Specification Directed Module Testing

Ian J. Hayes

Abstract— **If a program is developed from a specification in a mathematically rigorous manner, work done in the development can be utilized in the testing of the program. We can apply the better understanding afforded by these methods to provide a more thorough check on the correct operation of the program under test. This should lead to earlier detection of faults (making it easier to determine their causes), more useful debugging information, and a greater confidence in the correctness of the final product. Overall, a more systematic approach should expedite the task of the program tester, and improve software reliability.**

The testing techniques described in this paper apply to testing of abstract data types (modules, packages). The techniques utilize information generated during refinement of a data type, such as the data type invariant and the relationship between the specification and implementation states; this information is used to specify parts of the code to be written for testing. The techniques are illustrated by application to the implementation of a symbol table as an ordered list and as a height-balanced tree.

Index Terms— **Abstract data types, data type invariant, modules, module testing, packages, pre- and postconditions, retrieval function, software reliability, specification language—Z.**

INTRODUCTION

Rigorous program development, such as that advocated in Jones's excellent book [1], can do much to increase our confidence in software we produce. The development of a program starts from a high-level specification, which is then refined through one or more stages to produce the final program. Rigorous methods rely heavily on mathematics to specify the software to be developed, and to formalize the relationship between the specification and an implementation. The work done in formalizing these relationships can be of great benefit to program testers in developing a thorough testing strategy that will trap errors as early as possible and thus be an aid to debugging.

Given a rigorously developed program it is possible to prove that it meets its specification [2], [3]. If such a proof is performed mechanically (and we trust the verifier), then testing should not be required; given the current state of the art, however, complete mechanical verification is a rarity and is expensive in resources. If the proof is done by hand then there is still room for error and hence room for testing. Rigorous methods can help greatly to increase our understanding of the program that we are developing and hence reduce the number of errors in the initial version of the program. However, we are still prone to make mistakes through oversights and typographical errors and without mechanical verification we will still require testing, especially on larger, more complex programs where errors could more easily slip in unnoticed. By making use of rigorous methods in testing we can increase our confidence in the correctness of the final product in a relatively straightforward manner that requires more moderate resources than complete mechanical verification.

The testing techniques described in this paper apply to the testing of abstract data types (modules, classes, packages, clusters). An abstract data type consists of some data, which we will refer to as its state, and a set of operations on that state. It is a good unit for testing purposes because it represents a coherent whole and, because the operations are all working on the same state, parts of the testing code are common to all the operations; in many cases it would be difficult to test an operation without having the other operations of the data type available. Testing of abstract data types can make use of the data type invariant for checking the consistency of the state between operations, the precondition for distinguishing errors in the module under test from those in the test program, and the relationship between the specification and implementation states along with the individual input-output relations for testing the correctness of the operations. These conditions and relations become specifications for parts of the code written for testing.

The techniques presented here for testing abstract data types differ from those of Gannon *et al.* [4] in that Gannon uses an algebraic specification of a data type using axioms which interrelate the operations on the data types, while in this paper we use model-based specifications of the individual operations giving the effect of each operation acting upon an abstract state. The algebraic approach is more appropriate for more primitive data types (e.g., stacks, queues) while the model-based approach is more manageable for specifying larger modules (e.g., subsystems, application-oriented packages). As the number of operations on data types increases, the algebraic axioms interrelating all the operations become more difficult to devise and the definition of an individual operation becomes spread amongst a larger number of axioms. With the model-based approach there is a single specification of each operation and the specifications can be built making use of previously defined data types. For more complex data types, the model-based specification tends to be simpler.

We will illustrate the testing technique by following through the development and testing of a symbol table module. The notation used in this paper will be based on the specification language Z [5], [6]; programs will be given in a Pascal-like notation.

The author was with the Programming Research Group, Oxford University, Oxford OX1 3QD, England. He is now with the University of Queensland.

This paper was originally published in *IEEE Transactions on Software Engineering*, SE-12(1):124–133, January 1986. This version (dated June 19, 2006) uses more up to date Z notation, corrects a few errors, and has had a second appendix added to illustrate handling of nondeterministic operations. Significant changes are commented on via footnotes.

Symbol Table Specification

This example specifies a symbol table with an operation to update an entry. We will describe the table by a partial function from symbols (*SYM*) to values (*VAL*).

$$\begin{array}{|l}\hline \textit{ST} \\ st : SYM \nrightarrow VAL \\ \hline\end{array}$$

The arrow "$\nrightarrow$" indicates a function from *SYM* to *VAL* that is not necessarily defined for all elements of *SYM* (hence "partial"). The subset of *SYM* for which it is defined is its domain of definition

$$\mathrm{dom}(st).$$

If a symbol, s, is in the domain of definition of st ($s \in \mathrm{dom}(st)$), then $st(s)$ is the unique value associated with s ($st(s) \in VAL$). The notation $\{s \mapsto v\}$ describes a function which is only defined for that particular s:

$$\mathrm{dom}(\{s \mapsto v\}) = \{s\}$$

and maps s onto v:

$$\{s \mapsto v\}(s) = v.$$

More generally, we can use the notation

$$\{x_1 \mapsto y_1, x_2 \mapsto y_2, \cdots, x_n \mapsto y_n\}$$

where all the x_k's are distinct, to define a function whose domain is

$$\{x_1, x_2, \cdots, x_n\}$$

and whose value for each x_k is the corresponding y_k. For example, if we have the following mapping

$$st = \{\text{“}\texttt{John}\text{”} \mapsto v_1, \text{“}\texttt{Mary}\text{”} \mapsto v_2\}$$

which maps "`John`" onto v_1 and "`Mary`" onto v_2, then the domain of st is the set

$$\mathrm{dom}(st) = \{\text{“}\texttt{John}\text{”}, \text{“}\texttt{Mary}\text{”}\}$$

and

$$st(\text{“}\texttt{John}\text{”}) = v_1$$
$$st(\text{“}\texttt{Mary}\text{”}) = v_2$$

The notation

$$\{\}$$

is used to denote the empty function, whose domain of definition is the empty set.

We are describing a symbol table by modeling it as a partial function. This use of a function is quite different from the normal use of functions in computing where an algorithm is given to compute the value of the function for a given argument. Here we use it to describe a data structure. There may be many possible models that we can use to describe the same object. Other models of a symbol table could be a list of pairs of symbol and value, or a binary tree containing a symbol and value in each node. These other models are not as abstract because many different lists (or trees) can represent the same function. The list and tree models of a symbol table tend to bias an implementor working from the specification towards a particular implementation. In fact, both lists and trees could be used to implement such a symbol table. However, any reasoning we wish to perform involving symbol tables is far easier using the partial function model than either the list or tree model.

Initially the symbol table is empty.

$$st = \{\}$$

The update operation can change the symbol table. We represent the effect of such an operation by the relationship between the symbol table before the operation and the symbol table after the operation. We use

$$\begin{array}{|l}\hline \Delta ST \\ ST_0 \\ ST \\ \hline\end{array}$$

to represent the state before (ST_0) and the state after (ST). The above definition of ΔST is equivalent to the following one in which ST_0 and ST have been expanded.

$$\begin{array}{|l}\hline \Delta ST \\ st_0 : SYM \nrightarrow VAL \\ st : SYM \nrightarrow VAl \\ \hline\end{array}$$

We use the convention that the zero-subscripted symbol table (st_0) represents the state before an operation and the undecorated (st) the state after. (This convention is slightly different from the convention used in [1] and [6], both of which use undecorated variables for the state before (st) and primed variables for the state after (st'); the convention used in this paper allows some simplification of the assertions used in programs.)

The operation to update an entry in the table is described by the following schema.

$$\begin{array}{|l}\hline \textit{Update} \\ \Delta ST \\ s? : SYM \\ v? : VAL \\ \hline st = st_0 \oplus \{s? \mapsto v?\} \\ \hline\end{array}$$

A schema consists of two parts: the declarations (above the center line) in which variables to be used in the schema are declared, and a predicate (below the center line) containing predicates giving properties of and relating those variables. In the schema *Update* the second line declares a variable with name "$s?$" which is the symbol to be updated. The third line declares a variable with name "$v?$" to be the value to be associated with $s?$ in the symbol table. By convention names in the declarations ending in "?" are inputs and names ending in "!" are outputs; the "?" and "!" are otherwise just part of the name.

The predicate part of the schema states that it updates the symbol table (st_0) to give a new symbol table (st) in which the symbol $s?$ is associated with the value $v?$. Any previous value associated with $s?$ (if there was one) is lost. The operator "$\oplus$" (function overriding) combines two functions of the same type to give a new function. The new function $f \oplus g$ is defined at x is either f or g are defined, and has the value $g(x)$ if g is defined at x; otherwise it has the value $f(x)$.

$$\begin{array}{rcl} \mathrm{dom}(f \oplus g) & = & \mathrm{dom}(f) \cup \mathrm{dom}(g) \\ x \in \mathrm{dom}(g) & \Rightarrow & (f \oplus g)(x) = g(x) \\ x \notin \mathrm{dom}(g) \land x \in \mathrm{dom}(f) & \Rightarrow & (f \oplus g)(x) = f(x) \end{array}$$

For example,

$$\begin{array}{l} \{\text{"}\mathtt{Mary}\text{"} \mapsto v_1, \text{"}\mathtt{John}\text{"} \mapsto v_2\} \oplus \\ \quad \{\text{"}\mathtt{John}\text{"} \mapsto v_3, \text{"}\mathtt{George}\text{"} \mapsto v_1\} \\ = \{\text{"}\mathtt{Mary}\text{"} \mapsto v_1, \text{"}\mathtt{John}\text{"} \mapsto v_3, \text{"}\mathtt{George}\text{"} \mapsto v_1\} \end{array}$$

For the operation *Update*, above, the value of $st(x)$ is $v?$ if $x = s?$; otherwise it is $st_0(x)$, provided x is in the domain of st_0. In *Update* we are only using "$\oplus$" to override one value in our symbol table function; however, the operator "$\oplus$" is more general: its arguments may both be any functions of the same type.

For a symbol table module we would normally define further operations to look up and delete entries in the table. For the purposes of illustrating testing, however, we will only consider the update operation.

If we were not allowed to know the internal structure of the implementation of the symbol table, this specification would give us all the information we needed to test that implementation. At one level this provides a reasonable test strategy but, as will be demonstrated, if we are allowed knowledge of the implementation we can construct a more rigorous test of that implementation.

IMPLEMENTATION AS AN ORDERED SEQUENCE

We will first consider implementing a symbol table as an ordered sequence and later as a height-balanced binary tree. The testing techniques do not have as much to offer for the simpler, ordered-sequence implementation, but it will serve to illustrate the ideas involved before moving on to the more complicated balanced-tree implementation.

Each item in the ordered sequence will consist of a pair of symbol and corresponding value.

$$\begin{array}{|l} \textit{Item} \\ \hline sym : SYM \\ val : VAL \\ \hline \end{array}$$

We also define a constructor function

$$mkItem : SYM \times VAL \rightarrow Item$$

which given a symbol and a value constructs the item containing that pair.

The state is given by

$$\begin{array}{|l} \textit{SST} \\ \hline sst : \mathrm{seq}\, Item \\ \hline ordered(sst) \\ \hline \end{array}$$

where sst is a sequence of items, that is, a function from its indices (one upto the length of the sequence) to elements of type *Item*, and

$$\begin{array}{l} ordered(s : \mathbb{N} \nrightarrow Item) \mathrel{\widehat{=}} \\ \qquad (\forall\, i, j : \mathrm{dom}(s) \bullet i < j \Rightarrow s(i).sym <_S s(j).sym) \end{array}$$

where we are assuming there is some total order ($<_S$) on symbols. The state is modeled by a sequence of items, *sst*. The domain of the sequence, $\mathrm{dom}(sst)$, is the set of integers that are valid indexes into the sequence. The invariant states that *sst* is in strictly ascending order on symbols. Initially the sequence is empty.

$$sst = [\ ]$$

Before describing the *Update* operation on this state let us look at the relation between the ordered sequence model and the partial function model.

$$\begin{array}{|l} \textit{ST-SST} \\ \hline ST \\ SST \\ \hline st = \{it : \mathrm{ran}(sst) \bullet it.sym \mapsto it.val\} \\ \hline \end{array}$$

where the range of the sequence, $\mathrm{ran}(sst)$, is the set containing all the items in the sequence. *ST-SST* shows how, given a sequence representation, we can retrieve the partial function model of a symbol table by, for each item in the sequence, mapping its symbol to its value.

The update operation on the sequence model is given by

$$\begin{array}{|l} \textit{UpdateS} \\ \hline \Delta SST \\ s? : SYM \\ v? : VAL \\ \hline \mathrm{ran}(sst) = \mathrm{ran}(sst_0) \cup \{mkItem(s?, v?)\} \\ \hline \end{array}$$

where

$$\begin{array}{|l} \Delta SST \\ \hline SST_0 \\ SST \\ \hline \end{array}$$

The invariant on the states ensures that the final state *sst* is ordered; the predicate part of *UpdateS* ensures that the final sequence contains the correct values.

The following is a possible implementation written in a Pascal-like notation. It uses the simple scheme of appending the new pair to the sequence and then rippling it down the sequence into the correct place to maintain the ordering.[1]

[1]The invariant has been strengthened from that in the original paper which included the following two conjuncts instead of the conjuncts involving *ordered*,

$$ordered((1..i-1) \lhd sst) \land ordered((i..\#sst) \lhd sst).$$

```
$Update(s? : SYM, v? : VAL)$ :
    $\{sst = sst_0 \wedge ordered(sst)\}$
    $sst := sst \frown [mkItem(s?, v?)]$;
    $i := \#sst$;
    $\left\{\begin{array}{l} Inv : \mathrm{ran}(sst) = \mathrm{ran}(sst_0) \cup \{mkItem(s?, v?)\} \wedge \\ 1 \le i \le \#sst \wedge ordered(\{i\} \mathbin{⩤} sst) \wedge \\ ordered((i..\#sst) \lhd sst) \end{array}\right\}$
    while $i \neq 1$ cand $sst(i-1).sym >_S sst(i).sym$ do
        begin
            $swap(sst(i-1), sst(i))$;
            $i := i - 1$
        end
    $\{Inv \wedge (i = 1 \vee sst(i-1).sym \le_S sst(i).sym)\}$
```

where

- $s \frown t$ is concatenation of sequences,
- $[mkItem(s?, v?)]$ is a sequence containing a single item: that with symbol $s?$ and value $v?$,
- $\#s$ gives the length of a sequence s,
- $(i..j) \lhd sst$ is the sequence sst with its domain restricted ($\lhd$) to values in the subrange i to j inclusive,
- $\{i\} \mathbin{⩤} sst$ is the function from natural numbers to items corresponding to the sequence sst with i removed from its domain, and
- p **cand** q is the conditional "and" operator: it only evaluates its second argument if its first argument is true.

CHECKING THE INVARIANT

To test this implementation we will first write a procedure to check if the invariant holds. This will be used to check the invariant initially and then after every operation performed on the symbol table during testing. The invariant on the ordered sequence is

$$(\forall i, j : \mathrm{dom}(sst) \bullet i < j \Rightarrow sst(i).sym <_S sst(j).sym)$$

The following code should suffice to check this holds.

```
$k := 1$;
$\{Inv : ordered((1..k) \lhd sst) \wedge 1 \le k \le \#sst\}$
while $k < \#sst$ cand $sst(k).sym <_S sst(k+1).sym$ do
    $k := k + 1$
$\{Inv \wedge (k \ge \#sst \vee sst(k).sym \ge_S sst(k+1).sym)\}$
if $k < \#sst$ then
    $\{sst(k).sym \ge_S sst(k+1).sym\}$
    "report unordered sequence"
```

The above procedure is written solely for testing purposes. In this case the testing code is as complex as the update operation itself. For more sophisticated implementations the invariant check is generally (although not always) simpler and shorter than an operation. If the invariant check on a data structure is very simple and efficient then it is a good idea to leave the check on the invariant in the code when it is put into operation in order to aid earlier detection of faults that do occur in operational use. The generality of the specification language used here precludes automatic generation of the code to check invariants.

The strategy of checking the invariant after every operation on the symbol table will catch a violation of the invariant immediately after the operation that caused it. To aid in debugging, diagnostic information such as the point at which the sequence is out of order and the corresponding items, should be displayed if the invariant check fails.

It is possible that the invariant check fails to detect an invalid state because there is an error in the invariant check that "cancels out" the error in the operation. In the majority of cases, however, we hope that the extra redundancy of the invariant check will not be of the cancelling out form. Perhaps using different people to code the testing and the module may help avoid this problem and make full use of the redundancy in detecting errors.

If we now run a series of tests on the "ordered sequence" implementation we should discover that it is incorrect: if the same symbol is inserted into the table more than once, then the ordered sequence implementation will leave the first pair in the sequence when the second pair is inserted. This will cause our invariant check to fail because there will be will be two consecutive items with the same symbol whereas the invariant states that the sequence is in strictly ascending order (no duplicates). The invariant check will fail as soon as a symbol is inserted a second time. If we followed the advice given above and displayed the items which caused the invariant check to fail, it should be obvious that the problem is due to the duplicate entry.

If we did not perform the invariant check while testing, the error in the ordered sequence implementation would not be discovered immediately after the second insertion of the same symbol. The problem would probably be detected when we perform an operation that looks up the value associated with the duplicated symbol. This could happen at a point in the program far removed from the cause of the problem, and may not occur until a considerably time after the duplicate entry has been inserted; locating the cause of the problem could then be much more difficult. Furthermore, the value returned on look up could be either the (incorrect) first added or the (correct) second added depending on the look up algorithm and the other values stored in the symbol table.

CHECKING THE PRECONDITION

The invariant check in the above example failed because the implementation was incorrect. In general, the invariant check can fail either because of an incorrect implementation or because the testing program incorrectly used the operations of the module. In the latter case, a failure can be caused if the precondition of an operation does not hold when the operation is invoked. In our example, *UpdateS* has a precondition of true so the testing program can never use the operation incorrectly. At this stage let us not try to correct the implementation of *UpdateS*, but rather change the original specification to include the following precondition stating that the symbol to be updated is not already in the symbol table.

$$s? \notin \mathrm{dom}(st_0)$$

Having now changed our specification (a tactic widely used in practice but not really recommended as the most appropriate solution in general) it is the test program that is now incorrect if it calls *UpdateS* with a symbol that is already in the table. In order to distinguish between a failure of the implementation and a failure of the test program, we can insist (at least for testing purposes) that the operations should check that their preconditions hold and, if not, report an error. For our symbol table example, checking the precondition that the symbol to be inserted is not already in the table can be achieved by adding the following code at the end of the current implementation.

```
$\left\{\begin{array}{l} \mathrm{ran}(sst) = \mathrm{ran}(sst_0) \cup \{mkItem(s?, v?)\} \land \\ 1 \le i \le \#sst \land ordered(\{i\} \lhd sst) \land \\ ordered((i..\#sst) \lhd sst) \land \\ (i = 1 \lor sst(i-1).sym \le_S sst(i).sym) \end{array}\right\}$
**if** $i > 1$ **cand** $sst(i-1).sym = sst(i).sym$ **then**
    “report symbol already in table”
```

Note that the above check only discovers that the precondition does not hold after it has modified the data structure. This is reasonable if all we do on a precondition failure is to print a message and abort; we should not attempt to carry on testing any further.

If the precondition checks are inexpensive, then it is prudent to leave them in the code permanently. If they are too expensive to leave in, then we should at least have the ability to reintroduce them during the testing of any program that makes use of the module so that errors in its use of the module are detected as early as possible. A good rule is to design module interfaces in such a way that the precondition can always be checked efficiently. This is an essential requirement for public interfaces such as operating system calls or widely used packages; it can help sort out debates about which component is at fault.

Checking the Input-Output Relation

Checking invariants and preconditions is not a thorough test of an implementation; the implementation could be quite disastrously wrong and still maintain the invariant. To thoroughly check an algorithm we also need to check that it conforms to the input-output relation of the specification.[2]

To perform such checking by testing we need to compare results of two implementations of the same high-level specification. To illustrate the technique on our symbol table example let us assume that we have available a (very high-level) programming language with maps and operations on maps as primitives. (In practice, such programming languages are not generally available; when we consider the more involved example of testing balanced trees, we will make use of a simpler implementation, namely the ordered list implementation described above, to provide a cross-check.) The operation to update a symbol table can be coded in our very high-level programming language as

```
$Update(s? : SYM, v? : VAL)$ :
    $st := st \oplus \{s? \mapsto v?\}$
```

where the state for this implementation is identical to that in the original specification.

[2]Appendix II has been added to explore directly checking the postcondition for nondeterministic operations.

We now have two implementations, *Update* and *UpdateS*, of the operation to update a symbol table. The states that the two implementations work on are quite different—in one case a mapping and in the other an ordered sequence—so the two are not directly comparable. In order to perform a cross-check between the “mapping” implementation and the “ordered sequence” implementation, we need to implement a retrieval function that extracts a mapping from an ordered sequence. We can then compare the extracted mapping to that from the “mapping” implementation both initially and after every operation, each operation being performed on both implementations before the retrieval and comparison test.

The relation between the “mapping” and “ordered sequence” states is defined by the retrieval relation *ST-SST* given previously. The following code will retrieve the output mapping $st!$ from the input sequence $sst?$.

```
$ST\text{-}SST(sst? : \mathrm{seq}\, Item, st! : ST)$ :
    $i := 0;$
    $st! := \{\};$
    $\left\{\begin{array}{l} Inv : 0 \le i \le \#sst? \land \\ st! = \{it : \mathrm{ran}((1..i) \lhd sst?) \bullet it.sym \mapsto it.val\} \end{array}\right\}$
    **while** $i \ne \#sst?$ **do**
        **begin**
            $i := i + 1;$
            $st! := st! \oplus \{sst?(i).sym \mapsto sst?(i).val\}$
        **end**
    $\{st! = \{it : \mathrm{ran}(sst?) \bullet it.sym \mapsto it.val\}\}$
```

The retrieved mapping can then be compared directly with that used in the mapping implementation:

```
**if** $st! \ne st$ **then**
    “input-output relation check failed”
```

Any error detected by the comparison may indicate an error in either

- the “ordered sequence” implementation,
- the “mapping” implementation,
- the ordered sequence to mapping retrieval function, or
- the comparison itself.

The last three should normally be less likely because they should be somewhat simpler. However, they cannot be ruled out as possible causes of errors, and if an error is detected further investigation will be required in order to determine which of the above is the cause and to find the actual fault.

When we combine input-output relation checks with invariant and precondition checks we get a thorough test mechanism for operations on the “ordered sequence” symbol table implementation. It is almost certain that the redundancy incorporated into the above checks is sufficient to catch any fault manifested during testing. Furthermore, the fault will have been isolated to a particular operation and if appropriate diagnostics have been added to the checking code, the cause should be easily found. However, we are only dealing with a testing strategy and like all testing it does not exclude the possibility of latent errors: errors that did not occur on the test cases used but could occur on other cases. Such latent errors show the inherent weakness of program testing when compared to program verification.

To reduce the possibility of latent errors left after testing we should use our knowledge of the implementation to ensure that it is thoroughly exercised; all parts of the code should be tested. The selection of test cases is covered in other treatments of program testing [7] and will not be pursued further here.

Height-Balanced Binary Trees

In the "ordered sequence" implementation the procedures to test the invariant and retrieve the symbol table are both as complicated as the operation to update an item. We will now consider a more involved example in which the invariant testing and retrieval function are somewhat simpler than the operations.

Height-balanced binary trees were invented by Adel'son-Velskii and Landis [8] to provide a binary search tree with worst-case insert and delete times of $O(\log N)$, where N is the number of nodes in the tree. A binary tree is height balanced if at every node in the tree the heights[3] of its left and right subtrees differ by at most one. The beauty of a height-balanced tree is that its worst-case height is at most 45 percent greater than that of an equivalent perfectly-balanced tree,[4] and insertion and deletion of nodes can be performed by examining a path from the root to a node, unlike perfectly-balanced trees. Search, insert, and delete operations can all be performed in $O(\log N)$ time in the worst case, which should be compared with a worst-case time of $O(N)$ for these operations on an ordinary (unbalanced) tree.

The major disadvantage of balanced trees[5] is that the algorithms to manipulate them are considerably more complicated than those for an unbalanced tree. Fortunately, for the purposes of this paper we do not need to delve into the details of these operations in order to illustrate the approach to testing them. The interested reader is referred to one of the many books on algorithms that discuss operations on balanced trees in detail. One such book is Wirth's *Algorithms + Data Structures = Programs* [9]. To give a crude idea of the complexity of the operations on balanced trees, the Pascal versions given by Wirth consist of 63 lines for insertion (pp. 220–221) and 92 lines for deletion (pp. 223–225). These figures should be compared with those for unbalanced trees: 19 lines for insertion (p. 205) and 18 lines for deletion (p. 211). Not only are balanced tree operations considerably longer than their unbalanced tree counterparts, they are, in the opinion of the author, a good deal more subtle and more liable to erroneous implementation.

As promised earlier we do not need to look in detail at the implementation of the operations on balanced trees. What we do need to look at closely, however, is the state invariant for a balanced tree. A tree is given by

$$Tree \mathrel{\widehat{=}} Node \mid nil$$

[3]The height of a binary tree is the maximum number of nodes on a path starting at its root and descending down the tree.

[4]A perfectly-balanced tree is a binary tree in which at every node the number of nodes in its left and right subtrees differ by at most one.

[5]For the remainder of this paper we will abbreviate "height-balanced binary tree" to "balanced tree".

That is, a *Tree* is either a *Node* or it is the special value *nil*, where

Node

$$\begin{array}{|l}
sym : SYM \\
val : VAL \\
bal : -1..1 \\
left, \\
right : Tree \\
\hline
(\forall s : syms(left) \bullet s <_S sym) \land \\
(\forall s : syms(right) \bullet sym <_S s) \land \\
bal = height(left) - height(right)
\end{array}$$

where $syms : Tree \rightarrow \mathbb{P}\,SYM$ such that for $n : Node$

$$\begin{array}{rcl}
syms(nil) & = & \{\} \\
syms(n) & = & syms(n.left) \cup \{n.sym\} \cup syms(n.right)
\end{array}$$

and $height : Tree \rightarrow \mathbb{N}$ such that for $n : Node$

$$\begin{array}{rcl}
height(nil) & = & 0 \\
height(n) & = & max(height(n.left), height(n.right)) + 1
\end{array}$$

The trees are both ordered and balanced. A tree is ordered if at each node in the tree all the symbols in its left subtree are less than the symbol at the node, which is less than all the symbols in its right subtree. A tree is balanced if at every node the difference in heights between the left and right subtrees is equal to the *bal* field of the node (which can only take on values in the range -1..1).

The state of a balanced tree is given by the following.

BT

$$\begin{array}{|l}
t : Tree
\end{array}$$

The relation between a balanced tree and the high-level specification of a symbol table is given by

ST-BT

$$\begin{array}{|l}
ST \\
BT \\
\hline
st = \{node : nodes(t) \bullet node.sym \mapsto node.val\}
\end{array}$$

where $nodes : Tree \rightarrow \mathbb{P}\,Node$ such that for $n : Node$

$$\begin{array}{rcl}
nodes(nil) & = & \{\} \\
nodes(n) & = & nodes(n.left) \cup \{n\} \cup nodes(n.right)
\end{array}$$

Checking the Invariant

As before, we can write a procedure to check the state invariant: the tree is both balanced and ordered. A procedure to check that a tree is balanced follows. It performs a post-order traversal of a tree, checking that each subtree is balanced and returning the height of the tree so that the higher level checking that the tree is balanced can take place.

```
$Balanced(t? : Tree, h! : integer)$ :
  if $t? = nil$ then
      $h! := 0$
  else$\{t? \neq nil\}$
      begin
          var $hl, hr : integer$;
          $Balanced(t?.left, hl)$;
          $Balanced(t?.right, hr)$;
          $\left\{\begin{array}{l} hl = height(t?.left) \land \\ hr = height(t?.right) \end{array}\right\}$
          if $hl - hr \neq t?.bal$ then
              “report unbalanced tree”
          $h! := max(hl, hr) + 1$
      end
```

We have assumed here that the implementation of our programming language will trap any assignment of a value outside the range -1..1 to the *bal* field of a node; if this were not the case then a check that the *bal* field of each node is in this range should be added to the above procedure. The procedure to check that a tree is ordered is straightforward and omitted here.

For balanced trees, the invariant checking is far less complicated than the operations; it is more akin to the complexity of the operations on the simpler unbalanced trees, requiring only straightforward tree traversal algorithms. The great value of the invariant check is that if an operation otherwise works correctly but manages to corrupt the data type invariant, the fault will be detected immediately after the operation rather than at some indeterminate time in the future when an operation tries to access the corrupted part of the data structure. Not only is the detection in this latter case well after the fault, it may be on an operation other than the one that caused the corruption; other than detecting that there is an error, one has been given little help in diagnosing the fault.

Given this invariant check procedure, our testing can now check that the invariant holds initially and then after each operation during testing. The invariant checking above requires $O(N)$ time versus the $O(\log N)$ time for the operations themselves. Hence it is not sensible to leave the invariant check in the program after testing. After all, the point of using balanced trees was to take advantage of their worst-case $O(\log N)$ performance; if we were to leave the invariant check in the code the performance would always be $O(N)$ and hence worse than the unbalanced tree which, while being $O(N)$ worst case, is only $O(\log n)$ average case.

The invariant check given above is a far more stringent test that the state of a module is consistent than any that can be carried out purely from knowledge of the high-level specification, even if one is given a retrieval function to extract the abstract state. It is possible that the implementation could be incorrect in a way that does not affect the high-level correctness. For example, the implementation may correctly maintain an ordered tree but it may be incorrectly balanced. In this case the operations would appear to work correctly but in some cases would not be as efficient. Such a fault could only be detected externally by timing operations and would require the testing to generate a badly balanced tree. With knowledge of the internal operation of the algorithm in the invariant check it is far less likely that an incorrect implementation would go undetected.

Checking the Precondition

As with the “ordered sequence” implementation, a precondition check can be incorporated into the implementation using balanced trees. This will detect any incorrect use of the operations by the testing program. For balanced trees a simple constant-time check (which should be left in the code permanently) can be incorporated into the update operation. As this is quite simple to do, but to explain requires detailed knowledge of the update operation on balanced trees, we will not elaborate the precondition check for balanced trees here.

Checking the Input-Output Relation

As with the “ordered sequence” implementation, we need to check that the input-output relation is satisfied. For this example we will not assume that we have available a very high-level programming language with mappings as primitives. In order to cross-check the input-output relation we need a second (simpler) implementation of a symbol table. Fortunately, we have just that in our “ordered sequence” implementation. To perform the cross-check we need a retrieval function that extracts an ordered sequence from a balanced (ordered) tree. The relation between ordered sequences and ordered trees is given by

```
── SST-BT ──────────────────
│ SST
│ BT
├──────────
│ $\{node : nodes(t) \bullet node.sym \mapsto node.val\}$
│     $= \{it : \mathrm{ran}(sst) \bullet it.sym \mapsto it.val\}$
└──────────────────────────
```

Extracting an ordered sequence from an ordered tree can be achieved by the following tree traversal algorithm.

```
$TreetoSequence(t? : Tree, sst! : \mathrm{seq}\, Item)$ :
  if $t? = nil$ then
      $sst! := [\ ]$
  else$\{t? \neq nil\}$
      begin
          var $lsst, rsst : \mathrm{seq}\, Item$;
          $TreetoSequence(t?.left, lsst)$;
          $TreetoSequence(t?.right, rsst)$;
          $sst! := lsst \frown [mkItem(t?.sym, t?.val)] \frown rsst$
      end
```

The sequence retrieved by *TreetoSequence* is compared to the sequence maintained by the “ordered sequence” implementation after each operation is performed (on both implementations). The code for the comparison is straightforward and has been omitted here.

A note of warning is required here. In the example above there is a unique representation (as an ordered sequence) for every distinct abstract symbol table. This is not necessarily the case. For example, if the representation used unordered sequences there could be a number of possible representations of a single abstract symbol table. In such cases the code for the comparison needs to determine if the two representations correspond to the same abstract object (abstract equivalence) rather than if the two representations are identical.

For the height-balanced binary tree example, the procedures required to use the testing techniques outlined in this paper require only a fraction of the time necessary for a programmer to develop the somewhat more sophisticated balanced tree operations. The extra time is well spent in terms of increasing one's confidence in the correct operation of the algorithms, but furthermore the techniques are likely to actually save time if there are errors in the operations: the testing will isolate the errors quickly and provide useful diagnostics to aid in debugging.

Discussion

When implementing abstract data types in a programming language with facilities to support them (for example, Modula modules, Ada packages, or Clu clusters) the invariant check and retrieval procedures will both have to be part of the module as they need access to the internal data structure, which should not be accessible externally. This will probably imply that the person responsible for the module should write these when writing the module (although as mentioned earlier there are good reasons for having a separate person write them). In practice this probably represents a reasonable line of demarcation between the module writer and tester as these functions provide everything that the tester needs from the module internals to apply the testing techniques.

The author has used the techniques described above to test an implementation of B-trees [10]: balanced multi-way trees suitable for secondary storage databases. B-trees are more complicated data structures than height-balanced trees, and the algorithms to manipulate them have a number of special cases that can easily lead to errors in implementation. In the testing of the B-tree implementation, the techniques described above were able to isolate two errors (one omission and the other a swap of variable names) and to give good hints as to the nature of the fault; in this respect the invariant check, which for the B-tree is involved but not difficult to implement, was particularly useful in detecting faults as soon as possible after their prime cause. The use of these techniques certainly increased the author's confidence in the correctness of the final implementation—especially that the algorithms actually implemented B-trees rather than some other (strange) variety of multi-way tree.

Another technique that can be used in testing programs is to check assertions, such as loop invariants, at execution time. This could be useful if a fault is detected in an operation of an abstract data type but the cause is not obvious. Unfortunately, expanding such assertions is non-trivial; in some cases the code to check a loop invariant can be more complicated than the original loop. The tactic of testing at the abstract data type level seems to provide the most benefits for the amount of effort involved; coding up assertions can be left to aid in debugging when a non-obvious error is detected, although it is probably better to go back to the original reasoning about the program and find the fault there.

The testing procedures should not be discarded once a module has been tested; they will be useful to anyone responsible for making changes to the module (where introduction of errors is more likely due to lack of understanding). The invariant check procedure is of more general use if data are kept on permanent storage devices. It can be used to check the consistency of the data after a hardware or software failure has occurred. It cannot guarantee the correctness of the data, but it can find inconsistencies which imply the data are incorrect and it can ensure that the data are in a state suitable for running the system.

Appendix I
Notation

A. *Definitions and Declarations*

Let x, x_k be identifiers and T, T_k sets.

$LHS \mathrel{\widehat=} RHS$ — Definition of LHS as syntactically equivalent to RHS.

$x : T$ — Declaration of identifier x of type T.

$x_1 : T_1;\ x_2 : T_2;\ \ldots;\ x_n : T_n$ — List of declarations.

$x_1, x_2, \ldots, x_n : T \mathrel{\widehat=} x_1 : T;\ x_2 : T;\ \ldots;\ x_n : T$.

B. *Logical Symbols*

Let P, Q be predicates and D declarations.

$\neg P$ — Negation: "not P".

$P \vee Q$ — Disjunction: "P or Q".

$P \wedge Q$ — Conjunction: "P and Q".

$P \Rightarrow Q$ — Implication: "P implies Q" or "if P then Q".

$\exists x : T \bullet P$ — Existential quantification: "there exists an x of type T such that P".

$\forall x : T \bullet P$ — Universal quantification: "for all x of type T, P holds".

$\exists x_1 : T_1;\ x_2 : T_2;\ \ldots;\ x_n : T_n \bullet P$ — "There exist x_1 of type T_1, x_2 of type T_2, …, and x_n of type T_n, such that P holds."

$\forall x_1 : T_1;\ x_2 : T_2;\ \ldots;\ x_n : T_n \bullet P$ — "For all x_1 of type T_1, x_2 of type T_2, …, and x_n of type T_n, P holds."

C. *Sets*

Let S and T be subsets of X; t, t_k terms; P a predicate; and D declarations.

$t \in S$ — Set membership: "t is an element of S".

$t \notin S \mathrel{\widehat=} \neg (t \in S)$.

$S \subseteq T \mathrel{\widehat=} (\forall x : S \bullet s \in T)$ — Set inclusion.

$S \subset T \mathrel{\widehat=} S \subseteq T \wedge S \neq T$ — Strict set inclusion.

$\{\}$ — The empty set.

$\{t_1, t_2, \ldots, t_n\}$ — The set containing t_1, t_2, …, and t_n.

$\{x : T \mid P\}$ — The set containing exactly those x of type T for which P holds.

$(t_1, t_2, \ldots, t_n)$ — Ordered n-tuple of t_1, t_2, …, t_n.

$T_1 \times T_2 \times \cdots \times T_n$ — Cartesian product: the set of all n-tuples such that the kth component is of type T_k.

$\{x_1 : T_1;\ x : 2 : T_2;\ \ldots;\ x_n : T_n \mid P\}$ — The set of n-tuples $(x_1, x_2, \ldots, x_n)$ with each x_k of type T_k such that P holds.

$\{x_1 : T_1;\ x : 2 : T_2;\ \ldots;\ x_n : T_n \mid P \bullet t\}$ — The set of values of the term t such that given all the x_k of type T_k, P holds.

$\{D \bullet t\} \mathrel{\widehat{=}} \{D \mid true \bullet t\}$.

$\mathbb{P}\,S$ — Powerset: $\mathbb{P}\,S$ is the set of all subsets of S.

$\mathbb{F}\,S$ — Finite subsets of S.

$S \cup T \mathrel{\widehat{=}} \{x : X \mid x \in S \vee x \in T\}$ — Set union.

$S \cap T \mathrel{\widehat{=}} \{x : X \mid x \in S \wedge x \in T\}$ — Set intersection.

$\#S$ — Size (number of elements) of a finite set.

D. Relations and Functions

A relation is modeled by a set of ordered pairs. Hence operators defined for sets can be used on relations. A function is a relation with the property that for each element in its domain there is a unique element in its range related to it. As functions are relations, operators defined for relations also apply to functions. Let A and B be sets; S a set of the same type as A; R, R_1 and R_2 be relations between A and B; f be a function; and x, x_k, y, y_k be terms.

$A \leftrightarrow B \mathrel{\widehat{=}} \mathbb{P}(A \times B)$ — The set of relations from A to B.

$A \mathrel{+\!\!\!\rightarrow} B \mathrel{\widehat{=}} \{f : A \leftrightarrow B \mid (\forall a : A;\ b, b' : B \bullet (a, b) \in f \wedge (a, b') \in f \Rightarrow b = b'\}$ — The set of partial functions from A to B.

$A \rightarrow B \mathrel{\widehat{=}} \{f : A \mathrel{+\!\!\!\rightarrow} B \mid \mathrm{dom}(f) = A\}$ — The set of total functions from A to B.

$x \mapsto y \mathrel{\widehat{=}} (x, y)$.

$\{x_1 \mapsto y_1, x_2 \mapsto y_2, \ldots, x_n \mapsto y_n\}$ — The relation that maps x_1 to y_1, x_2 to y_2, ..., and x_n to y_n. It is equivalent to $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$.

$f(x)$ — The function f applied to x.

$\mathrm{dom}(R) \mathrel{\widehat{=}} \{a : A \mid (\exists\, b : B \bullet (a, b) \in R)\}$ — The domain of definition of a relation (or function).

$\mathrm{ran}(R) \mathrel{\widehat{=}} \{b : B \mid (\exists\, a : A \bullet (a, b) \in R)\}$ — The range of a relation (or function).

$S \lhd R \mathrel{\widehat{=}} \{a : A;\ b : B \mid (a, b) \in R \wedge a \in S\}$ — Domain restriction.

$S \ntriangleleft R \mathrel{\widehat{=}} \{a : A;\ b : B \mid (a, b) \in R \wedge a \notin S\}$ — Domain subtraction.

$R_1 \oplus R_2 \mathrel{\widehat{=}} (\mathrm{dom}(R_2) \ntriangleleft R_1) \cup R_2$ — Relational (or functional) overriding.

E. Numbers

$\mathbb{N}$ — The set of natural numbers (nonnegative integers).

$\mathbb{Z}$ — The set of integers (positive, zero and negative).

$m..n \mathrel{\widehat{=}} \{k : \mathbb{Z} \mid m \leq k \wedge k \leq n\}$ — The set of integers between m and n inclusive.

F. Sequences

Let X be a set; S be a sequence; and lowercase variables be terms.

$\mathrm{seq}\,X \mathrel{\widehat{=}} \{S : \mathbb{N} \mathrel{+\!\!\!\rightarrow} X \mid (\exists\, n : \mathbb{N} \bullet \mathrm{dom}(S) = 1..n)\}$ — The set of finite sequences whose elements are drawn from X.

$\#S$ — The length of sequence S.

$[\ ] \mathrel{\widehat{=}} \{\ \}$ — The empty sequence.

$S(i)$ — The ith element in the sequence S.

$[x_1, \ldots, x_n]$ — The sequence $\{1 \mapsto x_1, \ldots, n \mapsto x_n\}$.

$[s_1, \ldots, s_n] \frown [t_1, \ldots, t_m] \mathrel{\widehat{=}} [s_1, \ldots, s_n, t_1, \ldots, t_m]$ — Concatenation.

$\mathrm{ran}([s_1, s_2, \ldots, s_n]) \mathrel{\widehat{=}} \{s_1, s_2, \ldots, s_n\}$ — Range of a sequence: the set of items in the sequence; $\mathrm{ran}([\ ]) \mathrel{\widehat{=}} \{\ \}$.

G. Schema Notation

Schema definition:

$$\begin{array}{|l}
\multicolumn{1}{l}{\textit{SCH}} \\
\hline
a : A \\
b : B \\
\hline
\textit{predicate} \\
\hline
\end{array}$$

A schema groups together some declarations of variables and a predicate relating those variables. A schema can be used as a type, in which case for a variable s of type *SCH* its a and b fields can be referred to by $s.a$ and $s.b$. The following conventions are used for variable names in those schemas which represent operations:

Subscript "0"	State before the operation.
Undecorated	State after the operation.
Ending in a "?"	Inputs to the operation.
Ending in a "!"	Outputs from the operation.

A schema S may be included within a schema T, in which case the declarations of T are merged with the other declarations of S (variables declared in both S and T must be the same type) and the predicates of S and T are conjoined.

APPENDIX II
CHECKING NONDETERMINSITIC OPERATIONS

This appendix did not appear in the original paper but numerous people have suggested that it should have.

Consider the following nondeterministic operation that removes any symbol and value pair from a nonempty symbol table.

$$\begin{array}{|l}
\multicolumn{1}{l}{\textit{RemoveAny}} \\
\hline
\Delta ST \\
s! : SYM \\
v! : VAL \\
\hline
st_0 \neq \{\} \wedge \\
(s!, v!) \in st_0 \wedge \\
st = st_0 \setminus \{(s!, v!)\} \\
\hline
\end{array}$$

Because the entry to be removed is chosen nondeterministically, the scheme outlined for deterministic operations is not adequate. That scheme relied on running both the abstract and concrete implementations and comparing the results, but with a nondeterministic operation it is perfectly valid for the two implementations to remove different entries.

An approach that copes with nondeterminism is to implement code to check the postcondition of the abstract operation, and apply it to the states retrieved before and after the operation. To do this we

- retrieve and save the state of the system before the operation,
- call the implementation of the operation,

- retrieve the updated state of the system after the operation, and
- check that the postcondition of the operation is satisfied by the inputs to the operation, the outputs actually produced, and the retrieved before and after states.

As an example, consider checking the balanced-tree implementation against the ordered sequence level abstraction. For the sequence abstraction the postcondition of the *RemoveAny* operation is

$$\begin{array}{l} \mathit{mkItem}(s!, v!) \in \operatorname{ran}(\mathit{sst}_0) \wedge \\ \operatorname{ran}(\mathit{sst}) = \operatorname{ran}(\mathit{sst}_0) \setminus \{\mathit{mkItem}(s!, v!)\} \end{array}$$

We assume that the procedure $\mathit{RemoveItem}(s, it, t)$ returns the sequence t consisting of the sequence s with item it removed. The following code suffices to check the *RemoveAny* operation.

```
$\mathit{error} := \mathit{false}$;
$\mathit{TreetoSequence}(t, \mathit{sst}0)$;
$/*$ Call the implementation $*/$
$\mathit{RemoveAny}(s!, v!)$;
$/*$ Check the invariant holds $*/$
$\mathit{Balanced}(t, h)$;
$\mathit{TreetoSequence}(t, \mathit{sst})$;
$/*$ Check the postcondition holds for the
    retrieved sequences $*/$
if $\mathit{mkItem}(s!, v!) \notin \operatorname{ran}(\mathit{sst}0)$ then
    $\mathit{error} := \mathit{true}$;
$\mathit{RemoveItem}(\mathit{sst}0, \mathit{mkItem}(s!, v!), \mathit{sst}1)$;
if $\mathit{sst} \neq \mathit{sst}1$ then
    $\mathit{error} := \mathit{true}$
$\left\{ \begin{array}{l} \neg\, \mathit{error} \Leftrightarrow \\ \quad \mathit{mkItem}(s!, v!) \in \operatorname{ran}(\mathit{sst}0) \wedge \\ \quad \operatorname{ran}(\mathit{sst}0) = \operatorname{ran}(\mathit{sst}) \setminus \{\mathit{mkItem}(s!, v!)\} \end{array} \right\}$
```

The above checks not only that the entry removed was in the sequence retrieved before the call to *RemoveAny*, but that the only change to the sequence retrieved after the call to *RemoveAny* is the removal of that entry. Note that we have assumed the existence of the procedure *RemoveItem* operating on sequences, which unlike *RemoveAny* is deterministic. *RemoveItem* has an additional parameter which is the item to remove, which in this case is the item returned by *RemoveAny*. Of course, as before, errors in the implementation of the above code, the procedures *TreetoSequence*, or *RemoveItem* may invalidate the testing process.

The above only covers the case when the relationship between the implementation state and the abstract state is a (retrieve) function, i.e., there is a unique abstract state for any implementation state. If the relationship is not a function there may be multiple abstract states corresponding to a single implementation state. For example, if we use unordered sequences in the above example, then there are many possible unordered sequences for a given tree, and the above code is invalid if *TreetoSequence* returns an arbitrary sequence with the same elements as the tree. For practical testing purposes it is simpler to strengthen the invariant on the abstract state to ensure that there is a unique abstract state for each implementation state, and hence for the above example used ordered sequences rather than unordered sequences.

Acknowledgement

I would like to acknowledge the support of IBM (U.K.) Laboratories during the time in which this work was done. I would also like to thank the reviewers for some useful suggested improvements to the paper and K. A. Robinson for helping sow the seeds of the idea presented here and for his help in preparing the revised version of this paper. The work of C. B. Jones on data refinement forms the formal basis and main inspiration for the testing techniques described here.

References

[1] C. B. Jones, *Software Development: A Rigorous Approach*. Prentice-Hall International Series in Computer Science, 1980.

[2] A. L. Ambler, D. I. Good, J. C. Browne, W. F. Burger, R. M. Cohen, C. G. Hoch, and R. E. Wells, "Gypsy: A language for specification and implementation of verifiable programs," *ACM SIGPLAN Notices (Proc. Conf. Language Design for Reliable Software)*, vol. 12, no. 3, pp. 1–10, Mar. 1977.

[3] G. J. Popek, J. J. Horning, B. W. Lampson, J. G. Mitchell, and R. L. London, "Notes on the design of Euclid," *ACM SIGPLAN Notices (Proc. Conf. Language Design for Reliable Software)*, vol. 12, no. 3, pp. 11–18, Mar. 1977.

[4] J. Gannon, P. McMullin, and R. Hamlet, "Data abstraction implementation specification and testing," *ACM Trans. Program. Lang. Syst.*, vol. 3, no. 3, pp. 211–223, July 1981.

[5] J. R. Abrial, "The specification language Z: Basic library," Programming Research Group, Oxford University," Internal report, 1982.

[6] C. C. Morgan and B. A. Sufrin, "Specification of the Unix filing system," *IEEE Trans. on Software Engineering*, vol. SE-10, no. 2, pp. 128–142, March 1984.

[7] B. Beizer, *Software Testing Techniques*. Van Nostrand Reinhold, 1983.

[8] G. M. Adel'son-Velskii and Y. M. Landis, "An algorithm for the organization of information (English translation)," *Sov. Math. Dokl.*, vol. 3, pp. 1259–1262, 1962.

[9] N. Wirth, *Algorithms + Data Structures = Programs*. Prentice-Hall, 1976.

[10] R. Bayer and E. M. McCreight, "Organization and maintenance of large ordered indexes," *Acta Informatica*, vol. 1, no. 3, pp. 173–189, 1972.