Optimal Resource Allocation for Reliability of Modular Software Systems

K. WORWA, J. STANIK
kazimierz.worwa@wat.edu.pl

Institute of Computer and Information Systems
Faculty of Cybernetics, Military University of Technology
Kaliskiego Str. 2, 00-908 Warsaw, Poland

Considerable development resources are consumed during the software-testing phase. The software development manager has to decide how to use the testing-resources effectively in order to maximize the software quality and reliability. The paper discusses a management problem to achieve a reliable software system efficiently during the module testing stage by applying a software reliability growth model. This model both describes the software-error detection phenomenon and represents the relationship between the cumulative number of errors encountered by software testing and the time span of the testing. As testing cost and software reliability are both important factors in the testing-resource allocation problems an investigation is performed in the paper to search for the optimal solution for modular software system with the objectives of maximising system reliability and minimising testing cost.

Keywords: modular software systems, software testing, software reliability.

1. Introduction

It is well known that a software development process consists of the following four phases: specification (including user requirements definition and software functional and non-functional requirements specification), design (structural design and detailed design), coding and testing. The reciprocal relationships between these phases, including phase sequence and development activities range, depend on assumed software life cycle model.

Practically, a software testing process consists of several testing stages including module testing, integration testing, system testing and installation testing. During the testing phase, software faults can be detected and removed. The quality of the tests usually corresponds to the maturity of the software test process, which in turn relates to the maturity of the overall software development process. In general, most popular and commercial software products are complex systems composed of a number of modules. Typically, module testing is the most time-critical part of testing to be performed. All the testing activities of different modules should be completed within a limited time, and these activities normally consume approximately 40–50% of the total amount of software development resources [14]. Therefore, project managers should know how to allocate the specified testing-resources among all the modules and develop quality software with high reliability. During the unit-testing phase, all the testing activities of different modules are competing for the limited testing-resource, i.e. testing time and testing cost. Thus, a critical problem is how to allocate the total available program testing-resource among program modules in an optimal way. There are many papers that investigate this problem, e.g. [6] has presented an overview of the methods that have been developed for solving various reliability optimization problems including testing resource allocation problems; in paper [9] two resource allocation problems for software-module testing are formulated and solved, considering the mean number of remaining faults in the software modules; paper [7] studied a dynamic resource allocation strategy for software module testing; in paper [13] the problem of determining an optimal testing strategy is investigated.

The optimal testing-resource allocation problem is important, but difficult in project planning and management. It has been studied extensively in the literature (see e.g. [2, 6, 7, 9, 14]). When only limited resources are available during testing of a software system, it is important to allocate the testing resources efficiently, so that the maximum reliability of the complete system is achieved.

Testing cost and software reliability are both important factors in the testing-resource allocation problems. This paper discusses a management problem to achieve a reliable software system efficiently during module
testing stage by means of formulating and solving the testing-resource allocation optimization problem with the objectives of maximizing program reliability and minimizing the program testing cost. Section 2 introduces a formal model of a program testing process. This model considers the effect of testing-resources on the software reliability growth. Section 3 uses the program testing model to derive optimal allocation of the testing-resources for module testing. Finally, section 4 presents a numerical example to illustrate the method of determining an optimal program reliability structure.

2. Description of the Program Testing Process

We assume that the program under the testing process consists of several modules. Although there is no universally accepted definition of modularization, most programmers would conceive a module as a fragment of a program that carries out a specific function and may be used alone or combined with other modules of the same program. This usually implies that a module can be designed, implemented, and tested independently. Since a large-scale program always involves a substantial number of programmers working concurrently, a large program can be viewed as a collection of logically independent modules. A module is usually defined to perform a particular function. Let \( M \) mean a set of module numbers of the program under the testing, \( M = \{1, 2, 3, ..., M\} \). Let us define the reliability of a module as the probability that the module performs the function correctly, i.e., the module produces the correct output and transfers control to the next module correctly. When a set of user input is supplied to the program, a sequence of modules will be executed. The reliability of the output will depend on the sequence of modules executed and the reliability of each individual module. We first assume that the reliabilities of the modules are independent. This means that errors will not compensate each other, i.e., an incorrect output from a module will not be corrected later by subsequent modules. Since errors do not compensate each other, the result of the execution of the program is correct, if and only if the correct sequence of modules is executed and in every instance of module execution the module produces the correct result. The reliability of a module, in general, is a function of many factors, and the study of the reliability function of a module is beyond the scope of this paper. However, if no modification is made on the modules and the user environment does not change, the reliability function of a module should remain invariant. Let \( R_m \) stand for the \( m \)-th module reliability function. We will assume that the module reliability functions can be determined and the following vector \( R = (R_1, R_2, R_3, ..., R_M) \) will be called a program reliability structure.

We next assume that the transfer of control among program modules is a Markov process. This implies that the next module to be executed will depend probabilistically on the present module only and is independent of the past history. It is noteworthy that this assumption may not be valid for all types of programs. Although the assumption of Markovian behavior of control flow at the instruction level is questionable, experiments performed by researchers in memory management and scheduling have shown that this assumption may be valid at the macroscopic level, i.e. module level for a large number of programs [4, 8].

Let us represent the control structure of the program by a directed graph where every node \( i \) represents a program module and a directed branch \((i, j)\) represents a possible transfer of control from \( i \) to \( j \). To every directed branch \((i, j)\) we will attach a probability \( P_{ij} \) as the probability that the transition \((i, j)\) will be taken when control is at node \( i \). If \( P_{ij} = 0 \), the branch \((i, j)\) does not exist. This transition probability represents the branching characteristics of the decision point at the exit point of the module \( i \). Without loss of generality, let us assume that the program graph has a single entry node \( 1 \) and a single exit node \( M \). Let us consider every node in the graph as a state of the Markov process, with the initial state corresponding to the entry node of the program graph. Two states \( C \) and \( F \) are added as the terminal states, representing the state of correct output and failure, respectively. For every node \( i \), a directed branch \((i, F)\) is created with transition probability \( 1 - R_i \) representing the occurrence of an error in the execution of module \( i \). Since errors do not compensate each other, a failure in \( i \) will ultimately lead to an incorrect program output, regardless of the sequence of modules executed afterwards. This phenomenon is represented by the transition to the terminal state \( F \). The original transition probability between \( i \) and \( j \) is modified into \( R_i \cdot P_{ij} \), which represents the probability that the execution of module \( i \) produces the correct
result and control is transferred to module $i$. For the exit note $M$, a directed branch $(M, C)$ is created with a transition probability $R_m$ to represent correct termination at the exit node.

Assuming that the testing process is modelled by the above-described Markov process let $\overline{P}(R)$ mean the transition matrix of this process.

According to earlier assumptions the matrix $\overline{P}(R)$ has the form of (1). As we mentioned earlier vector $R$ is the so-called program reliability structure. The first two rows and two columns of the matrix $\overline{P}(R)$ represent absorbing states $C, F$ of the Markov process.

The reliability of the program is, therefore, the probability of reaching the terminal state $C$ from the initial state 1 of the Markov process. We can express the program reliability as a function of the reliability of the modules and frequency distribution of utilization of these modules. Using this approach, the reliability of the program is a function of both the deterministic properties of the structure of the program and the stochastic properties of the module failure behaviour and the utilization of these components by the user. Ideally, we would like a reliability model where the program structure, in terms of its modules, can be easily constructed, where the component reliabilities can be independently determined, and where a simple user profile can be measured easily by monitoring the dynamic behaviour of the program.

3. Optimization of the Testing-Resource Allocation Problem

The Markov reliability model uses the program flow graph to represent the structure of the system so that the structure can be easily obtained by analyzing the code. It uses functional modules as the basic components so that the component reliabilities are reasonably independent. It uses the branching characteristics among modules as the user profile so that they can be easily measured in the operational environment. It is noteworthy that similar structural models are often proposed in the literature, see e.g.[1, 3, 5, 8–10, 13]. The overall software reliability can be expressed in terms of individual module reliabilities and other parameters, such as visit statistics to each module.

Let $\mathbf{P}(R)$ be the matrix of order $M$ obtained from $\overline{P}(R)$ by deleting all the rows and columns corresponding to the absorbing states $C$ and $F$. According to earlier remarks reliability of the program is equal to the probability of reaching by the Markov process the terminal state $C$ from the initial state 1. Using a standard method of computing absorption probabilities in the Markov process reliability of the program can be computed as follows [1]

$$R(R) = S_{1M}(R)R_M,$$  \hfill (2)

where $S_{1M}$ is an element of the matrix $S = (I - P(R))^{-1}$ that lies on the intersection of the first row and the $M$-th column. $I$ is the identity matrix of order $M$ and $R_M$ is the reliability function of the $M$-th module.

It should be mentioned that this method of computing the program reliability coefficient as given in (1) can be quite inconvenient without the use of software that can perform the requisite
symbolic linear algebra, although packages such as Mathematica [12].

The cost of a module is associated with a number of parameters such as the design and development time, and testing time, and can be assumed to be monotonically related to the value of module reliability. The module cost – reliability relation can be linear, exponential, and even random while maintaining the monotonic property. We assume that the cost of the module consisting of a number of parameters such as the design and development time, and testing time, and can be assumed to be monotonically related to the value of module reliability. The module cost – reliability relation can be linear, exponential, and even random while maintaining the monotonic property. We assume that the cost function of module reliability is a monotonically related to the value of module reliability. The module cost – reliability relation can be linear, exponential, and even random while maintaining the monotonic property. We assume that the cost of the module consisting of a number of parameters such as the design and development time, and testing time, and can be assumed to be monotonically related to the value of module reliability.

Taking into account that values of objective functions \( R(R) \) and \( K(R) \) are inverse (in sense that if a value \( R(R) \) is increased, the value \( R(R) \) is increased, too) it is reasonable to expect that the dominate solution set will empty. In such a situation a practically recommended approach is to determine a nondominated solution set. If this set is very numerous we can narrow it down by determining a so-called compromise solution [11]. The program reliability structure \( R^* \) that to be obtained after the solving the optimization problem (4)–(8) will both maximize the value of the program reliability function \( R(R) \) and minimize the value of the program testing cost \( K(R) \).

4. Numerical Example

The methodology of determination of the program reliability structure that has been presented can be illustrated by the following numerical example. Let Fig. 1 be the directed graph representing the control structure of a program with five modules where \( I \) represents the input module and 5 is the output module. The set of a program component modules is \( M = \{1, 2, 3, 4, 5\} \) and the program reliability structure is a vector \( R = (R_1, R_2, R_3, R_4, R_5) \).

![Fig. 1. A program control graph](image)

Table 1 presents branching probabilities \( p_{ij} \) between the modules \( i \) and \( j \).

<table>
<thead>
<tr>
<th>Table 1. Values of branching probabilities between the modules ( i ) and ( j )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p_{11} = 0 )</td>
</tr>
<tr>
<td>( p_{21} = 0 )</td>
</tr>
<tr>
<td>( p_{31} = 0 )</td>
</tr>
<tr>
<td>( p_{41} = 0.3 )</td>
</tr>
</tbody>
</table>

Using (1) the value of the program reliability function can be determined as \( R(R) = S_{15}(R)R_5 \), where quantity \( S_{15} \) can easily be computed by means of software packages Mathematica [12] as follows.
Let the cost of the m-th module testing cost, and the following condition

\[ \text{complies with } \{0.95, 1.00, 0.95, 1.00, 1.00\} \]

As a result of the normalization (10)–(11) both normalized objective functions \( \overline{K(R)} \) and \( \overline{R(R)} \) have values belong to range \([0, 1]\).

It is easy to notice that in the example that is considered the ideal point \((\overline{K(R)}, \overline{R(R)})\) is of the form of \((0, 1)\).

Table 2 presents the results of determination of the optimisation problem (3)–(7).

The row that corresponds to the vector \(\textbf{R}^*=(1.00, 1.00, 0.95, 1.00, 1.00)\) is the compromise solution of this problem, i.e. such a vector that is nearest to the ideal point \((0, 1)\), where a distance function \(d((\overline{K(R)}, \overline{R(R)}), (0,1))\) is of the form of

\[ d((\overline{K(R)}, \overline{R(R)}), (0,1)) = \sqrt{[\overline{K(R)}]^2 + [\overline{R(R)} - 1]^2} \]

(15)

It is easy to notice that the vector \(\textbf{R}^*=(1.00, 1.00, 0.95, 1.00, 1.00)\) complies with the following condition

\[ d((\overline{K(R^*)}, \overline{R(R^*)}), (0,1)) = \min_{R \in R} d((\overline{K(R)}, \overline{R(R)}), (0,1)) \]

(16)

For the program control graph from Fig. 1 and parameter values in Table 1 we will solve the bicriterial optimization problem (4)–(8). A feasible solution set for this problem is a set of module reliability structure vectors \( \textbf{R}=(R_1, R_2, R_3, R_4, R_5) \).

If we assume that \( R_m \in [0.95, 1.00] \), \( m \in \{1, 2, 3, 4, 5\} \), a cardinality number of this set is \(6^5=7776\). It is easy to check that the dominated solution set of the problem (4)–(8) is empty (it is because of objective functions \(R(R)\) and \(K(R)\) are inverse in sense their values). In that case practically recommended approach is to determine a nondominated solution set. The nondominated solution set of the bicriterial optimization problem (4)–(8) is plotted in Fig.2. Every point among points that are situated on the shadow contour is a nondominated solution of the bicriterial optimization problem (4)–(8). In order to narrow down this a nondominated solution set we will determine a compromise solution of this problem, i.e. such a solution belongs to the nondominated solution set that is nearest (in the sense of Euclidean distance) to the so-called ideal point [11]. For this reason both objective functions \(R(R)\) and \(K(R)\) determined by (6) and (7), respectively, will be normalized by means of the following formulae:

\[ \overline{K(R)} = \frac{K(R) - K_{\min}(R)}{K_{\max}(R) - K_{\min}(R)}, \]  
\[ \overline{R(R)} = \frac{R(R) - R_{\min}(R)}{R_{\max}(R) - R_{\min}(R)}, \]

(11)  
(12)  

where:

\[ K_{\min}(R) = \min_{R \in R} K(R), \quad R_{\min}(R) = \min_{R \in R} R(R), \]  
(13)  

\[ K_{\max}(R) = \max_{R \in R} K(R), \quad R_{\max}(R) = \max_{R \in R} R(R), \]

(14)

where \( \textbf{R} \) is a set of all feasible program reliability structure, i.e. \( \textbf{R} = \{R_1, R_2, R_3, R_4, R_5\} : 0 \leq R_m \leq 1, m \in M = \{1, 2, 3, 4, 5\} \} \).

Table 2 presents the results of determination of the bicriterial optimization problem (4)–(8). The row that corresponds to the vector \(\textbf{R}^*=(1.00, 1.00, 0.95, 1.00, 1.00)\) is the compromise solution of this problem, i.e. such a vector that is nearest to the ideal point \((0, 1)\), where a distance function \(d((\overline{K(R)}, \overline{R(R)}), (0,1))\) is of the form of \((0, 1)\).

Table 2 presents the results of determination of the bicriterial optimization problem (3)–(7). The row that corresponds to the vector \(\textbf{R}^*=(1.00, 1.00, 0.95, 1.00, 1.00)\) is the compromise solution of this problem, i.e. such a vector that is nearest to the ideal point \((0, 1)\), where a distance function \(d((\overline{K(R)}, \overline{R(R)}), (0,1))\) is of the form of \((0, 1)\).

Table 2. Values of parameters \(KS_m, \alpha_m, \beta_m, R_{min}\) and \(K_{max}\) be as in Table 2.

<table>
<thead>
<tr>
<th>(m)</th>
<th>(KS_m)</th>
<th>(\alpha_m)</th>
<th>(\beta_m)</th>
<th>(R_{min})</th>
<th>(K_{max})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10000</td>
<td>50.0</td>
<td>5.0</td>
<td>0.8</td>
<td>160000</td>
</tr>
<tr>
<td>2</td>
<td>4000</td>
<td>35.0</td>
<td>2.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1500</td>
<td>25.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2500</td>
<td>30.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>9000</td>
<td>40.0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(m \in \{1, 2, 3, 4, 5\}\) and \(R_m \in [0.95, 1.00]\).
The values of the objective functions \( R(R) \) and \( K(R) \) for the compromise solution
\[
R^* = (1.00, 1.00, 0.95, 1.00, 1.00)
\]
are \( R(R^*) = 0.8711 \) and \( K(R^*) = 158,481.43 \).

![Fig. 2. The nondominated solution set of the optimization problem (4)–(8)](image)

Table 2. The compromise solution of the optimization problem (4)–(8)

<table>
<thead>
<tr>
<th>( R_1 )</th>
<th>( R_2 )</th>
<th>( R_3 )</th>
<th>( R_4 )</th>
<th>( R_5 )</th>
<th>( R(R) )</th>
<th>( K(R) )</th>
<th>( d(\overline{K(R)}, R(R), (0,1)) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>0.5888</td>
<td>0.77</td>
<td>0.876889</td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>0.99</td>
<td>0.99</td>
<td>1.00</td>
<td>0.5981</td>
<td>0.78</td>
<td>0.874058</td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>0.99</td>
<td>1.00</td>
<td>0.95</td>
<td>0.7705</td>
<td>0.88</td>
<td>0.909792</td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>0.99</td>
<td>1.00</td>
<td>0.96</td>
<td>0.7821</td>
<td>0.88</td>
<td>0.908320</td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>0.99</td>
<td>1.00</td>
<td>0.97</td>
<td>0.7936</td>
<td>0.88</td>
<td>0.907053</td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>0.99</td>
<td>1.00</td>
<td>0.98</td>
<td>0.8052</td>
<td>0.88</td>
<td>0.905995</td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>0.99</td>
<td>1.00</td>
<td>0.99</td>
<td>0.8168</td>
<td>0.89</td>
<td>0.905147</td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>0.99</td>
<td>1.00</td>
<td>1.00</td>
<td>0.8284</td>
<td>0.89</td>
<td>0.904515</td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.95</td>
<td>0.95</td>
<td>0.2088</td>
<td>0.51</td>
<td>0.942643</td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.95</td>
<td>0.96</td>
<td>0.2145</td>
<td>0.51</td>
<td>0.938661</td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.95</td>
<td>0.97</td>
<td>0.2201</td>
<td>0.52</td>
<td>0.934730</td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.95</td>
<td>0.98</td>
<td>0.2258</td>
<td>0.52</td>
<td>0.930854</td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.95</td>
<td>0.99</td>
<td>0.2315</td>
<td>0.52</td>
<td>0.927034</td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.96</td>
<td>0.2372</td>
<td>0.52</td>
<td>0.923273</td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.96</td>
<td>0.95</td>
<td>0.2720</td>
<td>0.59</td>
<td>0.939384</td>
</tr>
</tbody>
</table>

5. Conclusions

This paper proposes a method of determining the optimal program reliability structure for the purpose of the program testing stage realization. The optimal program reliability structure that is to be obtained after solving the bicriteria optimization problem (4)–(8) both maximizes the value of the program reliability function \( R(R) \) and minimizes the value of the program testing cost \( K(R) \).

For formulating and solving the optimal testing-resource allocation problem, the following assumptions are made, which are valid in most cases:

- a software program under testing is composed of several modules, which are developed and tested independently during the unit-testing phase
- in the unit-testing phase, each program module is subject to failures at random times caused by faults remaining in the module and the failures of the modules are independent
- a module structure of a program under testing is known and a transfer of control among program modules follows a Markov process and the exchanges of controls among these modules are characterized according to the rules of a Markov process;
the next transfer of control to be executed is independent of the past history and depends only on the present module

- the probability of transition from one module to another is determined from the operational profile of a system
- the reliability of a module-based program depends on the architecture of the program and the reliabilities of their component modules
- the cost of a module function is monotonically related to the value of the module reliability.

Knowledge of the optimal reliability structure of the program under testing allows the software testing manager to rationally decide how to use the limited testing resources effectively in order to maximize the program reliability and minimize the program testing cost. On the basis of the optimal program reliability structure the manager knows how to allocate the specified amount of testing-resource expenditures for each program module to achieve both the maximal program reliability growth and minimal cost of the program testing.

The method of determining the program reliability structure that has been proposed was illustrated by the simple numerical example. For assumed module structure of a program, the optimal program reliability structure has been obtained as a compromise solution of the bicriteria optimization problem (4)–(8).

6. Bibliography

Optimalizacja alokacji nakładów w procesie wytwarzania programu o znanej strukturze modularnej

K. WORWA, J. STANIK

W artykule przedstawiona jest metoda określania struktury niezawodnościowej programu, rozumianej jako wektor wskaźników niezawodności jego modułów składowych. Modelem rozpatrywanego programu jest graf przepływu sterowania, w którym prawdopodobieństwa uaktywniania poszczególnych modułów składowych w procesie wykonywania programu wynikają z tzw. profilu operacyjnego programu, charakteryzującego rzeczywiste środowisko jego pracy. Struktura niezawodnościowa wyznaczana jest w wyniku rozwiązania określonego zadania programowania matematycznego. Znajomość struktury niezawodnościowej programu umożliwia właściwe zaplanowanie nakładów czasowo-finansowych, wymaganych dla wytworzenia programu, spełniającego założone wymagania niezawodnościowe. Zastosowanie przedstawionej metody zilustrowane zostało przykładem liczbowym.

Słowa kluczowe: struktura modułowa programu, testowanie oprogramowania, niezawodność oprogramowania.