text{then } p
\]

\[
\text{else } \text{tl } p
\]

**remove-dummy-final: \( Z \rightarrow N_1^* \)**

**remove-dummy-final(p) \( \Delta \)** let

\[
\text{let } m = \text{len } p \text{ in}
\]

\[
\text{if } (p(m+1) > 0) \quad \text{then } p
\]

\[
\text{else } \text{subseq } (p, 1, m)
\]

Removal of the dummy initial is straightforward. If the first node has a positive label, the path received is returned without any change; otherwise the rest of the path is returned. Removal of the dummy final involves deleting the last node. This is achieved by the standard VDM function for subsequence.

After the removal of the dummy initial, the path may still have a negative node as a final. Hence the output of the function \( \text{remove-dummy-initial} \) is a sequence of integers. However, the removal of the dummy final will result in a sequence of positive integers, which is the reason for the output of the functions \( \text{remove-dummy} \) and \( \text{remove-dummy-final} \) to be \( N_1^* \).

Once the dummy nodes are removed from the path extracted, the path can be added to the output of \( \text{decompose} \) function.
4.5 Functions: Updating the Original Graph

Now we turn our attention to the problem of updating the graph by removing the edges of the path just extracted and the resulting isolated nodes.

update-RGraph: RGraph × N1 → Graph

\[\text{update-RGraph}(g, p) = \text{let } e = \text{edges}(g) \land n = \text{nodes}(g) \land e_1 = \text{new-edges}(e, p) \land n_1 = n - \text{isolated-nodes}(n, e_1) \land g_1 = \text{mk-RGraph}(n_1, e_1) \text{ in } g_1\]

Note that the selector functions \text{edges} and \text{nodes} are used to decompose \(g\). It is equivalent to \(\text{mk-RGraph}(n, e) = g\).

The function \text{new-edges}(e, p) is invoked to remove the edges included in the path \(p\). The set of all nodes that become isolated as a result of such a removal are returned by the function \text{isolated-nodes}(n, e).

\[\text{new-edges}: (N_1 \times N_1)\text{-set} \times N_1 \rightarrow (N_1 \times N_1)\text{-set}\]

\[\text{new-edges}(e, p) = \text{if } \text{len } p = 0 \text{ then } e \text{ else let } e_1 = e - <p(1), p(2)> \land p_1 = \text{tl } p \text{ in } \text{new-edges}(e_1, p_1)\]

The function is defined recursively. We remove the first edge of the path \(p\) from the edge set and the process is repeated with the tail of \(p\) until there is no edge for deletion.

\[\text{isolated-nodes}: N_1\text{-set} \times (N_1 \times N_1)\text{-set} \rightarrow N_1\text{-set}\]

\[\text{isolated-nodes}(n, e) = \text{if } (n = \{\}) \text{ then } \{\} \text{ else let } x \in n \text{ in } \text{elsif is-isolated-node}(x, e) \text{ then } \{x\} \cup \text{isolated-nodes}(n-\{x\}, e) \text{ else isolated-nodes}(n-\{x\}, e)\]

\[\text{is-isolated-node}: N_1 \times (N_1 \times N_1)\text{-set} \rightarrow B\]

\[\text{is-isolated-node}(x, e) = \text{if } (e = \{\}) \text{ then true else let } a, b \in e \text{ in } \text{elsif } x = a \land x = b \text{ then false else is-isolated-node}(x, e-<a, b>)\]

We again have two recursive functions. An arbitrary node of the node set \(n\) is tested to see whether it is isolated. If so, it is included along with the other such nodes to be collected from the rest of the nodes in \(n\); otherwise the process continues with the selection of another arbitrary node from \(n\). It is repeated until there is no more node in \(n\). The boolean function \text{is-isolated-node}(x, e) is used to test whether a node \(x\) is isolated by selecting an edge from \(e\) arbitrarily. If either end of the edge is the same as \(x\), then \(x\) is not isolated and the testing ends; otherwise another edge is selected arbitrarily and the test is repeated. If there is no more edge to be tried, then \(x\) is isolated and the testing terminates.

With this, the third sub-task of the \text{decompose()} function is completed. The problem of multiple initials and multiple finals remain to be tackled.

merge-initials: N1 × Graph → Graph

\[\text{merge-initials}(m, g) = \text{if is-unique-initial}(g) \text{ then } g \text{ else } g_1\]

If there is only one initial, nothing needs to be done. If not, all the initials are collected first. A dummy node with label \(-2m-1\) is created and included in the set of nodes. To the set of edges, we add new initial edges connecting this node with all the initial nodes. The new initial edges are created as follows:

\[\text{add-initial-edges}: Z \times N_1\text{-set} \rightarrow (Z \times N_1)\text{-set}\]

\[\text{add-initial-edges}(k, x) = \text{if } (x = \{\}) \text{ then } \{\} \text{ else let } a \in x \text{ in } \text{elsif } a = k \text{ then false else } \text{add-initial-edges}(k, x-{a})\]

The dummy initial \(k\) is connected to each of the initial nodes which are selected arbitrarily one by one. The resulting edges are collected recursively.

merge-finals: N1 × Graph → Graph

\[\text{merge-finals}(m, g) = \text{if is-unique-final}(g) \text{ then } g \text{ else } g_1\]

The dummy final \(k\) is connected to each of the final nodes which are selected arbitrarily one by one. The resulting edges are collected recursively.

add-final-edges: Z × N1 × Graph → Graph

\[\text{add-final-edges}(k, (m, y)) = \text{if } (y = \{\}) \text{ then } \{\} \text{ else let } a \in y \text{ in } \text{elsif } a = k \text{ then false else } \text{add-final-edges}(k, (m, y-{a}))\]

The dummy final \(k\) is connected to each of the final nodes which are selected arbitrarily one by one. The resulting edges are collected recursively.
add-final-edges: $Z \times N_1$-set $\rightarrow (Z \times N_1)$-set

\[
\text{add-final-edges}(k, x) \triangleq \\
\begin{cases} 
(x = \emptyset) & \text{then } a \\
\text{let } a \in x \text{ in } & \\
\text{else } <a, k> \in \text{add-final-edges}(k, x\setminus \{a\})
\end{cases}
\]

The two functions defined above work the same way as the corresponding functions for merging initials. This completes the VDM specifications and definitions of the decomposition algorithm.

5. Conclusion

In this paper we have developed an algorithm for decomposing a graph into disjoint paths and represented the process in VDM. Our primary aim is to demonstrate the use of VDM in representing a process without ambiguity. Two composite objects, required for the VDM specification, have been defined and discussed. Necessary function specifications and definitions have been developed.

Representation of the algorithm using VDM has brought in rigour and robustness, facilitating easy translation of the algorithm into a programming language.

Acknowledgements

The authors would like to thank Middle East College of Information Technology, Sultanate of Oman for the support provided in making this research possible and the referees for their valuable comments and suggestions.

References


Abdul Huq received his PhD from the University of Madras, India. He is currently Head of the Department of Computing, Mathematics and Applied Sciences at Middle East College of Information Technology, Sultanate of Oman. His research interests include Artificial Intelligence, Software Engineering and Syntactic Pattern Recognition. He has published several papers and authored books on Computing.

Narayanan T Ramachandran received his PhD from the University of Madras, India. He is currently Dean of Middle East College of Information Technology, Sultanate of Oman. His research interests include Simulation studies, Optimisation techniques and Software Engineering. He has published several papers and authored books on Computing.