Evaluation of Verification and Testing Tools for FORTRAN Programs

Kathryn A. Smith

JULY 1980

NASA
Evaluation of Verification and Testing Tools for FORTRAN Programs

Kathryn A. Smith
Langley Research Center
Hampton, Virginia
SUMMARY

The traditional procedure to ensure the provision of quality, reliable software involves extensive testing. In recent years, several automated software verification and testing systems (V&T tools) have been developed to discover errors in computer programs, saving computer and analyst time and improving the quality of programs. Two such tools which may be used in the analysis of FORTRAN programs have been installed and evaluated at the National Aeronautics and Space Administration Langley Research Center. These are DAVE, a static analysis tool, and PET (Program Evaluator and Tester), a dynamic analysis tool. These V&T systems were evaluated using a series of test cases. The results of this evaluation indicate that, despite some limitations, both of these tools are significantly beneficial in the development and testing of FORTRAN programs. Based on this analysis, these tools are recommended for general use.

INTRODUCTION

Computers are becoming larger and faster. As a result, computer programs are becoming more complex, and software tools for the development and testing of programs are becoming increasingly important. A number of tools are emerging for use in the software development and testing process. Included in this number are several software tools which may be used in the verification and testing of FORTRAN programs. Two such tools which have been applied to computer programs at the National Aeronautics and Space Administration (NASA) Langley Research Center (LaRC) are a static analyzer, DAVE (refs. 1 and 2), and a dynamic analyzer, PET (refs. 3 and 4). This paper presents our evaluation of these tools and our experiences with them.

Use of trade names or names of manufacturers in this report does not constitute an official endorsement of such products or manufacturers, either expressed or implied, by the National Aeronautics and Space Administration.

SOFTWARE DEVELOPMENT PROCESS

The software development process consists of several steps. The first is the definition of program requirements. Figure 1 illustrates the basic development phases which follow the establishment of requirements. Although these steps may appear disjointed, they overlap in many instances. For example, documentation should be developed concurrently with the program. Once the program requirements are established, the design of the software can begin. After completion of the design phase, the program is coded and several steps of successive program testing begin. The first step is the use of a compiler to perform a syntax analysis which checks that the program is compatible with the grammatical structure of the programming language. Any syntax errors should be corrected before going to the next phase of testing.
Frequently, a programmer will consider the testing completed at this point. However, careful, disciplined testing includes both static analysis and execution time testing. Static analysis takes a subject program as input and tests the logic flow and data flow through the program. After the logic- and data-flow errors are found and corrected, the execution time testing begins. Part of this testing is a dynamic analysis which identifies the portions of the subject program which have been executed. Static analysis and dynamic analysis are difficult and time consuming without the aid of software tools. At LaRC, the DAVE system is recommended for static analysis and PET is recommended for execution time analysis of FORTRAN programs. When the execution testing is completed, the documentation of the program can be completed, after which the program is ready for productive use. Many of the steps used in software development are also used in the acceptance of programs developed under contract and in the maintenance and revision of existing programs. These steps are indicated by the solid boxes in figure 1.

DAVE

Description

DAVE is a static software verification tool which was developed by the Department of Computer Science, University of Colorado, under a grant from the National Science Foundation. LaRC obtained DAVE from the University of Colorado in 1976. It is a system tool for gathering information about global data flow in syntactically correct ANSI FORTRAN\(^1\) programs and for identifying the anomalous use of data in these programs. DAVE is called a static analysis tool, because the FORTRAN program is neither compiled nor executed; instead, information is gathered about the subject program by analyzing the source code. DAVE uses global data flow analysis to reveal suspicious or erroneous use of data; that is, it examines all possible data paths. Thus, DAVE offers a multimodule analysis capability not available in the Control Data (CDC) FORTRAN Extended (FTN) compiler at LaRC. In general, DAVE does not require any modification of the subject program, nor does it require any intervention by the user during execution. The steps in DAVE system execution are shown in figure 2. The appendix describes the procedure to use DAVE at LaRC.

DAVE issues three types of diagnostics - "errors," "warnings," and "messages." "Error" diagnostics indicate likely execution errors. These errors include such things as variables uninitialized on all paths, a constant subroutine argument assigned a value in the subroutine, and a subprogram FUNCTION name never assigned a value. "Warning" diagnostics are possible execution errors. Examples of warnings include: variables uninitialized on some paths and variables unused on some paths. The "message" diagnostics are used to give information about COMMON blocks and COMMON variables. For example, a message would identify those COMMON variables in a module which are initialized in BLOCK DATA. The user has the option to select the diagnostics he desires.

\(^1\)American National Standard FORTRAN, ANSI X3.9-1966.
Resources

The DAVE system consists of four programs (phases) which are executed sequentially. Information is passed between phases by intermediate files. (See fig. 2.) The largest phase of DAVE (PHASE2) requires 207 000 words of memory at LaRC. The execution time depends on the length (i.e., number of lines of code) and complexity of the subject program (e.g., number of subroutines and variables). Timing information for several test cases is provided in table I. For these test cases, the average execution time/line of input code was 0.09 sec on the Control Data CYBER 175 (or 0.45 sec/line on a CYBER 173) at LaRC. Execution time/line increases as the size of the subject program increases. There is no direct relationship between DAVE execution time and FORTRAN compilation time. In separate test cases, DAVE time ran from 25 times the compilation time (800 lines of code) to 130 times (7200 lines). A comparison of FORTRAN compilation time and DAVE execution time is shown in figure 3.

Experiences

DAVE was first installed at LaRC in 1976. An improved version, DAVE 8.0, was installed in 1977. Although this second version of DAVE performed well, its execution time was greater than desired. With the help of the University of Colorado, changes were made to DAVE, in the fall of 1977, that decreased execution time by as much as 47 percent. The new version works well on small and medium programs (i.e., less than 2000 lines of code). DAVE was originally written to accept only ANSI FORTRAN, a limitation which was not entirely satisfactory. While improving DAVE execution time, changes were made to handle some non-ANSI FORTRAN. The LaRC version of DAVE will now process the following non-ANSI input/output (I/O) features: NAMELIST, ENCODE/DECODE, and BUFFER IN/BUFFER OUT. In addition, most other non-ANSI statements now are treated as comments. An updated version of DAVE, which analyzes many widely used non-ANSI features of FTN, has become available. It is anticipated that this version will be installed at LaRC.

Approximately 15 people have used DAVE since its installation at LaRC. It has been used on several applications programs, including a cross-assembler, an ozone model program, and ITRAN, a structured preprocessor. DAVE typically found three to five errors in test programs that were syntactically correct, including some in production use. Generally, these errors were characterized by incorrect use of subroutine parameters. This particular type of error is difficult to detect by ordinary means. DAVE has found errors that are several levels deep. For example, a literal parameter was passed through three subroutine calls and then changed at the lowest level. A simplified example follows:

```
PROGRAM TEST
...
CALL SUB1(1.)
...
END
SUBROUTINE SUB1(A)
B=10.0
CALL SUB2(A,B,C)
```
RETURN
END
SUBROUTINE SUB2 (X, Y, Z)
  Z = Y**2
  X = 5.0*Z - Y
RETURN
END

In this example, the literal parameter (.) is passed from the main program to subroutine SUB1. SUB1 in turn passes it as the first parameter (A) in its call to subroutine SUB2. In SUB2, the first parameter (X) is not used as an input, but rather X is set to the value of a computed expression. This computed value will replace the present value of X, which is passed back to SUB1 and the main program. Thus, the value of the literal (.) in the main program is destroyed. This could cause some unknown and untraceable errors in the main program. An error of this type generally is not detected by the LaRC FTN compiler.

Other errors which have been detected include incorrect number of subroutine parameters in calling program, uninitialized variable passed as an input parameter to a subroutine, and disagreement on subroutine parameter type.

Evaluation

Most of our experiences with DAVE have been favorable. DAVE can be a cost effective verification tool whether used as an aid in the development or acceptance of new programs or in the maintenance of existing programs. Our experiences have shown DAVE to be an effective tool when used with small to medium sized programs. Execution time required for programs of this size is not unreasonable. As can be seen in figure 3 however, execution time increases significantly with program size for programs over 2000 lines long. Thus, DAVE is not recommended for use with long programs. One other drawback of DAVE (LaRC) is that some non-ANSI constructs still cause DAVE to halt. Among these are octal constants, variable names with more than six characters, and OVERLAY statements.

From the user's point of view, DAVE is easy to use. It requires only a few control cards and does not require user intervention or interaction during execution. On the surface, DAVE appears to give the same sort of diagnostics as a compiler. Actually, it has a multimodule capability not available on the FTN compiler and, thus, can detect errors not easily found by other means.

The overall evaluation is that DAVE is a useful analysis tool when used in the development and maintenance of small to medium FORTRAN programs. DAVE is easy to use and provides analysis diagnostics not available from other sources, thus saving analyst and computer time.
PET

Description

PET (Program Evaluator and Tester) is a package of programs designed as an automated aid to assist in the debugging, testing, and documentation of FORTRAN programs. PET is a proprietary program which was acquired by LaRC from the McDonnell Douglas Astronautics Company in the spring of 1977 as part of the MUST (Multipurpose User-Oriented Software Technology) program.

PET is a two-step process consisting of a preprocessor and a postprocessor. The preprocessor instruments the user's FORTRAN program based on the options specified. That is, PET translates the user's options into code which is inserted into the program. When the instrumented program is executed with test data, the PET inserted code generates an intermediate file of run-time statistics. These statistics include execution counts and information on assignment variables and DO-loop parameters. After execution, the postprocessor sorts through the run-time statistics and generates the requested reports. Figure 4 is an illustration of the way PET interacts with the user's program. The procedures to use PET at LaRC are described in the appendix.

PET generates several types of reports which can be used to determine the thoroughness of test cases. These reports are generated for each instrumented module and in summary for the whole program. The module reports include a syntactic profile, an operational profile, an assertion summary, and a detailed execution report. The syntactic profile indicates such things as the number of executable statements, the number of nonexecutable statements, and the number of comments. The operational profile gives such information as how many statements were executed and what percentage of the total executables were used. Syntactic and operational profiles are shown in figure 5. The detailed execution report gives execution counts for all lines of code and optionally reports the first/last and minimum/maximum values of assignments and first/last values of DO-loop parameters. An example of the execution report with these options is shown in figure 6. Summary reports include syntactic and operational profiles which are composites of the modular profiles, a module operational profile, a module execution summary, and a module timing summary with an optional relative timing histogram.

Figure 7 is an illustration of the way probes are inserted into the subject programs to keep track of execution counts. The PET options are controlled by PET directives which are inserted into the source code. These directives are implemented to look like FORTRAN comments. For example, selection of the option which reports the minimum/maximum values of assignment variables is accomplished by inserting the following option card into the user's program:

\[ \text{COPT=1} \]

This can be selected for all or part of a program. To change options, the user would simply insert a new COPT card. A complete description of the PET directives can be found in the PET user's manual (ref. 4).
Among the many features of PET are a modular capability and an assertion capability. The modular feature allows the user to optionally instrument all or any part of his program. This option allows existing instrumented code to be combined with new or modified modules. In the modular mode, only the changed modules are input to the preprocessor, resulting in reduced preprocess and compilation time. This particular feature is useful when testing only the modified parts of large programs.

The assertion capability allows user inserted statements to assert that computed values meet given criteria. Local (at a given line of code) and global (over a module) assertions are available. The assertion checking option is useful in debugging and in verifying that a program is operating correctly. PET will report any violations of these assertions and will give a summary of violations for each module. For example, the user may assert that a variable $T$ is greater than $-290.0$, whenever it is computed in a module by inserting the following comment card at the beginning of the module:

\[
\text{CASSERT VALUE}(T) (\text{GT.-290.0})
\]

The PET preprocessor will insert a test into the user's program wherever variable $T$ appears on the left side of an assignment statement. Whenever this assertion is violated during execution, a counter will be incremented. The postprocessor will sort these violations and will report the number and location of violations.

Resources

The PET preprocessor uses 145 000\text{g} words of memory and the postprocessor 62 000\text{g} words of memory for execution on a CYBER 173 at LaRC. Execution time for test cases ranged from 5 to 40 seconds for the preprocessor and from 3 to 10 seconds for the postprocessor on a CYBER 173. Run-time field length for the subject programs was increased by about 30 percent. This may vary slightly depending on the options selected. The execution time of most test programs was increased by about 20 to 50 percent. Table II is a summary of the resources used by our test cases.

Experiences

PET was installed at LaRC in 1977. The initial version of PET did not recognize certain FORTRAN constructs, such as certain sequences of multiple statements. These problems have since been corrected. The few limitations to PET are offset by its effectiveness as a software tool. These limitations are as follows:

(1) Increased execution time of the subject program

(2) Logical IF statements are rewritten as logical IF/NOT statements in instrumented code
Since its installation at LaRC, PET has been used by about 20 people. PET has provided some outstanding results on a variety of programs. For example, the use of PET output as an aid to pinpoint problem areas enabled a programmer to streamline one program extensively, reducing the cost/run by over 80 percent. In another example, PET was used as an aid to pinpoint those parts of a program executed most frequently. By making appropriate program changes, the programmer was able to cut execution time by a factor of 2.5 from 90 seconds to 36 seconds. The execution counts of PET were used to aid in debugging a program. These counts quickly revealed that an entire section of code was never executed, although it should have been. The assertion capability has been used as an aid in the conversion of an old program to FORTRAN Extended. In this instance, PET was used to pinpoint the range of a variable exponent causing an underflow condition and some extraneous error messages.

PET also can be used to check the effectiveness and completeness of test cases in determining if all possible branches and paths have been tested. This feature can be used during the development of new programs, in the acceptance of programs developed under contract, and during the testing of modifications made to existing programs to provide test coverage results.

Evaluation

Experiences with PET have shown it to be an effective analysis tool when used in the verification and testing of FORTRAN programs. PET is useful in both the development and maintenance of programs, providing an analysis capability not otherwise available. This includes such things as: the assertion capability, the relative timing of modules, and the run-time statistics. The flexibility of the modular mode allows instrumentation of new or changed modules as they are added to a program. The assertion capability is useful in detecting computation of erroneous data based upon user specified criteria. PET also provides testing coverage assessment with its run-time statistics, which allows a user to determine the effectiveness of test cases.

The PET system can be easy to use, depending on the options selected. The PET directives, which look like FORTRAN comments, are inserted into the user's code, requiring a minimal amount of program modification. Since the execution of the PET preprocessor and postprocessor is controlled by NOS procedure files, only a few extra control cards are needed by PET users. PET does have a few drawbacks. These include the increased execution time and storage requirements of the subject program, and the special handling required for OVERLAY programs. The modular mode, however, may be used to ease these problems.

CONCLUDING REMARKS

In this paper we have discussed the automated verification and testing tools, DAVE and PET, used in the analysis of FORTRAN programs at Langley Research Center (LaRC). These tools form an effective, complementary set of tools for testing FORTRAN programs. The overhead cost of using these tools may seem high, but this cost is offset by the effectiveness of DAVE and PET. In fact, the overhead incurred when using these tools is not unreasonable
since each of these tools need only be used a few times. Some of the types of errors found by these tools are not readily detected by other means. Resulting improvements made on test programs at LaRC have shown that these tools are cost effective aids in improving software quality.

Langley Research Center
National Aeronautics and Space Administration
Hampton, VA 23665
June 26, 1980
APPENDIX

PROCEDURES FOR USE OF DAVE AND PET AT

LANGLEY RESEARCH CENTER

DAVE

The DAVE system is simple to use. It requires only two inputs: (1) the source programs to be analyzed and (2) a file specifying the DAVE options desired. Available options include simulation of missing subprograms (SI) and suppression of specified diagnostics (SU). Through the SU parameter, the user can optionally suppress selected diagnostics or whole category of diagnostics. The maximum memory required by the DAVE system is 207,000 words. The DAVE system is accessed on the Network Operating System (NOS) at Langley Research Center (LaRC) by calling procedure file DAVE, using the following control cards:

GET, DAVE/UN=82487ON.
CALL(DAVE(INPUT=infl,OPTIONS=opfl)

where

infl = name of file containing the subject program
opfl = options file

The procedure file controls the execution of the four phases of DAVE and the transfer of files between these phases. A sample deck setup using DAVE would be:

GET, MYFILE.
GET, DAVE/UN=82487ON.
CALL(DAVE(OPTIONS=input, INPUT=MYFILE)
END OF RECORD
SI=ON
END OF FILE

In the example, the user's FORTRAN program is on file MYFILE. The user has selected the option SI=ON, which means that missing subroutines will be simulated by DAVE. Also, no diagnostics will be suppressed. Additional information is available in The DAVE System User's Manual (ref. 2).

PET

Usage of the PET system is controlled by two procedure files, PREPET and PSTPET. PREPET activates the preprocessor and, based on the options selected, the preprocessor inserts code into the programs which is used to generate an intermediate file of run-time statistics. After instrumenting the code, PREPET also controls compilation of the FORTRAN source code and, if specified, handles
APPENDIX

the modular features. Upon completion of PREPET, control is returned to the user for execution. After execution, the user calls PSTPET to activate the postprocessor. PSTPET automatically sorts the run-time statistics and generates the PET reports.

PET options are selected by inserting PET directives into the user's program. These directives look like FORTRAN comments with a C in column 1. There are three types of PET directives: (1) the COPT card which controls the type of run-time statistics, (2) the CONENDCASE card which controls when run-time statistics are to be collected, and (3) the CMOD card used with the modular mode.

The most frequently used directive is COPT which can be used to select or change the run-time statistics. The options which can be selected are:

<table>
<thead>
<tr>
<th>Option</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>min/max on assignment statements</td>
</tr>
<tr>
<td>2</td>
<td>min/max on DO-loop variable parameters</td>
</tr>
<tr>
<td>16</td>
<td>first/last on assignment statements</td>
</tr>
<tr>
<td>32</td>
<td>timing on modules</td>
</tr>
</tbody>
</table>

These options may be selected singly or in combination (by summing the options selected). For example, COPT=17 would give min/max and first/last values on assignment parameters. A complete description of the PET directives and the control cards needed to run PET is given in the PET user's manual (ref. 4).

An example of a nonmodular (sequential) case would be:

```
GET,PREPET,PSTPET/UN=824870N.
CALL(PREPET(Compile=INPUT))
```

user load and execute sequence
APPENDIX

```
CALL(PSTPET)
EOR
COPT=51

user's source program

EOR

program data

EOF
```

This example represents a sequential run in which the entire program is to be instrumented for PET option COPT=51. This is the combination of options 1, 2, 16, and 32.
REFERENCES


TABLE I.- DAVE TIMING STUDY

<table>
<thead>
<tr>
<th>Computer</th>
<th>DAVE test case</th>
<th>No. of blocks(^a)</th>
<th>Execution time, sec</th>
<th>Time/line of code, sec</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYBER 175</td>
<td>1</td>
<td>228</td>
<td>12.438</td>
<td>0.0545</td>
</tr>
<tr>
<td>CYBER 175</td>
<td>2</td>
<td>500</td>
<td>46.250</td>
<td>0.0925</td>
</tr>
<tr>
<td>CYBER 175</td>
<td>3</td>
<td>715</td>
<td>61.021</td>
<td>0.0853</td>
</tr>
<tr>
<td>CYBER 175</td>
<td>4</td>
<td>704</td>
<td>76.016</td>
<td>0.107</td>
</tr>
<tr>
<td>CYBER 173</td>
<td>5</td>
<td>704</td>
<td>290.075</td>
<td>0.412</td>
</tr>
<tr>
<td>6400(^b)</td>
<td>6</td>
<td>228</td>
<td>57.946</td>
<td>0.254</td>
</tr>
<tr>
<td>6400</td>
<td>7</td>
<td>965</td>
<td>592.301</td>
<td>0.613</td>
</tr>
<tr>
<td>6400</td>
<td>8</td>
<td>1579</td>
<td>2068.779</td>
<td>1.31</td>
</tr>
</tbody>
</table>

\(^a\)Note blocks are lines of executable code.
\(^b\)Control Data series 6400 computer system at LaRC.
### TABLE II. - PET RESOURCES STUDY

<table>
<thead>
<tr>
<th>PET test case</th>
<th>WITH PET</th>
<th></th>
<th>WITHOUT PET</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PREPET compile time, execution time, sec</td>
<td>FORTRAN compile time, execution time, sec</td>
<td>Test case compile time, execution time, sec</td>
<td>PREPET compile time, execution time, sec</td>
</tr>
<tr>
<td>2</td>
<td>15.639</td>
<td>17.366</td>
<td>8.335</td>
<td>1.94</td>
</tr>
<tr>
<td>3</td>
<td>38.942</td>
<td>29.94</td>
<td>1976.</td>
<td>4.703</td>
</tr>
<tr>
<td>4</td>
<td>5.553</td>
<td>11.13</td>
<td>.643</td>
<td>9.226</td>
</tr>
<tr>
<td>5</td>
<td>35.772</td>
<td>23.613</td>
<td>256.611</td>
<td>4.829</td>
</tr>
</tbody>
</table>

Note: Times are in central processing unit (CPU) seconds on a CYBER 173.
Figure 1.- Software tools used in verification and testing of programs.
Figure 2.- Steps in DAVE execution.
Figure 3.— Comparison of FORTRAN compilation time and DAVE execution time.
Figure 4.— PET interaction with user's program.
PROGRAM EVALUATOR AND TESTER REPORT

FOR MODULE PCH

SYNTACTIC PROFILE

TOTAL NUMBER OF SOURCE STATEMENTS 91

<table>
<thead>
<tr>
<th>TYPE OF STATEMENT</th>
<th>NUMBER</th>
<th>PERCENT OF TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXECUTABLE SOURCE</td>
<td>52</td>
<td>57.1</td>
</tr>
<tr>
<td>NONEXECUTABLE SOURCE</td>
<td>31</td>
<td>34.1</td>
</tr>
<tr>
<td>COMMENT</td>
<td>8</td>
<td>8.8</td>
</tr>
<tr>
<td>ASSERTION DIRECTIVE</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>NONSTANDARD</td>
<td>15</td>
<td>16.5</td>
</tr>
<tr>
<td>BRANCH</td>
<td>5</td>
<td>N/A</td>
</tr>
<tr>
<td>CALL</td>
<td>2</td>
<td>N/A</td>
</tr>
<tr>
<td>UNFORMATTED I/O</td>
<td>2</td>
<td>N/A</td>
</tr>
<tr>
<td>FORMATTED I/O</td>
<td>15</td>
<td>N/A</td>
</tr>
</tbody>
</table>

OPERATIONAL PROFILE

TOTAL EXECUTION COUNT 13755

<table>
<thead>
<tr>
<th>TYPE OF STATEMENT</th>
<th>NO EXECUTED</th>
<th>PERCENT EXECUTED</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXECUTABLE SOURCE</td>
<td>51</td>
<td>98.1</td>
</tr>
<tr>
<td>BRANCH</td>
<td>5</td>
<td>100.0</td>
</tr>
<tr>
<td>CALL</td>
<td>2</td>
<td>100.0</td>
</tr>
<tr>
<td>UNFORMATTED I/O</td>
<td>2</td>
<td>100.0</td>
</tr>
<tr>
<td>FORMATTED I/O</td>
<td>14</td>
<td>93.3</td>
</tr>
</tbody>
</table>

Figure 5.- PET syntactic and operational profiles.
Figure 6.- Sample PET output.
BEGIN
READ x,y
COUNT(1)
Z = 1
COUNT(2)

COUNT(8)

y = y/2
x = x*x
COUNT(7)

COUNT(5)

y mod 2 = 1

COUNT(3)

COUNT(6)

y = 0
COUNT(4) END

END

Figure 7.- Insertion of PET counters.
### Abstract

In recent years, several automated software verification and testing systems have been developed for use in the analysis of computer programs. Such tools not only discover errors in computer programs, but they save computer and analyst time and help to improve the quality of programs. This paper discusses two such tools, the static analyzer DAVE and the dynamic analyzer PET, which are used in the analysis of FORTRAN programs on the Control Data (CDC) computers at NASA Langley Research Center. The report describes the evaluation of these systems based on a set of test cases. These tools were found to be effective and complementary, and they both are recommended for use in testing FORTRAN programs.