Software Reliability Prediction Model Analysis

L. Mirtskhulava, M. Khunjgurua, N. Lomineishvili, B. Bakuria

Abstract—Software reliability prediction gives a great opportunity to measure the software failure rate at any point throughout system test. A software reliability prediction model provides with the technique for improving reliability. Software reliability is very important factor for estimating overall system reliability, which depends on the individual component reliabilities. It differs from hardware reliability in that it reflects the design perfection. Main reason of software reliability problems is high complexity of software. Various approaches can be used to improve the reliability of software. We focus on software reliability model in this article, assuming that there is a time redundancy, the value of which (the number of repeated transmission of basic blocks) can be an optimization parameter. We consider given mathematical model in the assumption that in the system may occur not only irreversible failures, but also a failure that can be taken as self-repairing failures that significantly affect the reliability and accuracy of information transfer. Main task of the given paper is to find a time distribution function (DF) of instructions sequence transmission, which consists of random number of basic blocks. We consider the system software unreliable; the time between adjacent failures has exponential distribution.

Keywords—Exponential distribution, conditional mean time to failure, distribution function, mathematical model, software reliability.

I. INTRODUCTION

An important quality attribute of a computer system is the degree to which it can be relied upon to perform its intended function. Evaluation, prediction, and improvement of this attribute have been of concern to designers and users of computers from the early days of their evolution. Until the late 1960’s, attention was almost solely on the hardware related performance of the system. In the early 1970’s, software also became a matter of concern, primarily due to a continuing increase in the cost of software relative to hardware, in both development and the operational phases of the system [1]-[3].

First serious effort on software reliability started at Bell Laboratories in 1964. Nonetheless, some of the important works that appeared in the 1960s were by Haugk, Tsang, and Zimmerman, Floyd, Hudson, Barlow and Scheuer, London, and Sauter [2].

Software is essentially an instrument for transforming a discrete set of inputs into a discrete set of outputs. It comprises of a set of coded statements whose function may be to evaluate an expression and store the results in a temporary or permanent location, decide which statement to execute next, or to perform input/output operations [1].

It is well known that fixing a fault in a program becomes increasingly expensive in later phases of software development. It is much more cost effective to fix as many faults as possible before releasing a program. Unfortunately, because it becomes harder to detect a fault as the software becomes more reliable, the cost of testing also increases [2], [3]. Moreover, as the lines of code increase, the point, testing is no longer cost-effective and the software has to be released. It has also been observed that modular testing is a testing effort required to fix a fault grows super-linearly [2]. Hence, modular testing with fewer lines of code would significantly reduce the overall effort required for testing.

Conventional approaches to software reliability growth modeling are black-box based, i.e. the software is treated as a black-box and its interactions with the external world are modeled [4], [5]. Tests are generated from the specified functional properties of the program based on its operational profile [6]-[8]. The internal structure of the program is not taken into account while generating the test cases. A stochastic model is calibrated using the failure data collected during the functional testing of the software, and this model is then used to predict the reliability of the software, and to determine when to stop testing. Thus the black-box approach relies on the assumption that the software is tested as per its operational profile.

Software reliability is affected by many factors during the life cycle of a software product, from the definition of the product to the operation and maintenance. All the activities within the software development life cycle are prone to introduce faults [9]-[11].

II. SOFTWARE RELIABILITY METRICS

- Reliability metrics are units of measure for system reliability
- System reliability is measured by counting the number of operational failures and relating these to demands made on the system at the time of failure
- A long-term measurement program is required to assess the reliability of critical systems [11]-[14].

Software reliability has been defined as the probability of a software product to insure operating without failure in a
specified environment, for a given amount of time. Based on this definition, software reliability prediction has been defined as a forecast of how reliable a software product will be in the future, based on data available so far. Software reliability tends to improve over testing and operating time, due to errors being removed. This is the reason for the models being named reliability growth models [15]-[17].

For software products, continuous availability is necessary, and reliability is an important component of this. Even so, there can be defects in software products that cause system failure. In order to avoid these situations and to decrease support expenses, companies want to deliver to the client’s reliable software. Developing reliable software is one of the hardest problems of the IT industry. Pressure brought by schedule, resource limitations and unrealistic demands can negatively impact the reliability. Knowing the reliability of a delivered product is a difficult issue. After reaching the clients, the reliability is indicated by the feedback coming from them, under the form of reports, complaints, or compliments. But, by this time, it is too late to change anything: that is why companies selling software want to know ahead of time the product reliability. Reliability models try to do this.

The most important cause of defects in software is bugs, which means incorrect implementations. Even the most talented programmers produce software with defects. The software products complexity is too big, at this moment, to be handled by people. With all the progress in programming techniques, such as splitting the programs in small modules, using evolved programming languages and complex developing tools, results are still far away from perfection, and the programming productivity has not increased significantly in the past two decades.

The most unpredictable defects in software manifest only after a specific combination of values for input data or certain external events that were not predicted by the programmer. Such combinations appear with low probability during normal testing procedures, so they often make it to the operating phase. Also, new versions are built on older versions, fixing defects found and adding new functionality. Even so, the process of fixing defects often introduces new defects, because the effects of a fix have unpredictable consequences. Defect prediction deals with estimating the number of defects. Although other terms have been used to describe it, such as estimation, fault estimation, we should clarify the difference between the two notions. Defect estimation has been defined by Nayak as a process of identifying different types of defects of a software product, aiming to reach high quality. However, defect prediction helps in estimating the quality of a product before it is released [2].

III. SOFTWARE REQUIREMENTS

A. Functional and Nonfunctional Requirements

Functional requirements describe the functions that the software is to execute; for example, formatting some text or modulating a signal. They are sometimes known as capabilities. Nonfunctional requirements are the ones that act to constrain the solution. Nonfunctional requirements are sometimes known as constraints or quality requirements. They can be further classified according to whether they are performance requirements, maintainability requirements, safety requirements, reliability requirements, or one of many other types of software requirements. These topics are also discussed in the Software Quality KA [8].

B. Quantifiable Requirements

Software requirements should be stated as clearly and as unambiguously as possible, and, where appropriate, quantitatively. It is important to avoid vague and unverifiable requirements which depend for their interpretation on subjective judgment ("the software shall be reliable"; "the software shall be user-friendly"). This is particularly important for nonfunctional requirements. Two examples of quantified requirements are the following: a call center’s software must increase the center’s throughput by 20%; and a system shall have a probability of generating a fatal error during any hour of operation of less than 1 * 10−8. The throughput requirement is at a very high level and will need to be used to derive a number of detailed requirements. The reliability requirement will tightly constrain the system architecture.

C. System Requirements and Software Requirements

In this topic, system means “an interacting combination of elements to accomplish a defined objective. These include hardware, software, firmware, people, information, techniques, facilities, services, and other support elements.” as defined by the International Council on Systems Engineering (INCOSE00). System requirements are the requirements for the system as a whole. In a system containing software components, software requirements are derived from system requirements. The literature on requirements sometimes calls system requirements “user requirements.” The Guide defines “user requirements” in a restricted way as the requirements of the system’s customers or end-users. System requirements, by contrast, encompass user requirements, requirements of other stakeholders (such as regulatory authorities), and requirements without an identifiable human source.

Reliability is defined as the probability that a device will perform its required function under stated conditions for a specific period of time. Predicting with some degree of confidence is very dependent on correctly defining a number of parameters. For instance, choosing the distribution that matches the data is of primary importance. If a correct distribution is not chosen, the results will not be reliable. The confidence, which depends on the sample size, must be adequate to make correct decisions. Individual component failure rates must be based on a large enough population and relevant to truly reflect present day normal usages. There are empirical considerations, such as determining the slope of the
failure rate and calculating the activation energy, as well as environmental factors, such as temperature, humidity, and vibration. Lastly, there are electrical stressors such as voltage and current.

IV. RELIABILITY DEMONSTRATION TESTING

A reliability demonstration test is a test to determine if a component or system has achieved a specified level of reliability. Typically, a test plan specifies a test environment, operational profile, test duration and the number of permissible failures. The system or component is then operated according to the plan and the number of observed failures recorded. The component is rejected if the number of observed failures is above the number of permissible failures. The system or component is then operational profile, test duration and the number of observed failures recorded. The component is rejected if the number of observed failures is above the number of permissible failures [15].

V. MODEL DESCRIPTION

The overall system reliability can be estimated from the individual component reliabilities by modelling the interactions between the components using one or more of the building blocks.

In order to estimate the reliability of software, we consider a mathematical model in the assumption that in the system may occur not only irreversible failures, but also a failure that can be taken as self-repairing failures that significantly affect the reliability and accuracy of information transfer. We focus on software reliability model in this article assuming that there can be taken as self-repairing failures that significantly affect the reliability and accuracy of information transfer. We focus on software reliability model in this article assuming that there is a time redundancy, the value of which (the number of repeated transmission of information) is a parameter that allows to optimize.

We introduce the following notations. Assume:

- $n$ is the number of the blocks. A sequence of instructions forms a basic block if: a) the instruction in each position dominates, or always executes before, all those in later positions, and b) no other instruction executes between two instructions in the sequence.
- $\Phi_j(t)$ - the probability distribution function of time of the instruction sequences transmission, which consists of $n$ basic blocks if transmission is started from $j$-th block ($j = \overline{1,n}$);
- $F_i(u)$ - the distribution function of transmitted block length; $r$-1-the quantity of basic block transmission repetition before repair
- $\alpha$-error rate
- $G(u)$- the distribution function of recovery.

Main task of the given paper is to find a time distribution function (DF) of instructions sequence transmission, which consists of random number of basic blocks.

When creating the given model, we used the following assumptions (A):

- We consider the system software unreliable; the time between adjacent failures has exponential distribution;
- Block Coverage: It is the Number of the basic blocks that have been executed by the test cases. A basic block, or simply a block, is a sequence of instructions that. The instructions in any basic block are either executed all together, or not at all.

Where the block length distribution low has the form:

$$F_j(u)=\Phi_1(t(t-\tau_j)), \ j=\overline{1,n}, \text{ where } \tau_j \cdot \text{ time of transmission of } j\text{-th block;}$$

Control of correctness of information is made after the reception of the next block and the time spent on control is considered negligible.

Under assumptions (A) in case of failure origin block transmission is repeated until receiving accurate block but no more than $r$-1 times. System need to be repaired after $r$ - times repetition. The block transmission is renewed after the repair from distorted block repetition.

System software model, starting with $j$-th block transmission is described by the following integral equations:

$$\Phi_{j+1}(t) = \Phi_{j}(t) + \int_{0}^{t} \alpha \Phi_{j}(t-u) du + \int_{0}^{t} \alpha \Phi_{j}(t-u) du$$

$$\Phi_{j+1}(t) = \Phi_{j}(t) + \int_{0}^{t} \alpha \Phi_{j}(t-u) du + \int_{0}^{t} \alpha \Phi_{j}(t-u) du$$

where $j = \overline{1,n}$

The boundary condition has the following form:

$$\Phi_{n+1}(t) = 1$$

(3)

Taking the Laplace-Stieltjes transform to (1), (2) u (3), we obtain:

$$\tilde{\Phi}_j(s) = \tilde{\Phi}_{j+1}(s) + \int \left[ s + \alpha \right] \tilde{\Phi}_{j+1}(s)$$

(4)

$$\tilde{\Phi}_{j+1}(s) = \tilde{\Phi}_{j+1}(s) + \int \left[ s + \alpha \right] \tilde{\Phi}_{j+1}(s)$$

(5)

Denote:

$$\tilde{\Phi}_j(s) = \tilde{\Phi}_j(s) = \int e^{-\alpha t} d\Phi_j(t)$$

$$\tilde{\Phi}_j(s) = \int e^{-\alpha t} d\Phi_j(t) dt$$
\[
\bar{g}(s) = \int_{0}^{\infty} e^{-st}dG(t);
\]

Let \(\alpha = \alpha = \alpha; \quad F_j(u) = F_j(u) = F_j(u)\) and for \(j = n:\)

\[
\Phi_n^{(r-1)}(S) = \frac{\gamma(s + \alpha)}{S} + \left[\gamma(S) - \gamma(s + \alpha)\right]g(s)\Phi_1^{(i)}(S);
\]

for \(i = r-2:\)

\[
\Phi_n^{(r-2)}(S) = \frac{\gamma(s + \alpha)}{S} + \left[\gamma(S) - \gamma(s + \alpha)\right]g(s)\Phi_1^{(i)}(S);
\]

In (6) instead of \(\Phi_n^{(r-1)}(S)\) substituting its value, we obtain:

\[
\Phi_n^{(r-3)}(S) = \frac{\gamma(s + \alpha)}{S} + \left[\gamma(S) - \gamma(s + \alpha)\right]g(s)\Phi_1^{(i)}(S);
\]

For \(i = r-3:\)

\[
\Phi_n^{(r-3)}(S) = \frac{\gamma(s + \alpha)}{S} + \left[\gamma(S) - \gamma(s + \alpha)\right]g(s)\Phi_1^{(i)}(S);
\]

Finally:

\[
\Phi_n^{(i)}(S) = \frac{\gamma(s + \alpha)\left[\gamma(S) - \gamma(s + \alpha)\right]^{r-i}}{\gamma(S) - \gamma(s + \alpha) - 1} + \left[\gamma(S) - \gamma(s + \alpha)\right]g(s)\Phi_1^{(i)}(S);
\]

For \(j = n-1:\)

\[
\Phi_n^{(i)}(S) = \frac{\gamma(s + \alpha)\Phi_{n-1}^{(i)}(S) + \gamma(S) - \gamma(s + \alpha)\Phi_{n}^{(i+1)}(S)}{\gamma(S) - \gamma(s + \alpha) - 1} + \left[\gamma(S) - \gamma(s + \alpha)\right]g(s)\Phi_1^{(i)}(S);
\]

For \(i = r-2:\)

\[
\Phi_n^{(r-2)}(S) = \frac{\gamma(s + \alpha)\Phi_{n-1}^{(i)}(S) + \gamma(S) - \gamma(s + \alpha)\Phi_{n}^{(i+1)}(S)}{\gamma(S) - \gamma(s + \alpha) - 1} + \left[\gamma(S) - \gamma(s + \alpha)\right]g(s)\Phi_1^{(i)}(S);
\]

as a result, we obtain:

\[
\Phi_n^{(i)}(S) = \frac{\gamma(s + \alpha)\left[\gamma(S) - \gamma(s + \alpha)\right]^{r-i}}{\gamma(S) - \gamma(s + \alpha) - 1} \times \Phi_{n}^{(i+1)}(S) + \left[\gamma(S) - \gamma(s + \alpha)\right]g(s)\Phi_1^{(i)}(S);
\]

For \(i = 1:\)

\[
\Phi_1^{(1)}(S) = a(S)\Phi_{n}^{(i)}(S) + b(S)\Phi_1^{(i)}(S)\ldots
\]

where

\[
a(S) = \frac{\gamma(S) - \gamma(s + \alpha)}{\gamma(S) - \gamma(s + \alpha) - 1};
\]

first-order difference equation (8) is solved by successive substitution:

for \(j = n:\)

\[
\Phi_n^{(i)}(S) = \frac{\gamma(s + \alpha)\Phi_{n-1}^{(i)}(S) + \gamma(S) - \gamma(s + \alpha)\Phi_{n}^{(i+1)}(S)}{\gamma(S) - \gamma(s + \alpha) - 1} + \left[\gamma(S) - \gamma(s + \alpha)\right]g(s)\Phi_1^{(i)}(S);
\]

under \(j = n-1:\)

\[
\Phi_{n-1}^{(i)}(S) = \frac{\gamma(s + \alpha)\Phi_{n-2}^{(i)}(S) + \gamma(S) - \gamma(s + \alpha)\Phi_{n-1}^{(i+1)}(S)}{\gamma(S) - \gamma(s + \alpha) - 1} + \left[\gamma(S) - \gamma(s + \alpha)\right]g(s)\Phi_1^{(i)}(S);
\]

under \(j = n-2:\)

\[
\Phi_{n-2}^{(i)}(S) = \frac{\gamma(s + \alpha)\Phi_{n-3}^{(i)}(S) + \gamma(S) - \gamma(s + \alpha)\Phi_{n-2}^{(i+1)}(S)}{\gamma(S) - \gamma(s + \alpha) - 1} + \left[\gamma(S) - \gamma(s + \alpha)\right]g(s)\Phi_1^{(i)}(S);
\]

under \(j = n-1:\)

\[
\Phi_{n-1}^{(i)}(S) = \frac{\gamma(s + \alpha)\Phi_{n-2}^{(i)}(S) + \gamma(S) - \gamma(s + \alpha)\Phi_{n-1}^{(i+1)}(S)}{\gamma(S) - \gamma(s + \alpha) - 1} + \left[\gamma(S) - \gamma(s + \alpha)\right]g(s)\Phi_1^{(i)}(S);
\]

Finally, we obtain:

\[
\Phi_1^{(i)}(S) = \frac{\gamma(s + \alpha)\Phi_{0}^{(i)}(S) + \gamma(S) - \gamma(s + \alpha)\Phi_{1}^{(i+1)}(S)}{\gamma(S) - \gamma(s + \alpha) - 1} + \left[\gamma(S) - \gamma(s + \alpha)\right]g(s)\Phi_1^{(i)}(S);
\]

Conditional mean time to failure:

\[
-\tau_m = a\Phi_1^{(i)}(S)_{\|} = b\Phi_1^{(i)}(S)_{\|} = \frac{a^n(s)[a(s) - 1]}{[a(s) - 1][a^n(s) - a^n(s - 1)]}b(s)_{\|}
\]
where
\[ a(0) = 1 - b(0) ; \quad b(0) = \left[ 1 - f(a) \right]^{-1} \]

After solving (12):
\[ -\frac{T_m}{T} = \frac{r_b + \left( 1 - \frac{1}{a(0)} \right)^{-1} \left[ r_a f(a) - r_b \right]}{f(a) b(0)} \]
\[ \times \left( \frac{1}{a(0)} - 1 \right) r \geq 3 \]

where
\[ r = -g'(0) ; \quad r_b = f'(0) \]

VI. CONCLUSION

In this paper, we have investigated a queuing-based model for software analysis, which allows investigating the problem of determination of the probability characteristic of time of the blocks performance with account of failure origin. We focused on software reliability model assuming that there is a time redundancy for recovery of the system, the value of which (the number of repeated transmission of the sequences) is a parameter that allows to optimize.

We have found a time distribution function (DF) of instructions sequence transmission, which consists of random number of basic blocks.

VII. BACKGROUND AND RELATED WORK

This section provides a brief overview of relevant background and related studies in the analysis of software reliability prediction.

Dhillon in his book “Applied Reliability and Quality: Fundamentals, Methods and procedures”, Chapter 7 presents some aspects of software reliability evaluation models. In this chapter two mathematical models: Mills and Musa models are presented. First model was developed by arguing that the faults remaining in a given software program can be estimated through a seeding process that makes an assumption of a homogeneous distribution of a representative class of faults. Thus, both seeded and unseeded faults are identified during reviews or testing and the discovery of seeded and unseeded faults permits an assessment of remaining faults for the fault type in question. Second model is based on the premise that reliability assessments in the time domain can only be based upon actual or real execution time, as opposed to elapsed or calendar time, because only during execution a software program really becomes exposed to failure-provoking stress. Some of the important assumptions pertaining to this model are as follows: Failure intervals follow a Poisson distribution and are statistically independent; failure rate is proportional to the remaining defects; execution times between failures are piecewise exponentially distributed [1]-[10].

There is a definite need for reliability and quality professionals working in diverse areas to know about each other’s work activities because this may help them, directly or indirectly, to perform their tasks more effectively. At present to the best of author’s knowledge, there is no book that covers both applied reliability and quality within its framework. More specifically, at present to gain knowledge of each other’s specialties, these specialists must study various books, articles, or reports on each of the areas in question. This approach is time consuming and rather difficult because of the specialized nature of the material involved [1].

My previous work is devoted to study and evaluation of data transmission through unreliable wireless channel, subjected to distortions on the physical layer. The time between neighboring failures is distributed according to Erlang ratio. The method of enhance of reliability of transmission through unreliable wireless channel (WCH) is suggested [18]. My previous work describes the study of special Erlang distribution model in wireless networks and mobile computing. Demonstrates the applicability of the Erlang distribution, where queueing model is considered as wireless channel where the interarrival times between failures have the Erlang Distribution [19]. In my previous works Modeling based approach is described for analyzing and evaluating Internet server system reliability and availability in this paper. In the given model the states are defined by the different kinds of failures of the server system. The Markov model evaluates the probability of jumping from one known state into the next logical state. The probabilities between transitions of the states are a function of the failure rates of the transitional probabilities from one to another state. The number of first-order differential equations is equal the number of the states of the servers. First-order differential equations are developed by describing the probability of being in each state in terms of states of the model [20].

REFERENCES


