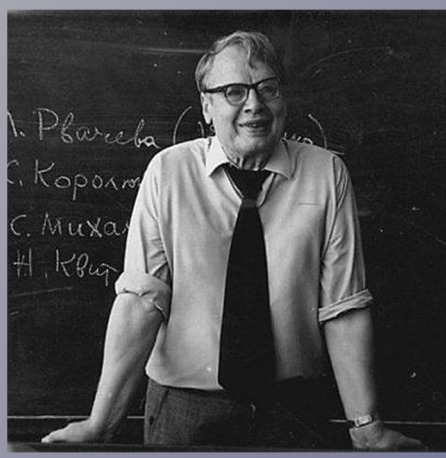
ELECTRONIC JOURNAL
OF INTERNATIONAL GROUP
ON RELIABILITY

Gnedenko Forum Publications



JOURNAL IS REGISTERED
IN THE LIBRARY
OF THE U.S. CONGRESS

•

ISSN 1932-2321

•

VOL.7 NO.3 (26)
SEPTEMBER, 2012

RELIABILITY:
THEORY&APPLICATIONS

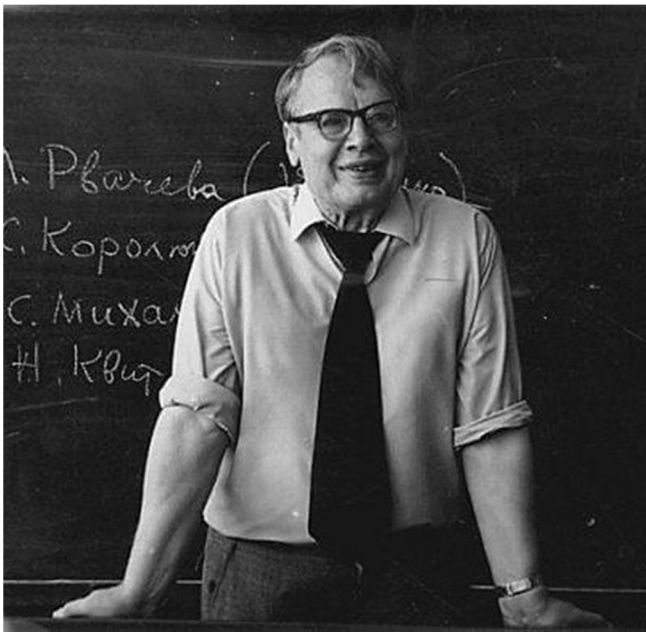
San Diego

Special Issue 4

ISSN 1932-2321

http://www.gnedenko-forum.org/Journal/index.htm





RELIABILITY:
THEORY & APPLICATIONS

Vol.7 No.3 (26),
September, 2012

Special Issue 4

100th anniversary of Boris Vladimirovich Gnedenko's birthday

San Diego
2012

Send your paper

e-Journal ***Reliability: Theory & Applications*** publishes papers, reviews, memoirs, and bibliographical materials on Reliability, Quality Control, Safety, Survivability and Maintenance.

Theoretical papers have to contain new problems, finger practical applications and should not be overloaded with clumsy formal solutions.

Priority is given to descriptions of case studies.

General requirements for presented papers

1. Papers have to be presented in English in MSWord format. (Times New Roman, 12 pt , 1.5 intervals).
2. The total volume of the paper (with illustrations) can be up to 15 pages.
3. A presented paper has to be spell-checked.
4. For those whose language is not English, we kindly recommend to use professional linguistic proofs before sending a paper to the journal.

* * *

The Editor has the right to change the paper title and make editorial corrections.

The authors keep all rights and after the publication can use their materials (re-publish it or present at conferences).

Publication in this e-Journal is equal to publication in other International scientific journals.

Papers directed by Members of the Editorial Boards are accepted without referring.

The Editor has the right to change the paper title and make editorial corrections.

The authors keep all rights and after the publication can use their materials (re-publish it or present at conferences).

Send your papers to

the Editor-in-Chief ,
Igor Ushakov
igusha22@gmail.com

or

the Deputy Editor,
Alexander Bochkov
a.bochkov@gmail.com

Table of Contents

An approach to ensure the reliability of engineering systems at design stage is considered in this paper. This approach is associated with construction of an acceptable region inside the system parameter space. A model that describes an acceptable region constructed on the basis of multidimensional grid is offered. The methods for reducing amount of data with respect of resource limitations and particulars of data decomposition for its parallel processing are described.

In this paper a problem of asymptotic estimate for connectivity probability of non oriented connected graph with fold and low reliable arcs is solved. An algorithm of a calculation of asymptotic constants with cubic complexity by a number of nodes is constructed. This algorithm is based on Kirchhoff`s theorem for a calculation of a number of spanning trees and relative characteristics.

This paper is a review integrating, amending, and developing the approach applied in authors' previous works devoted to the tensile and fatigue reliability analysis of unidirectional composite material considered as a series system the links of which are, in general case, complex parallel systems with redistribution of load after failure of some items. By processing experimental data it is shown that the models based on the Markov chains (MCh) theory allow (1) to describe connection of cdf of tensile strength of fibers (strands) and a composite specimen, (2) to perform nonlinear regression analysis of fatigue curve and prediction of its changes due to a change of tensile strength characteristics of the composite components, (3) to predict the fatigue life at a program loading, (4) to estimate the cdf of the residual strength and residual life after a preliminary fatigue load.

This paper summarizes the result of an effort to develop a unified approach to design-driven testability evaluation of avionic systems. These systems include both internal diagnostic equipment referred to as built-in-test (BIT) and external off-line test equipment. At the designing stage an adequate database to evaluate the quality of the BIT is the failure mode and effect analysis. In the paper various mathematical indices are suggested and constructed to quantify testability of avionic systems. The indices provide the needed flexibility for representing structural and reliability properties of the controlled system. Analytical model for evaluation BIT performance impact on the system's reliability is discussed.

In this paper limit theorems for closed queuing networks with excess of servers are formulated and proved. First theorem is a variant of the central limit theorem and is proved using classical results of V.I. Romanovskiy for discrete Markov chains. Second theorem considers a convergence to chi square distribution. These theorems are mainly based on an assumption of servers excess in queuing nodes.

The Method of Universal Generating Functions (U-functions) was introduced in [Ushakov, 1986]. Since then the method has been developed, in first order, by my friends and colleagues – Gregory Levitin and Anatoly Lisnianski. They actively and successfully apply the method of U-function to optimal resources allocation, to multi-state system analysis and other problems. Frankly, now I feel like a hen sat on duck eggs and then wanders how hatched chicks fearlessly swim so far from shore. I decided to remind you a Russian folk proverb: "new is well forgotten old". What is U-function? It is, first of all, generalization of a classical Generation Function (GF) permitting perform more general transforms. From technical side, this method represents a modification of the Kettelle's Algorithm conveniently arranged for calculations with the use of computer.

Dependability assessment is typically based on complex probabilistic models. Markov and semi-Markov models are widely used to model dependability of complex hardware/software architectures. Solving such models, especially when they are stiff, is not trivial and is usually done using sophisticated mathematical software packages.
We report a practical experience of comparing the accuracy of solutions stiff Markov models obtained using well known commercial and research software packages. The study is conducted on a contrived but realistic cases study of computer system with hardware redundancy and diverse software under the assumptions that the rate of failure of software may vary over time, a realistic assumption. We observe that the disagreement between the solutions obtained with the different packages may be very significant. We discuss these findings and directions for future research.

SOFTWARE RELIABILITY. NON-PROBABILISTIC APPROACH

Dmitry A. Maevsky, Helen D. Maevskaya, Alexander A. Leonov
•
Odessa National Polytechnic University, Odessa, Ukraine
e-mail: toe-onpu@ukr.net

ABSTRACT

The article describes the main provisions of the new theory of software reliability, which is not based on probability theory and the theory of non-equilibrium processes. Emerging from the operation of software systems, defects are considered as the result of the forward and reverse defect flows. Relations are developed to predict the number of identified and introduced to system defects and they are opening the possibility of modeling the reliability of software systems, taking into account the secondary defects. It is shown that the majority of existing software reliability models can be derived from the provisions of the dynamics of software systems.

1 INTRODUCTION

Software reliability is the most confusing and intriguing area of general reliability theory. On the early stages of its development this theory was based on the probabilistic reliability concepts. The main features of software reliability are: stochastic nature of failures, time dependence of failures and independence of failures from other ones. However, various attempts to create a single universal model that describes defects exposure law on this conception have failed. Now there are more than twenty different models that are trying to describe the same physical process – software defects exposure. Naturally, such diversity shows that this theory requires a thorough revision.

One of the most influencing reliability experts Igor Ushakov wrote (Ushakov 2006): “Errors caused by software have no stochastic nature: they will repeat as soon as some conditions will be repeated. Errors of software, in a sense, are not “objective” – they depend on type of operations, type of inputs and, at last, on type of users”. And later: “… attempts to put “hardware reliability shoes” on “software legs” are absolutely wrong and, moreover, will lead only to a logical dead end”.

Other opinion through software reliability is: “It should be stressed that so far the theory of software reliability can’t be regarded as an established science. … one can ascertain the presence of a substantial gap between theory (mathematical models and methods) and practice” (Kharchenko at al. 2004).

Six years ago, in 2012, Igor Ushakov wrote (Ushakov 2012): “One thing is clear: software reliability specialists should distinguish their reliability from hardware reliability, develop their own non-probabilistic and non-time dependent mathematical tools”.

This article is devoted to the new non-probabilistic approach to the software reliability problem.

2 DYNAMIC THEORY OF THE SOFTWARE SYSTEMS: FUNDAMENTALS

Theory of the Software System Dynamics (SSD) considers a software system (SS) as an open non-equilibrium system that interacts with the environment. Subject area of the SS is considered as the environment. Non-equilibrium system is a system which has gradients of certain properties of

the system, such as concentration, temperature, etc. In SSD the number of defects in the systems at any given time is considered as such property. In the general theory of non-equilibrium processes of physical nature of the subject matter of these properties, which are called "thermodynamic potentials" (Onsager 1931), does not matter. The only important thing is that their gradients that play a role of forces exist in the system. Under the influence of these forces there are flows that are designed to bring the system to equilibrium with its environment. The dynamics of such a system are determined by the spatio-temporal distribution of these flows, with their values at each physical point.

For the SS the concept of "space area" is possible only in the sense of "within" or "outside" the system and the notion of a physical point generally cannot be used. Therefore, with respect to the SS one can only talk about the patterns of distribution of flows over time. The openness of the SS is determined by the nature and extent of its relationship with the environment, which serves as the subject area of the system, and the level of equilibrium is determined by the number of defects contained in the system. In this case subject area itself is accepted as the etalon, that is, by definition it does not contain defects.

Let's denote the number of defects contained in the SS at the specific time t as f or $f(t)$.

SSD is based on the following hypothesis:

1. SS is an open non-equilibrium system that interacts with its subject area according to the laws of the non-equilibrium processes.
2. The state of the SS is characterized by a special state function – the number of the defects contained in it.
3. Disappearance and appearance of defects in the SS is the result of the joint action of the direct (output) and backward (incoming) defect flows.
4. The intensity of each flow is proportional to the number of defects, forming the flow.
5. All defects are equal and participate in the formation of the flow in the same way, regardless of the causes, location, and type of defect and the possible consequences of its manifestation (the principle of equality).
6. Function $f(t)$ is differentiable on the whole domain (the principle of continuity).

The basic concept SSD is the concept of software defect flows. Each defect is seen as an integral part of the total flow, which obeys not the laws of the theory of probability but the laws of emergence and evolution of the flows in non-equilibrium systems. Emergence of the defect flows in the SS is shown at Figure 1.

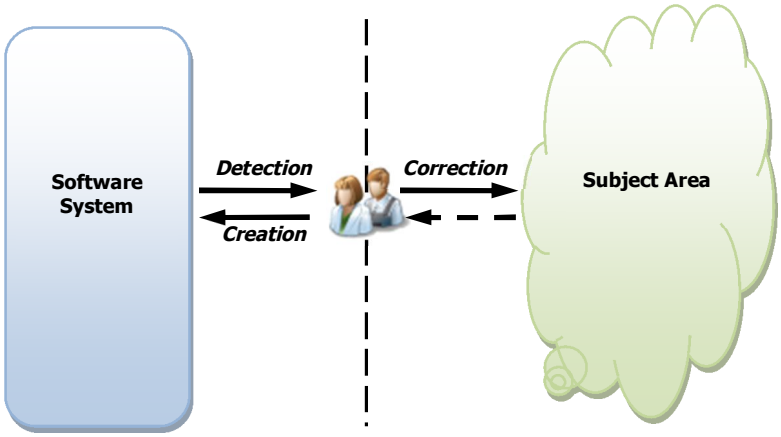


Figure 1. Defect emergence in the SS.

During operation of the SS defects lead to the fact that result, which produces its software, does not meet the outcome that is expected by subject area. This discrepancy is detected by the user which is in contact on the one hand with the SS, on the other – with its subject area. Thus, the user acts as the first, error detector, and secondly – a sort of "contact surface" between the SS and its

subject area. We assume that the user is ideal, that is, detects and records each defect at the time its manifestation.

In the process of fixing the defect disappears from the SS due to changes made in its code. This loss can be considered as a result of the removal of defects from the SS. Considering this process in time, we have the flow of defects from the SS out through the "contact surface" - the user. These streams are shown in Figure 1 are shown by arrows "Detection" and "Correction".

In the process of fixing defects in the SS is possible to introduce additional "secondary" defects. The process of introduction of the secondary defect may be regarded as the second, counter-flow of defects, which operates in the direction from the subject area to the SS.

The flow of defects will be numerically characterized by the speed (intensity) of the flow, which can be determined by hypothesis 6 (principle of continuity). Taking into account only the output stream, SS is characterized by a number of defects, which are contained in the system – coordinate *f(t)*. It can be considered as having only one degree of freedom and is described by the differential equation of first order. In the case of taking into account of the second process - introduction secondary defects, its coordinate is their current number – $f_2(t)$. In general, taking into account both processes we obtain two coordinates, which characterize the effect of defects in a software system – $f_1(t)$ and $f_2(t)$. SS in this case should be considered as a system with two degrees of freedom and described by differential equations of second order. On this basis, we introduce the concept of the order of the SS model.

Definition. The order of the SS model is the order of a differential equation, which in this model describes the variation of the number of defects over time.

In the non-equilibrium dynamics by the flux vector ***j*** of some value *f* we consider a vector, whose modulus is equal to the value *f* transferred during a unit of time through the unit area *dS* perpendicular to the direction of the transfer *x*:

$$j = \frac{df}{dt \cdot dS},$$

and the direction – the same as the direction of transport [6]. The very flow of value *f* in this case corresponds to the integral

$$J = \int_S j \cdot dS = \frac{df}{dt}.$$

This expression makes it possible to avoid the use of the concept of area, indeterminable for the SS.

In reliability theory, the value of *j* has a corresponding failure rate λ:

$$\lambda = \frac{df}{dt}.$$

3 DEFECT FLOW IN THE FIRST-ORDER SS MODEL

The one-dimensional system – is the simplest case SSD. It is assumed by almost all currently available software reliability models, based on the traditional theory of reliability. Therefore, walkthrough of the dynamics of the SS will start with such one-dimensional case.

We assume that the software system has only the direct flow of defects, ie, the flow is directed from the SS. The statement of the uniqueness of the flow is equivalent to the following two assumptions of most well-known software reliability models (Lyu 1996):

- when an error occurs it is corrected before the discovery of the next;
- new defects are not introduces during the fixing of existing ones.

Indeed, the first assumption is actually equivalent to the presence of the defects flow (they are fixed, that is, cease to exist, are removed from the PC), and the second assumption says that there is no flow of secondary defects.

The flow of any scalar value in a non-equilibrium system only occurs by the action of the driving forces behind this flow (Prigogine 1991). As a driving force in continuous systems there is a gradient of the potential of the corresponding value, and in discontinuous – the difference of potentials at the contact boundary. SS, as shown in Figure 1, should be seen as a discontinuous - there are defects within the system, and in the environment they are absent. From this it follows that the potential at the contact boundary changes abruptly. Given the lack of defects in the external environment, we can take the potential of defects of this medium to be zero. Then, according to [9], the flow of primary defects, being the value of f_1, can be represented as:

$$\frac{df_1}{dt} = -G_1 \cdot \varphi_1 , \tag{1}$$

where G_1 – aspect ratio, and φ_1 – the potential of the defects in the SS. "Minus" sign in the formula (1) says that the flow is directed toward decreasing the potential, that is, from the SS to the external environment.

Between potential φ_1 and its corresponding value of f_1 there is a relation

$$f_1 = C_1 \cdot \varphi_1 , \tag{2}$$

where ratio C_1 will be called as a defect capacity of the system regarding to the value f_1. Therefore, considering (2), defined

$$A_1 = \frac{G_1}{C_1},$$

the equation (1) can be represented as

$$\frac{df_1}{dt} = -A_1 \cdot f_1 . \tag{3}$$

Let us explain the physical meaning of the coefficients G_1 and C_1 for the SS. In the theory of non-equilibrium processes the coefficient G_1 is called the conductivity of the system with respect to value f_1. From equation (1) it follows that for a constant value φ_1 the rate of detection of the primary defects is directly proportional to size G_1. In the real SS rate of detection of defects is directly proportional to the frequency of calls to the system. Therefore, the conductivity G_1 in (1) can be interpreted as the frequency of user calls to the SS. In this case we mean an "ideal user", each time specifying a different, in general case random set of input data. In fact, approaching the ideal can be considered as staff members, each member of which works with its narrow set of data sets. Thus, as the conductivity G_1 in the SS we take the rate of the access, and conductivity itself has a dimension of s^{-1}. Potential φ_1, because of this, must be dimensionless.

Defect capacity C_1 shows how a number of the defects in the SS should increase to that their potential φ_1 grows by one. Keeping in mind that φ_1 is dimensionless; defect ration is dimensionless as well. Defect ration of the SS can be understood as the maximum possible defect number which can be contained in the analyzed system.

Equation (3) is a homogeneous linear differential equation. Its solution can be obtained in the form

$$f_1(t) = F_0 \cdot e^{-A_1 \cdot t} , \tag{4}$$

where F_0 – initial number of defects in SS at the start of research.

According to the formula (4) the number of defects that remain to SS at the time t shall be calculated. As shown (Lyu 1996), the most convenient for experimental determination of the dependence of the total number defects identified in the system at the same time (cumulative number of defects μ). To calculate μ we can use

$$\mu(t) = \int_0^t \lambda(\tau) d\tau, \tag{5}$$

therefore:

$$\mu(t) = F_0 - F_0 \cdot e^{-A_1 t}. \tag{6}$$

Note that the formula (4) and (6) fully comply with similar expressions for the most famous models of software reliability. This suggests that these models are consistent with the theory of first-order SSD.

4 DEFECT FLOWS IN THE SECOND-ORDER SS MODEL

In the case taking into account the secondary flow of defects, SS has two degrees of freedom and is characterized by two coordinates – the number of defects f_1, which will be removed from the system and the number of the secondary defects f_2. The connection between the flows of primary and secondary defects is represented by the system of equations:

$$\begin{cases} \dfrac{df_1}{dt} = -G_{11} \cdot \varphi_1 - G_{12} \cdot \varphi_2 \\ \dfrac{df_2}{dt} = -G_{21} \cdot \varphi_1 - G_{22} \cdot \varphi_2 \end{cases}. \tag{7}$$

In this system, φ_1 – the potential of removed defects, and φ_2 – potential for insertion of the secondary. Ratios G_{11} and G_{22} characterize the influence of potentials φ_1 and φ_2 on the flows related to them. By analogy with previous statements, these factors play a role of conductivity and characterize the frequency of accesses to the system. The frequency of entering the secondary defects into software system tends to be lower than the frequency of detection of the primary. On this basis, it can be said that $G_{11} > G_{22}$. We call these ratios the intrinsic conductivities of the SS.

Ratios G_{12} and G_{21} characterize the influence of potentials φ_1 and φ_2 on the flows related to them. According to the Onsager symmetry principle (Onsager 1931), these cross-effects are the same, which leads to the equality $G_{12}=G_{21}$. Ratios G_{12} are G_{21} will be called mutual conductivities.

Potentials φ_1 and φ_2 are associated with the corresponding values of f_1 and f_2 by the relations of the form (2):

$$f_1 = C_1 \cdot \varphi_1; f_2 = C_2 \cdot \varphi_2,$$

where C_1 – defect capacity of the SS related to the primary defects, and C_2 – defect capacity of the same system related to the secondary ones. Obviously, if we are talking about the same system, then these two should be equal: $C_1=C_2$.

Using the relation between the number of defects and the corresponding potential, taking into account the equality $G_{12}=G_{21}$ and defining

$$A_1 = \frac{G_{11}}{C_1} = \frac{G_{22}}{C_2}, A_2 = \frac{G_{12}}{C_1} = \frac{G_{21}}{C_2},$$

system (7) can be re-written as:

$$\begin{cases} \dfrac{df_1}{dt} = -A_1 \cdot f_1 - A_2 \cdot f_2 \\ \dfrac{df_2}{dt} = -A_2 \cdot f_1 - A_1 \cdot f_2 \end{cases}. \tag{8}$$

The system (8) is an autonomous system of differential equations whose solution is fully determined only by the initial conditions and to determine the time variation of the existing substation primary and secondary defects.

It should be noted that the flow described by the first equation of the system is the flow of defects **carried out** from this SS, rather than the primary flow. In fact, trapped in a system of secondary defects are indistinguishable but the primary and along with them are removed from SS (Lyu 1996). In this sense we can say that the division of defects existing in the SS to primary and secondary is purely arbitrary. They differ only in the moment of introduction, but impact on the state of SS in the same way.

The solution of (8) for the outgoing stream of defects is an expression

$$f_1 = F_0 \cdot e^{-A_1 t} \cdot cosh(A_2 t). \quad (9)$$

Comparing (9) with (4) obtained for the output stream of defects without a countering input flow, we can see that it differs by the presence of the factor $cosh(A_2 t)$, whose role is to adjust the output stream of defects by the countering flow of the secondary ones.

To interpret and analyze the results, Figure 2 shows plots of the number of defects that remain in the system from time to time for different ratios $k=A_2/A_1$. These curves are plotted for a hypothetical software system with the following parameters: initial number of defects $F_0=100$, value of the ratio $A_1=100\ days^{-1}$, ratio k varies from 0 to 1,1. Here $k = 0$ corresponds to the complete absence of the secondary flow of defects, and the value $k = 1$ – case, where the correction of one of the primary defect is accompanied by the introduction of a second. For values of $k>1$, the number of secondary defects exceeds the number of fixed ones.

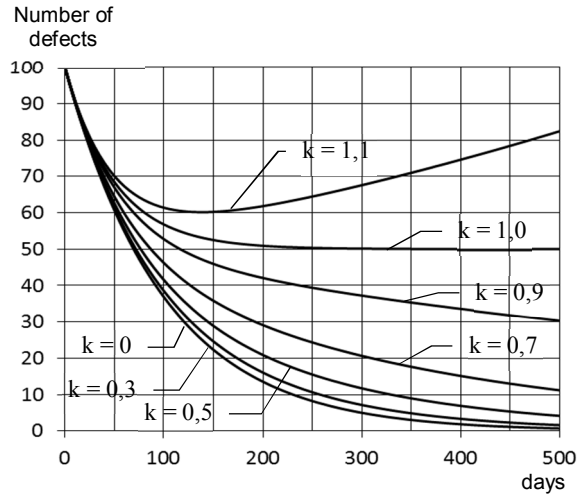


Figure 2. The relation of the number of defects in the SS through time

Analyzing the relations following conclusions can be made:

– In the absence of secondary defects (k = 0), formula (9) coincides with formula (4), obtained without regard to their influence. This coincidence with reality, which may indicate the correctness of the basic statements of the SSD.

– The influence of secondary defects reduces to increasing the decay time of the output flow. Thus, the SSD theory confirms intuitive assertion that in case of introduction of the secondary defects to SS, the total time of their identification increases.

– With $k=1$ the number carried out from SS defects stabilizes and tends to the value of $F_0/2$. Non-evident interpretation of this fact we will give later.

The solution of (7) for output flow of defects is an expression

$$f_2 = -F_0 \cdot e^{-A_1 t} \cdot sinh(A_2 t). \quad (10)$$

The sign of "minus" in (10) can be explained on the basis of differences in directions of output and input flows. However, given the fact that the number of defects cannot be negative, in the future, when determining the number of secondary defects "minus" sign will be omitted.

Figure 3 shows plots of relation $f_2(t)$, built for the same hypothetical SS for different values of the coefficient k. Analyzing the relations presented in Figure 3, the following conclusions can be made:

– With $k=0$ there is no secondary defects flow.

– With $0<k<1$ the number of secondary defects being introduces in the SS has a maximum, which is expressed the more the bigger value of k is.

– The rate of increase of the number of secondary defects is most important at the initial stage, before reaching the maximum. After that, the number of secondary errors tends to zero, but with a much slower rate.

– With $k=1$ number of secondary of defects decreases with time, which corresponds to processes in the real SS, and can serve as a confirmation of the SSD.

– With $k=1$ the number of defects introduced into the SS is stabilizing and tends to the value $F_0/2$.

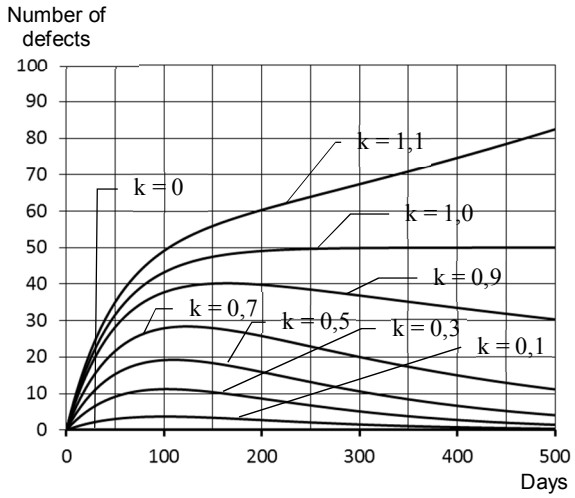


Figure 3. The relation of the number of secondary defects in the SS through time

At any arbitrary point in time the number present in SS of defects can be calculated as the sum of the number of defects that will be removed from it (f_1) and of the number of the already introduced secondary defects (f_2). For a plot of this relation we have to simply sum the corresponding curves from the plots in Figure 2 and 3. The result of this adding is shown in Figure 4. As can be seen from Figure 4, provided k = 1, ie, when the number of introduced secondary defects equals the number of corrected, the residual amount of defects in the SS remains unchanged. Now it is clear why, when k = 1, the values of f1 (t) and f2 (t) tend to the value $F_0/2$. Indeed, in this case their sum is at any given time is equal to the initial number of defects – F_0, which fully corresponds to the physical representations of the processes that must occur in SS at a given condition.

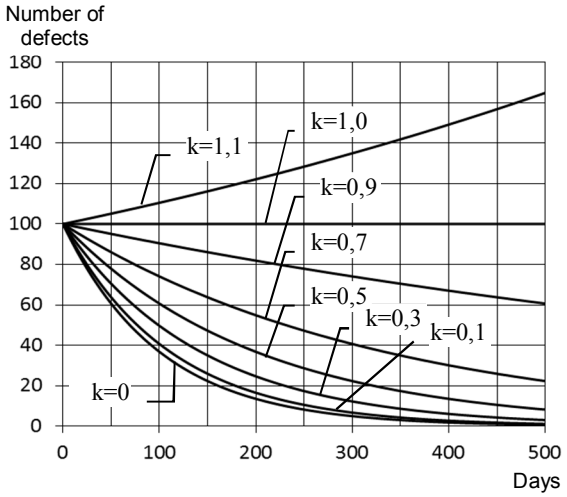


Figure 4. The relation of the total number of defects in the SS through time

5 RELIABILITY MODEL BASED ON THE SSD

For the Software Reliability Model (SRM) creation we have to define a set of input data, write the mathematical relationships that define the reliability and bring the method of determining the coefficients in these dependences.

As input data of the model developed on the basis of SSD time series giving the cumulative number of detected defects are adopted. Despite the fact that the defects form flows with well-defined laws, the process of their identification has a significant uncertainty (Kharchenko at al. 2004). Therefore, consideration for modeling of each individual defect complicates the analysis of the results by having a considerable "noise". Due to this, SRMs, input data of which are points of identification of every defect, cannot ensure the accuracy of the simulation, due to the fact that their input data are already inaccurate. Time series, forming a cumulative number of defects is more accurate, because a random registration or non-registration of each specific defect cannot affect the overall trends in this series. In fact, the time series formed by the cumulative number of defects is relieved of the random component and is a trend.

On the basis of foregoing, for the construction of the model mathematical relationships derived from the theory of SSD should be converted to operate with cumulative defects trends.

On the basis of the formula (5), for a cumulative trend of output flow we obtain the expression:

$$\lambda_1(t) = \int_0^t \frac{F_0}{2}\left[(A_1 + A_2)\cdot e^{(A_2 - A_1)t}\right] + \int_0^t \frac{F_0}{2}\left[(A_1 - A_2)\cdot e^{-(A_2 + A_1)t}\right]. \qquad (11)$$

It is not difficult to see that having $A_1=A_2$ primitive of $\lambda_1(t)$ does not exist, since the difference between the A_1 and A_2 is zero. Therefore, finding the cumulative of the number of defects of the original flow we consider separately for the two cases.

Case 1. $A_1 \neq A_2$.

In this case, the primitive for $\lambda_1(t)$ always exists, therefore after integration we obtain:

$$\mu_1(t) = \frac{F_0}{2}\left[\left(\frac{A_2 + A_1}{A_2 - A_1}e^{A_2 t} + \frac{A_2 - A_1}{A_2 + A_1}e^{-A_2 t}\right)e^{-A_1 t}\right] - F_0 \cdot \frac{A_1^2 + A_2^2}{A_2^2 - A_1^2}. \qquad (12)$$

Case 2. $A_1 = A_2$.

In this case, before finding a primitive we transform the expression (11), given that $A_1=A_2$. Therefore:

$$\lambda_1(t) = F_0 \cdot A_1 ,$$

and:

$$\mu_1(t) = F_0 \cdot A_1 \cdot \int_0^t dt = F_0 \cdot A_1 \cdot t . \quad (13)$$

Expression (13) having $A_1=A_2$ correlates well with the expected result. Indeed, with each defect, which is removed from SS, there is one secondary defect that is introduced into it. Therefore, the total number of defects in the SS remains unchanged and the frequency of making defects, because of this, too, remains unchanged. Thus, when $A_1=A_2$ linear relationship of the cumulative number of defects removed through time is expected, and is derived from the SSD.

For the cumulative trend of input flow (secondary defects), we obtain:

$$\mu_2(t) = \frac{F_0}{2}\left(\frac{A_2 + A_1}{A_2 - A_1} e^{A_2 t} - \frac{A_2 - A_1}{A_2 + A_1} e^{-A_2 t}\right) e^{-A_1 t} - 2F_0 \cdot \frac{A_1 \cdot A_2}{A_2^2 - A_1^2} \quad (14)$$

having $A_1 \neq A_2$ and

$$\mu_2(t) = F_0 \cdot A_1 \cdot t \quad (15)$$

having $A_1=A_2$.

Comparing (13) and (15) can be seen that for $A_1=A_2$ cumulative trends in output and input flows are the same. This result also corresponds to the physical representations. If $A_1=A_2$ then the number of defects, which are introduced into the system equals the number of removed ones. From this it follows that their cumulative trends, too, must be the same.

Thus, the mathematical model of software reliability are the expressions (12), (13) and (14), (15) for the outgoing and input flow, respectively. For practical application of reliability models it is necessary to develop a methodology for calculating the parameters of the model based on experimental data.

In the experimental determination of parameters of the model cumulative trend of defects identified at a certain time interval acts as the experimental data. Subject to determination of model parameters are the influence coefficients A1 and A2, as well as the initial number of defects in the system F_0.

Determination of the parameters will be carried out in two stages. The first step is a preliminary assessment of the parameters, while the second - is their clarification.

For a preliminary assessment, we assume that the input stream of defects is absent, so the coefficient $A_2=0$. In this case, the cumulative relationship of these defects will be exponential

$$\mu_1(t) = F_0 - F_0 \cdot e^{-A_1 t} . \quad (16)$$

In of this relation the inverse of the coefficient A1, is called time constant of the process

$$\tau = \frac{1}{|A_1|} ,$$

and is the length of sub-tangent exponentially. In turn, the value of sub-tangent can be defined as shown in Figure 5.

From Figure 5 it follows that magnitude of sub-tangent τ can be defined as:

$$\tau = \frac{F_0 - f_1}{tg\alpha},$$

where the slope of the $tg\alpha$ can be determined from the formula

$$tg\alpha \approx \frac{f_1 - f_2}{t_1 - t_2}.$$

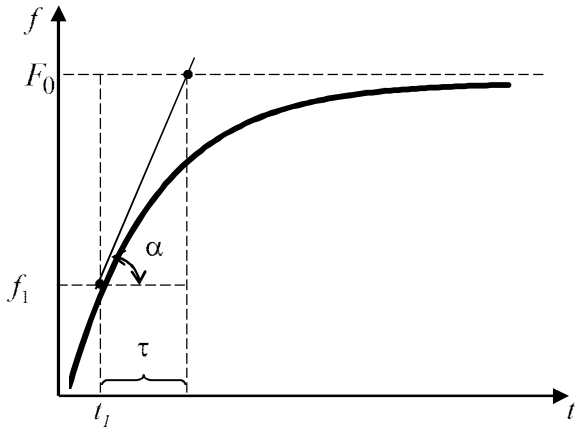


Figure 5. For the A_1 definition

To improve the accuracy of the calculations one should determine the slope of the $tg\alpha$ for every two consecutive points of the experimental cumulative curve. As an unknown quantity F_0 it is possible to use the last point of the cumulative curve f_n. Values of τ defined in this manner for each successive pair of points must be averaged. From the average of τ we find an approximation for the coefficient of influence A_1:

$$A_1 = \frac{1}{\tau}. \tag{17}$$

Approximation for the F_0 can be derived from the expression:

$$F_0 = \frac{f_i}{1 - e^{A_1 t_i}}. \tag{18}$$

To improve the accuracy values found in this way F_0 are averaged over all points of the experimental cumulative curve.

Please note that setting the coefficient $A_2=0$, that is, excluding the impact of input flow, we have inflated estimates for the F_0 and A_1.

After defining the approximations for F_0 and A_1 we must iteratively validate them. It should be kept in mind that there is a dependence of $\mu_1(t)$ on the coefficients A_1 and A_2, which is shown in Figure 6.

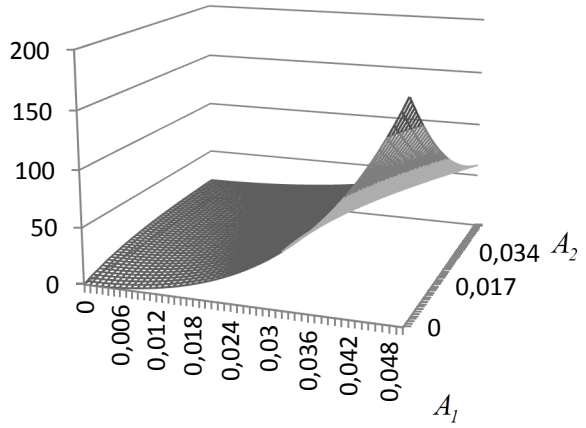


Figure 6. Relation $\mu_1(t)$ on the influence coefficients

From Figure 6 it follows that with the change of the coefficient A_2 values of $\mu_1(t)$ change only slightly. Therefore, clarification the model parameters should start with the coefficient A2: defects admitted here have negligible impact on the accuracy of the process. To assess the accuracy of determining the parameters you should use the criterion of standard deviation (SD), calculated as:

$$SD = \frac{\sum_{i=1}^{n} (f_{io} - f_{ic})^2}{n}, \quad (19)$$

where n – number of points in the experimental cumulative curve; f_{io} – observed value of i-th point of the curve; f_{ic} – value calculated with given parameters.

To clarify the parameters of the model the following algorithm is proposed:

Step 1. Definite initial value of CKO_b for approximations of A_1 and F_{10} obtained through (12) and (13). Assume $A_2 = 0$.

Step 2. Changing A_2 by step

$$\Delta A_2 = \frac{1.5 \cdot A_1}{10},$$

while current value of CKO_x is less, than CKO_b obtain clarified value for A_2. Assume $CKO_b = CKO_x$.

Step 3. Changing F_{10} in range from 0,5 to 1,5 from approximation obtained in step 1, while current value of the CKO_x is less, than CKO_b, find clarified value for F_{10}. Assume $CKO_b = CKO_x$.

Step 4. Changing A_1 in range from 0,5 to 1,5 from approximation obtained in step 1, while current value of the CKO_x is less, than CKO_b, find clarified value of A_1. Assume $CKO_b = CKO_x$.

Step 5. Repeat step 2 – step 4 for the next stage of refinement. Refinement process is considered complete if at the next stage the value $CKO_b \le \varepsilon$ is achieved, where ε – required calculation precision.

6 EVALUATION OF THE ACCURACY OF SIMULATION IN THE SSD

To assess the accuracy of modeling the reliability of the model (SSD model) and its comparison with existing models we used data on the identified defects in twenty different software systems (Android, Lyu 1996) To increase the accuracy of modeling, each series of observations was divided into intervals during which the time variation of the cumulative curves of these defects remained unchanged. 123 total numbers of observations processed. To compare the accuracy of modeling, in addition to the described SRMs there have been taken well-known reliability models,

covering all existing classes of models. They are based on different concepts, which make it possible to objectively evaluate and compare the accuracy of models between each other. The following models were taken into the research: Jelinsky-Moranda's (Moranda & Jelinski 1972), nonhomogeneous Poisson process (Goel & Okumoto 1979), Schneidewind's (Schneidewind 1993), Musa's (Musa 1979), Weibull's (Quadri & Ahmad 2010), S-shaped model (Yamada at al.1983), Duan's model (Duan 1964), Moranda's geometric model (Moranda 1979) and logarithmic model of Musa-Okumoto (Musa & Okumoto 1984). 1230 total estimations of the reliability modeling were made. Modeling results are shown at Figure 7.

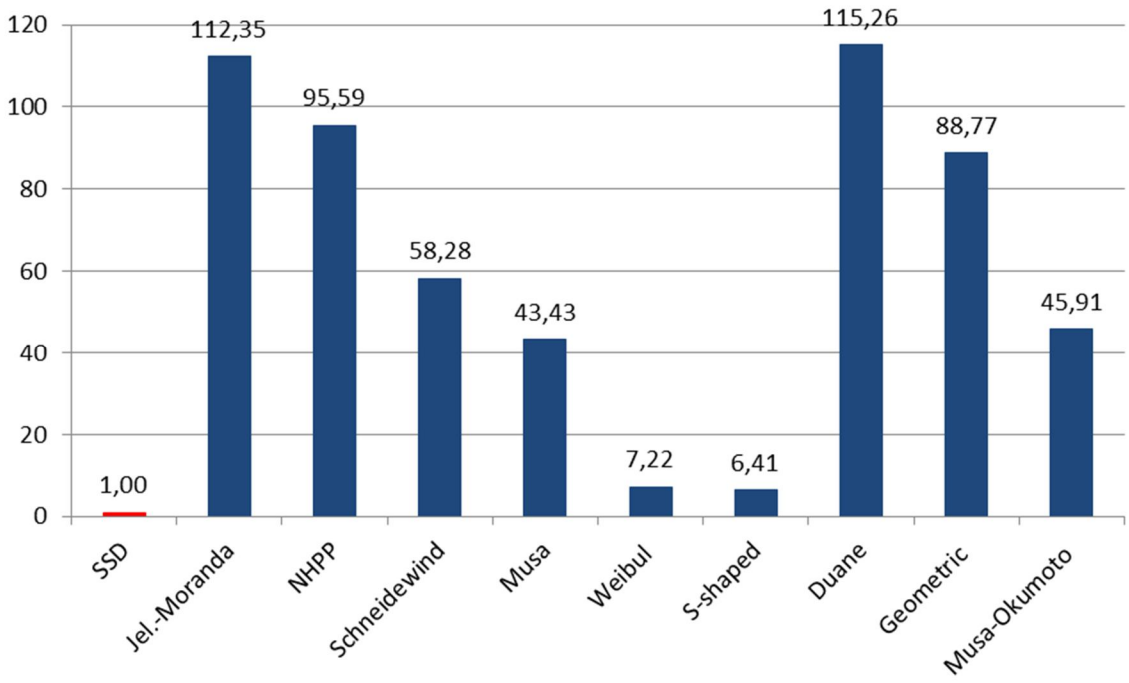


Figure 7. Reliability models' comparison results

In the diagram shown in Figure 7 there are SD values shown which were obtained from the comparison. For convenience, standard deviation of the proposed model of SSD is assumed to be equal to one. The diagram shows that for all classes of the test software model of the SSD showed a result with accuracy of more than six times superior to the result of the best known reliability models – S-shaped model.

7 CONCLUSION

The comparative analysis of different SRMs showed that the model based on the SSD proposed in this article consistently shows the best results in comparison with other known models. The stability of the simulation results is essential to avoiding phase selection model of reliability for each software product.

Therefore, the results are a practical confirmation of the SSD, and reliability model developed based on them can be used for modeling and predicting a wide class of reliability indices of software systems.

The new theory of software reliability proposed in the article opens a wide scope for research. In particular, it is interesting to find out the physical nature of the defect capacity index and suggest ways to define it, and to explore other effects, whose existence is a consequence of the new theory.

8 REFERENCES

Android – An Open Handset Alliance Project. *http://code.google.com/p/android/issues/list*

Duan J.T. (1964). Lourning Curve Approach to Reliability Monitoring. *IEEE Trans. on Aerospace.* Vol. 2: 563-566.

Goel, A.L., K. Okumoto. (1979). Time-Dependent Error-Detection Rate Model for Software and Other Performance Measures. *IEEE Transactions on Reliability,* v. R-28, № 5: 206-211

Kharchenko V.S., Sklar V.V., Tarasyuk O.M. (2004). *Methods for modelling and evaluating the quality and reliability of the programming security*. Kharkov: Nat. Aerospace Univ. "HAI"

Lyu, M.R. (1996). *Handbook of Software Reliability Engineering*. London: McGraw-Hill.

Moranda, P.B. (1979). *Event-Altered Rate Models for General Reliability Analysis* IEEE Transactions on Reliability. Vol.R-28. No. 5: 376-381

Moranda, PL., Jelinski, Z. (1972). *Final Report on Software Reliability Study*, McDonnell Douglas Astronautics Company, MADC Report Number 63921.

Musa J.D. (1979). *Validity of Execution time theory of software reliability*. IEEE Trans. on reliability. No 3: 199-205.

Musa J.D. Okumoto K. A. (1984). *Logarithmic Poisson Time Model for Software Reliability Measurement*. Proc. Sevent International Con-ference on Software Engineering. – Orlando, Florida: 230-238.

Onsager, L. (1931) *Reciprocal relations in irreversible processes*, «Physical Review», 1931, v. 38, № 12: 405-4026

Prigogine, I. (1991) *The Behavior of Matter Under Nonequilibrium Conditions: Fundamental Aspects and Applications*. Progress Report for Period April 15,1990 – April 14, 1991. Center for Studies in Statistical Mechanics and Complex Systems at the University of Texas-Austin, United States Department of Energy-Office of Energy Research.

Schneidewind, N.F. (1993). *Software Reliability Model with Optimal Selection of Failure Data.* IEEE Transactions on Software Engineering, Vol. 19, No. 11: 1095-1104.

Quadri, S. M. K. Ahmad N. (2010). *Software Reliability Growth Modeling with New Modified Weibull Testing–effort and Optimal Release Policy*. International Journal of Computer Applications Vol. 6. № 12: 1-10

Ushakov, I. (2006). *Reliability: past, present, future*. Reliability: Theory & Applications, No.1: 10 – 16.

Ushakov, I. (2012). *Reliability theory: history & current state in bibliographies*. RT&A # 01 (24) Vol.1: 8 – 35

Yamada, S. Ohba M., Osaki S. (1983). *S-Shaped Reliability Growth Modeling for Software Error Detection*. IEEE Trans-actions on Reliability, vol. R-32, no. 5: 475-478.

ANTITERRORISM RESOURCES ALLOCATION UNDER FUZZY SUBJECTIVE ESTIMATES

A. Bochkov, I. Ushakov

e-mail: a.bochkov@gmail.com, igusha22@gmail.com

1. PRELIMINARY

The problem of optimal resources allocation for antiterrorism preventive measures is naturally based on subjective estimates made by experts in this field. Relying on expert estimates is inevitable in this case: there is no other possibility to get input data for the system survivability analysis. There is no such phenomenon like "collecting real data", moreover, there is no "homogenous samples" for consistent statistical analysis of observations, since any case is unique and non-reproducible. Nevertheless, quantitative analysis of necessary level of protection has to be performed.

What are the subjects of such expertise? It seems to us that they are:

- possibility of terrorist attacks on some object or group of objects,
- possible time of such attack,
- expected consequences of the attack and possible losses,
- possible measure of protection and related expenses

Since expert estimates of such complex things are fuzzy due to lack of common understanding the same actions and counter-actions within a group of experts, the question arises: is it possible at all to make any reasonable prognosis and, moreover, say about "optimal allocation of protection resources"?

First of all, we should underline that concept of "optimal solution" relates only to mathematical models. In practice unreliable (and even inconsistent) data and inevitable inaccuracy of the model (i.e. difference between a model and reality) allow us to say only about "rational solutions".

Nevertheless, in practice the problem exists and in any particular case has to be solved with or without using mathematical models. Our objective is to analyze stability of solutions of the optimal resources allocation under fuzziness of experts' estimates.

2. ANALYSIS OF SOLUTION STABILITY: VARIATION OF THE EXPENSE ESTIMATES

First, let us analyze how variation of expenses estimates influence on the solution of the problem on the level of a single object which has to be protested against terrorist attacking. For transparency of explanation, we avoid to consider the influence of defense on the Federal and State levels.

Let us consider some conditional object (Object-1) which can be a subject of a terrorist attack. It is assumed that there may be three different types of enemy's actions (Act.-1, Act.-2 and Act.-3). Defending side can choose several specific protection measures against each type of action {M(i, j), where "i" corresponds to the action type, and "j" corresponds to the type of undertaken protective measure.

Assume that we have three variants of estimates of protection measures costs: lower, middle and upper as it presented in the table below. Here the lower estimates are about 20% lower of the corresponding middle estimates, and the upper ones are also about 20% higher.

There are three types of expert estimates: “optimistic”, “moderate” and “pessimistic”. First ones assume that success in each situation can be reached by low expenses for protective measures; the last group requires larger expenses for protection in the same situation; and the middle group gives date in between.

Table 1. Case of optimistic estimates

OBJECT-1		γ_1	C
Act-1	M(1, 1)	0.25	0.8
	M(1, 2)	0.2	2
	M(1, 3)	0.1	2.5
	M(1, 4)	0.01	3.8
Act-2	M(2, 1)	0.2	1.6
	M(2, 2)	0.16	2.8
	M(2, 3)	0.07	3.2
	M(2, 4)	0.02	5.6
Act-3	M(3, 1)	0.11	0.4
	M(3, 2)	0.1	2
	M(3, 3)	0.05	2.4
	M(3, 4)	0.04	3.6
	M(3, 5)	0.01	5.6

Table 2 Case of moderate estimates

OBJECT-1		γ_2	C
Act-1	M(1, 1)	0.25	1
	M(1, 2)	0.2	2.5
	M(1, 3)	0.1	3
	M(1, 4)	0.01	4
Act-2	M(2, 1)	0.2	2
	M(2, 2)	0.16	3
	M(2, 3)	0.07	4
	M(2, 4)	0.02	7
Act-3	M(3, 1)	0.11	0.5
	M(3, 2)	0.1	2.5
	M(3, 3)	0.05	3
	M(3, 4)	0.04	5
	M(3, 5)	0.01	7

Table 3 Case of pessimistic estimates

OBJECT-1		γ_3	C
Act-1	M(1, 1)	0.25	1.2
	M(1, 2)	0.2	3
	M(1, 3)	0.1	6
	M(1, 4)	0.01	7.8
Act-2	M(2, 1)	0.2	2.4
	M(2, 2)	0.16	3.2
	M(2, 3)	0.07	4.8
	M(2, 4)	0.02	8.4
Act-3	M(3, 1)	0.11	2
	M(3, 2)	0.1	3
	M(3, 3)	0.05	3.6
	M(3, 4)	0.04	6
	M(3, 5)	0.01	8.4

How the data of the tables are interpreted?

Let us consider a possibility of action-1 against the object. With no protection at all, the object’s vulnerability equals 1 (or 100%). If one would have spent we spent $\Delta E = 0.8$ conditional cost units (c.c.u.) and undertook measure M(1, 1) the object’s vulnerability decreases to 0.25. If one does not satisfy such a level of protection, next protective measure (M1, 2) is applied; that leads to decreasing the object vulnerability from 0.25 to 0,3 and costs 2 c.c.u.

Now let us consider all three possible terrorists’ actions. In advance nobody knows what kind of action will be undertaken against the defending object. In this situation the most reasonable strategy is providing equal defense levels against all considered types of terrorist attacks, suggested in [1]. It means that if one needs to ensure a level of protection equals □ than one has to consider only such measures against each action that delivers vulnerability level not less than □□□ For instance, in the considered case, if the required level of vulnerability had to be not higher than 0.1, one has to use simultaneously the following measures of protection against possible terrorists’ attacks: M(1, 3), M(2, 3) and M(3, 2).

Method of equal protection against the various types of hostile attacks appears to be quite natural. If dealing with natural or other unintended impacts, one can speak about the subjective probabilities of impacts of some particular typef course, in a case of intentional attack from a reasonably thinking enemy, such approach is not appropriate. The fact is, that as soon as the enemy knows about your assumptions about his possible. actions, he takes advantage of this knowledge and choose the hostile action that you expect least of all.

In the example considered above if one chose measures M(1,2) with $\gamma_1 = 0.2$, $M(2, 3)$ with $\gamma_2 = 0.07$ и and M(3, 4) with $\gamma_3 = 0.04$, guaranteed level of the object protection is

$$\gamma_{\text{Object}} = \max(\gamma_1, \gamma_2, \gamma_3) = \max(0.2; 0.07; 0.04) = 0.2.$$

For choosing required (or needed) level of object protection, one can compile a function reflecting the dependence of vulnerability of protection cost.

Table 4. Case of optimistic estimates

Object 1			
Step Number	**Undertaken measures**	**Resulting γ_{Object}**	**Total Expenses, C_{Object}**
1	M(1, 1), M(2, 1), M(3, 1)	max {0.25, 0.2, 0.11}=0.25	0.8+1.6+0.4=2.8
2	M(1, 2), M(2, 1), M(3, 1)	max {0.2, 0.2, 0.11}=0.2	2+1.6+0.4=4
3	M(1, 3), M(2, 2), M(3, 1)	max {0.1, 0.16, 0.11}=0.16	2.5+2.8+0.4=5.7
4	M(1, 3), M(2, 3), M(3, 1)	max {0.1, 0.07, 0.11}=0.11	2.5+3.2+0.4=6.1
5	M(1, 3), M(2, 3), M(3, 2)	max {0.1, 0.07, 0.1}=0.1	2.5+3.2+2=7.7
6	M(1, 4), M(2, 3), M(3, 3)	max {0.01, 0.07, 0.05}=0.07	3.8+3.2+2.4=9.4
7	M(1, 4), M(2, 4), M(3, 3)	max {0.01, 0.02, 0.05}=0.05	3.8+5.6+2.4=11.8
8	M(1, 4), M(2, 4), M(3, 4)	max {0.01, 0.02, 0.04}=0.04	3.8+5.6+3.6=13
9	M(1, 4), M(2, 4), M(3, 5)	max {0.01, 0.02, 0.01}=0.02	3.8+5.6+5.6=15

This function is depicted in Figure 13.8.

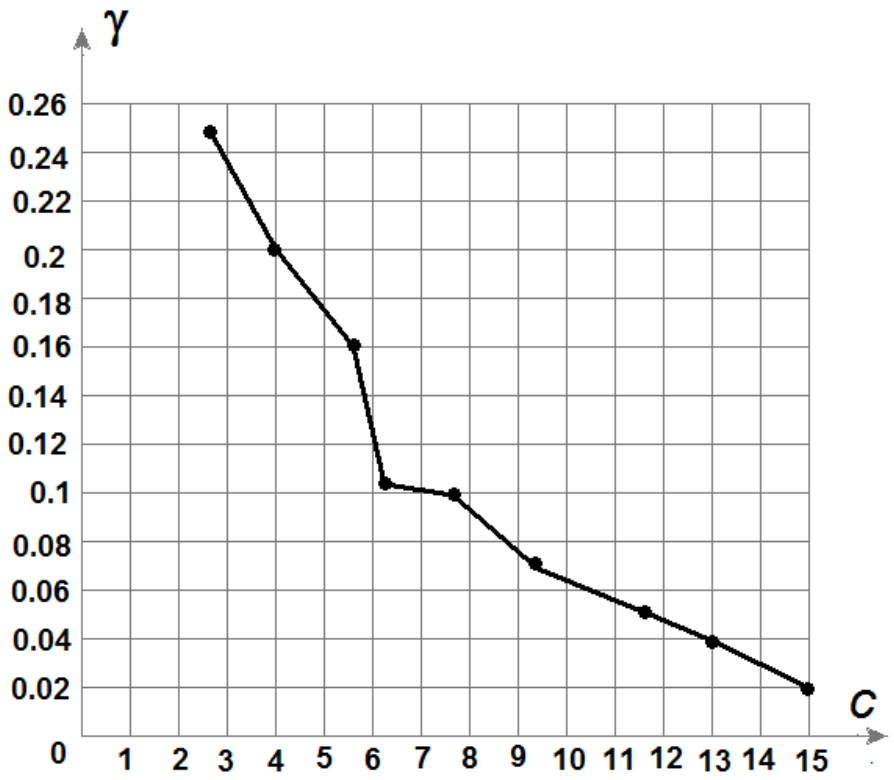


Figure 1. Dependence of the object survivability on cost of protection measures (for "optimistic" estimates).

Without detailed explanations we present numerical results for the cases of "moderate" and "pessimistic" estimates.

Table 5. Case of moderate estimates

Object 1			
Step Number	**Undertaken measures**	**Resulting γ_{Object}**	**Total Expenses, C_{Object}**
1	M(1, 1), M(2, 1), M(3, 1)	max {0.25, 0.2, 0.11}=0.25	1+2+0.5=3.5
2	M(1, 2), M(2, 1), M(3, 1)	max {0.2, 0.2, 0.11}=0.2	2.5+2+0.5=5
3	M(1, 3), M(2, 2), M(3, 1)	max {0.1, 0.16, 0.11}=0.16	3+3+0.5=6.5
4	M(1, 3), M(2, 3), M(3, 1)	max {0.1, 0.07, 0.11}=0.11	3+4+0.5=7.5
5	M(1, 3), M(2, 3), M(3, 2)	max {0.1, 0.07, 0.1}=0.1	3+4+2.5=9.5
6	M(1, 4), M(2, 3), M(3, 3)	max {0.01, 0.07, 0.05}=0.07	4+4+3=11
7	M(1, 4), M(2, 4), M(3, 3)	max {0.01, 0.02, 0.05}=0.05	4+7+3=14
8	M(1, 4), M(2, 4), M(3, 4)	max {0.01, 0.02, 0.04}=0.04	4+7+5=16
9	M(1, 4), M(2, 4), M(3, 5)	max {0.01, 0.02, 0.01}=0.02	4+7+7=18

Data of Table 5 are depicted in Figure 2.

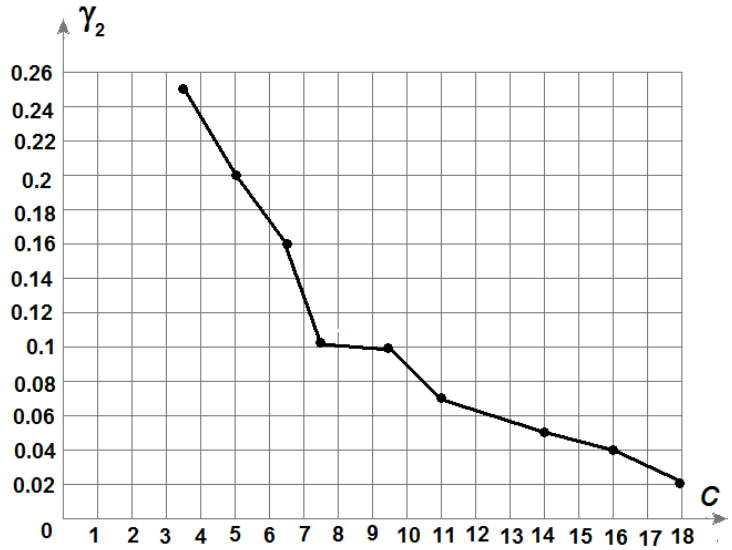


Figure 2. Dependence of the object survivability on cost of protection measures (for "moderate" estimates).

Table 13.6. Case of pessimistic estimates

Object 1			
Step Number	**Undertaken measures**	**Resulting** □Object	**Total Expenses, C_{Object}**
1	M(1, 1), M(2, 1), M(3, 1)	max {0.25, 0.2, 0.11}=0.25	1.2+2.4+2=5.6
2	M(1, 2), M(2, 1), M(3, 1)	max {0.2, 0.2, 0.11}=0.2	3+2.4+2=7.4
3	M(1, 3), M(2, 2), M(3, 1)	max {0.1, 0.16, 0.11}=0.16	3+3.2+2=8.2
4	M(1, 3), M(2, 3), M(3, 1)	max {0.1, 0.07, 0.11}=0.11	3+4.8+2=9.8
5	M(1, 3), M(2, 3), M(3, 2)	max {0.1, 0.07, 0.1}=0.1	3+4.8+3=10.8
6	M(1, 4), M(2, 3), M(3, 3)	max {0.01, 0.07, 0.05}=0.07	4+4.8+3.6=12.4
7	M(1, 4), M(2, 4), M(3, 3)	max {0.01, 0.02, 0.05}=0.05	4+8.4+3.6=16
8	M(1, 4), M(2, 4), M(3, 4)	max {0.01, 0.02, 0.04}=0.04	4+8.4+6=18.4
9	M(1, 4), M(2, 4), M(3, 5)	max {0.01, 0.02, 0.01}=0.02	4+8.4+8.4=20.8

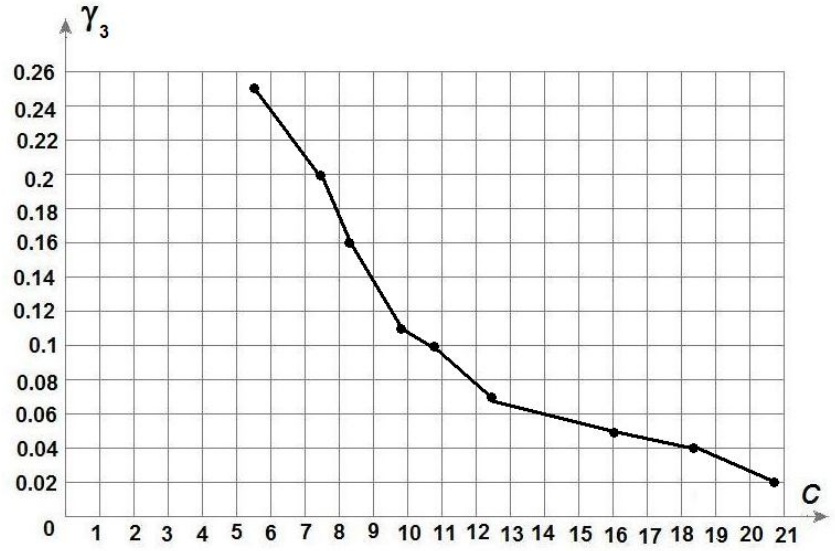


Figure 3. Dependence of the object survivability on cost of protection measures (for "pessimistic" estimates).

Such analysis gives a possibility to find what measures should be undertaken for each required level of protection (or admissible level of vulnerability) and given limited resources.

The final "trajectory" of the dependency "Expenses vs. Vulnerability" presented below.

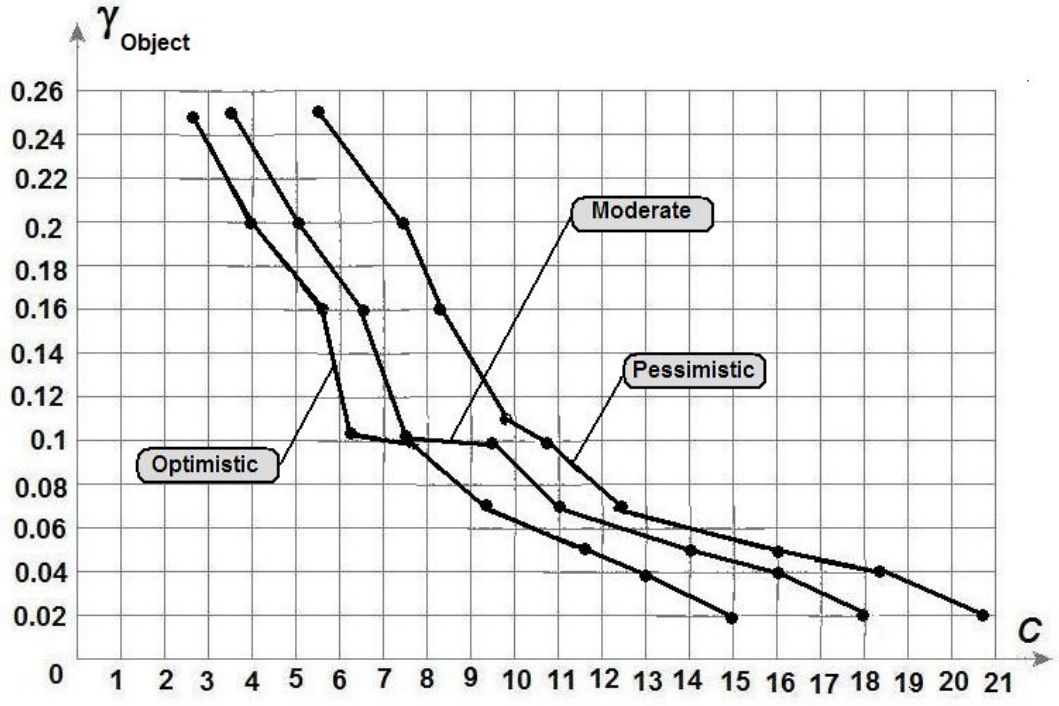


Figure 4. Comparison of solutions for three types of estimates.

Using the tables above, consider two solutions of the Direct Problem with required levels of vulnerability 0.1 and 0.02.

However, decision makers are interested mostly in correctness of undertaken measures, rather than in difference in absolute values of the estimated costs of the object protection. In other

words, in other words, he is interested in how, for example, the solution adopted for the optimistic scenario, will be wealthy in case, if in fact the situation is better described pessimistic scenarios.

It is obvious that if the goal is to reach some given level of vulnerability the vector of solution (i.e. set of undertaken measures for protecting the object against terrorist attacks) in the frame of considered conditions the vector of solution will be the same, though will led to different expenses.

Consider solutions for required level of object vulnerability not higher than 0.1 and not higher than 0.02. They are:

Table 7. Comparison of solutions for three types of scenarios.

Type of scenario	**Undertaken protection measures**	
	$\gamma_{Object} \leq 0.1$	**$\gamma_{Object} \leq 0.02$**
Optimistic	M(1, 3), M(2, 3), M(3, 2)	M(1, 4), M(2, 4), M(3, 5)
Moderate	M(1, 3), M(2, 3), M(3, 2)	M(1, 4), M(2, 4), M(3, 5)
Perssimistic	M(1, 3), M(2, 3), M(3, 2)	M(1, 4), M(2, 4), M(3, 5)

One can see that solutions for all three scenarios coincides for both levels of object protection! Of course, such situation occurs not always, however we should underline that vectors of solution for minimax criterion γ_{Object}=max (γ_1,γ_2,γ_3) is much more stable than vector for probabilistic criterion $1 - \gamma_{\mathrm{Object}} = 1 - \prod_{1 \leq k \leq n}(1 - \gamma_k)$.

CONCLUSION

The presented analysis shows that presented model of optimal allocation of counter-terrorism resources, suggested in the previous section, is working stably enough.

Development of improved computer model will allow analyzing more realistic situations, including random instability of input data. However, it seems that such "one-side biased" expert estimates should lead to more serious errors than random variations of the parameters.

REFERENCES

2006. Ushakov, I.A. Counter-terrorism: Protection Resources Allocation. Part I. Minimax Criterion. *Reliability: Theory and Applications* , No.2.

2007. Bochkov, A.V., and Ushakov, I.A. Sensitivity analysis of optimal counter-terrorism resources allocation under subjective expert estimates. Reliability: Theory & Applications, No. 2

2007. Ushakov, I.A Counter-terrorism: Protection Resources Allocation. Part III. Fictional "Case Study". *Reliability: Theory and Applications*, No.1.

SAFETY CASE-ORIENTED ASSESSMENT OF HUMAN-MACHINE INTERFACE FOR NPP I&C SYSTEMS

A. Orekhova, V. Kharchenko

•

National Aerospace University "KhAI", Kharkiv, Ukraine
e-mail: nastya.orehova@rambler.ru, V.Kharchenko@khai.edu

V. Tilinskiy

•

LTD "Westron", Kharkiv, Ukraine
e-mail: tilinsky@westron.kharkov.ua

ABSTRACT

A safety assessment approach for human-machine interfaces (HMI) of Nuclear Power Plant (NPP) instrumentation and control systems (I&Cs) based on the Safety Case methodology is proposed. I&C assessment model is described taking into account human factor impact. Normative profile based on harmonization and standard requirements selection for choice of HMI safety assessment methods is developed. Ranking of major design principles of safe HMI is provided. Set of methods for comprehensive human machine interface safety assessment at life cycle stages is analyzed and adopted taking into consideration features of HMI safety attribute.

1 INTRODUCTION

To guarantee safety of nuclear power plants (NPP) it is required the modernization and development of new instrumentation and control systems (I&Cs). I&Cs functionality, reliability and effectiveness of human activities depend heavily on human-machine interfaces (HMI) [1, 2].Here's what's relevant now in the HMI field:

- human factor studies in order to reduce the likelihood of errors;
- analysis of the reliability of operator's actions associated with the risk assessment and taking into account the possible consequences;
- development of techniques for evaluation of safety [3].

An approach based on the Safety Case methodology [4] in assessing the security of critical systems has been extended. It involves a comprehensive assessment of the system and its software safety. As a result of work related analysis, the safety assessment of HMI in the Safety Case was not considered.

The work objective is an adaptation of the Safety Case methodology for the development of the integrated technique for the safety assessment of the HMI of critical systems. The paper is structured in the following way. The Section 2 represents general model of HMI as a object of safety assessment. Elements of Safety Case methodology are described in the Section 3. Regulatory requirements and normative base related to HMI safety are analyzed in the Section 4. The safety assessment methods are compared and adopted for HMI in the Section 5. The Section 6 concludes the paper.

2 FORMALIZATION OF OBJECT OF AN ASSESSMENT

Modern I&Cs of the NPP are complex systems of the distributed information processing, where HMI implementation is usually based on workstations. The main purpose of these HMI is to provide staff with the information on the status of the power unit systems, as well as an interface to control the actuators. Information is provided on the monitors of the Main Control Room (MCR) and workstations for personnel. Figure 1 presents a model of the human-machine system. Its

interface consists of two parts: hardware (HW) and software (SW). Besides monitors, HMI hardware may include a standard keyboard with a trackball or a mouse and a functional keyboard.

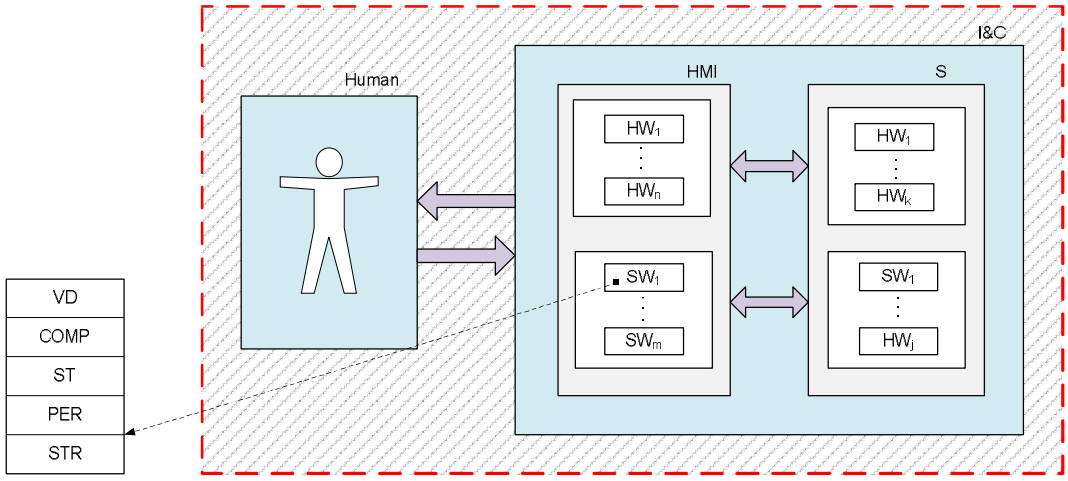


Figure 1. Model of object of an assessment.

Process displays are the core software components of HMI of the I&Cs. They represent plant process data grouped mostly by systems and organized in a multilevel hierarchical structure that allows to navigate among hierarchical levels, as well as within the levels and between the systems. In addition, process displays can be invoked via menu or functional keyboard.

Process displays provide all technological information to the operator in real time in form of symbolic circuits (animated snippets of technological schemes or equipment drawings), diagrams, histograms, tables, graphs, etc.

Detailed structure of the display system is provided at the design stage. HMI software model can have lots of levels (Fig. 1):

$$\text{HMI} = \{ \text{STR}, \text{PER}, \text{ST}, \text{COMP}, \text{VD} \},$$

where STR – strategy; PER – capabilities; ST – structure; COMP – layout; VD – visual design.

The level of strategy (STR) defines objectives of the interface and the user needs; functional specifications and information requirements are determined at the level of Capabilities (PER), the level of ST is for interaction design and informational architecture; Layout (COMP) and visual design (VD) levels define the levels of information and visual design interface. As noted in [1], the main condition for achieving high quality HMI is to follow the standards.

3 METHODOLOGY OF SAFETY CASE

The safety assessment, based on Safety Case methodology includes a formal presentation of evidence, arguments and assumptions aimed at providing assurance that the HMI meets safety requirements, and safety requirements are adequate. At the same time attention should be paid to the logical arguments that will be used to demonstrate that the system is safe to use.

Purpose, which can be interpreted as testing requirement, is divided into sub-goals until one can identify tools, confirming that the sub-goal is achieved (Fig 2). Then these tools are used to verify the safety during development of the system.

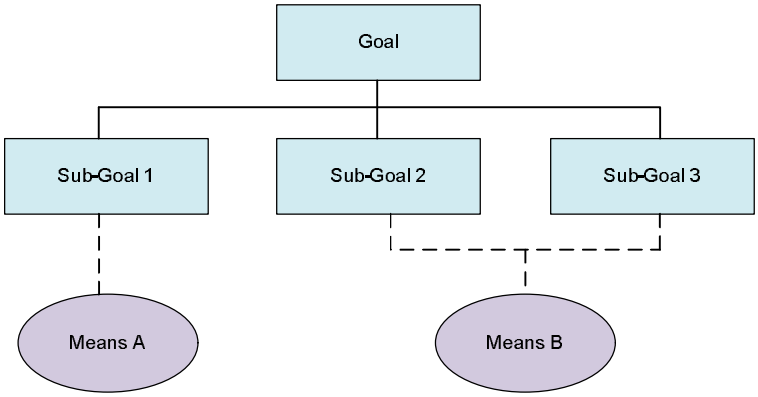


Figure 2. Structure of the objectives.

It is important to plan for a Safety Case at the very beginning of the design process. Firstly, this will determine which evidence is necessary to collect and secondly, what should be used to support them in various stages of the life cycle. One problem is the choice of the depth and rigor of evidence. Some items of evidence may be more persuasive than others, and it must be considered when evaluating the effectiveness of the safety case as a whole.

Safety Case Report should contain all necessary information to assess the safety of HMI. The higher safety requirements the more details are required. Good quality Safety Case provides information to the extent and form that make the work of the expert comfortable in terms of reliability, availability, and ease of use. Typical content of the Safety Case includes:

System Description - defines the purpose of the evaluation, describes the system under consideration (the objectives, functions, structure, components, context of use) and its interaction with other systems. Quality Management Report - gives evidence that the requirements for the process of quality assurance have been met.

Safety management report suggests that an actions, defined in the safety plan, had been implemented. It should include the results of the various analyses, as well as a list of all identified hazards (Journal of Hazards).

Technical Safety Report – it explains technical principles, which provide safety. It should include reports to verify each component, including HMI.

Related Safety Cases – a document that contains references to any Safety Cases for other vital systems, related to the system under consideration.

Findings should be presented in the form of analysis of activities carried out by the developer, and why system attributes are sufficient.

To adapt this approach to the assessment of HMI, elements of Safety Case must be defined as part of the design, development and production, used for HMI of NPP I&Cs.

Figure 3 shows a conceptual model of the system safety assessment of HMI of NPP I&Cs [5]. The solution of the safety assessment problems of HMI of NPP I&Cs is complex and directly related to the modeling and analysis of the design process, specification requirements, the context of use and design.

The HMI safety model is constructed by analysis (profiling) of the regulatory framework. The choice of assessment methods directly depends on the safety profile and the stage of the life cycle of the HMI. Before using of different assessment methods, it is important to formalize the process of the upcoming evaluation. This will help to determine the best approach to effectively assess and select the most appropriate method or methods. Selecting of assessment methods should be

preferred to those methods which have tool support. Evaluation results have a direct impact on improving of the safety of HMI of NPP I&Cs.

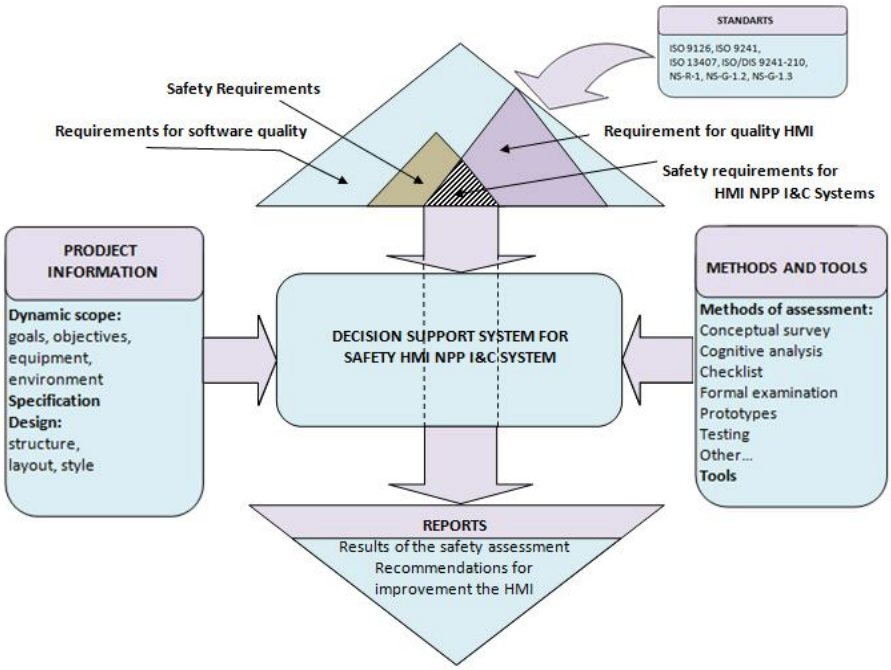


Figure 3. Conceptual model of the safety assessment HMI I&C system in nuclear power plants.

General procedure of the Safety Case-oriented assessment is the following. At the first stage HMI safety requirement profile is developed (specified). The profile includes international and industry standards, and regulatory documents developed for various industry domains. The next stage is to determine the goals, objectives and characteristics for the HMI safety evaluation. There is an analysis and a choice of methods of an assessment which directly depends on a design stage, and also from earlier formulated purposes and problems of estimation. The most exact and reliable assessment can be obtained by applying several methods at the same time. The next stage is evaluation of HMI by tools implementing the chosen method. Finally, in the final stage we obtain the results of the evaluation in the form of certain reports and recommendations to improve the HMI. For this an expert combines the results obtained by different methods at the different stages of evaluation. The end result is highlighted in the safety case document, prepared for the evaluated system and HMI.

REGULATORY FRAMEWORK ANALYSIS

Safety assessment of HMI in the Safety Case is a multi-disciplinary problem. Its scientific rationale and solution requires knowledge of disciplines such as systems design, ergonomics and usability, human factors engineering, software engineering, safety and risk management. There is its own regulatory framework in each of these areas, which regulates approaches, processes, methods and tools for design and evaluation, which may be useful to create an effective methodology for integrated safety assessment of HMI. Fig. 4 shows possible profile-forming database of standards for the choice of methods and processes of NPP I&Cs safety assessment.

Basic design principles and requirements for HMI of NPP I&Cs are given in [6-7]. The same principles can be used as criteria for assessing the safety of I&Cs HMI. Nevertheless all these principles are important for proper HMI design, some of them may contradict with another, so a

compromise between different principles should be reached to ensure effective system design. That is why it is important to identify relative weight of the principles in comparison with other principles. Results of the expert analysis and ranking of these principles/criteria are given below.

Personnel Safety - this principle is ambiguous. In the broad sense, PS is a consequence of the implementation of its main purpose – to provide the safety of NPP. In this sense, it is an integral characteristic, which is inapplicable as a basic design principle. In a narrow sense - as an independent criterion - this principle can be attributed to the safety of I&Cs HMI only, which depends mostly on hardware components of the HMI and cannot deviate significantly under condition that I&Cs is built on modern technical means (for example a workstation monitor can affect user's vision, but all modern LCD monitors are rather similar from this point of view), so relative weight of this principle is rather low in comparison to other principles.

Cognitive Compatibility and *physiological compatibility* - these principles require physiological and psychological capabilities of the operator and the level of his training to be taken into account, when designing HMI. As main criteria, these principles allow us to estimate the quality of information, as well as ease of its perception, analysis and understanding. This is very important criteria for the human factor.

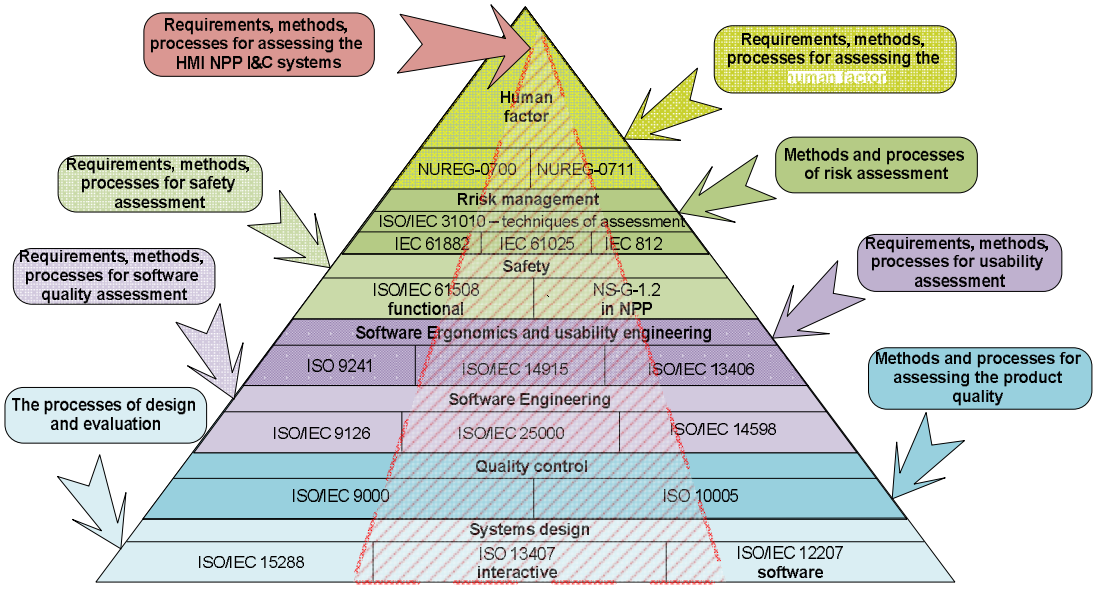


Figure 4.. Profile-forming base of standards.

Consistency is among high priority principles/criteria. Only mutual coherence feedback to the operator through different channels of information can allow him to make right decisions. Hierarchy of priorities of the informational sources must be clearly defined in case of conflicting data.

Situation Awareness is one of the most important principles, because it describes the ability of HMI to perform its basic function - to provide an understanding of the situation by the operator by providing him accurate information on the status of the systems.

Task Compatibility indicates that the system should meet users' requirement. This feature also is one of the most important, because the system must conform to its destination.

Error Tolerance and Control – priority of this principle depends on the class of the System. For systems important to safety, this characteristic has very high priority, because it can directly affect the safety of NPP.

Organization of HSI Elements - this principle ensures provision of the information to personnel in accordance with the distribution of roles in the power unit control, the most important

information relating to security should be available to all operational staff. This principle is important enough, but not critical.

The low-priority design principles include: *Cognitive Workload , User Model Compatibility, Timeliness, Logical Structure, Controls Compatibility, Flexibility, Feedback, Simplicity of Design*. All of these principles should be considered when designing HMI of the I&Cs, however, because the real HMI is a solution based on a compromise, which doesn't satisfy the above criteria completely, the greatest attention should be given to the high priority principles. There are some results of the safe HMI design principles ranking on Fig.5.

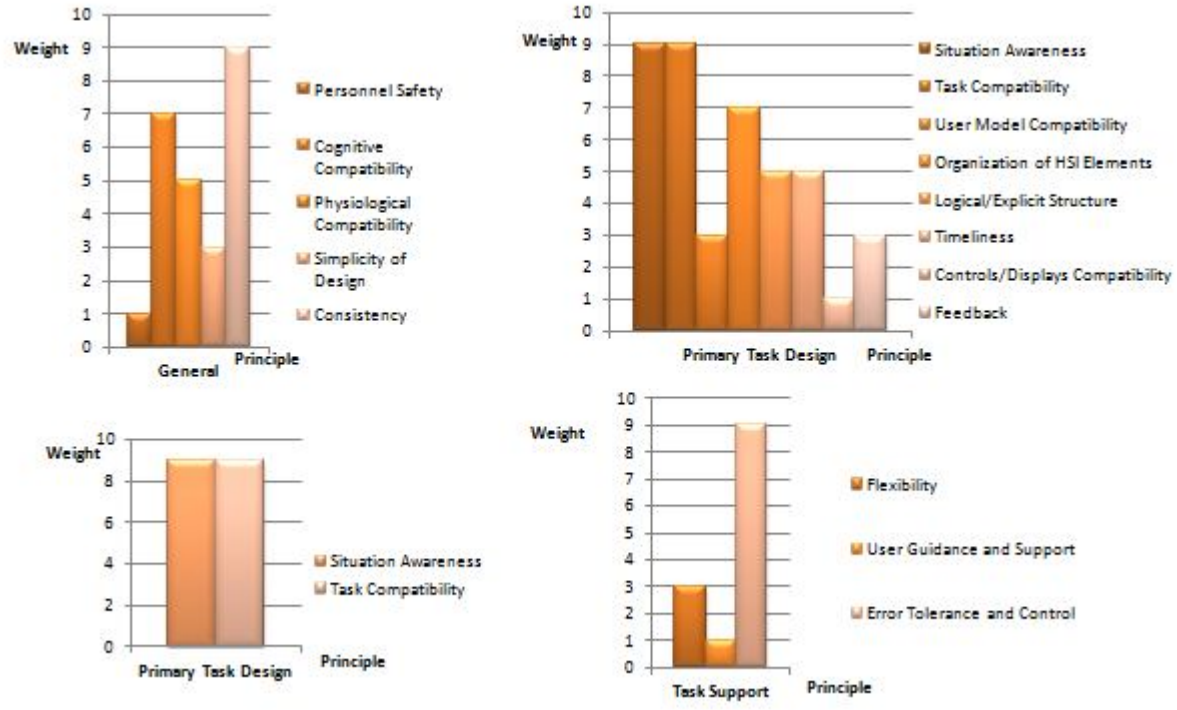


Figure 5. Design principles ranking.

4 CHOICE OF METHODS

To date, the task of choosing methods for safety assessment in the Safety Case was complicated by the large number of techniques of varying degrees of formality, complexity, ability to use of the life cycle stages, etc.

Since we discuss in this paper HMI software only, one can significantly limit the range of the analyzed approaches and methods. As part of UCD-design process of user-centered interactive systems, there is large number of methods relevant to usability [8, 9].

We believe these methods are the most effective at the pre-design gathering stage, at the stage of analysis of the use context (task analysis), as well as at the stage of verification and validation of the finished product (usability testing). Processes and methods of safety HMI evaluation, developed within a software engineering, are mainly focused on the metric evaluation of the finished product.

Methods of risk assessment are given in [10]. Risk assessment can be carried out with varying degrees of depth and detail. The use of one or more methods is possible. When selecting methods, the rationale for their suitability should be presented.

Methods must have the following features:

- to be scientifically sound;
- conform to the system under study;

- to give an understanding of nature and the nature of risk, how to control and process.

Method selection can be implemented based on the following factors:

- purpose of the evaluation;
- system development ;
- type of system;
- resources and opportunities;
- nature and degree of uncertainty;
- complexity of methods;
- ability to obtain quantitative data output;
- the applicability of the method;
- availability and accessibility of information for the system;
- needs of decision makers.

Table 1 shows the results of a comparative analysis of several method-candidates for Safety Case. Recommendations and the applicability of a specific technique throughout the risk assessment process of HMI have been considered when selecting methods.

Table 1. A comparative analysis of risk assessment methods.

Type of risk assessment methods	**Relevance of influencing factors**			**Possibility of the use of the HMI**
	Resources, and capability	**Nature and degree of uncertainty**	**Complexity**	
Checklists	Low	Low	Low	+
Preliminary analysis of the hazards	Low	High	Average	–
Scenario Analysis	Average	High	Average	–
Fault tree analysis (FTA)	High	High	Average	_
Analysis of the "tree" of events	Average	Average	Average	_
Analysis of the causes and consequences	High	Average	High	_
The analysis of types and the consequences of failures (FMEA and FMECA)	Average	Average	Average	+
Hazard and Operability Study (HAZOP)	Average	High	High	+
Reliability assessment of the operator (HRA)	Average	Average	Average	+
Multi-criteria decision analysis (MCDA)	Low	High	Average	+
“+” - applicable; “– “ - no data				

A possible profile of methods for Safety Case and the process of integrated safety assessment of HMI of NPP I&Cs at all stages of the life cycle is shown on Fig. 6.

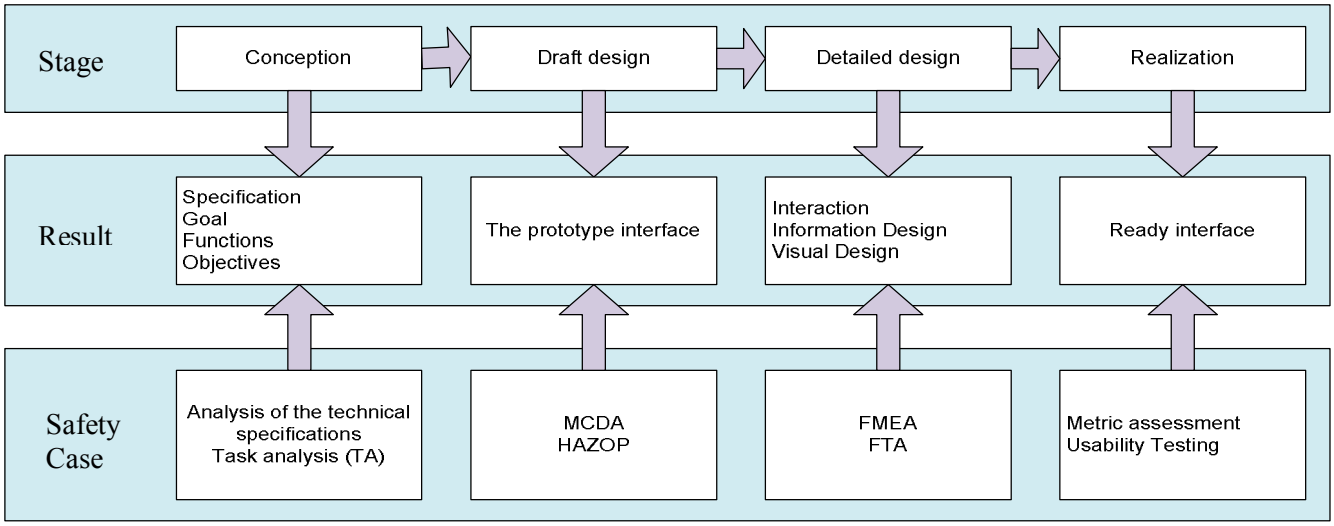


Figure 6. Method of assessment of HMI I&Cs safety.

5.1. Multi-criteria decision analysis (MCDA).

The purpose of this analysis is to estimate the variety of options by applying a set of criteria. In the HMI case many prototypes can serve as such set of options.

The result of the analysis is to establish the order of preference of options available. While analysing, matrix of options and ranked and combined criteria are prepared, to provide an assessment for each option. This method is particularly useful at the early stages of design under uncertainty.

Analysis of the safety criteria has shown that, as a rule, they have the interval based on a quality and a character. The values of the most of them can be described by linguistic variables. Therefore, the safety assessment problem and selection of the best option in terms of safety criteria of the HMI of NPP I&Cs can be formulated as the problem of a fuzzy multi-criteria analysis of options [11].

Suppose we are given many options for HMI $P=\{P_1,P_2,...,P_k\}$ and safety criteria – set $G=\{G_1,G_2,...,G_n\}$, then the problem of multi-criteria analysis is to reorder the elements P based on a set of criteria G. HMI version $P_j \in P$ is evaluated by criteria $G_i \in G$ by the number $\mu_{G_i}(P_j)$ in the range $[0,1]$. The higher the number $\mu_{G_i}(P_j)$ the better option P_j for the criterion G_i, $i=\overline{1,n}$, $j=\overline{1,k}$. Then, the criterion G_i can be represented by a fuzzy set G_i on the universal set of options P [11]:

$$\widetilde{G}_i = \left\{\frac{\mu_{G_i}(P_1)}{P_1}, \frac{\mu_{G_i}(P_2)}{P_2}, \dots, \frac{\mu_{G_i}(P_k)}{P_k}\right\}, \quad (1)$$

where $\mu_{G_i}(P_j)$ is the degree of affiliation of the element P_j to the fuzzy set $\widetilde{G}_i$.

Building affiliated functions based on the twin comparisons is very convenient in finding the degree of affiliation of fuzzy set (1). When using this method, one must generate a matrix of twin comparisons of the results for each criterion. The total number of such matrices is equal to the number of criteria. The best option is the one that's best for all criteria simultaneously.

Fuzzy solution $\widetilde{D}$ is the intersection of partial criteria (formula 3).

According to the fuzzy set $\widetilde{D}$, the best option is the one with the highest degree of affiliation:

$$D = \operatorname{argmax}\left(\mu_D(P_1),\ \mu_D(P_2), \dots, \mu_D(P_k)\right)$$

When the criteria of non-equilibrium degree of affiliation of fuzzy sets $\widetilde{D}$ are found with the formula:

$$\mu_D(P_j) = \min_{i=\overline{1,n}}(\mu_{G_i}(P_j))^{\alpha_i}, j = \overline{1,k} \ , \quad (2)$$

where α_1- coefficient of the relative importance of the criterion $G_i, \alpha_1 + \alpha_1 + \ldots + \alpha_n = 1$

$$\widetilde{D} = \widetilde{G_1} \cap \widetilde{G_2} \cap \ldots \cap \widetilde{G_n} = \left\{\frac{\min_{i=\overline{1,n}} \mu_{G_i}(P_1)}{P_1}, \frac{\min_{i=\overline{1,n}} \mu_{G_i}(P_2)}{P_2}, \ldots, \frac{\min_{i=\overline{1,n}} \mu_{G_i}(P_k)}{P_k}\right\} \quad (3)$$

Saaty method has become widely spread method to find the rank of criteria based on the matrix of the twin comparisons [11]. This approach campaign is about finding the approximate values of the vector of ranks, as the geometric mean values of each row of the matrix of twin comparisons. Thus obtained geometric mean values of the eigenvector are normalized by dividing by the sum of the geometric means.

5.2 Hazard and Operability Study (HAZOP)

The HAZOP method is a procedure for identifying potential or unforeseen hazards in the object due to a lack of information at the design/project stage or hazards manifested by abnormalities in the functioning of the system. The main objectives of this method are:

– making a complete description of an object or a process;
– systematic check of each part of the object or process in order to detect deviations from the project objectives;
– decision making on the possibility of hazards or problems, associated with these deviations.

The HAZOP process is a methodology of a good quality, based on control words, like questions about how design tasks or conditions of functioning may not be met at each stage of the project, process, procedures or a system. The composition and the interpretation of the control words can vary, depending on the object of analysis. The process usually has a team of specialists from different areas in the course of several meetings.

HAZOP method can be applied to HMI at various stages of a design or while the system is functioning. The possibility of using the HAZOP method for the HMI risk assessment is based on the fact that control words can be applied to the physical parameters and transmission of information. HAZOP allows you to explicitly take into account the causes and consequences of errors.

HAZOP study involves the following stages:

– identification of goals, objectives and scope of the study;
– acquisition of HAZOP study group;
– gathering the necessary documentation, drawings and descriptions of the technological process;
– dividing of the object of analysis into smaller elements and analysing them by using the collected documents and control words. Guidewords stimulate individual thinking and encourage brainstorming;
– it documents any abnormalities and related conditions. In addition, identification of the ways to find and/or to prevent rejection is detected. It's documented on the HAZOP worksheets. Examples of the guidewords deviation are shown in the table.

Guidewords can be applied to parameters such as:

– physical parameters;
– transfer of information;
– aspects of the operation.

Examples of deviations and the respective control words for HMI are shown in table 2.

HAZOP study can identify abnormalities that require the development of mitigation measures. In cases where mitigation measures are not obvious or very expensive, HAZOP study results allow identification of initiating events necessary for the further risk analysis. The HAZOP process allows

determination of different failure types, their causes and consequences. If deviations cannot be corrected, the risk of each such deviation should be evaluated.

HAZOP process can be applied to all types of the design goal deviations due to shortcomings of the project, the component(s), planned procedures, and personnel actions.

The method can be applied to various systems of mechanical, electronic, software, control systems for safety-critical facilities and computer systems (CHAZOP, Hazard and Operability Study management or Hazard and Operability study of computer-assisted tools).

Table 2. Deviations and associated guidewords for the HMI.

Type of deviation	**Guidewords**	**HMI example**
Negative	No	Data or signals do not pass
Quantitative deviations	More	Data is transmitted with higher speed than required
	Less	Data is transmitted with lower speed than required
Qualitative deviations	As well as	Error signal
	Part of	Incomplete data or signals
Reverse	Reverse	Inappropriate signals or data
	Other than	Incorrect signals or data
Time	Early	Signals come too soon
	Late	Signals come too late
Order or sequence	Before	Signals come earlier than required
	After	Signals come later than required

The advantages of the HAZOP method include the following:

− provides tools for the systematic and comprehensive research system, process or procedure;
− is conducted by experts from various fields;
− allows solution development and processing risks;
− applicable to a variety of systems, processes and procedures;
− can explicitly take into account the causes and consequences of human errors.

Successful use of HAZOP methodology applied to a complex object like an I&Cs HMI, with multiple links and relationships to other complex object, like an NPP, is very dependent on proper identification of goals, objectives and scope of the study.

Scope of study shall be very clearly idefined; any relationships with interfacing objects should be formalized to minimize number of study cases to a lowest reasonable number. Without such limitations number of study cases tends to increase beyond any reasonable limits and ambiguously cover safety study of related objects.

For example, from point of isolated HMI estimation in many cases it makes no sense to trace a single operator's fault in reading of specific process point value to its possible consequences on the plant side, because number of such consequences may be unlimited or undefined. Instead it makes sense to split process points into limited number of groups depending on their importance for plant safety in specific operation mode and establish a formal definition of hazards for every specific group.

5.3. Failure mode and effects analysis (FMEA)

FMEA methodology allows you to identify the nature of failures, mechanisms for their occurrence and impact. FMEA can be accompanied by a critical analysis, when the significance of each type is determined (FMECA). FMEA analysis is applicable to both systems, and their component, including software, SFME(C)A.

HAZOP process is similar to the FMEA. It allows failure modes identification, their causes and consequences. The difference is that HAZOP is carried out in reverse order of unwanted results and deviations to the possible causes and failure types, whereas FMEA starts with the failure type determination (Fig. 7).

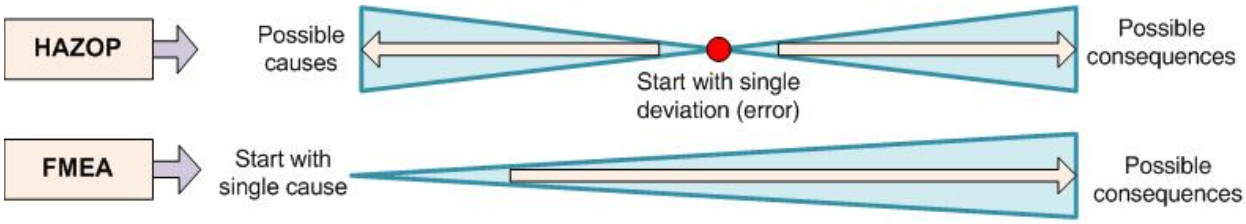


Figure 7. Comparison of HAZOP and FMEA.

5.4. Recommendations to joint application of safety assessment methods of HMI

Obviously, no one of the above methods do not guarantee the accuracy of safety assessment in the process of designing HMI at all stages of the life cycle. It is also clear that the right choice and the sharing of different methods can increase completeness, confidence and cost-effectiveness of the analysis. Integration of qualitative and quantitative methods in the Safety Case should be performed on the basis of individual evaluation techniques with using selected criteria, and development the methodology for their joint use. Let us consider the profile of the methods proposed above in terms of compatibility and application to evaluate the safety HMI. Compatibility of methods is caused by their supplement and using the results of each other. At the initial stage of designing the concept of HMI is represented as a list of tasks and system requirements.

The problems of safety arising at this stage are caused by incompleteness and inconsistency of requirements with the principles recommended by normative documents. Assessment of specifications is carried out by an expert way, recommendations about revision of some important requirements for safety can be formulated as result of estimation. In the second stage of the design several prototypes of the HMI are developed, that implement the proposed concept. At this stage the assessment consists of a choice of the safest HMI, that in the best way takes into account the human factor and meets safety criteria. For HMI components (video frames) which provide the solution of critical tasks, in addition carried out a qualitative analysis of studies using HAZOP, which resulted in the identification of a possible deviation from the requirements and assessment of possible consequences. Features of interaction, configuration, information and visual design are specified at the stage of detailed design. Analysis of HMI can be supplemented by the methods of FMEA and FTA, which will allow to receive probabilities of a deviation (error) and gravity of consequences. Elements of the analysis are video frames and information flows. At the final stage for the ready HMI the usability testing is performed by the operator using simulator. A metric assessment of safety is also used. The results of these assessments must confirm the quality and safety of the finished product and design process.

5 CONCLUSION

Safety assessment of the I&C HMI is based on the Safety Case methodology, which allows us to improve the completeness and reliability of the integrated assessment at all stages of the life cycle from concept to finished product.

Rationale and methods selection is done by multidisciplinary profile-forming regulatory framework, which let us to combine the Safety Case methods in software engineering, risk assessment, human factor engineering and usability.

Rank of the design principles of the HMI safety has been implemented as a result of their analysis. Ranking was conducted with the participation of Westron company experts that has twenty years of experience in the developing safety critical I&C systems of NPPs. Nowadays these results are used at carrying out experiment to assessment HMI I&C system "Vulkan" on compliance to safety principles. Application of the techniques discussed above is planned to be used to evaluate the quality and safety of human-machine interfaces in the process of modernization and development of new NPP I &C.

Future research will be focused on the development of tools, means and techniques for evaluating HMI safety.

PREFERENCES

1. *A. Anokhin, N. Nazarenko Designing interfaces / / Biotehnosphere, 2010, № 2 (8), P. 21-27.*
2. *Xiaojun Wua, Qin Gaoa, Fei Songb, Pengbo Liub, Zhizhong Lia, Xiaolu Donga Evaluating FBTA-Based User Interface Design for digital Nuclear Power Plants // PSAM11/ESREL2012, Scandic Marina Congress Center, Helsinki, Finland, 2012.*
3. *N. Orekhova, V. Kharchenko Analysis of the requirements for interfaces NPP I&Cs / / Bulletin KNTU Named after P. Vasilenko. Engineering. Issue 102. "The problems of energy and Saving energy in agriculture of Ukraine." - Kharkov: KhNTUA, 2010. - P.109-111.*
4. *Andrashov A., Kharchenko V., Netkachova K.,et.al. Safety Case-Oriented Assessment of Critical Software: Several Principles and Elements of Techniques. Monographs of System Dependability. Dependability of Networks, Wroclaw, OWPW, 2010. – p. 11-25.*
5. *N.Orekhova,V. Kharchenko Safety assessment of NPP I&C HMI based on fuzzy multi-criteria analysis of options / / Computational Intelligence. Materials of first ISTC. - Cherkasy: Maklaut, 2011. - P. 219-220.*
6. *The safety assessment and independent verification for nuclear power plants. Manual / Series Safety Standard № NS-G-1.2 ,Vienna, 2004. – 99 p.*
7. *Human-System Interface Design Review Guidelines, NUREG-0700, U.S. Nuclear Regulatory Commission Office of Nuclear Regulatory Research Washington, 2002, 659 p.*
8. *Bevan, N. International Standards for HCI and Usability. International Journal of Human-Computer Studies, 55 (4), 2006.*
9. *Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11. Guidance on usability: ISO 9241, First edition, 15.03.1998. – 28 c.*
10. *Risk management – Risk assessment techniques: ISO/IEC 31010:2000.*
11. *D. Shtovba Design of fuzzy systems by means of MATLAB. - M.: Hot line - Telecom, 2007. – 288 p.*

LOGICAL ANALYSIS OF FAILURES GRAPH

G.Tsitsiashvili
●
*IAM FEB RAS, Vladivostok, Russia,
guram@iam.dvo.ru

ABSTRACT

In this paper a problem to define direct and inverse sets of nodes connected with failed node is considered. This problem is solved by a calculation of connectivity matrix. To simplify initial network a problem of a minimization of its numbers of nodes and arcs is solved also. Calculation complexity of this solution is approximately cubic by a number of nodes.

1. INTRODUCTION

Failures graphs or failures trees are widely used in an analysis of different engineering networks: communication, internet, software, in parallel and distributed systems and in technology processes control, etc. (see for example [1]-[3]). This concept is developed presumably in engineering literature and so needs some mathematical interpretation and generalization by modern methods of discrete mathematics: mathema-tical logic, graph theory, general algebra, theory of algorithms and discrete optimization. A purpose of such generalization is to decrease dimension of failures graphs and trees without a change of their main characteristics: direct and inverse sets of nodes connected with failed node. A combination of such approaches allows to obtain new results in this classical area of the reliability theory.

2. DIRECT AND INVERSE SETS OF NODES CONNECTED WITH FAILED NODE

Our aim is to analyze how a failure in some node of oriented graph leads to failures in another nodes. This problem is connected with a necessity to investigate a spread of failures in technological networks. The problem is formulated analogously to a problem of failures tree analysis. But in a case of failures graph which appeared in manifold applications to systems of energy supplement this problem is significantly more complicated. Later we assume that if there is the arc (i,j) in the failures graph then a failure in the node i spreads to the node j.

Consider oriented graph Gwithout fold arcs and without loops. Suppose that its nodes set I consists of $n < \infty$ elements and its arcs set V consists of m elements. This graph may be described by incidence matrix $\|g_{ij}\|_{i,j\in I}$where $g_{ij} = 1$ if the arc$(i,j) \in V$ and $g_{ij} = 0$ if the arc $(i,j) \notin V$. Further we assume that $g_{ii} = 1$, $i \in I$.

Introduce on the nodes set I of the graph G binary relation [4] " $\geq$ ".

Definition1 .Say that $i_1 \geq i_2$ if in the graph G there is a way from the node i_1 to the node i_2.

It is obvious that this binary relation is reflexive: $i \geq i$, $i \in I$ and transitive: if $i_1 \geq i_2$, $i_2 \geq i_3$ then $i_1 \geq i_3$, $i_1, i_2, i_3 \in I$.

Denote by $I_1(i) = \{j: j \geq i\}$, $I_2(i) = \{j: i \geq j\}$ direct and inverse sets of nodes connected with failed node i. To define the family of sets $\{I_1(i),\ I_2(i),\ i \in I\}$ we suggest the following algorithm.Suppose that the graph$\|g_{ij}^k\|_{i,j\in I}$ is defined by the following conditions: $g_{ij}^k = 1$ if in the graph G there is a way from the node i to the node j with a length not larger than k. In opposite case $g_{ij}^k = 0$. It is obvious that $g_{ij}^1 = g_{ij}$.

Theorem 1. The following equalities are true

$$g_{ij}^{k+s} = \max_{r\in I} \min\left(g_{ir}^{k}, g_{rj}^{s}\right),\ i \in I, j \in I, k \geq 1, s \geq 1. \quad (1)$$

Proof. Suppose that $g_{ij}^{k+s} = 1$ then there is a way from the node i to the node j with a length not larger than $k+s$. This way may be divided into two ways: from the node i to the node r with a length not larger than $k$and from the node r to the node j with a length not larger that . Consequently $g_{ir}^{k} = 1$, $g_{rj}^{s} = 1$ and the formula (1) is true.

Vice-versa assume that $g_{ij}^{k+s} = 0$ and suppose that $\max\limits_{r\in I} \min\left(g_{ir}^{k}, g_{rj}^{s}\right) = 1$. Then there is $r \in I$ so that $g_{ir}^{k} = 1$, $g_{rj}^{s} = 1$ and consequently there is a way from the node i to the node j with a length not larger than $k+s$ and so $g_{ij}^{k+s} = 1$. This contradiction proves the equality $\max\limits_{r\in I} \min\left(g_{ir}^{k}, g_{rj}^{s}\right) = 0$. The equality (1) is true.

Theorem 2. The relation $i \geq j,\ i,j \in I$ is true if and only if

$$g_{ij}^{2^l} = 1,\ l = \min(t: 2^t \geq n). \quad (2)$$

Proof. The formula (2) leads to the inequalities $g_{ij}^{1} \leq g_{ij}^{1} \leq \cdots$. Any way from the node i to the node j with a length not larger than n contains cycles because the number of nodes in the graph Gequals n. So by a deletion of all cycles it is possible to transform a way between the nodes i,j with a length larger than n to a way with a length not larger than n.Consequently the equalities $g_{ij}^{n} = g_{ij}^{n+1} = \cdots$are proved and so $i \geq j$ is true if $g_{ij}^{n} = g_{ij}^{2^l} = 1$.

Corollary 1. From Theorems 1,2 we obtain that to calculate the matrix $\left\| g_{ij}^{2^l} \right\|_{i,j\in I}$ it is possible to use the following algorithm

$$g_{ij}^{2^{k+1}} = \max_{r\in I} \min\left(g_{ir}^{2^k}, g_{rj}^{2^k}\right), i \in I, j \in I, 1 \leq k < l \quad (3)$$

with calculation complexity $2ln^3$ where $l \sim log_{\,2}\, n, n \to \infty$.

Corollary 2. The following equalities are true:

$$I_1(i) = \left\{j: g_{ji}^{2^k} = 1\right\}, I_2(i) = \left\{j: g_{ij}^{2^k} = 1\right\}, i \in I. \quad (4)$$

Remark 1. A specific of the algorithm described by the formulas (3), (4) is that analogously to the Floid and Steinberg algorithm [5] we calculate complete family $\{I_1([i]),\ I_2([i]),\ i \in I\}$ not its representatives. Given algorithm contains logic product of matrices described by the formula (3) and permitting a paralleling by well known methods [6]. To economy of computer memory it is worthy to describe elements of all matrices by logic variables not decimal ones.

3. REPRESENTATIN OF THE SETS $I_1(i), I_2(i)$ BY CLASSES OF NODES

Consider now a structure of results obtained from the formula (4).

Definition2. On the set I introduce binary relation " ≡ "by the condition $i_1 \equiv i_2$ if and only if $i_1 \geq i_2, i_2 \geq i_1$.

Lemma 1. The relation" ≡ " is equivalence relation.

Proof. From Definition 2 we have that binary relation " ≡ " is reflexive and transitive and symmetric.

Consequently the set I may be divided into equivalence classes

$\hat{\imath} = \{i': i' \equiv i\} = \left\{i': g_{i'i}^{2^l} = g_{ii'}^{2^l} = 1\right\}$.

Denote by $\mathcal{I} = \{\hat{\imath}: i \in I\}$ the set of such equivalence classes.

Definition 3. On the set $\mathcal{I}$ introduce binary relation $\hat{\imath}_1 \geq \hat{\imath}_2 \Leftrightarrow \exists i_1' \in \hat{\imath}_1, i_2' \in \hat{\imath}_2 : i_1' \geq i_2'$.

Lemma 2. Binary relation " $\geq$ "on the set $\mathcal{I}$ is a relation of partial order and is defined by the condition $\hat{\imath} \geq \hat{\jmath} \Leftrightarrow g_{ij}^{2^l} = 1$.

2. If $\hat{\imath}_1 \geq \hat{\imath}_2$ then for any $i_1' \in \hat{\imath}_1, i_2' \in \hat{\imath}_2$ the formula $i_1' \geq i_2'$ is true.

Proof. 1. From Definitions 2, 3 we have that partial order " $\geq$ " on the set $\mathcal{I}$ of equivalence classes is refle-xive and transitive and antisymmetric and consequently is the relation of partial order. The formula $\hat{\imath} \geq \hat{\jmath} \Leftrightarrow g_{ij}^{2^l} = 1$ is obvious.

2. Assume that $\hat{\imath}_1 \geq \hat{\imath}_2$ consequently for any $i_1' \in \hat{\imath}_1, i_2' \in \hat{\imath}_2$ it is possible to construct in the graph G a way from the node i_1' to the node i_2' and so $i_1' \geq i_2'$.

Corollary 3. The following formulas are true

$$I_1(i) = \cup_{\hat{\jmath} \geq \hat{\imath}} \hat{\jmath}, \; I_2(i) = \cup_{\hat{\imath} \geq \hat{\jmath}} \hat{\jmath}, i \in I. \tag{5}$$

On the set $\mathcal{I}$ of equivalence classes using the relation of partial order " $\geq$ " it is possible to construct the oriented graph $\mathcal{G}$ with the set of arcs $\mathcal{V}$ using the following procedure. Suppose that $\hat{\imath} \geq \hat{\jmath}$ and there are nodes $i' \in \hat{\imath}, j' \in \hat{\jmath}$ so that $(i', j') \in V$ then the arc $(\hat{\imath}, \hat{\jmath}) \in \mathcal{V}$. It is obvious that the graph $\mathcal{G}$ is acyclic and
without fold arcs. It is possible to restore the relation " $\geq$ "using an analog of Definition1. Denote by $\left\| g_{\hat{\imath}\hat{\jmath}} \right\|_{\hat{\imath}\hat{\jmath} \in \mathcal{I}}$ incidence matrix of the graph $\mathcal{G}$ and define $\left\| g_{\hat{\imath}\hat{\jmath}}^{2^l} \right\|_{\hat{\imath}\hat{\jmath} \in \mathcal{I}}$ as connectivity matrix of the graph $\mathcal{G}$ nodes. It is obvious that a calculation of the matrices $\left\| g_{\hat{\imath}\hat{\jmath}} \right\|_{\hat{\imath}\hat{\jmath} \in \mathcal{I}}$, $\left\| g_{\hat{\imath}\hat{\jmath}}^{2^l} \right\|_{\hat{\imath}\hat{\jmath} \in \mathcal{I}}$ by known matrices $\left\| g_{ij} \right\|_{ij \in I}$, $\left\| g_{ij}^{2^l} \right\|_{ij \in I}$ demands not larger than $2n^3$ arithmetical operations.

4. MINIMIZATION OF ARCS NUMBER IN GRAPH $\mathcal{G}$

Consider the graph $\mathcal{G}$ with the set $\mathcal{I}$ of nodes and the relation of partial order " $\geq$ " and the set of arcs $\mathcal{V}$. Our problem is to remove from the set $\mathcal{V}$ maximal possible subset of arcs so that the matrix of nodes connectivity $\left\| g_{\hat{\imath}\hat{\jmath}}^{2^l} \right\|_{\hat{\imath}\hat{\jmath} \in \mathcal{I}}$ does not change and consequently the sets $\{\hat{\jmath}: \hat{\jmath} \geq \hat{\imath}\}$, $\{\hat{\jmath}: \hat{\imath} \geq \hat{\jmath}\}$, $\hat{\imath} \in \mathcal{I}$ from the formula (5) do not change also. This procedure of a minimization of arcs number is necessary to make failures graph maximally transparent and compact.

Assume that the set $\mathcal{V}$ contains the arc $(\hat{\imath}, \hat{\jmath})$.

(A). Suppose that there is the node $\hat{k} \in \mathcal{I}$, $\hat{k} \neq \hat{\imath}$ such that $g_{\hat{\imath}\hat{k}}^{2^l} g_{\hat{k}\hat{\jmath}}^{2^l} = 1$. Then the arc $(\hat{\imath}, \hat{\jmath})$ is to be deleted from the set $\mathcal{V}$. As the graph $\mathcal{G}$ is acyclic so a way $\hat{\imath}, \hat{k}, \hat{\jmath}$ does not contain the arc $(\hat{\imath}, \hat{\jmath})$ and consequently a deletion of this arc from the set $\mathcal{V}$ does not change the connectivity matrix $\left\| g_{\hat{\imath}\hat{\jmath}}^{2^l} \right\|_{\hat{\imath}\hat{\jmath} \in \mathcal{I}}$. If the condition (A) is not true then the arc $(\hat{\imath}, \hat{\jmath})$ remains in the set $\mathcal{V}$ and the matrix $\left\| g_{\hat{\imath}\hat{\jmath}}^{2^l} \right\|_{\hat{\imath}\hat{\jmath} \in \mathcal{I}}$ does not change also.

As a deletion of different arcs from the set $\mathcal{V}$ is realized independently so we obtain minimal possible number of arcs in the set $\mathcal{V}$ and this solution is unique. Algorithm of the number of arcs in the graph $\mathcal{G}$ minimization may be easily parallelized because a rejection of arcs is realized independently. It is not difficult to obtain that this procedure demands not larger than $2n^3$ arithmetic operations.

REFERENCES

1. Tseng Y.C., Chang S.H., Sheu J.P. 1997. *Fault-tolerant ring embedding in a star graph with both link and node failures.* IEEE Transactions on Parallel and Distributed Systems. Volume 8, issue 12. P. 1185 – 1195

2. Datta P., Frederick M.T., Somani A.K.2003. *Sub-graph routing : a novel fault-tolerant architecture for shared-risk link group failures in WDM optical networks*. Proceedings of Fourth International Workshop on Design of Reliable Communication Networks. P. 296 - 303.

3. Yastrebenetskiy M.A., Ivanova G.M. 1989. *Reliability of automatized systems for technology processes control.* M.: Energoatomizdat.

4. Kurosh A.G. 1962. *Lectures on general algebra.* M.: Nauka. (In Russian).

5. Floid R. W., Steinberg L.1975. *An adaptive algorithm for spatial greyscale.* SID 75 Digest. P.36--37.

6. Valkovsky V.A. 1989. *Paralleling of algorithms and programs. Structural approach.* M.: Radio and sviaz. (In Russian).

REGIONS OF ACCEPTABILITY APPROXIMATION IN RELIABILITY DESIGN[1]

O.V. Abramov, D.A.Nazarov

•

Institute of Automation and Control Processes,
Far Eastern Branch of Russian Academy of Sciences, Vladivostok, Russian Federation

e-mail: abramov@iacp.dvo.ru, nazardim@iacp.dvo.ru

ABSTRACT

An approach to ensure the reliability of engineering systems at design stage is considered in this paper. This approach is associated with construction of an acceptable region inside the system parameter space. A model that describes an acceptable region constructed on the basis of multidimensional grid is offered. The methods for reducing amount of data with respect of resource limitations and particulars of data decomposition for its parallel processing are described.

1 INTRODUCTION

The task of feasible parameter region exploration often arises at engineering systems design. This task is associated with a set of specific procedures such as parameters tolerancing and choosing their nominal values, estimation of system sensitivity and parametric reliability. As a rule, a feasible parameter region (region of acceptability) is a domain comprised of parameter vectors which yield proper system performance. Obtaining the region characteristics or its approximation significantly facilitates solutions of design tasks associated with reliability control. Essential difficulty of the region approximation consists in high dimension of parameter space, incomplete prior information and only pointwise exploration of parameter space with system performances calculation.

There are different methods for constructing acceptable region. The method of approximation with a hyper-parallelepiped is in general use (Abramov et al. 2007, Jess et al. 2003). The methods of approximation with ellipsoids and polytopes are more advanced, but more complicated (Conti et al. 1994, Stehr 2005). One more approach consists in approximation with discrete set of elementary figures. Usually, hyper-parallelepipeds are used as elementary figures (Abramov & Nazarov 2011). In this paper, this method of approximation is applied to determine a region of acceptability.

The process of a multidimensional region construction with a discrete set of elementary parallelepipeds (boxes) is associated with two problems. The first problem consists in large amount of data to be processed and stored. The second problem arises from the first one and consists in high computational requirements. The second problem can be solved with application of parallel computing technologies (Abramov et al. 2007).

The application of parallel computing requires decomposition of a task. In this work, it is shown that the decomposition of the process of acceptable region construction can be carried out by splitting data to be processed in parallel. The model of the region approximation allows splitting the data into various parts of various volumes dynamically.

[1] This work is supported by FEB RAS: Grants 12-I-OEMMITY -01 (Basic Research Program of PMMCP Branch RAS № 14) and 12-III-B-03-023.

The problem of storing large amounts of data is solved by optimization of the data structure, described by the model. The optimization consists in eliminating redundant data using algorithms of lossless data compression. Generally computer memory limitations do not allow keeping all the data describing the region of acceptability. In addition, to increase access speed when using compressed data it is required data splitting with their index ranges assigning. Thus, within the framework of acceptable region construction with respect of data compression and parallel processing, the task of file storage organizing arises.

This work is devoted to solving the problem of storage and processing region of acceptability data with account of computer memory limitations, application of data compression algorithms and parallel processing.

2 A REGION OF ACCEPTABILITY

2.1 A region of acceptability definition

From a consumer point of view, a system has its performance characteristics (gain, temperature, voltage, etc.). The performances are given as m-vector (1):

$$\mathbf{y} = (y_1, y_2, ..., y_m). \tag{1}$$

From a design point of view, any system consists of elements/components that perform their functions. These elements are considered to be atomic. Thus, system parameters are considered as the n-vector:

$$\mathbf{x} = (x_1, x_2, ..., x_n). \tag{2}$$

System performances (1) depend on parameters (2) of system elements (system parameters). System topology is defined by the model (3) which relates system parameters (2) to the system performances (1):

$$\mathbf{y} = \mathbf{y}(\mathbf{x}). \tag{3}$$

System components are influenced by different factors like ambient temperature, supply voltage, radiation, etc. These factors are usually taken into account in the model (3) as operational parameters and cannot be controlled by the designer. Operational parameters and aging factor cause deviations of system parameters which, consequently, cause system performances deviations. Usually, system performances (1) are constrained by performance specifications (4):

$$\mathbf{y}_{\min} \le \mathbf{y}(\mathbf{x}) \le \mathbf{y}_{\max} \tag{4}$$

The deviations of system parameters may cause violation of performance specifications (4) that means system failure. The task of parametric synthesis (Abramov et al. 2007) as one of design stages consists in nominal parameters choosing to meet the performance specifications (4) with the account of system parameter deviations during operating cycle. The solution of this task is often associated with determination of a region of acceptability as defined in (5):

$$D_x = \{\mathbf{x} \in \mathbf{R}^n \mid \mathbf{y}_{\min} \le \mathbf{y}(\mathbf{x}) \le \mathbf{y}_{\max}\}. \tag{5}$$

The region of acceptability and schematic illustration of system parameter deviation from nominal values $\mathbf{x}^0$ at the moment t_0 to gradual parametric failure at the moment t_2 are presented in Figure 1.

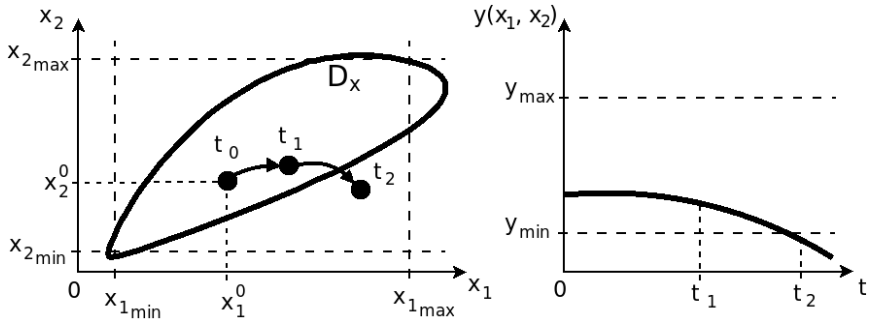


Figure 1. The region of acceptability and gradual parametric failure.

2.2 Grid Approximation of a Region of Acceptability

As it was said before, the approximation of a n-dimensional region with a discrete set of elementary hyper-parallelepipeds (boxes) is used in this work. The basis of such representation of a region is a n-dimensional regular grid inside a bounding box (circumscribed box (Abramov et al. 2007) or parameter tolerances box) defined by the constraints (6):

$$\mathbf{x}_{\min} \le \mathbf{x} \le \mathbf{x}_{\max}. \tag{6}$$

Grid nodes (7) define corners of the elementary boxes:

$$x_{ij} = x_{i0} + j \cdot h_i, \tag{7}$$

where i=1,2,…,n is an index of a parameter, j=0,1,…,Q_i is the node index for i-th parameter (the first node $x_{i0} = x_{i\ min}$), $h_i = (x_{i\ max} - x_{i\ min}) / Q_i$ is the grid spacing for i-th parameter, Q_i – is the amount of "quanta" – the atomic intervals into which the range $[x_{i\ min}, x_{i\ max}]$ is divided. For each x_i inside a "quantum" $\partial\mathbf{y}/\partial x_i = 0$ is supposed. Every "quantum" is indexed with k_i = 1,2,…,Q_i , thus the set of indices (k_1,k_2,…,k_n) identifies an elementary box. It is supposed that every point $\mathbf{x}$ inside an elementary box yields the same performances as its central point $\mathbf{x}_c(k_1,k_2,\ldots,k_n)$ with the coordinates defined in (8):

$$x_{i_c}^{k_i} = x_{i0} + k_i \cdot \frac{h_i}{2},\ \forall i = 1,2,...,n. \tag{8}$$

Every point $\mathbf{x}_c(k_1,k_2,\ldots,k_n)$ acts as a sampling point for elementary box identified by the indices (k_1,k_2,…,k_n). System performances (1) are calculated for every elementary box's sampling point using the model (3). Then these performances are compared with performance specifications (4). Thus, for every elementary box, the binary function (9) is calculated:

$$F_{D_x}(k_1,k_2,...,k_n) = \begin{cases} 1,\ \mathbf{y}_{\min} \le \mathbf{y}(\mathbf{x}_c(k_1,...,k_n)) \le \mathbf{y}_{\max} \\ 0,\ otherwise \end{cases}. \tag{9}$$

The function (9) determines the membership of a sampling point in the region D_x. Let us denote the set of elementary boxes B_g, then the function (9) defines a partitioning (10) of this set:

$$B_g = B_g^0 \cup B_g^1,\ B_g^0 \cap B_g^1 = \varnothing. \tag{10}$$

The subset B_g^1 is the approximation of the region of acceptability D_x, constructed with a discrete set of elementary boxes, defined with a regular grid. The example of 2-dimensional sections of 8-dimensional region of acceptability for an amplifier parameters choosing is illustrated in Figure 2.

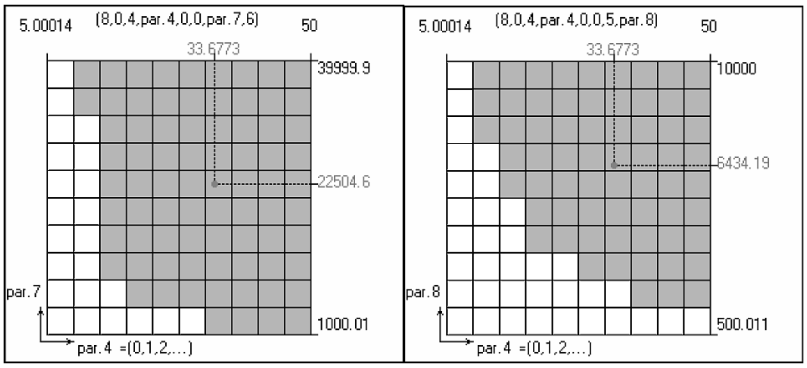


Figure 2. 2D sections of 8D region of acceptability for nominal sizing.

The region of acceptability approximation with a grid is defined with the model (11):

$$G_R = (n, B, Q, S), \tag{11}$$

where n is the amount of designable system parameters (2), $B = \{(x_{i\ min}, x_{i\ max}), \forall i=1,2,\ldots,n\}$ is a bounding box, defined by the constraints (6) in system parameter space, $S = (s_1, s_2, \ldots, s_n)$ is a set of membership indicators that store results of membership function (9). Every indicator $s_p \in \{0,1\}$ displays the membership of the corresponding elementary box in subset B_g^1 or B_g^0, $R=Q_1 \cdot Q_2 \cdot \ldots \cdot Q_n$ is the amount of elementary boxes and, consequently, the amount of membership indicators. The one-to-one correspondence between indices $(k_1,k_2,\ldots,k_n)$ and the index p of an indicator in the set S is defined in (12). It is evident, that zero-based indices are preferable.

$$\begin{aligned} p = k_1 + Q_1 \cdot (k_2 - 1) + Q_1 \cdot Q_2 \cdot (k_3 - 1) \times \ldots \\ \times Q_1 \cdot Q_2 \cdot \ldots \cdot Q_{n-1} \cdot (k_n - 1) \end{aligned} . \tag{12}$$

The process of the region of acceptability construction on the basis of the model (11) was described in the work (Abramov & Nazarov 2011). Briefly, this process consists in complete enumeration of the values of index p with calculation of corresponding indices $(k_1,k_2,\ldots,k_n)$, calculation its sampling point (8), calculation of membership function (9) and assigning its result to the indicator s_p. The illustration of the result of this process and indicators assignment is presented in Figure3.

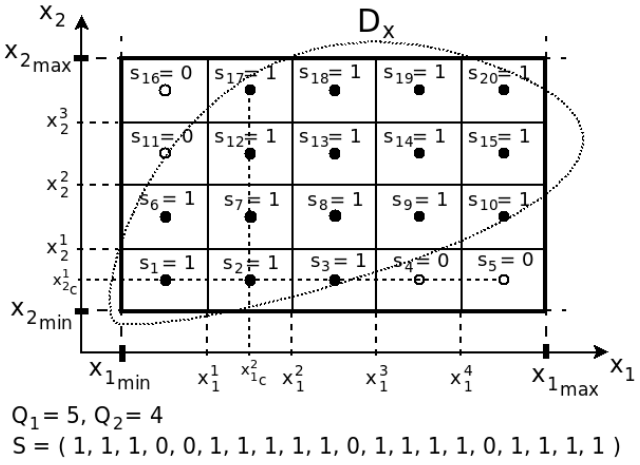


Figure 3. A region of acceptability approximation with a regular grid.

The usage of 1-dimensional structure for storing indicators is explained both by the convenience of the data storage and transmission and by its flexibility for task decomposition for parallel processing. The flexibility consists in the opportunity to splitting of the indicators array into arbitrary amount of portions of various volumes (e.g. for load balancing).

3 DECOMPOSITION OF THE DATA

The problem of the region D_x approximation with the model (11) consists in large amount of data. In addition, the indicators array data are redundant. This redundancy is induced by the way of its consecutive storage and its inner binary representation. The redundancy associated with the array data storage consists in occurrence of long sequences of indicators of the same value on the back of the way of consecutive storage of 1-dimensional sections of a multidimensional region (Fig. 3). The redundancy related to inner array data representation consists in the follows. Every indicator requires only one binary bit, but the basic addressable memory element is a byte, that consists of several binary bits (usually, eight bits). Thus when using usual byte array to store one byte per an indicator, a kind of "rarefaction" in the data occurs, that is highly undesirable when storing large amount of data. The solution to this problem is utilizing of every binary bit of a continuous byte array that can be achieved with the use of binary arithmetic (Abramov & Nazarov 2011).

Within the framework of the array data decomposition task, consider reducing of redundancy related to long sequences of the same indicator values. The most evident solution is storing lengths of the sequences of the same values, e.g. the array of characters BBBBWWWBBWWWWW will be reduced to B4W3B2W5. This is well-know algorithm called Run-Length Encoding (RLE) and usually it is used in computer graphics and communication technologies. The advantage of this algorithm is speed of compression and decompression, but only when enumerating sequentially (Salomon 2007).

With respect to storing of indicators array, the algorithm of RLE is described as follows. The array of membership indicators $S = (s_1, s_2, \ldots, s_n)$, $s_p \in \{0,1\}$ $\forall p = 1, 2, \ldots, R$ is stored as a set of pairs like (13):

$$S_{RLE} = ((c_1, l_1), (c_2, l_2), \ldots, (c_{R^*}, l_{R^*})), \tag{13}$$

where $c_i \in \{0,1\}$ is an indicator value, l_i is the amount of its repetitions (the length of the sequence), R^* is the amount of pairs (c_i, l_i) in the array S_{RLE} that can not be evaluated without enumeration over all the elements of initial indicators array.

The algorithm of RLE reduces the data size of indicator array from 10 to 1000 times (Fig.4). The compression ratio depends on the grid spacing ("quanta" amount) and the actual region D_x configuration.

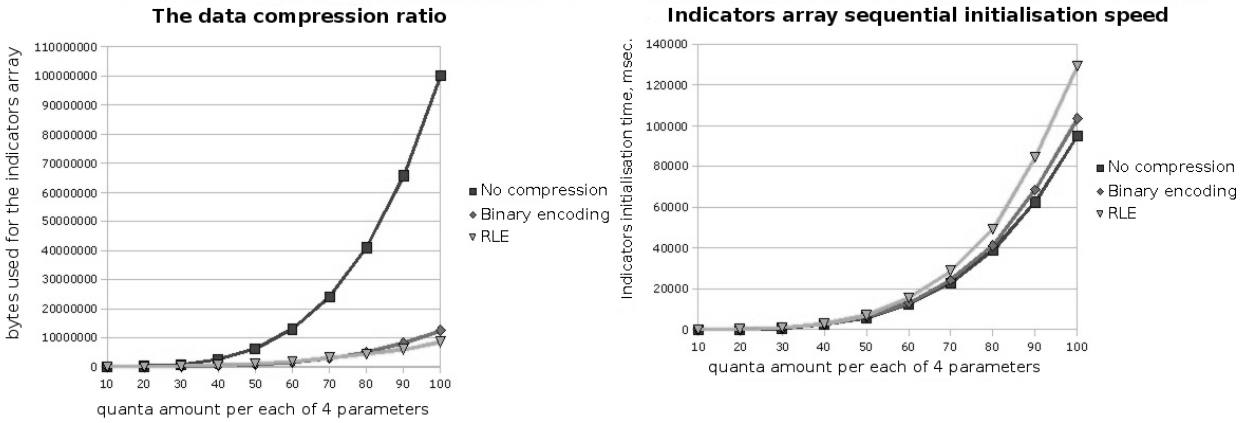


Figure 4. The indicators array compression ratio with RLE algorithm.

The problem of RLE algorithm is low speed of random indicator access. Since the array (13) includes only sequences codes, it is required to obtain the index q of the pair (c_q, l_q) which encodes

the indicator s_p identified by random index p. The searching for the pair (c_q, l_q) requires sequential addition of the lengths until p satisfies (14):

$$L_{q-1} < p \le L_q \,, \quad L_q = \sum_{i=1}^{q} l_i \,, \; L_0 = 0 \,. \tag{14}$$

Then q is the index of the desired pair (c_q, l_q), and $s_p = c_q$. As can be seen, the access to random indicator takes q operations of addition. That is the problem of using this algorithm to reduce data redundancy. One of the solutions to this problem is offered in this work. The amount of addition operations when searching q in (14) can be reduced if the search is started from pair (c_r, l_r) and previously calculated sum $L_r < p$. Thus it is proposed to split the array S_{RLE} into T portions with assigning corresponding ranges (15) of indicator indices:

$$(p_{i_{\min}}, p_{i_{\max}}), \; \forall i = 1,2,...,T \,, \tag{15}$$

Moreover, $p_{i\ min} = p_{(i-1)\ max} + 1$, $\forall i = 2,3,\ldots,T$ and $p_{T\ max} = R$. In this case, the search of the desired pair (c_q, l_q) starts from the search of the portion t, which encodes the indicators with the indices from the range (15) where index p falls:

$$p_{t_{\min}} \le p \le p_{t_{\max}} \,. \tag{16}$$

After the portion t of pairs is found, the search of desired pair (c_q, l_q) is performed according to (14) as it was described before, but with significant improvement in the form of $L_0 = p_{t\ min}$ -1. If the algorithm of binary search is used for finding the portion according to (16), it takes $O(\log_2 T)$ operations, that is much less than linear sum (14).

Another advantage of indicator array splitting is the possibility of its parallel processing. As was said before, high computational requirements are another problem of the region of acceptability approximation. Thus, the solving of this task is almost incogitable without parallel computations. The algorithm (Abramov & Nazarov 2011) of region of acceptability approximation on the basis of the model (11) represents the same instructions performed for every elementary box (SPMD – Single Process, Multiple Data). This fact allows for task decomposition on the basis of indicators array splitting.

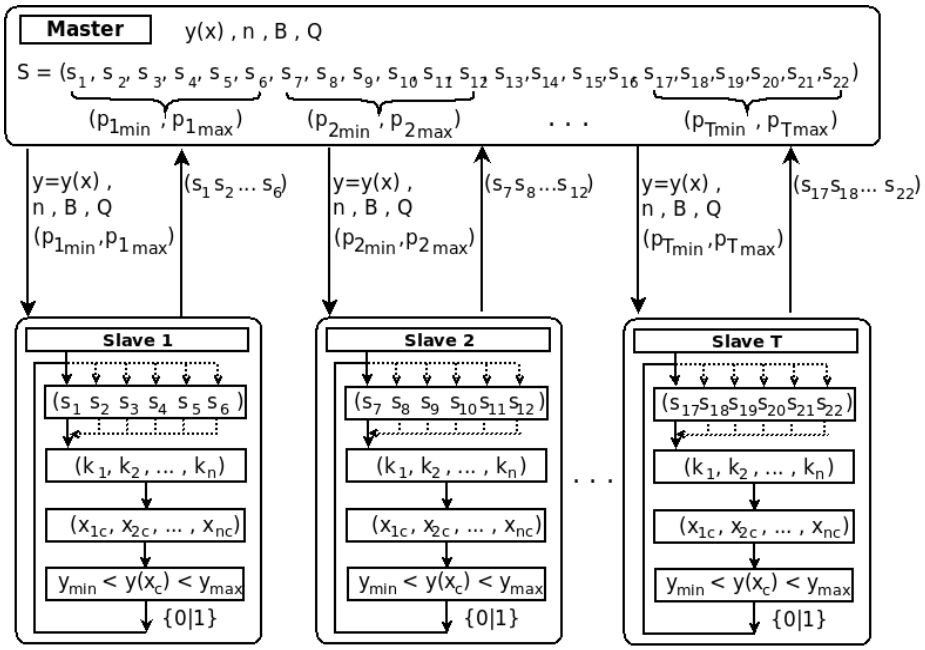


Figure 5. The indicators array decomposition and its parallel processing.

The construction of the region of acceptability approximation on the basis of the model (11) in parallel processes requires passing the model parameters n, B, Q and the range (15) to every parallel process. Using these parameters, every process is able to restore univalent indices $(k_1,k_2,\ldots,k_n)$ on the every iteration of indicators enumeration inside the range, and, consequently, to calculate the performances (1) and membership function (9). This process is illustrated in Figure 5.

4 CONCLUSIONS

The problem of large amount of data in the framework of the region of acceptability construction on the basis of approximation with a discrete set of elementary boxes is considered in this work. The ways of reducing of the data redundancy and corresponding problem of data access are considered. The methods of increasing random access speed on compressed data are offered. The efficiency of indicators array partitioning both for access speed and for the task decomposition for using of parallel computations technique is presented in this work.

5 REFERENCES

O.V. Abramov, Y.V., Katueva, and D.A. Nazarov, "Reliability-Directed Distributed Computer-Aided Design System," Proc. Of the IEEE International Conference on Industrial Engineering and Engineering Management. Singapore, 2007, pp. 1171 – 1175.

O.V. Abramov, D.A. Nazarov, "Regions of Acceptability for Reliability Calculation and Creation," Proceedings of the 7th International Conference on "Mathematical Methods in Reliability: Theory. Methods. Applications. (MMR2011)" / edited by Lirong Cui & Zian Zhao. - Beijing: Beijing Institute of Technology Press, 2011, pp. 600 - 606.

M. Conti, S. Orcioni, C. Turchetti, "Parametric Yield Optimization of MOS VLSI circuits based on simulated ennealing and its parallel implementation," IEE Proc. Circuits Devices Syst., vol. 141, No 5, October 1994. pp. 387 – 398.

J.A.G. Jess., K. Kalafala, S.R. Naidu, R.H.J.M. Otten, C. Visweswariah, "Statistical timing for parametric yield prediction of digital integrated circuits," Proc. Of the 40th conference on Design automation, June 02-06, 2003, Anaheim, CA, USA. – ACM, New York, NY, USA, 2003, pp.932 - 937.

D. Salomon, "Data compression: the complete reference, Volume 10," Springer, 2007.

G. Stehr, "On the Performance Space Exploration of Analog Integrated Circuits," Ph.D Thesis. Verlag Dr. Hut, Munich, 2005.

RT&A # 03 (26)
(Vol.7) 2012, September

ASYMPTOTICS OF CONNECTIVITY PROBABILITY OF GRAPH WITH LOW RELIABLE ARCS

G. Tsitsiashvili

IAM FEB RAS,
690041, Radio str., 7, Vladivostok, Russia

guram@iam.dvo.ru

ABSTRACT

In this paper a problem of asymptotic estimate for connectivity probability of non oriented connected graph with fold and low reliable arcs is solved. An algorithm of a calculation of asymptotic constants with cubic complexity by a number of nodes is constructed. This algorithm is based on Kirchhoff`s theorem for a calculation of a number of spanning trees and relative characteristics.

1. INTRODUCTION

A problem of connectivity probability calculation for random graphs with low reliable arcs is considered in manifold articles. In [1], [2] upper and lower bounds for connectivity probability (reliability polynomial) in general type networks are constructed using maximal systems of non intersected skeletons. Networks for which these bounds equal zero are constructed. In [3] connectivity probability for multi graphs (graphs with fold arcs) is analysed. For relatively small number of arcs in [4] accelerated algorithms of reliability polynomial coefficients calculation are constructed. These algorithms showed sufficiently good results in a comparison with Maple 11. In [5] this problem is solved using Monte-Carlo method with some specific combinatory formulas.

But for large number of arcs these calculations become sufficiently complicated. So a problem of a construction of sufficiently convenient asymptotic formulas for a calculation of connectivity probability in graphs with high reliable or low reliable arcs becomes actual. First problem is solved in [6] for planar graphs using Burtin-Pittel asymptotic formula [7], Whithney theorem about a relation between cross sections in planar graphs and cycles in dual graphs [8] and Harary formulas [9] for numbers of simple (without self intersection) cycles with length not larger than 5 in non oriented connected graph.

In this paper a problem of asymptotic estimate for connectivity probability of non oriented connected graph with fold and low reliable arcs is solved. An algorithm of a calculation of asymptotic constants with cubic complexity by a number of nodes is constructed. This algorithm is based on Kirchhoff`s theorem for a calculation of a number of spanning trees [11] and relative characteristics. It is worthy to note that last years this theorem and its manifold generalizations are widely used in different disciplines: physics of elementary particles, electromagnetism, acoustics, elasticity theory, sociology, biology etc. [10] – [12]. In this paper Kirchhoff`s theorem is used in reliability theory.

2. MAIN RESULTS

Assume that non oriented and connected and simple (without loops and fold arcs) graph G has the nodes set $U = \{1, \dots, n\}$ and the arcs set V. Each arc $v = (i,j)$, $1 \le i \ne j \le n$, of the graph G works with the probability $p(v)$ and fails with the probability $1 - p(v)$ and arcs work independently. We shall analyse connectivity probability $p(G)$ of the graph G with randomly working and low reliable arcs. Define Kirchhoff`s matrix $K = \|k(i,j)\|_{i,j=1}^{m}$ of the graph G :

$$k(i,j) = \begin{cases} degree\ of\ \ node\ i, & i = j, \\ -1, & (i,j) \in V, \\ 0, & else. \end{cases}$$

Here node`s degree is a number of its incident arcs. Spanning tree of the graph G is a tree with the nodes set U and with the arcs set which contains in V. Denote m the number of spanning trees of the graph G.

Theorem 1. If $p(v) = h,\ h \in V$ then for $h \to 0$ we have the equality

$$P(G) = mh^{n-1}\big(1 + O(h)\big) \qquad (1)$$

where calculation complexity of the coefficient m definition is $O(n^3)$.

Proof. Formula (1) is proved in [2, Formula (5)]. From Kirchhoff`s-Trent theorem (see for example [10]) algebraic complements of all elements of Kirchhoff`s matrix K are equal to the number m of all spanning trees in the graph G. It is well known (see for example [13]) that a calculation of a determinant with the order $n-1$ and consequently a calculation of the coefficient m by the Gauss method demands $O((n-1)^3)$ arithmetical operations.

Corollary 1. If for some natural $l(v)$ the probability $p(v) = h^{l(v)}$ then each arc in the graph G may be replaced by $l(v)$ serial connected arcs , $v \in V$, and then Theorem 1 may be applied to this graph .

This corollary may be used to analyse random time to a failure of network connectivity if analogous characteristic for an arc is $p(v) = P(\tau(v) > T)$, $T \to \infty$.

Theorem 2. If for some positive $s(v)$ the probabilities $p(v) \sim s(v)h,\ h \to 0$, $v \in V,$ then

$$P(G) = m_1 h^{n-1}\big(1 + O(h)\big)\,,\ h \to O\,, \qquad (2)$$

Here m_1 is algebraic complement of each element (which are equal) of the matrix $K_1 = \|k_1(i,j)\|_{i,j=1}^{m}$ where

$$k_1(i,j) = \begin{cases} \sum\limits_{t \in V, (i,t) \in V} s(i,t)\,, i = j \\ -s(i,t), (i,j) \in V \\ 0, else. \end{cases}$$

Proof. Denote $G_1, \dots, G_m$ spanning trees of the graph G. Each of them has $n-1$ arcs. Put A_k the event that all arcs in the tree G_k work, $1 \le k \le m$. Then we have: $P(G) = P(\bigcup_{k=1}^{m} A_k)$ and so $\sum_{k=1}^{m} P(A_k) - \sum_{1 \le k_1 < k_2 \le m} P\big(A_{k_1} A_{k_2}\big) \le P(G) \le \sum_{k=1}^{m} P(A_k)$.

From the conditions on $p(v)$ we obtain that

$\sum_{k=1}^{m} \prod_{v \in G_k} hs\,(v) - O(h^n) \le P(G) \le \sum_{k=1}^{m} \prod_{v \in G_k} hs\,(v)$,

$$\sum_{k=1}^{m} \prod_{v \in G_k} hs\,(v) = h^{n-1} \sum_{k=1}^{m} \prod_{v \in G_k} s\,(v)$$

Designate $m_1 = \sum_{k=1}^{m} \prod_{v \in G_k} s\,(v)$ and obtain Formula (2). From a generalization of Kirchhoff's-Trent theorem (see for example [11, Theorem 1]) we obtain that m_1 coincides with algebraic complement of each element of the matrix K_1 and calculation complexity of its definition is $O(n^3)$.

Remark 1. Theorem 2 statement may be spread onto multi graph constructed from the graph G by a replacement of each arc $v \in V$ by $s(v)$ parallel and independently working copies of this arc. This parallel connection has working reliability $\sim s(v)h$, $h \to 0$.

REFERENCES

1. Lomonosov M.V., Polesskiy V.P. 1971. Upper bound for reliability of information networks. *Problems of information transmission.* vol.7, No. 4. P. 78-81. (In Russian).
2. Lomonosov M.V., Polesskiy V.P. 1972. Lower bound of network reliability. P*roblems of information transmission.* vol. 8. No 2. P. 47-53. (In Russian).
3. Popkov V.K. 2006. Mathematical models of connectivity. *Novosibirsk: Institute for Computing Mathematics and Mathematical Geophysics, Siberian Branch of RAS.* (In Russian).
4. Rodionov A. 2011. To question of acceleration of reliability polynomial coefficients calculation in random graph *Automatics and remote control,* no. 7: P. 134-146. (in Russian).
5. Gertsbakh I., Shpungin Y. 2010. Models of Network Reliability. *Analysis, Combinatorics and Monte-Carlo. CRC Press. Taylor and Francis Group.*
6. Tsitsiashvili G.Sh. 2012. Complete calculation of disconnection probability in planar graphs. *Reliability: Theory and Applications.* Vol. 1. No 1. P. 154-159.
7. Burtin Yu., Pittel B. 1972. Asymptotic estimates of complex systems reliability. *Automatics and remote control.* No. 3. P. 90-96. (In Russian).
8. Whithney H. 1932. Nonseparable and planar graphs. *Transactions of American Mathematical Society.* Vol. 34. P. 339-362.
9. Harary F. 1973. Graph theory. Moscow: Mir. (In Russian).
10. Priezjev V.B. 1985. Problem of dimers and Kirchhoff's theorem. UFN, vol. 147. P. 747-765.
11. Chebatarev P.Yu, Shamis E.V. 1997. Matrix theorem about forests and measurement of connections in small social groups. *Automatics and remote control.* No. 9. P. 125-137. (In Russian).
12. Kozma R., Puljic M., Balister P., Bollobas B., Freeman W. 2005. Phase transition in the neuropercolation model of neural populations with mixed local and non-local interactions. *Biological Cybernetics.* vol. 92. P. 367-379.
13. Ilyin V.A., Pozniak A.G. 2004. Linear algebra. Moscow: Phyzmatlit. (In Russian).

MARKOV MODELS FOR TENSILE AND FATIGUE RELIABILITY ANALYSIS OF UNIDIRECTIONAL FIBER COMPOSITE

Yuri Paramonov[1], Rafal Chatys[2], Janis Andersons[2], Viacheslavs Cimanis, Martinsh Kleinhofs[1]

[1]Aviation Institute, Riga Technical University, Lomonosova 1, Riga LV 1019, Latvia
[2]Institute of Polymer Mechanics, University of Latvia, Aizkraukles 23, Riga LV 1006, Latvia
[1]Corresponding author e-mail: yuri.paramonov@gmail.com

Abstract

This paper is a review integrating, amending, and developing the approach applied in authors' previous works devoted to the tensile and fatigue reliability analysis of unidirectional composite material considered as a series system the links of which are, in general case, complex parallel systems with redistribution of load after failure of some items. By processing experimental data it is shown that the models based on the Markov chains (MCh) theory allow (1) to describe connection of cdf of tensile strength of fibers (strands) and a composite specimen, (2) to perform nonlinear regression analysis of fatigue curve and prediction of its changes due to a change of tensile strength characteristics of the composite components, (3) to predict the fatigue life at a program loading, (4) to estimate the cdf of the residual strength and residual life after a preliminary fatigue load.

Keywords: Composite, Fiber, Fatigue life, Strength

1. INTRODUCTION

The distribution of static strength, the fatigue curve, and the accumulation of fatigue damage under a program loading are often described by poorly interconnected hypotheses. The distribution of static strength is usually analyzed by the Weibull or lognormal distributions, while the fatigue curve is described by formal regression dependences. The accumulation of fatigue damage under a program loading, as a rule, is carried out by using the Palmgren-Miner hypothesis or its modifications. Here we consider the application of the Markov chain (MCh) theory for a unified approximate solution of the mentioned problems. Application of the Markov process theory for specific problems is discussed in several publications (see, for example, [1]-[2] as the most interesting) but the idea of connection of cdfs of tensile strength, fatigue life, residual strength and residual fatigue life (after some preliminary fatigue loading) with the cdf of tensile strength of a composite material component is relatively new. First steps in that direction were made in 1980 [3]. This paper is a review integrating, amending, and developing the approach applied in authors' previous works [4-7] "furnished" with examples of solution of the above-mentioned problems for unidirectional fibrous composite within the framework of some specific case of unified mathematical model.

Actually this paper is an addition to [4]. The main new idea of the paper is to show that nearly the same type of MCh model can be used for the case of tensile strength, as well as for the case of fatigue life analysis of a composite material. A new idea of using random Daniels sequence is discussed also.

2. MODEL OF A UNIDIRECTIONAL FIBER-REINFORCED COMPOSITE MATERIAL. MAIN IDEAS

We consider the composite specimen as a series system with n_L links, a random number of which, K_L, $1 \le K_L \le n_L$ have defects. We call them weak micro-volumes (WMV). Contrary to a general set of probability structure (ps) for a single fiber (strand), described in [4], in which the failure of both types of links (with defect and without defect) can be the cause of failure of

specimens here we suppose that failure of the specimens can be only as the consequence of failure of some WMV. This assumption is equivalent to the assumption that the strength of links without defects is equal to infinity or very large.

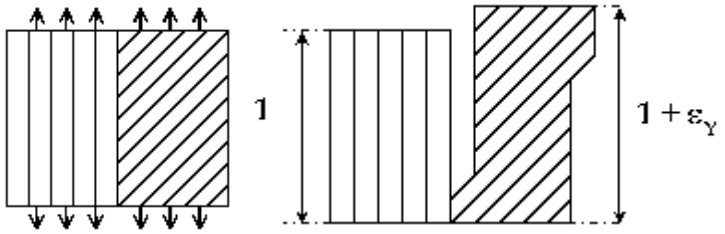


Fig.1. Model of the weak microvolume of a composite under a load and after removal of the load .

We suppose that in general case the WMW consists of n_C perfectly elastic (brittle) longitudinal items (LI) (fibers or strands) and a matrix where plastic deformations are accumulated if cyclic loading takes place (Fig.1). We assume that, except for the LIs, the plastic part includes all other composite components, i.e. the matrix itself and all the layers with stacking different from the longitudinal one! And we assume finally that, if the number of LIs in the WMV decreases by r_R items, the elastic part of the specimen breaks down, which is followed by the failure of the specimen as a whole. The total number of LIs in one WMV, n_C, in general case can be more than r_R but we suppose that failure of (r_R / n_C)-th part of elastic LIs of WMV is considered as failure of elastic part and the whole WMV also by definition. The value of $\overline{f}_C = r_R / n_c$ is a parameter of the model. The slanted hatching in Fig.1 symbolically points to the possibility of accumulating an irreversible plastic strain. If it exceeds some critical level ε_{YC}, the failure of the WMV and the specimen as a whole takes place also. We emphasize that this graphic image, as applied to a composite, should be understood symbolically. It is more suitable for metals, where the accumulation of plastic strains is associated with some “act of flow” (for metals - a shift over slip planes). We assume that one act of flow leads to the appearance of a constant plastic strain ε_{Y1}. The failure of WMV takes place after the accumulation of a “critical” number of such acts, r_Y, i.e., after the accumulation of a critical plastic strain, ε_{YC}, for which the relation $\varepsilon_{YC} = \varepsilon_{Y1} r_Y$ is valid, where ε_{YC} and r_Y are model parameters. Since the elastic and plastic parts are integrated in a unit, the accumulation of plastic strains (irreversible deformation of the plastic part) leads to the appearance of residual stress: tension in the elastic and compression in the plastic part of the specimen [4].

For description of the process of failure of WMV using MCh theory we should provide the description of space of states of MCh and its connection with the structure of the composite WMV, the corresponding structure of the matrix of transition probabilities and its connection with the cdf of mechanical characteristic of the component of WMV, the process of loading.

2.1. Probability description of WMV

2.1.1. Description of space of states and transition probability matrix

As it was already mentioned, a set of probability structures (ps) for description of **fiber (or strand) as series system** is considered in [4]). Now we consider probabilities structure of specimen but, first, ps of one WMV.

Let us, in general case, associate the process of gradual failure of a WMV with an absorbing MCh the set of states of which is determined by the number of broken LIs and the

number of acts of flow. The matrix of transition probabilities is presented as a totality of (r_Y+1) blocks with (r_R+1) states within each of them. Then, the indices of input and output states, i and j, can be expressed in terms of the corresponding local indices i_Y, i_R, j_Y and j_R: $i=(r_R+1)(i_Y-1)+i_R$; $j=(r_R+1)(j_Y-1)+j_R$.

Table 1 shows an example of (symbolic) filling of the matrix for the case where $r_Y=r_R=2$ for independent failure of LI and act of flow. In this case the destruction of a specimen occurs if two LIs fail (event A), or two acts of flow are accumulated (event B), or events A and B take place simultaneously. The absorbing states of the MCh correspond to these events. In the example considered, there are $(r_Y+1)(r_R+1)=9$ states. The symbols p_{R0}, p_{R1}, … designate the probabilities of failure of the corresponding numbers of elastic (rigid) elements; p_{Y0}, p_{Y1}, … are the probabilities of the corresponding numbers of acts of flow (yielding).

TABLE I
EXAMPLE OF THE TRANSITION PROBABILITY MATRIX

		j_Y	*1*			*2*			*3*		
		j_R	*1*	*2*	*3*	*1*	*2*	*3*	*1*	*2*	*3*
i_Y	i_R	i \ j	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1	*1*	**1**	$p_{R0}p_{Y0}$	$p_{R1}p_{Y0}$	$p_{R2}p_{Y0}$	$p_{R0}p_{Y1}$	$p_{R1}p_{Y1}$	$p_{R2}p_{Y1}$	$p_{R0}p_{Y2}$	$p_{R1}p_{Y2}$	$p_{R2}p_{Y2}$
	2	**2**	0	$p_{R0}p_{Y0}$	$p_{R1}p_{Y0}$	0	$p_{R0}p_{Y1}$	$p_{R1}p_{Y1}$	0	$p_{R0}p_{Y2}$	$p_{R1}p_{Y2}$
	3	**3**	0	0	1	0	0	0	0	0	0
2	*1*	**4**	0	0	0	$p_{R0}p_{Y0}$	$p_{R1}p_{Y0}$	$p_{R2}p_{Y0}$	$p_{R0}p_{Y1}$	$p_{R1}p_{Y1}$	$p_{R2}p_{Y1}$
	2	**5**	0	0	0	0	$p_{R0}p_{Y0}$	$p_{R1}p_{Y0}$	0	$p_{R0}p_{Y1}$	$p_{R1}p_{Y1}$
	3	**6**	0	0	0	0	0	1	0	0	0
3	*1*	**7**	0	0	0	0	0	0	1	0	0
	2	**8**	0	0	0	0	0	0	0	1	0
	3	**9**	0	0	0	0	0	0	0	0	1

Since the local order number of state is defined by the number of failed LIs, it is connected also with the local stress in intact LIs. The set of states of MCh can be connected also not only with the number intact LIs but with the set of corresponding values of local stress. For tensile test it is more convenient to use the connection with the intact (or failed) LIs. For fatigue test it is more convenient to consider the connection with the local stress (see the definition of Daniels' sequence in the following).

Consider now the simplest **special case** (the most interesting for **tensile** test of a unidirectional composite) when there are only n_C LIs (fibers or strands). Equality $r_R=n_C$ is used for the definition of failure of WMV, and the existence of composite matrix is not taken into account. In this case the WMV is a parallel system with (n_C-K_C) LIs, where n_C is a constant (initial number of LIs without any defect), $K_C, 0\le K_C\le n_C$, is the number of failures of LIs in the link. Note that in this case the equality $K_C=n_C$ means the failure of link (WMV) and the specimen also.

Note also, that for this type of WMV there is only one block in the transition probability matrix P corresponding to $r_Y = 0$, $r_R = n_C$. The number of states of MCh is equal to $(r_R + 1)$.

Note again, that it is not necessary to connect the failure of WMV with equality $K_C = n_C$. We can in general case to define the failure of unidirectional WMV as the event when the intact part of LIs becomes less than some critical value $\bar{f}_C$: $(n_C - K_{Ci})/n_c < \bar{f}_c$. In this case $r_R = [\bar{f}_C n_C] + 1$, where $[x]$ is the integer part of x.

For the considered type of WMV the four main versions (hypotheses) of the structure of matrix P, denoted as P_a, P_{an_C}, P_b and P_c, are considered in [4]. Matrix P_a corresponds to the assumption that, in one step of MCh, only one LI can fail, and it is the nearest one to the already failed LIs (in some way this corresponds to development of a crack); P_{an_C} corresponds to the failure of the weakest item in the considered WMV(cross section); P_b corresponds to a binomial distribution of the number of failures at every step of MCh; P_c corresponds to the case when the stress concentration function is known.

Total initial load of this type of WVM is equal to Sn_C, where S is the initial stress (load of one LI). For the three first types of matrix a uniform distribution of load between intact LIs is supposed. Then, if the number of failed LIs is equal to i, the stress in the still intact LIs will be equal to $Sn_C/(n_C - i)$. For the matrix of type P_c the function of stress distribution across the cross section of WMV should be known (see the details in [4]). This connection of the local stress and the number of states of MCh should be taken into account for calculation of matrix of transition probabilities.

The corresponding set of states can be used also for modeling of **fatigue** test (again, see details in [4]). But in [6,7] the set of MCh states is connected with the random Daniels' sequence (RDS). The components of RDS, $\{s_0, s_1, s_2, ...\}$, correspond to the **random** process, a realization of which has the following form: $s_{i+1} = S/(1 - \hat{F}_{X_s,n_C}(s_i))$, $i = 0,1,2,...,n_C$, where $s_0 = S$, S is the initial stress (initial load of one LI), $\hat{F}_{X_s,n_C}(.)$ is the estimate of cdf of strength of a LI, which is defined by some sample $(x_{s1},...,x_{sn_C})$ of observations of n_C random variables (random strength of n_C LIs) with the same cdf $F_{X_s}(x)$. Here we use the following definition: $\hat{F}_{X_s,n_C}(x) = k(x)/n_C$, where $k(x)$ is the number of observations which are lower than x. or equal to x.So here $(x_{s1},...,x_{sn_C})$ is a vector of observations of random strengths, $X_{s1},...,X_{sn_C}$, of components of some WMV, $s_{i+1} = S/(1 - k(s_i)/n_C)$, $i = 0,1,2,...,n_c - 1$. In following we suppose that $(x_{s1},...,x_{sn_C})$ is the vector of ordered statistics: $x_{s1} \le x_{s2} \le ... \le x_{sn_C}$.

The increase of local stress corresponds to decreasing of local cross section (reduction of the number of intact components of WMV). Let us again define that the failure of WMV takes place if local cross section become less than some value $\bar{f}_C$ (initial total cross section area of WMV is equal to one). Here $\bar{f}_C$ is a constant, a parameter of the considered model. Then critical local stress corresponding to this event, S^*_{UT}, is defined as minimum of stress, s, for which the part of LIs with strength less than s is more than $\bar{f}_C$: $S^*_{UT} = \min\{s : \hat{F}_{X_s}(s) \ge \bar{f}_C,\ s \in \{s_0, s_1, s_2, ...\}\}$, where $\{s_0, s_1, s_2, ...\}$ is RDS. The random number $N_{RDS} = \max\{i : s_i < S^*_{UT},\ s_i \in \{s_0, s_1, s_2, ...\}\}$, we call as RDS fatigue life (RDSFLf) at stress S. Here s_i is an item of RDS (for specific S, for specific realization $x_{s,1-n_C} = (x_{s1},...,x_{sn_C})$ of random vector of ordered statistics $X_{s,1-n_C} = (X_{s1},...,X_{sn_C})$).

Let us define the random function $S_{DFLm}(k_S, x_{s,1-n_C})$ the value of which is equal to the maximum value of S for which in RDS with $s_0 = k_S S$ and specific $x_{s,1-n_C}$ there is some i for which $s_i = s_{i+1} = s_{i+2} = ... = \max(s_0, s_1, s_2, ...) = s^*(S, k_S, x_{s,1-n_C}) < \infty$. Growth of stress in RDS is stopped after it reaches $s^*(S, k_S, x_{s,1-n_C})$ which is the solution of the equation $x = k_S S / (1 - F_{X_s}(x))$ or $k_S S = x(1 - F_{X_s}(x))$. We call $S_{DFLm}(k_S, x_{s,1-n_C})$ the k_S-RDS-fatigue-limit (k_S-RDSFLm)) because if initial stress, S, is lower its value then corresponding k_S-RDSFLf (for $s_0 = k_S S$ and specific $x_{s,1-n_C}$) is equal to infinity. Existence of k_S-RDSFLm explains the phenomenon of random fatigue limit.

Solution of the equation $k_S S = x(1 - F_{X_s}(x))$ exists if $k_S S \le \max x(1 - F_{X_s}(x))$. In accordance with Daniels [8,9] the strength of a parallel system of n_C LIs is random variable $S_D = \max x(1 - F_{X_s}(x))$ with asymptotically normal distribution $N(\mu_D, \sigma_D^2)$, where parameters μ_D and σ_D are defined by the cdf of strength of LI . By fitting the fatigue life data by k_S-RDS-MCh model we can find an appropriate estimate parameters of the model including k_S. Then we can make estimate the probability that k_S-RDSFLf is equal to infinity: $p_{\inf} = P(T_C = \infty \mid S) = P(S \le S_{DFLm}(k_S, x_{s,1-n_C})) = \Phi\left((k_S S - \mu_D)/\sigma_D\right)$.

Example of Monte Carlo calculation of "normalized" RDS, $((1/k_S)\{s_0, s_1, s_2, ...\}$, $s^*(S, k_S, x_{s,1-n_C})$ and $p_{\inf}$ for different k_S and 10 random realizations of $x_{s,1-n_C}$ was provided in [7] . Similar result for the same initial data is shown in Fig. 2.

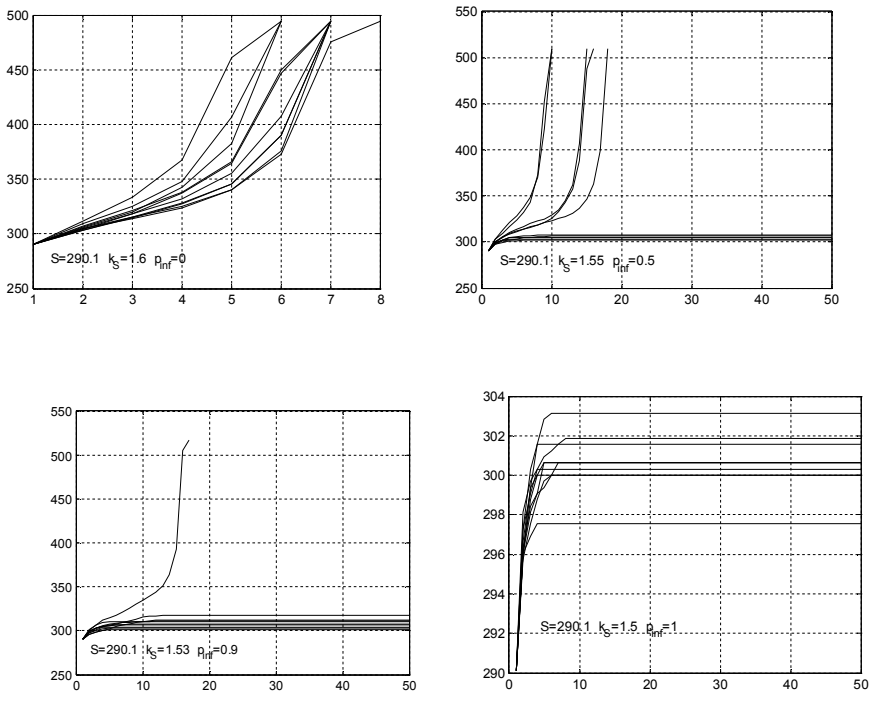


Fig. 2. Examples of 'normalized' RDSs and estimates of $p_{\inf}$ for fatigue test of carbon-fiber composite [5] for S = 290.1 MPa and different k_S (see [7])

If for calculation RDS we use $k_S = 1$ then RDSFLfs are very small, RDSFLm is very high (see [4,6,7]). So although the use of RDS provides a qualitative explanation of fatigue failure of a composite material and can also explain the phenomenon of fatigue limit, the quantitative prediction is very poor.

But the possibility of explanation of the existence of phenomenon of fatigue limit is very attractive. The very high value of RDSFLm can be explained by existence of local stress concentration, i.e. instead of equality $s_0 = S$ (see [6]) the initial stress in RDS should be defined by equality $s_0 = k_S S$, where k_S is a local stress concentration coefficient. The probability characteristics of fatigue life and appropriate description of fatigue curve in the framework of this model can be fitted to the real characteristics of fatigue life using the theory of MCh with space of states based on RDS.

For this type of MCh the first r states of state space are related with the items of RDS, $\{s_0, s_1, ..., s_{r-1}\}$, s_r is connected with the absorbing state . In Fig.2 we see two types of RDS. For RDS of first type items of RDS grow up to infinity. For this type of RDS-MCh model, absorbing state is connected with the event that the local stress is equal or larger than S^*_{UT}. For the second type of RDS there is a final limit and absorbing state should be connected with $s^*(S, k_S, x_{s,1-n_C})$. In the simplest case it can be assumed that only transitions to the nearest 'senior' state can take place. So the following simple matrix of transition probabilities can be considered:

$$P = \begin{bmatrix} q_1 & p_1 & 0 & & & \cdots & 0 \\ 0 & q_2 & p_2 & 0 & & \cdots & 0 \\ 0 & 0 & q_3 & p_3 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & & & \cdots & 0 & q_r & p_r \\ 0 & & & \cdots & 0 & 0 & 1 \end{bmatrix},$$

where $p_i = (k(s_i) - k(s_{(i-1)}) / (n_C - k(s_{(i-1)}))$, $q_i = 1 - p_i$. Note that here P is a realization of random matrix $P = P(x, X_{s,1-n_C})$. It is a function of load x and random vector of strength of n_C LIs, $X_{s,1-n_C} = (X_{s,1}, ..., X_{s,n_C})$. So all results of calculation using this matrix will be random if n_C is not large enough. In order to get mean results the Monte Carlo method can be used. But if n_C is large enough then, for example, if there is normal distribution, $N(\theta_0, \theta_1^2)$, of logarithm of strength of LI then the items of matrix P , approximately, can be defined in following way : $p_1 = \Phi((\log(k_S S) - \theta_0) / \theta_1)$; $s_2 = k_S S / (1 - p_1)$; $p_i = (p_{ic} - p_{(i-1)c}) / (1 - p_{(i-1)c})$,

where $p_{ic} = \Phi((\log(s_i) - \theta_0) / \theta_1)$, , $s_{i+1} = k_S S / (1 - p_{ic})$, $i = 1, 2, ..., r$.

and the corresponding nonrandom matrix P can be used. A numerical example for this special case is considered in [6].

2.1.2. Description of the process of loading. CDF of tensile strength and fatigue life of WMV

By renumbering the states, the matrix of transition probabilities of any absorbing MCh can be reduced to the form

$$P = \begin{bmatrix} Q & R \\ 0 & I \end{bmatrix},$$

where I is the unit matrix and 0 is the matrix consisting of zeros.

As it was told already the matrix P is a function of **loading** stress, x. **Loading process** in **tensile** test is described by an ascending up to infinity sequence $x1\infty = \{x_1, x_2, ..., x_t, ...\}$. The cdf of the number of steps up to absorption, T_A, is defined by matrix P, a priori probability distribution on MCh states, π, and by an sequence of loads, $x1\infty$:

$$F_{T_A}(t) = \pi(\prod_{j=1}^{t} P(x_j))u, \; x_j \in \{x_1, x_2, ..., x_t, ...\} \tag{1}$$

where vector-column $u = (0,0,...,1,...,1)'$ has only zeros and units (for absorbing states). By the choice of π we can model different levels of a priori quality of tested specimens (for example, the modified binomial distribution of initial number of intact LIs in one WMV can be modeled).

The load corresponding to the time to absorption, x_{T_A} is a random tensile strength. If $x_t = g(t)$, where $g(.)$ is monotonicaly increasing function for which there is the inverse function $g^{-1}(.)$, then cdf of random variable $X = x_{T_A}$ is

$$F_X(x) = P(g(T_A) \le x) = F_{T_A}(g^{-1}(x)), \quad x \in x1\infty. \tag{2a}$$

For **fatigue** test for estimation of fatigue life at a stress level x, all items in the sequence $x1\infty = \{x_1, x_2, ..., x_t, ...\}$ are equal: $x_1 = x_2 = x_3 = ... = x$ where x is some parameter of cycle (for example, x is a maximum stress of pulsating cycle). Then fatigue life (cycle number up to failure) is equal to $T_C = k_m T_A$ cycles, where k_m, $k_m > 0$, is a scale factor, i.e. it is the number of cycles corresponding to one MCh step. In this case $\prod_{j=1}^{t} P(x_j) = P^t(x)$ and $F_{T_A}(t) = \pi P^t(x)u$, $t = 0,1,2,...$. Cdf of the number of cycles up to failure, T_C, is defined by equation

$$F_{T_C}(n) = F_{T_A}(n / k_m) = \pi P^{n/k_m}(x)u, \quad n = 0, k_m, 2k_m, 3k_m, ... \tag{2b}$$

Note again that we have random results of equations (1, 2a, 2b) if we have random matrix $P = P(x, X_{s,1-n_C})$ (if MCh state space is defined by RDS) but using Monte Carlo method (or if n_C is large enough) we can get approximately determined results.

Examples of equations for calculation of P for fatigue loading by pulsating cycle taking into account presence of the plastic part of WMV are given in [4]. A version of "translation" of any cycle with any other asymmetry into a pulsating cycle is given in [5].

Let us denote by S_n the **conditional** fatigue limit at n cycles of load. Then cdf of S_n is

$$F_{S_n}(x) = P(T_C \le n \mid S = x) = \pi P^{[n/k_m]}(x)u,$$

where P is defined by stress x, $[x]$ is the integer part of x, π and u are the same as previously.

It is worth to note that probability that fatigue life is larger than n at stress S is defined by equation $P(T_C > n \mid S = x) = 1 - \pi P^{[n/k_m]}(x)u$. This probability we can estimate relatively easy. But it is very difficult to estimate function $F_{S_n}(x)$.

As it was shown already, using RDS models we can estimate the probability that fatigue life at a stress level S is equal to infinity.

Let us denote $P^t = \begin{bmatrix} Q_t & R_t \\ 0 & I \end{bmatrix}$. Different columns of the matrix $B_t = (I - Q_t)R_t$ define the probabilities of absorption (failure) in different absorbing states for different initial states (rows).

For example, matrix element in the right upper corner of B_t corresponds to probability of failure of the WMV as consequence of failure of all LIs (or, in general case, for the composite matrix the relation $\varepsilon_{YC} = \varepsilon_{Y1}\, r_Y$ is reached) at MCh step number t if the initial state $i = 1$ (see Table 1).

But if we do not need this detailed information then instead of matrix P we can use the matrix P_U in which all the absorbing states are united in one absorbing state. The number of states of matrix P_U is equal to $h = r_Y r_R + 1$. The matrix P_U is useful for analysis of fatigue life in program loading. Cdf of fatigue life for a program loading defined by the sequence $x1\infty$ corresponding to some specific program of loading can be calculated again using equations (2a) and (2b). Consider specific two-stage fatigue loading: $x1t = \{x1t_1, xt_{1+1}\infty\} = \{\{x_1, x_2, ..., x_{t_1}\}, \{x_{(t_1+1)}, x_{(t_1+2)}, ..., x_{(t_1+\Delta t)}, ...\}\}$; $x_1 = x_2 = , ..., = x_{t_1} = x^I$, $x_{(t_1+1)} = x_{(t_1+2)} = , ..., = x_{(t_1+\Delta t)} = ... = x^{II}$, x^I is the stress in the first fatigue loading stage, x^{II} is the stress in the second fatigue loading stage. After preliminary loading $x1t_1 = \{x_1, x_2, ..., x_{t_1}\}$ an a priori distribution π_0 is transformed into „a posteriori" distribution $\pi_{t_1} = \pi_0 \prod_{j=1}^{t_1} P(j)$. Using π_{t_1} instead of π in equations (2a) and (2b) we can get the cdf of both residual strength and residual fatigue life ΔT in two-stage fatigue loading.

It is necessary to note that usually we are interested in these characteristics only for the specimens which are still intact after the preliminary fatigue loading. The components of „a posteriori" conditional distribution for them are: $\pi_{t_1 c}(k) = \pi_{t_1}(k) / (1 - \pi_{t_1}(h))$ for $k = 1, 2, ..., (h-1)$, $\pi_{t_1 c}(h) = 0$. This distribution, matrix P_U and the second stage loading $xt_{1+1}\infty = \{x_{(t_1+1)}, x_{(t_1+2)}, ..., x_{(t_1+\Delta t)}, ...\}$ should be used in (2a) and (2b) instead of π, P and $x1\infty$ for calculation of cdf of residual step number ΔT_A up to absorption (WMV failure). Now the residual strength is rv $x_{(t_1+\Delta T_A)}$ (see also in [4] the definition of the so called k_m-residual strength which defines the stress of fatigue cycle which produces the fatigue failure in k_m fatigue cycles which are equivalent to one step of MCh. Its cdf is defined by equation

$$F_{S_{x1t_1}}(x_j) = \pi_{t_1 c} P(x_j) u ,$$

where $x_j \in \{x_{t_1+1}, x_{t_1+2}, ...\}$, $x_{t_1+1} \geq x^I$, $\{x_{t_1+1}, x_{t_1+2}, ...\}$ is an ascending up to infinity sequence of stresses in "residual" tensile strength test) ,

2.2. Probability description of a specimen

We suppose that in the simplest case, neglecting the presence of composite matrix in a unidirectional composite (sequence of links), there are two types of links: there are K_L, $1 \leq K_L \leq n_L$, links with defects and $(n_L - K_L)$ without defects. In damaged links, we call them as WMV, there are only $(n_C - K_C)$ LIs, $1 \leq K_C \leq n_C - 1$; K_C LIs are failed. So now WMV is a parallel system with $(n_C - K_C)$ items. Recall that the equality $K_C = n_C$ means the failure of WMV and the specimen also. There are a priori probability mass functions (pmf) of random variables K_L and K_C : p_{K_L} and p_{K_C}.

In general case some WMVs can appear before but another during tensile or fatigue loading.

2.2.1. All WMV appear before test

Let $K_{Ci}(t)$, $0 \leq K_{Ci} \leq n_C$, be the number of failures of LIs in the *i*-th link at the tensile load x_t, $x_t \in \{x_1, x_2, ..., x_t, ...\}$, $x_1 < x_2 < x_3 ... < x_t < x_{t+1} < ...$ Load increases up to infinity. Then the number of steps of load increasing up to tensile failure of the *i*-th WMV

$$T_{Ai} = \max(t : n_C - K_{Ci}(t) \geq 0), \quad (3)$$

Two approaches for describing the second stage of total specimen failure can be considered: (1) the failure process development takes place in every WMV; (2) the failure process development takes place only in one WMV in which there is the maximum value of a priori failed LIs due to technological defects.

For the first hypothesis, which is studied in [4] (see also reference in [4]), we have the following definitions of the number of steps of tensile loading up to failure of the specimen

$$T_A = \min_{1 \leq i \leq K_L} T_{Ai} = \min_{1 \leq i \leq K_L} \max_t (t : n_C - K_{Ci}(t) \geq 0).$$

For corresponding cdf we have

$$F_{T_A}(x) = \sum_{k=1}^{n_L} p_{K_L}(k)(1-(1-F_{T_{A1}}(x))^k),$$

where $F_{T_{A1}}(x)$ is cdf of T_A of one WMV; p_{K_L} is pmf of rv K_L (modified ($K_L \geq 1$) binomial or Poisson distribution; see [4]). In the following we call this hypothesis as MinMax hypothesis and the corresponding family of cdf (for different versions of MChs) as MinMax cdf family. In the simplest case, if $P(K_L = n_L) = 1$, we have $F_{T_A}(x) = (1-(1-F_{T_{A1}}(x))^{n_L}$.

In this paper we consider the second hypothesis: the failure process development takes place only in one WMV in which there is a minimum of intact LIs, $N_{Cmn} = \min_i (n_C - K_{Ci}(0))$, where $K_{Ci}(0)$ is the initial number of (technological) defects in i-th WMV. Then

$$T_A = \max(t : \min_i (n_C - K_{Ci}(0)) - K_{Ci^*}(t) \geq 0) = \max(t : N_{Cmn} - K_{Ci^*}(t) \geq 0)$$

where $i^* = \arg\min_i (n_C - K_{Ci}(0))$ is the index of link corresponding to minimum intact LIs. This hypothesis we call as MaxMin hypothesis (MaxMin distribution family can be introduced also).

Obviously, instead of calculation of cdf of rv N_{Cmn} it is more convenient to calculate cdf of rv $K_{Cmx} = n_C - N_{Cmn} = \max_{1 \leq i \leq K_L} (K_{C_i}(0))$ for which we have

$$F_{K_{Cmx}}(m) = \sum_{k=1}^{n_L} p_{K_L}(k) F^k_{K_C(0)}(m), \; m = 1,2,...,n_C - 1,$$

where $F_{K_C(0)}(m) = \sum_{k=1}^{m} p_{K_C}(k)$, $p_{K_C}(.)$ is a priori pmf of rv $K_C(0)$. Then for $N_{Cmn} = n_C - K_{Cmx}$ we have the cdf $F_{N_{Cmn}}(m) = P(n_C - K_{Cmx} < m) = P(n_C - m < K_{Cmx}) = 1 - F_{K_{Cmx}}(n_C - m)$
and pmf $p_{N_{Cmn}}(m) = F_{N_{Cmn}}(m) - F_{N_{Cmn}}(m-1), \; m = 2,...,n_C - 1, \; p_{N_{C0}}(1) = F_{N_{C0}}(1)$.

But for $T_A = \max(t : N_{Cmn} - K_{Ci^*}(t) > 0)$ we have $F_{T_A}(t) = \sum_{m=1}^{n_C - 1} p_{N_{C0}}(m) F_{T_A | n_C = m}(t)$,
where $F_{T_A | n_C = m}(t)$ is cdf of T_A of one WMV for $n_C = m$ (see (1)).

Of course, we can reach the tensile failure of any specimens by increasing of load. So in every specimen there is at least one WMV and rv K_L is an integer which is larger or equal to one, more exactly: $1 \leq K_L \leq n_L$. Let rv K have a binomial or Poisson cdf (if $n_L \to \infty$). Then for rv K_L the conditional cdf of K under condition $K > 0$ or definition $K_L = 1 + K$ can be used.

Recall that connections between T_A, tensile strength, X, and number of cycles up to fatigue failure, T_C, are defined by (2a) and (2b).

2.2.2. The initiation of WMVs takes place during the process of loading

For tensile test it can be assumed that the number of WMV depends on the load. So the parameter of the binomial distribution can be taken equal to $F(x)$ where $F(.)$ is cdf of tensile strength of one LI, x is load (see details in [4]).

For fatigue test in [5] it is supposed that WMVs do not originate simultaneously but in accordance with a Poisson process. Then the number of WMVs is a random function of time. It increases during fatigue loading with intervals X_i, $i = 1,2,3,\ldots$. So X_1, $X_1 + X_2$, $X_1 + X_2 + X_3$, ... are the time moments of initiation of new WMVs. Let us denote by T_j fatigue life of j-th specific WMV. Then the life of specimen

$$Y = \min(T_1, T_2 + X_1, T_3 + X_1 + X_2, \ldots).$$

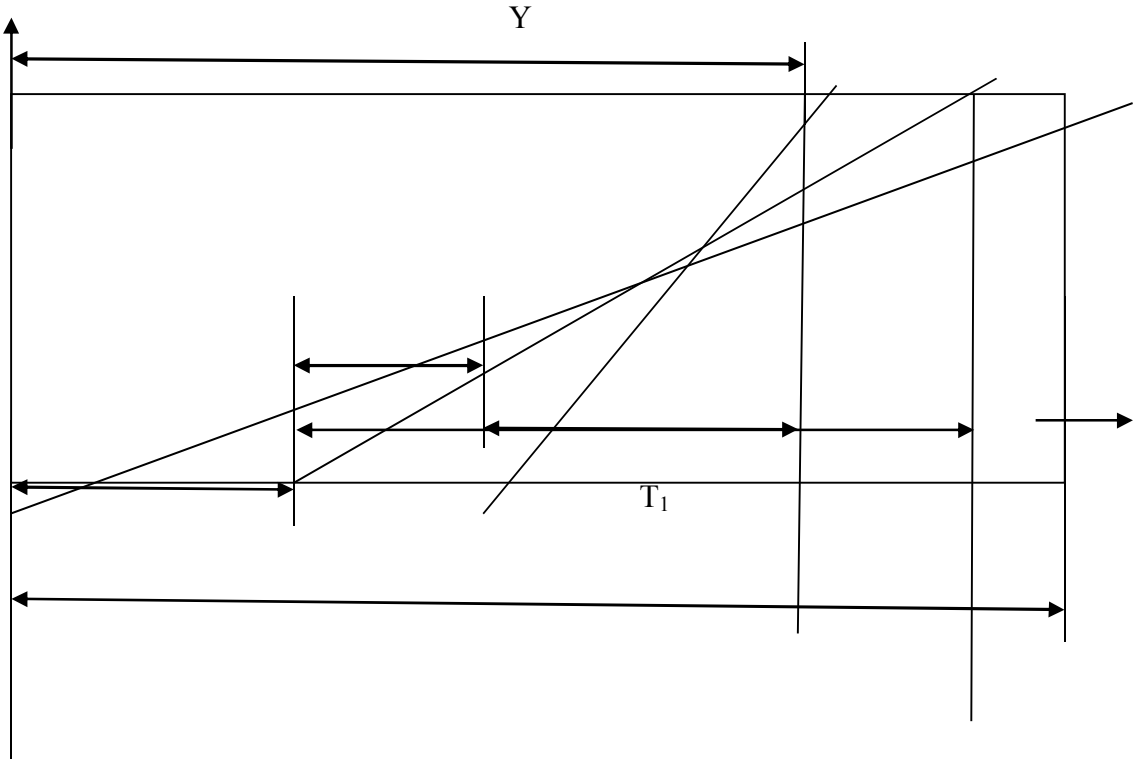


Fig. 3 . Definition of Y.

This equation can be written in the form $Y = \min(T_1, X_1 + Y_1)$, where $F_Y(y) = F_{Y_1}(y)$.

We have the following solution of this equation for the exponential distribution with a parameter μ of all independent random variables $X_1, X_2, X_3, \ldots$ (mean value of X is equal to $1/\mu$)

$$F_Y(y) = 1 - (1 - F_{T_C}(y)) \exp(-\mu \int_0^y F_{T_C}(t) dt).$$

where $F_{T_C}(t)$ is cdf of fatigue life of one WMV (see (2b), time unit is one cycle).

In [5] the example of using this approach for processing of test data is given.

Finally, we should mention that using results of fatigue test of glass fiber composite material example of the reasonably successful „ translation" of cycle with some positive assymetry into the pulsating cycle is also given in [5].

3. PROCESSING THE TEST DATA

In [4] there are examples of test data processing connected with MinMax approach. More exactly, there are examples of processing of results of fatigue tests (S-N curve and residual life) of carbon-fiber reinforced composite specimens using probability transition matrix of the form of Table 1 under assumption that there is only one WMV (see also reference in [4]). In [5] there is an example of processing glass-fiber reinforced composite specimen test data under assumption that different WMVs do not appear simultaneously but in accordance with the Poisson process. In both cases we have reasonable results of fitting test data and some prediction for different length and different stress ratio and examples of prediction of residual strength for two different preliminary loadings (S_i, n_i), $i = 1,2$: (292.53 MPa, 60 000), (390.05 MPa, 900).

In this paper a specific case of MaxMin approach is used for processing fatigue test and tensile strength test results of composite specimens made of unidirectional glass-fibre composite (Udo UD ES 500/300 - SGL epo GmbH c LH 160 of „Composites HAVEL"; [0/45/0]) for L= 60.

Because of the specific structure of specimens, for processing of tensile and fatigue data we use specific structure of matrix P_a (see [4]) with $r_R = 40$, r_Y=0 (there are only LIs); n_C=50. Lognormal distribution of LI tensile strength was assumed. For description of cdf of the rv $K_C(0)$ the conditional binomial cdf under condition that $K_C(0) < n_C$ (recall, equality $K_C(0) = n_C$ means failure of specimen) was used :

$$F_{K_{Ci}(0)}(k) = P(K_{Ci}(0) \le k \mid K_{Ci}(0) < n_C) =$$

$$\sum_{j=0}^{k} b(j, p_C, n_C) / (1 - b(n_C, p_C, n_C)),\ k = 1,...,n_C - 1,\ b(j, p_C, n_C) = p_C^i (1 - p_C)^{n_C - j} n_C! / j!(n_C - j)!.$$

For simplicity, only the case $K_L = n_L$ =100 was studied. So cdf

$F_{Cmx}(m) = \sum_{k=1}^{n_L} p_{K_L}(k) F_{K_C(0)}^k(m) = F_{K_C(0)}^{n_L}(m)$, $w = 1,2,...,n_C - 1$, was used.

The reasonable fitting of tensile (Fig. 4) and fatigue test results (Fig. 5) for k_m=0.150 was reached at $\theta_0 = 6.59$, $\theta_1 = 0.2$. These parameter do not differ too significantly from their estimates $\hat{\theta}_0 = 6.5869$, $\hat{\theta}_1 = 0.3008$ which are obtained processing tensile tests of strands (1200 fibers).

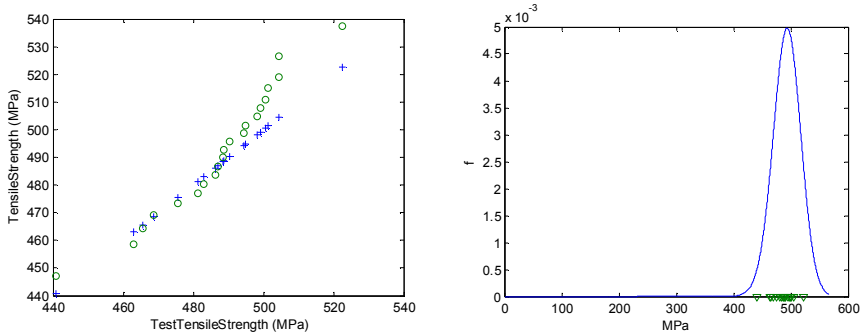


Fig. 4. Fitting of results of tensile strength test of specimens (+ and ∇) and predicted tensile strength pmf .

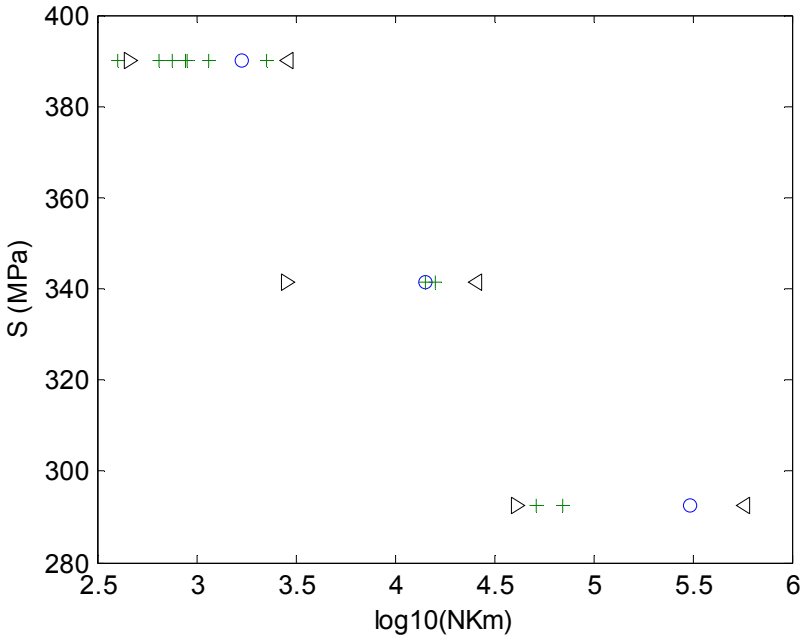


Fig. 5. Fitting of test fatigue life (+) (symbols > and < correspond to two sigma interval)

CONCLUSION

For different types of composite material the different versions of general model are applicable. As we see in Fig.3-Fig.5 and in conclusions of [4-7], by processing of test results it is shown that the considered versions of models, based on using MCh theory and both MinMax and MaxMin approaches (for modeling of the scale effect), can be used for nonlinear regression analysis in order to get fitting and some prediction of tensile and fatigue test results. For a general type of model a large number of parameters and specific type of a priori information should be known for corresponding numerical calculation. Clearly, for different types of composite material the different versions of general model are appropriate. Corresponding comparison analysis should be made and the best specific components of the general MCh model can be chosen for specific material (it is the subject of following papers). The components of the general MCh model are :

1.Type of cdf of mechanical characteristics (tensile strength, Young's modulus, …) and its parameters (for example, θ_0 and θ_1 for cdf with location and scale parameter) for LIs and matrix.

2. Parameters of composite structure (number of LIs in one WMVs, LIs part of cross section of WMV, …) .

3. Definition of failure of WMV (choice of $\overline{f}_C$, ε_{YC},…).

4. Definition of distribution of initial state of WMV.

5. Definition of distribution of number of WMV (links) in specimen as a series system.

6. Definition of process of loading $x1\infty$ for tensile or for fatigue test.

7. Criterion of quality of fitting and prediction of test data.

Some model parameters can be equated with (or can be taken approximately equal to) the parameters of cdf of the tensile strength of composite components and the parameters of composite structure (for example, relative total cross section area of LIs) or just can be chosen a priori (for example, the value of $\overline{f}_C$ in definition of failure of WMV) . Simultaneous fitting of results of both tensile and fatigue test of specimens allows estimation of other model parameters.

Of course, the considered models are too simple to describe the actual basic physical process. The influence of some model components (for example, details of description of tensile loading) should be studied carefully before final statistical analysis conclusion should be made. But the set of the model versions can be used as a wide field of study of composite material strength and fatigue life not only for graduate work of students but and for some engineering applications: for approximate prediction of the effect of not too drastic changes of mechanical characteristics of composite material components and type of load process $x1\infty$.

ACKNOWLEDGEMENTS

Part of this work was supported by project 2009/0209/1DP/1.1.1.2.0/09/APIA/VIAA/114.

REFERENCES

1. J. Bogdanoff , F. Kozin , *Probabilistic models of cumulative damage (Russian translation)*. (Moscow: Mir; 1989).
2. K. Goda , T. Okabe , N. Takeda , A strength reliability model of unidirectional fiber-reinforced ceramic matrix composites by Markov process. *Adv. Composite Mater*. Vol. 15 , n. 3, pp. 263-285.2006..
3. M. Kleinhofs, Investigation of static strength and fatigue of composite material used in aircraft structure. Candidate degree thesis, Riga, 1983.
4. Yu. Paramonov, A. Kuznetsov , Kleinhofs M. *Reliabilty of fatigue-prone airframes and composite materials.* Riga: RTU, 2011.(*http://gnedenko-forum.org/library/Paramonov/*Reliability_Paramonov.pdf)
5. Yu. Paramonov , R. Chatys , J. Andersons , M. Kleinhofs, Markov model of fatigue of a composite material with Poisson process of defect initiation. *Mech. Compos. Mater*. 2012, v. 48, No 2, pp. 315-330.
6. V. Cimanis , Yu. Paramonov Yu., Fatigue curve approximation using Daniels' sequence and Markov chain. *In: XXIV International Research and Practice Conference "Theory and practice in physical, mathematical and technical sciences". International Academy of Science and Higher Education (IASHE; Great Britain) , May 03 to 13, 2012.*
7. Yu. Paramonov , R. Chatys , J. Andersons , M. Kleinhofs. Studying fatigue curve approximation using daniels' sequence and markov chain theory. Presented to 12th Annual International Conference "Reliability and Statistics in Transportation and Communication (RelStat`12)", Riga, Latvia, 17-20 October, 2012
8. H.E. Daniels. The statistical theory of the strength of bundles of threads. *I. Proceedings of the Royal Society of London*, Series A, 1945; 183(995), pp. 405-435.
9. H.E. Daniels.The maximum of a Gaussian process whose mean path has a maximum, with an application to the strength of bundles of fibers. *Advances in Applied Probability*, 1989; 21(2), pp. 315-333.

DESIGN TESTABILITY ANALYSIS OF AVIONIC SYSTEMS

Igori B. Spiridonov
•
IRKUT Corporation, Moscow, 125315, Russia
Igori.Spiridonov@irkut.com

Armen S. Stepanyants, Valentina S. Victorova
•
Institute of Control Sciences (IPU RAN), Moscow, 117997, Russia
lfvrk@ipu.ru, ray@ipu.ru

ABSTRACT

This paper summarizes the result of an effort to develop a unified approach to design-driven testability evaluation of avionic systems. These systems include both internal diagnostic equipment referred to as built-in-test (BIT) and external off-line test equipment. At the designing stage an adequate database to evaluate the quality of the BIT is the failure mode and effect analysis. In the paper various mathematical indices are suggested and constructed to quantify testability of avionic systems. The indices provide the needed flexibility for representing structural and reliability properties of the controlled system. Analytical model for evaluation BIT performance impact on the system's reliability is discussed.

1 BIT PERFORMANCE IMPACT ON SYSTEM RELIABILITY

Evaluation of the technical condition of the avionic systems is ensured by the presence of built-in diagnostic functions and monitoring tools – BIT. BIT performance defines testability of the systems or its adaptation to detect and isolate failures to the replaceable assembly level. Operational integrated BIT monitoring of system's components provides effective usage of spares, reconfiguration and graceful degradation, ensuring thereby the fault-tolerance and safety of avionic system. However, the BIT is not ideal – first, it can refuse to act, and, secondly, not all failures and events can be recognized by BIT. Therefore, in order to ensure high levels of reliability and safety of avionic systems it is required to conduct a thorough reliability analysis, taking into account many factors, one of which is characteristics of BIT (Victorova et al. (2007), Victorova&Stepanyants (2008)).

So, the role of diagnostic systems should be judged by the impact of their characteristics on the probability indices of reliability and safety. We will study the reliability of a recoverable system comprising two identical sub-systems that are parallel in terms of reliability. We assume absolute BIT reliability but not absolute fault coverage. BIT can identify only part of sub-system failures and recovery is possible only after failure detection by BIT. Under this assumption the parallel system can be represented by controlled/recoverable and uncontrolled/unrecoverable series parts as shown in Figure 1. The percentage of all sub-system faults or failures that BIT can detect is denoted as η. We make assumption about exponential failure and repair distribution with parameters λ and μ respectively. Then failure rates of controlled part of the sub-system and uncontrolled part are equal $\eta\lambda$ and $(1-\eta)\lambda$ respectively.

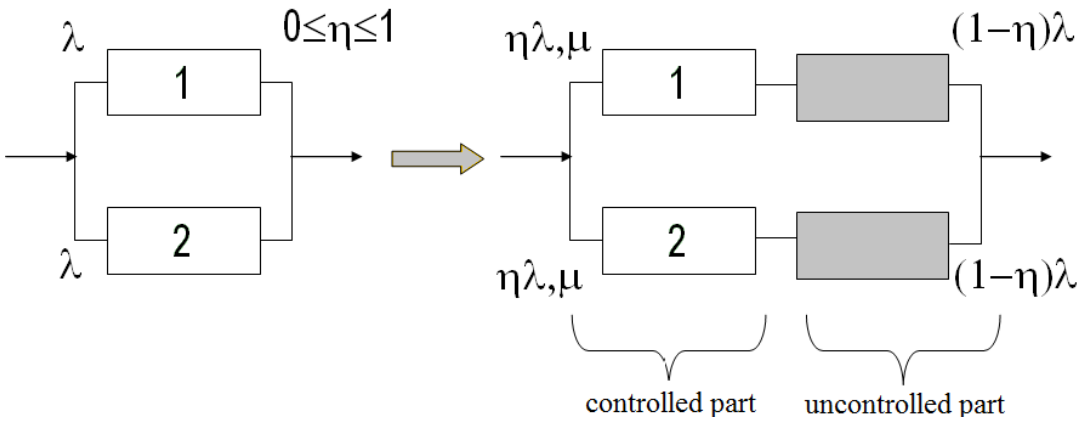


Figure 1. Duplicate repairable system with partial BIT

Let us consider the special treatment of system recovery when repairable actions start only after mission termination. This mode is typical for avionic systems when restoration is carried out only on the ground. Markov reliability model for this case is shown in Figure 2.

To determine the transition rate from state 2 to state 5 it is necessary to calculate conditional mean time to subsystem failure provided that the failure occurred during mission time interval $(0,t_m)$ $T_{av/tm}$:

$$T_{av/t_m} = M\{T/T<t_m\} = \int_0^{t_m} tf(t/t<t_m)dt = \int_0^{t_m} td\frac{F(t)}{F(t_m)} = \frac{1/\lambda - e^{-\lambda t_m}(t_m + 1/\lambda)}{1-e^{-\lambda t_m}}, \qquad (1)$$

where F(t) and f(t) are distribution function and distribution dencity of stochastic time to failure T. Derivation of Eq.(1) was done in Gnedenko et al.(1969).

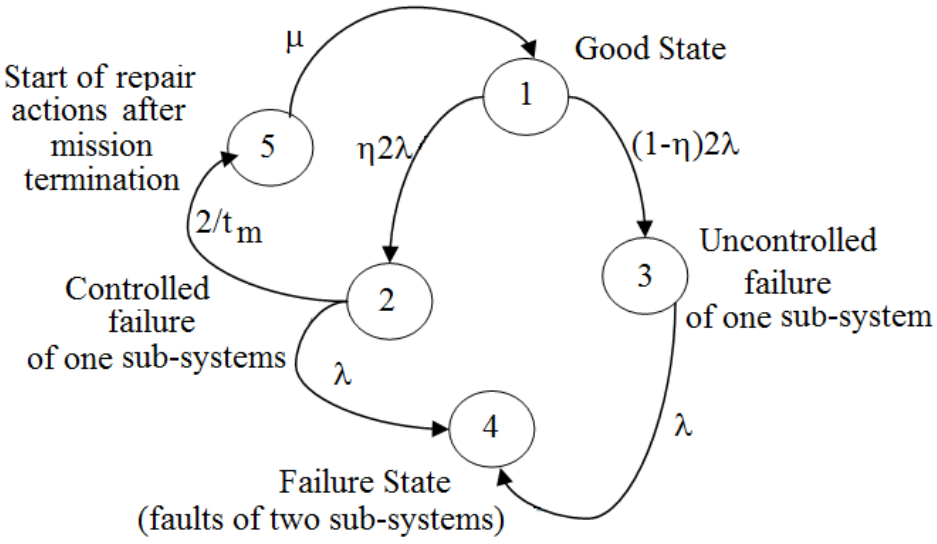


Figure 2. Markov reliability model

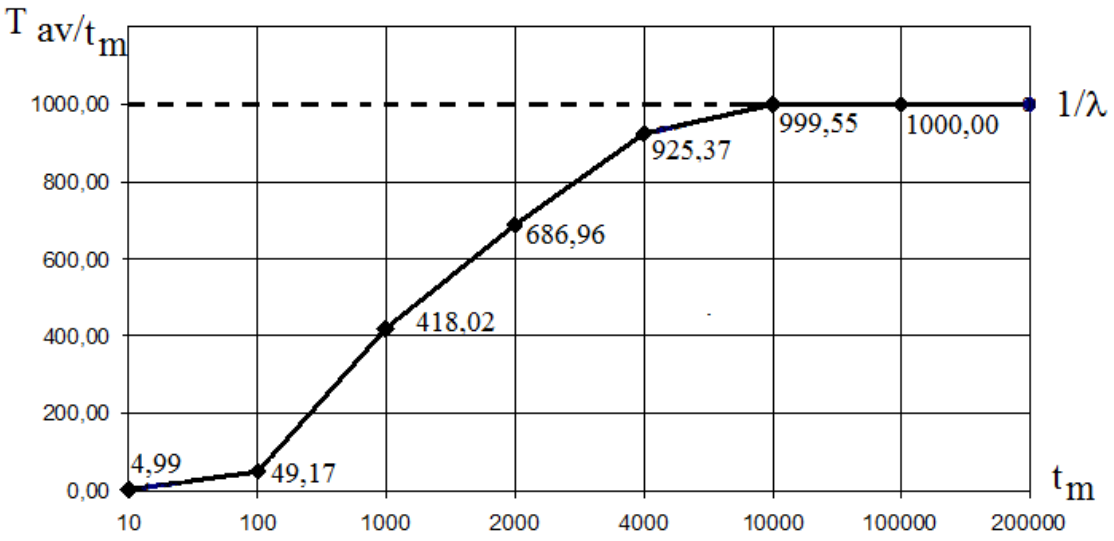


Figure 3. Conditional mean time to failure curve

If $t_m << 1/\lambda$ then $T_{av/tm} \approx t_m/2$. If $t_m > 1/\lambda$ then $T_{av/tm} \to 1/\lambda$. Chart in Figure 3 confirms these relations. Therefore, for avionic systems, which mission (flight) time is not more than a few hours, one can assume that the failure of subsystem occurs in the middle of the flight interval, and hence the transition rate from state 2 to state 5 is equal to $2/t_m$.

Reliability markov model of the duplicate repairable system with partial BIT was calculated at time interval (t = 0÷8760 hours). Probability of the system failure Q (probability of state 4) was calculated varying percent detection from 0 to 1. Normed $Q(\eta)$ curve from Figure 4 shows that even 50% failure detection reduces the probability of the system failure in hundreds of times. When η is close to 100% failure probability is reduced in more than thousand of times.

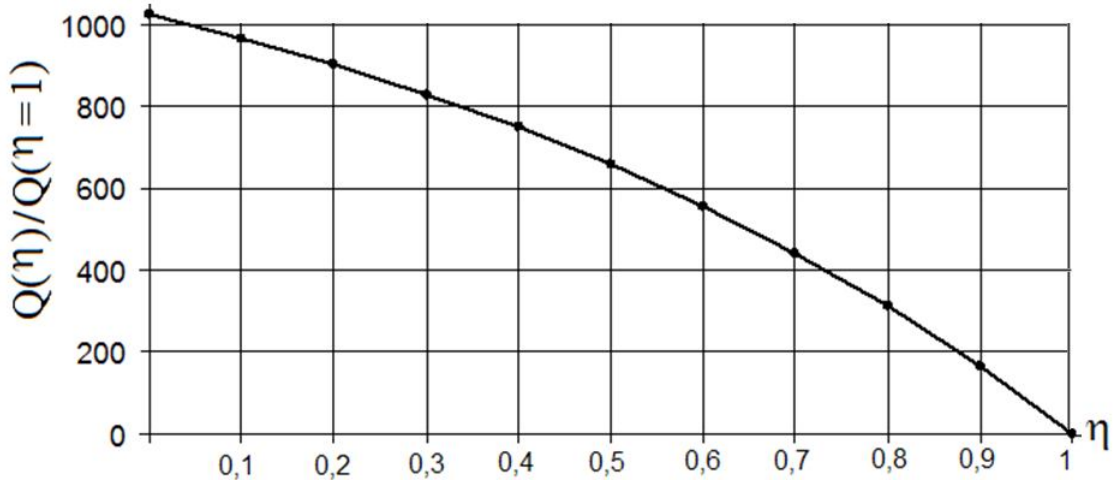


Figure 4. Normed curve of the system failure probability

2 TESTABILITY MODELS AND INDICES

The main characteristics of BIT equipment are percent detection and percent isolation.

2.1 Percent Detection Definition

Percent detection describes completeness of system's monitoring by BIT. In general, the quality of BIT is determined by the list of elements (modules), which failures are detected. Therefore, percent detection could be defined by the ratio of the number of controlled items to the total number of items in the system. However, for the joint reliability&testability modeling we should include some probabilistic constituent in percent detection definition. The usefulness of such approach consists in splitting the total failure flow into two components – the failures detected by the BIT and latent failures. Percent detection in this case can be defined as the conditional probability of failure detection, provided that the failure occurred:

$$\eta=\text{Prob}\{\text{failure is detected/failure occurred}\} = \frac{1-e^{-\int_0^t \Lambda_c(\tau)d\tau}}{1-e^{-\int_0^t \Lambda(\tau)d\tau}}, \quad (2)$$

where Λ - total failure rate of the system, Λ_c – total failure rate of detected by BIT failures.

After averaging the failure rates on the interval (0, t), we have

$$\eta = \frac{1-e^{-\Lambda_{cav}t}}{1-e^{-\Lambda_{av}t}}, \quad (3)$$

where $\Lambda_{av} = \frac{1}{t}\int_0^t \Lambda(\tau)d\tau, \quad \Lambda_{av}t << 1.$

Common percent detection Eqs. (2,3) in the case of exponential distribution is most useful to set as a ratio of the total failure rates of controlled components to total failure rates of all system components, i.e.

$$\eta = \frac{\sum_{j\in K}\lambda_j}{\sum_{i=1}^{n}\lambda_i}, \quad (4)$$

where n is the total number of elements of the system; K-subset of the controlled components; λ_i is failure rate of the i^{th} component. In this case percent detection is defined as stationary conditional probability of failure detection, provided that the failure occurred.

2.2 Percent Isolation Definition

Percent isolation characterizes BIT resolving ability. Percent isolation is diverse feature. For example, you can understand the percent isolation as the resolution of fault location in the hierarchy of the failed system components: subsystem, assembly, part. In this paper percent isolation will be determined through LRUs - Line Replaceable Unit as follows. If, in the event of a failure, the BIT points to a subset of elements that might be failed, then these items simultaneously removed (may be including not failed items) and replaced with a good LRUs (this is the specific of maintenance services). Similar to the detection isolation can be defined as percent of faults or failures that BIT system will isolate to a specified level (for example, to 1 LRU, 2 LRU, 3 LRU…). Therefore,

percent isolation can be represented by a discrete distribution. Stationary probabilities γ_k of this distribution are calculated as

$$\gamma_k = \frac{\sum_{j \in G_k} \lambda_j}{\sum_{i=1}^{n} \lambda_i}, \quad (5)$$

where G_k – subset of detected fault or failures results in removal of k LRUs.

Another stochastic characteristic of percent isolation γ may be suggested as ratio of mathematical expectation of numbers of detected failures ($n_f(t)$) to mathematical expectation of number of component removals ($n_r(t)$) for a specified time interval (0÷t):

$$\gamma = \frac{M\{n_f(t)\}}{M\{n_r(t)\}} \quad (6)$$

The advantage of the last definition is that percent isolation, calculated according to Eq. (6), may be associated with known logistics measure MTBUR (mean time between unscheduled removals). MTBUR is calculated as $t/M\{n_r(t)\}$.

2.3 Complex Measure of BIT Quality

In this section we will present complex performance measure of BIT, taking into account both considered detection and isolation characteristics and two modes of BIT possible failures.

Let us denote the following stochastic events:

A – good state of the controlled system

$\overline{A}$ - failure state of the controlled system

B - BIT indicates controlled system state as good state

$\overline{B}$ - BIT indicates failure of controlled system

Then we can formally define the following results of interaction between the system and BIT:

$A \wedge B$ - the system is good and BIT indicates good state of the system

$A \wedge \overline{B}$ - the system is good, but BIT indicates fault of the good system. This type of BIT failure is known as false alarm.

$\overline{A} \wedge B$ - the system is in failure state, but BIT does not detect fault and indicates good state of the system.

$\overline{A} \wedge \overline{B}$ - the system is in failure state and BIT detects fault and indicates failure state of the system.

Let us define quality measure, named BIT certainty or integrity, as the sum of the probabilities of events $A \wedge B$ and $\overline{A} \wedge \overline{B}$:

$$D = P(A \wedge B) + P(\overline{A} \wedge \overline{B}) \quad (7)$$

Then BIT uncertainty $\overline{D}$ is

$$D = P(A \wedge \overline{B}) + P(\overline{A} \wedge B) \quad (8)$$

Detailed expressions of the terms of BIT uncertainty Eq. (8) are

$$P(A \wedge \overline{B}) = P(A)P(\overline{B}/A), \; P(\overline{A} \wedge B) = P(\overline{A})P(B/\overline{A}) \quad (9)$$

Where $P(A)$ – probability of good state of controlled system, $P(\overline{A})$ – probability of failure state of controlled system, $P(\overline{B}/A)$ conditional probability of BIT failure indication on condition that system is good, $P(B/\overline{A})$ conditional probability of BIT indication good system state on condition that system is in failed state.

To calculate these conditional probabilities we will use event tree model (Kumamoto&Henley (2000)) and will take into account BIT percent detection, false alarms and "nonoperate" BIT's failure mode (detailed description of this approach is presented in Victorova (2009)).

We denote possible BIT events as

C – BIT is in good state

$\overline{C}_{no}$ – BIT is in failure state and failure mode is "nonoperate"

$\overline{C}_{fa}$ – BIT is in failure state and failure mode is "false alarm"

Figure 5 presents event tree for calculation the conditional probability of BIT failure indication under good system $P(\overline{B}/A)$.

Figure 6 presents event tree for calculation the conditional probability of BIT indication good system state on condition that system is in failed state $P(B/\overline{A})$.

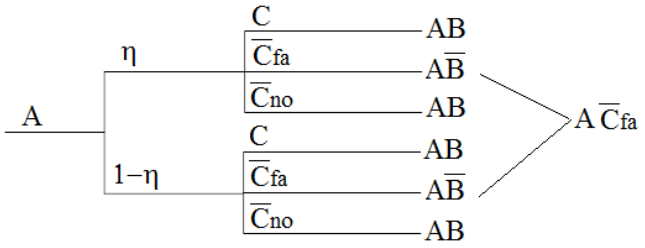


Figure 5. Event tree model for calculation $P(\overline{B}/A)$

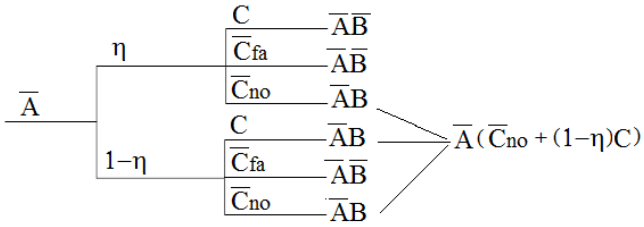


Figure 6. Event tree model for calculation $P(B/\overline{A})$

After calculation of the required conditional probability we will get the following expression for BIT uncertainty:

$$\overline{D} = P(A)P(\overline{C}_{fa}) + P(\overline{A})(P(\overline{C}_{no}) + (1-\eta)P(C)) \quad (10)$$

If BIT isolates two or more LRUs when only one LRU has failed, then probability of this event should be included in $\overline{D}$

$$\overline{D} = P(A)P(\overline{C}_{fa}) + P(\overline{A})(P(\overline{C}_{no}) + (1-\eta)P(C)) + P(\overline{A})\eta(1-\gamma_1)P(C) \quad (11)$$

3 TESTABILITY ORIENTED FAILURE MODE AND EFFECT ANALYSIS

Failure Mode and Effect Analysis (FMEA) is one of the most widely used tools for developing quality design. For the purpose of testability assessment we have used design detailed FMEA, applying some provisions of US MIL-STD-1629. Task 101, Task 102. Analysis was performed using inductive bottom-up approach starting the analysis with the failure modes at the LRU level and then successively iterating throw the levels of functional subsystems ending at the system level.

The main fields of FMEA worksheets, constructed for testability analysis, are presented in Table 1. Structured in such a way FMEA data were used to calculate the above indices Eqs. (5,6,11). The indices are calculated for each functional avionic subsystems and aircraft in general. Field FDM corresponds to the list of methods – CBIT (continuous BIT), PBIT (power-on BIT) and so on. Inclusion in this field of "none" item (the failure mode is not detected) makes it possible to calculate percent detection index Eq. (4).

Table 1. Main fields of FMEA worksheets

Field Name	Description
ID	LRU identifier. for example ATA code.
Name	LRU name and description
MTTF	LRU Mean time to failure
FM	LRU Failure Modes
FMP	Percent of each failure modes
FDM	Failure detection method
FID	Fault isolation descriptor – the list of LRUs ID, isolated by BIT
FMS	The LRU's failure mode severity
FMM	Mission Phase

4 CONCLUSION

We have presented unified approach to testability analysis of avionic systems at design stage. FMEA information was used as input data for testability evaluation. Calculation equations for computing BIT percent detection and isolation are described. Complex measure of BIT performance, viz BIT certainty, was suggested. This measure takes into account both fault detection and isolation characteristics and false alarm and "nonoperate" modes of BIT possible failures. Modification of standard FMEA worksheets was done for adaptation for the testability indices calculation. Markov reliability model for imperfect fault coverage and special strategy of avionic systems repair was constructed. It was shown that BIT behavior is a very important factor, which has a tremendous impact on the reliability of the avionic systems.

Described approach was applied in the study of testability of functional systems of Russian aircraft Sukhoi Superjet 100, developed by the Sukhoi Civil Aircraft Company, Moscow.

5 ACKNOWLEDGEMENT

Authors would like to thank and acknowledge the support of this work provided by management and staff of Sukhoi Civil Aircraft Company (CJSC), especially Pimenov A.V., Vedernikov B.I., Umashev V.G., Gangan Y.G. Authors appreciate Petrova N.A. from Tesabiliy Department of IRCUT for her great work on FMEA data structuring.

6 REFERENCES

Gnedenko, B.V., Belyaev, Yu. K., Solovyev, A.D. (1969). *Mathematical. Methods of Reliability Theory*. New York: Academic Press.

Kumamoto H., Henley E.J. (2000) *Probablistic Risk Assessment and Management for Engineers and Scientists.* John Wiley & Sons.

Victorova, V.S., Vedernikov B.I., Spiridonov I.B., Stepanyants A.S. (2007). Testability simulation and analysis of on-board aircraft systems. *Reliability Journal*, vol. 3 (22), pp. 62-71.

Victorova, V.S., Stepanyants A.S. (2008). Models for test integrity evaluation of on-board systems. *Proceedings of the 8th International Scientific School "Modeling and Analysis of Safety and Risk of Complex Systems" (MASR 2008)*, (pp. 357-362), June 24-28, StPb.

Victorova, V. S., (2009). *Aggregation of reliability and safety models of complex technical systems*. Doctoral dissertation. Institute of Control Sciences (IPU RAN), Moscow, Russia. http://www.dissercat.com.

LIMIT THEOREM FOR CLOSED QUEUING NETWORKS WITH EXCESS OF SERVERS

G.Tsitsiashvili
●
IAM FEB RAS, Vladivostok, Russia
guram@iam.dvo.ru

ABSTRACT

In this paper limit theorems for closed queuing networks with excess of servers are formulated and proved. First theorem is a variant of the central limit theorem and is proved using classical results of V.I. Romanovskiy for discrete Markov chains. Second theorem considers a convergence to chi square distribution. These theorems are mainly based on an assumption of servers excess in queuing nodes.

1. INTRODUCTION

In [1], [2] problems of a formulation and a proof of the central limit theorem for queuing systems and networks are considered. At the international conference "Probability theory and its applications$" the author of this manuscript had useful and productive discussion with A.A. Nazarov stimulated an interest to this problem.

In this paper a model of closed queuing network with an excess of servers in its nodes and with singular service time or geometrically distributed service time is considered. For loads of this model nodes the central limit theorem is formulated and is proved using classical results of V.I. Romanovskiy [4] for discrete Markov chains. It is suggested to define parameters of limit normal distribution using Monte-Carlo simulations of single customer motion along nodes of the network. This approach allows to decrease calculation time significantly because it is not necessary to simulate process of loads in nodes of this network.

The central limit theorem is complemented by a fact of a convergence to chi-square distribution for large number of customers.

2. FORMULATION AND PROOF OF MAIN RESULTS

In this paper we consider closed queuing network S with the set of nodes $N = \{1, \dots, n\}$ and with m customers. Assume that in each node there are m servers. The network works in discrete time $0,1, \dots$, and service time of each customer equals unit. Suppose that $\Theta = \left\|\theta_{i,j}\right\|_{i,j=1}^{n}$ is route matrix of the network S which satisfies the condition

(A) for any $i,j \in N$ there are $i_1, \dots, i_k \in N$ so that the product $\theta_{i,i_1} \cdot \theta_{i_1,i_2} \cdot \dots \cdot \theta_{i_{k-1},i_k} \cdot \theta_{i_k,j} > 0$

Consider discrete Markov chain x_t, $t = 0,1, \dots$ with the set of states $N$and with the matrix of transition probabilitiesΘ . From Condition (A) we have that this chain is ergodic [3, chapter XV], denote its limit distribution by π_i , $i \in N$. This distribution does not depend on initial state x_0 of the chain x_t , $t = 0,1, \dots$ Designate $\tau_i(T) = \#\{t: x_t = i,\ t = 0, \dots, T\}$ sojourn time of the chain x_t in the state i on time interval $t = 0, \dots, T$. Introduce n - dimensional random

vector $\frac{\tau_i(T)-\pi_i T}{\sqrt{T}}, i \in N$. In [4, chapters IV, V] it is proved that there is n - dimensional and normally distributed random vector $R = (r_i),\ i \in N$ with zero mean and with covariance matrix $\mathcal{B}$ so that for any real numbers $\tau_1, \dots, \tau_n$ independently on initial state x_0 for $T \longrightarrow \infty$ we have the convergence

$$P\left(\frac{\tau_i(T)-\pi_i T}{\sqrt{T}} < \tau_i\ ,\ i \in N\right) \to P(r_i < \tau_i\ ,\ i \in N). \quad (1)$$

Return now to closed queuing network S and enumerate customers of the network by $1, \dots, m$. Denote $x_t^j\ ,\ t \geq 0\ , j = 1, \dots, m$ trajectories of the network S customers along its nodes.
These trajectories are independent Markov chains with the set of states N and with matrix of transition probabilities Θ. Assume that $\tau_i^j(T) = \#\left\{t :\ x_t^j = i\ ,\ t = 0, \dots, T\right\}$ is sojourn time of the customer j in the node i on time interval $0, \dots, T$. Introduce random vectors $\frac{\tau_i^j(T)-\pi_i T}{\sqrt{T}}, i \in N$, which are independent because trajectories of different customers are independent also. Analogously with Formula (1) obtain for arbitrary real $\tau_1, \dots, \tau_m$ for $T \longrightarrow \infty$

$$P\left(\frac{\tau_i^j(T)-\pi_i T}{\sqrt{T}} < \tau_i\ ,\ i \in N\right) \to P\left(r_i^j < \tau_i\ ,\ i \in N\right), \quad (2)$$

where n - dimensional random vectors $r_i^j, i \in N,$ are independent and have normal distribution with zero mean and covariance matrix $\mathcal{B}$. Consequently for $T \longrightarrow \infty$ we have

$$P\left(\sum_{j=1}^m \frac{\tau_i^j(T)-\pi_i T}{\sqrt{T}} < \tau_i\ ,\ i \in N\right) \to P(R_i < \tau_i\ ,\ i \in N), \quad (3)$$

where $(R_i\ ,\ i \in N)$ is n - dimensional and normally distributed random vector with zero mean and with covariance matrix $m\mathcal{B}$. Formula (3) may be rewritten for $T \longrightarrow \infty$ as follows

$$P\left(\frac{T_i(T)-m\pi_i T}{\sqrt{T}} < \tau_i\ ,\ i \in N\right) \to P\left(r_i < \sqrt{m}\tau_i\ ,\ i \in N\right), \quad (4)$$

Here random variables $T_i(T) = \sum_{j=1}^m \tau_i^j(T)\ ,\ i \in N$ designate total loads on $i \in N$ of the network S on time interval $t = 0, \dots, T$. It is clear that $T_i(T) = \sum_{t=0}^T m_i(T)$, where $m_i(t)$ is a number of servers of the node i, busy by customers at moment t.

So the central limit theorem for discrete Markov chain with finite set of states may be transferred onto random vector $(T_i(T),\ i \in N)$ consisted of loads of the network S nodes. This result may be generalized in different directions. Assume that random service time η_i in the node i has geometrical distribution $P(\eta_i = k) = (1 - a_i)a_i^{k-1}, k = 1,2, \dots$. Then redefining route matrix Θ by the formulas:

$$\theta_{i,i} := \frac{\theta_{i,i}+a_i}{1+a_i}\ ,\ \ \theta_{i,j} := \frac{\theta_{i,j}}{1+a_i},\ j \neq i\ , j \in N\ , \quad (5)$$

it is possible to obtain results represented by Formula (4). Moreover Formula (4) may be transformed into formula which characterizes loads in some but not all nodes $1, \dots, n_1 < n$ of the network S for $T \longrightarrow \infty$:

$$P\left(\frac{T_i(T)-m\pi_i T}{\sqrt{T}} < \tau_i\ ,\ 1 \leq i \leq n_1\right) \to P\left(r_i < \sqrt{m}\tau_i\ ,\ 1 \leq i \leq n_1\right), \quad (6)$$

To estimate covariance matrix $\mathcal{B}$using Formulas (1) – (4), (6) it is possible to estimate the matrix $\left\|cov\left(\tau_i(T), \tau_l(T)\right)\right\|_{i,l=1}^n$ by Monte-Carlo simulations with sufficiently large T. This estimate is based on independent realizations of Markov chain $x_t,\ t = 0,1, \dots, T$. Formulas (4), (6) are constructed for total loads of the network S nodes and are not connected with a motion of single customer.

Consider now an aggregation of nodes in this model. An aggregation of nodes in closed queuing network leads to very complicated procedure because of difficult symbolic calculations. So it is interesting to consider this problem from a view of the central limit theorem. For a simplicity of a consideration divide the set of nodes $N = \{1, \dots, n\}$ into two subsets $N_1 = \{1, \dots, n_1\}$, $N_2 = \{n_1 + 1, \dots, n\}$, $1 \le n_1 \le n$. Then total loads $T^1(T)$, $T^2(T)$ on the sets N_1 , N_2 of nodes are defined by the equalities

$$T^1(T) = \sum_{i \in N_1} T_i(T), T^2(T) = \sum_{i \in N_1} T_i(T).$$

Consequently covariance matrix $\left\| cov\left(T^k(T), T^r(T)\right) \right\|_{k,r=1}^2$ may be calculated by covariance matrix $\left\| cov\left(T_i(T), T_l(T)\right) \right\|_{i,l \in N}$ using simple equalities

$$cov\left(T^k(T), T^r(T)\right) = \sum_{i \in N_k, k \in N_r} cov\left(T_i(T), T_k(T)\right) , \qquad k, r = 1,2 .$$

And in a case of two nodes we have

$$cov\left(T^1(T), T^1(T)\right) = cov\left(T^2(T), T^2(T)\right) = -cov\left(T^1(T), T^2(T)\right) = DT^1(T) .$$

Consequently an aggregation of nodes in closed network leads to simple and clear formulas for covariances of loads in aggregated nodes. And the central limit theorem for nodes of the network S are transformed into the central limit theorem for aggregated nodes of this network: for any real τ^1 , τ^2 and $T \longrightarrow \infty$

$$P\left(\frac{T^k(T) - m \sum_{i \in N_k} \pi_i}{\sqrt{T}} < \tau^k , \ k = 1,2 \right) \to P\left(\sum_{i \in N_k} r_i < \sqrt{m}\, \tau^k \ , \ k = 1,2 \right) .$$

Assume that service times in customers of all nodes are unit and $\pi_i , i \in N$ is the distribution of ergodic and stationary Markov chain x_t, $t = 0,1, \dots$ Introduce random variables

$$b_m(0) = \sum_{i=1}^{n} \frac{\left(\frac{m_i(0)}{m} - \pi_i \right)^2}{m \pi_i}$$

Then from [5, chapter III, pp. 169-172] we obtain the following statement. For any positive τ and for $m \to \infty$

$$P(b_m(0) < \tau) \to P(\chi^2_{n-1} < \tau).$$

Here $m_i(0)$ is the number of customers at the node i at the moment 0 and χ^2_{n-1} is random variable with chi square distribution and with n-1 degree of freedom. This statement may be generalized onto arbitrary stationary stochastic sequence x_t, $t = 0,1, \dots$, with stationary distribution $\pi_i , i \in N$.

The author thanks A.A. Nazarov for useful discussions.

REFERENCES

1. Nazarov A.A., Moiseeva S.P. Method of asymptotic analysis in queuing theory. Tomsk: edition of scientific and technique literature. 2006. (In Russian).
2. Moiseeva S.P., Nazarov A.A. Method of sifted flow for investigation of non Markovian queuing systems with unbounded number of servers. Abstracts of International conference "Probability theory and its applications$",\;$ in commemoration of the centennial of B.V. Gnedenko. M: LENAND. 2012. P. 200 - 201. (In Russian).

3. Feller W. Introduction to prabability theory and its applications. T. 1. Ì.: Mir. 1984. (In Russian).
4. Romanovsky V. 1949. *Discrete Markov chains.* Moscow-Leningrad: State Publishing House of Technical-Theoretical Literature (in Russian).
5. Rozanov Yu.A. Probability theory, stochastic processes and mathematical statistics. M.: Nauka. 1985. (In Russian).

U- FUNCTION IN APPLICATIONS

Igor Ushakov
Sun Diego, California
e-mail: igusha22@gmail.com

The Method of Universal Generating Functions (U-functions) was introduced in [Ushakov, 1986]. Since then the method has been developed, in first order, by my friends and colleagues – Gregory Levitin and Anatoly Lisnianski. They actively and successfully apply the method of U-function to optimal resources allocation, to multi-state system analysis and other problems. Frankly, now I feel like a hen sat on duck eggs and then wanders how hatched chicks fearlessly swim so far from shore. ☺

I decided to remind you a Russian folk proverb: "new is well forgotten old". What is U-function? It is, first of all, generalization of a classical Generation Function (GF) permitting perform more general transforms. From technical side, this method represents a modification of the Kettelle's Algorithm conveniently arranged for calculations with the use of computer.

U-FUNCTION

One can represent any discrete distribution of random variable (r.v.) x_k as a set of pairs:

$$S_k = \left\{(x_k^{(1)}, p_k^{(1)}), (x_k^{(2)}, p_k^{(2)}),...\right\} \tag{1}$$

This distribution can be represented in the form of polynomial (common generating function)

$$\varphi_k(z) = \sum_{1 \le j < n_k} p_k^{(j)} z^{x_k^{(j)}} \quad . \tag{2}$$

where n_k, in principle, can be infinite.

The polynomial representation allows one to obtain the distribution of sums of random variables, using the convolution procedure.

If instead of the distribution of x_1+x_2 one wish to obtain the distribution of an arbitrary function $f(x_1, x_2)$, for any combination of realizations $x_1 = X_1^{(1)}$ and $x_2 = X_2^{(1)}$, the realization of the function $f(x_1, x_2)$ takes the value of $f(X_1^{(1)}, X_2^{(1)})$ with probability $p_1^{(1)} \cdot p_2^{(1)}$. Having the discrete distributions of two random variables x_1+x_2 in the form $S_1 = \left\{(X_1^{(1)}, p_1^{(1)}),...,(X_1^{(n_1)}, p_1^{(n_1)})\right\}$ and $S_2 = \left\{(X_2^{(1)}, p_2^{(1)}),...,(X_2^{(n_3)}, p_2^{(n_2)})\right\}$ one can obtain the discrete distribution S of the function

$f(x_1, x_2)$ as a "Descartes composition" of two sets S_1 and S_2 (here we use the term "interaction" because the operation reminds Descartes product but does not coincide with it) .

$$\begin{aligned}
&S_f = \{(f(X_1^{(1)}, X_2^{(1)}); p_1^{(1)} \cdot p_2^{(1)}),...,(f(X_1^{(n_1)}, X_2^{(n_2)}); p_1^{(1)} \cdot p_2^{(n_2)}),\\
&(f(X_1^{(2)}, X_2^{(1)}); p_1^{(2)} \cdot p_2^{(1)}),...,(f(X_1^{(2)}, X_2^{(n_2)}); p_1^{(2)} \cdot p_2^{(n_2)}),\\
&...\\
&(f(X_1^{(n_1)}, X_2^{(1)}); p_1^{(n_1)} \cdot p_2^{(1)}),...,(f(X_1^{(n_1)}, X_2^{(n_2)}); p_1^{(n_1)} \cdot p_{n^{(2)}}^{(n_2)})
\end{aligned} \tag{3}$$

This distribution can be conveniently obtained using the "composition procedure" over functions $\varphi_1(z)$ and $\varphi_2(z)$:

$$\varphi_f(z) = \underset{f}{\otimes}(\varphi_1(z), \varphi_2(z)) = \left(\sum_{1 \le j < n_1} p_1^{(j)} z^{X_1^{(j)}} \right) \underset{f}{\otimes} \left(\sum_{1 \le i < n_2} p_2^{(i)} z^{X_2^{(i)}} \right)$$

$$= \sum_{1 \le j < n_2} \sum_{1 \le i < n_2} p_1^{(j)} \cdot p_2^{(i)} \, z^{f(X_1^{(j)}, X_2^{(i)})}. \quad (4)$$

The composition operator $\underset{f}{\otimes}$ possesses commutative property, i.e.

$$\underset{f}{\otimes}(\varphi_1(z), \varphi_2(z)) = \underset{f}{\otimes}(\varphi_2(z), \varphi_1(z)) \quad (5)$$

and associative property, i.e.

$$\underset{f}{\otimes}(\varphi_1(z) \underset{f}{\otimes}(\varphi_2(z)), \varphi_3(z)) = \underset{f}{\otimes}((\varphi_2(z), \underset{f}{\otimes}((\varphi_1(z), (\varphi_3(z))) \underset{f}{\otimes}(\varphi_3(z), \underset{f}{\otimes}(\varphi_1(z), \varphi_2(z)). \quad (6)$$

if the function f possesses these properties. In the most applications this is the case, though some exceptions exist (see, for example, [Levitin, 2005]).

USING U-FUNCTION FOR SOLVING THE OPTIMAL REDUNDANCY PROBLEMS

Let us consider a series system consisting of n subsystems. Each of subsystem k contains at least one unit with probability of failure-free operation (PFFO) p_k and cost c_k . For increasing the system reliability, one can introduce redundancy in each subsystem. Denote the number of redundant units of subsystem k as x_k. All the possible PFFOs and costs of any subsystem k for different levels of redundancy can be represented by the set of triplets

$$S_k = \{(P_k(x_k), C_k(x_k), x_k), 1 \le x_k < \infty\}$$

where $C_k(x_k)$ is the total cost of x_k redundant units (usually, a linear function); and $P_k(x_k)$ PFFO or availability coefficient) of the subsystem when it contains x_k redundant units. It is well-known that for loaded redundancy (the so-called "hot standby") of group including one main and x_k identical redundant units takes the form

$$P_k(x_k) = 1 - (1 - p_k)^{x_k + 1};$$

and for an unloaded redundant (the so-called "cold standby") units takes the form

$$P_k(x_k) = \sum_{0 \le j \le x_k} \frac{(\lambda_k t)^j}{j!} \exp(-\lambda_k t);$$

The set of triplets S_k can be represented in the GF form as

$$\varphi_k(z) = \sum_{1 \le x_k < \infty} P_k(x_k) z^{C_h(x_k)} y^{x_k}. \quad (7)$$

Now consider a general procedure of optimal redundancy with the use of U-function. First take units 1 and 2 and arrange the Descartes interaction procedure between sets S_1 and S_2. To distinguish interaction procedure from common product of two generating function, let us introduce symbol $\otimes$.

$$\varphi_{1,2}(z,y)=\bigotimes_{+,\cup}\left(\varphi_1(z,y),\,\varphi_2(z,y)\right)=\bigotimes_{+,\cup}\left(\sum_{1\le x_1<\infty}P_1(x_1)\,z^{C_1(x_1)}y^{\{x_1\}},\ \sum_{1\le x_2<\infty}P_2(x_2)z^{C_2(x_2)}y^{\{x_2\}}\right)$$

$$=\sum_{\substack{1\le x_1<\infty\\1\le x_2<\infty}}P_1(x_1)\cdot P_2(x_2)\,z^{\otimes_+(C_1(x_1),C_2(x_2))}y^{\otimes_\cup(x_1,x_2)}=\sum_{\substack{1\le x_1<\infty\\1\le x_2<\infty}}P_1(x_1)\cdot P_2(x_2)\,z^{C_1(x_1)+C_1(x_1)}y^{\{x_1,x_2\}} \qquad (8)$$

$$=\sum_{1\le X_{1,2}<\infty}P_{1,2}(X_{1,2})z^{C_{1,2}(X_{1,2})}y^{X_{1,2}}.$$

From (8) clear that composition operator $\overset{\otimes}{}_{+}$ means summation corresponding components and operator $\overset{\otimes}{}_{\cup}$ means a collection of corresponding components.

Thus the obtained U-function $\varphi_{1,2}(z, y)$ represents the set of related PFFO, costs and numbers of redundant units for different configurations of the series connection of subsystems 1 and 2. The group of subsystems 1 and 2 now can be treated as an equivalent "aggregated" subsystem. Notice that solving optimal redundancy problem, one has to make some kind of "sifting" of function $\varphi_{1,2}(z, y)$. One has to order all terms of the final expression in (8) by increasing values of $C_{1,2}$ and exclude all terms of $\varphi_{1,2}(z, y)$ that have value of $P_{1,2}$ equal or smaller than previous terms. (If two terms have the same values C_{12} and $P_{1,2}$ one leaves an arbitrary single of them.). In the remaining terms, all $X_{1,2}$ are numbered by natural numbers in accordance with their order by increasing cost. This procedure is equivalent to deleting dominated terms

At the next step one obtains the UGF for the group of three subsystems 1, 2 and 3 applying the same convolution operator over U-function $\varphi_{1,2}(z)$ representing the aggregated subsystem and $\varphi_3(z)$ representing the subsystem 3:

$$\varphi_{1,2,3}(z,y)=\bigotimes_{+,\cup}\left(\varphi_{1,2}(z,y),\,\varphi_3(z,y)\right)=\bigotimes_{+,\cup}\left(\sum_{1\le X_{1,2}<\infty}P_{1,2}(X_{1,2})\cdot z^{C_{1,2}(X_{1,2})}y^{\{X_{1,2}\}},\ \sum_{1\le x_3<\infty}P_3(x_3)\,z^{C_3(x_3)}y^{\{x_3\}}\right)$$

$$=\sum_{\substack{1\le X_{1,2}<\infty\\1\le x_3<\infty}}P_{1,2}(X_{1,2})\cdot P_3(x_3)\,z^{\otimes_+(C_{1,2}(X_{12}),C_3(x_3))}y^{\otimes_\cup(X_{1,2},x_3)}=\sum_{\substack{1\le X_{1,2}<\infty\\1\le x_3<\infty}}P_{1,2}(X_{1,2})\cdot P_3(x_3)\,z^{C_{1,2}(X_{1,2})+C_3(x_3)}y^{\{X_{1,2},x_3\}}$$

$$=\sum_{1\le X_{1,2,3}<\infty}P_{1,2,3}(X_{1,2,3})z^{C_{1,2,3}(X_{1,2,3})}y^{X_{1,2,3}}.$$

(9)

This procedure continues until necessary final UGF representing the entire system is generated. Instead of further abstract presentation of the procedure, let us turn to a simple illustrative numerical example.

Consider a series system of two units. Let unit-1 and unit-2 are characterized by strings

$$S_1=\{M_1^{(1)},M_2^{(1)},...,M_{n_1}^{(1)}\}$$

and

$$S_2=\{M_1^{(2)},M_2^{(2)},...,M_{n_2}^{(2)}\}$$

Each multiplet M is a set of parameters $M_j^{(k)}=\{\alpha_{1j}^{(k)},\alpha_{2j}^{(k)},...,\alpha_N^{(k)}\}$ where N is the number of parameters in each multiplet.

"Interaction" of these two strings is an analogue of the Cartezian product whose memberts fill the cells of the following table:

Table 1.

	$M_1^{(1)}$	$M_2^{(1)}$	...	$M_{n_1}^{(1)}$
$M_1^{(2)}$	$M_1^{(1)} \otimes M_1^{(2)}$	$M_2^{(1)} \otimes M_1^{(2)}$	...	$M_{n_1}^{(1)} \otimes M_1^{(2)}$
$M_2^{(2)}$	$M_1^{(1)} \otimes M_2^{(2)}$	$M_2^{(1)} \otimes M_2^{(2)}$	...	$M_{n_1}^{(1)} \otimes M_2^{(2)}$
...	...	...	...	...
$M_{n_2}^{(2)}$	$M_1^{(1)} \otimes M_{n_2}^{(2)}$	$M_2^{(1)} \otimes M_{n_2}^{(2)}$	...	$M_{n_1}^{(1)} \otimes M_{n_2}^{(2)}$

Interaction of multiplets consists of iteractions of their similar parameters, for instance,

$$M_j^{(k)} \otimes M_i^{(h)} = \{(\alpha_{1j}^{(k)} \underset{f_1}{\otimes} \alpha_{1i}^{(h)}), (\alpha_{2j}^{(k)} \underset{f_2}{\otimes} \alpha_{2i}^{(h)}), ..., (\alpha_{Nj}^{(k)} \underset{f_N}{\otimes} \alpha_{Ni}^{(h)})\} \tag{3}$$

Operator $\otimes$, as well as each operator $\underset{f_s}{\otimes}$, in most natural practical cases possesses the commutativity property, i.e.

$$\underset{f}{\otimes}(a, b) = \underset{f}{\otimes}(b, a), \tag{4}$$

and the associativity property, i.e.

$$\underset{f}{\otimes}(a, b, c) = \underset{f}{\otimes}(a \underset{f}{\otimes}(b, c)) = \underset{f}{\otimes}((a \underset{f}{\otimes} b), c). \tag{5}$$

U-FUNCTION IN GENERAL CASE

Of course, operator $\underset{f_s}{\otimes}$ depends on the physical nature of parameter α_s and the type of structure, i.e. series or parallel.

Table 2.

Type of parameter	Type of structure	Result of interaction
A) α is unit's PFFO	series	$\alpha_{Aj}^{(k)} \underset{f}{\otimes} \alpha_{Ai}^{(h)} = \alpha_{Aj}^{(k)} \times \alpha_{Ai}^{(h)}$
	parallel	$\alpha_{Aj}^{(k)} \underset{f}{\otimes} \alpha_{Ai}^{(h)} = 1 - (1 - \alpha_{Aj}^{(k)}) \times (1 - \alpha_{Ai}^{(h)})$
B) α is number of units in parallel	series	$\alpha_{Bj}^{(k)} \underset{f}{\otimes} \alpha_{Bi}^{(h)} = (\alpha_{Bj}^{(k)}; B_{Ri}^{(h)})$
	parallel	$\alpha_{Bj}^{(k)} \underset{f}{\otimes} \alpha_{Bi}^{(h)} = (\alpha_{Bj}^{(k)}; B_{Ri}^{(h)})$
C) α is unit's cost (weight)	series	$\alpha_{Aj}^{(k)} \underset{f}{\otimes} \alpha_{Ai}^{(h)} = \alpha_{Aj}^{(k)} + \alpha_{Ai}^{(h)}$
	parallel	$\alpha_{Aj}^{(k)} \underset{f}{\otimes} \alpha_{Ai}^{(h)} = \alpha_{Aj}^{(k)} + \alpha_{Ai}^{(h)}$
D) α is unit's ohmic resistance	series	$\alpha_{Aj}^{(k)} \underset{f}{\otimes} \alpha_{Ai}^{(h)} = \alpha_{Aj}^{(k)} + \alpha_{Ai}^{(h)}$
	parallel	$\alpha_{Aj}^{(k)} \underset{f}{\otimes} \alpha_{Ai}^{(h)} = \left[(\alpha_{Aj}^{(k)})^{-1} + (\alpha_{Ai}^{(h)})^{-1}\right]^{-1}$

Type of parameter	Type of structure	Result of interaction
E) α is unit's capacitance	series	$\alpha_{Aj}^{(k)} \underset{f}{\otimes} \alpha_{Ai}^{(h)} = \left[(\alpha_{Aj}^{(k)})^{-1} + (\alpha_{Ai}^{(h)})^{-1}\right]^{-1}$
	parallel	$\alpha_{Aj}^{(k)} \underset{f}{\otimes} \alpha_{Ai}^{(h)} = \alpha_{Aj}^{(k)} + \alpha_{Ai}^{(h)}$
F) α is pipeline unit's capacitance	series	$\alpha_{Aj}^{(k)} \underset{f}{\otimes} \alpha_{Ai}^{(h)} = \min\left\{\alpha_{Aj}^{(k)}, \alpha_{Ai}^{(h)}\right\}$
	parallel	$\alpha_{Aj}^{(k)} \underset{f}{\otimes} \alpha_{Ai}^{(h)} = \alpha_{Aj}^{(k)} + \alpha_{Ai}^{(h)}$
G) α is unit's random time to failure	series	$\alpha_{Aj}^{(k)} \underset{f}{\otimes} \alpha_{Ai}^{(h)} = \min\left\{\alpha_{Aj}^{(k)}, \alpha_{Ai}^{(h)}\right\}$
	parallel	$\alpha_{Aj}^{(k)} \underset{f}{\otimes} \alpha_{Ai}^{(h)} = \max\left\{\alpha_{Aj}^{(k)}, \alpha_{Ai}^{(h)}\right\}$

In the problem of optimal redundancy, one deals with triplet of type "*Probability-Cost-Number of units*" for each redundant group: $M_j = \{\alpha_{1j}, \alpha_{2j}, \alpha_{3j}\}$. If there is a system of *n* series subsystems (single elements or redundant groups) , one has to use a procedure almost completely coincided with the procedure of compiling the dominating sequences at the Kettelle's algorithm.

In accordance with the description given above, the block diagram of the using U-functions, for example, for four subsystems can be presented as follows (see Figures 1 and 2).

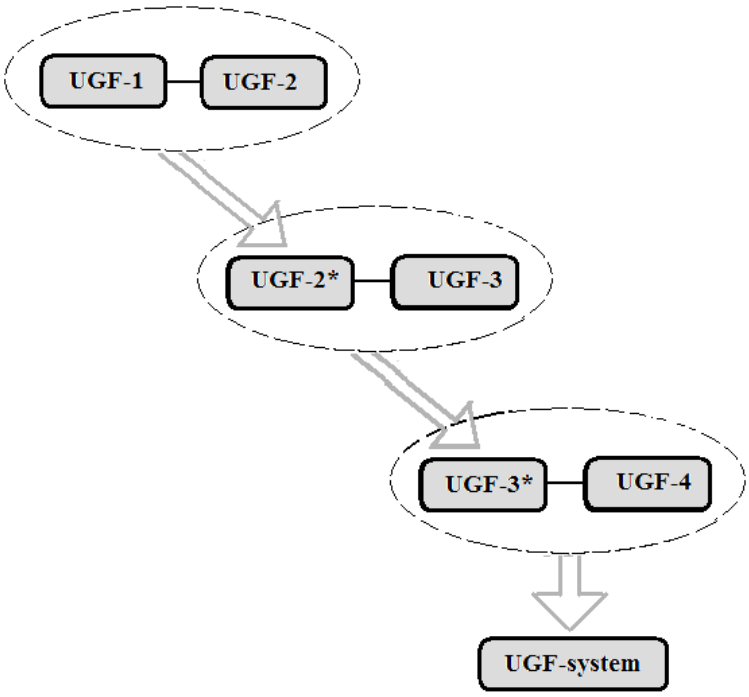


Figure 1. Block-diagram of the sequential procedure of solving.

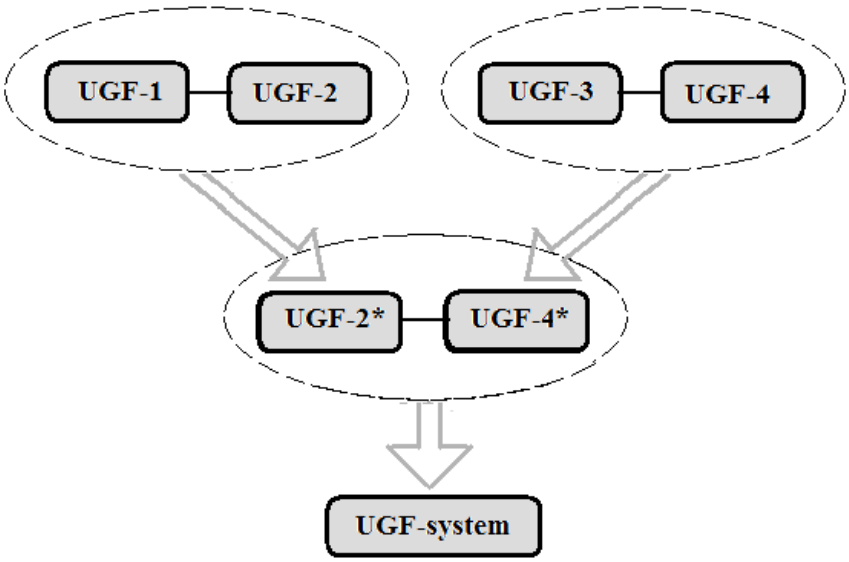


Figure 2. Block-diagram of the dichotomy procedure of solving.

CONCLUSION

The method of U-function is one of the methods of directed enumerating. It showed its effectiveness for solving a number of practical problems concerning with optimal resources allocation and analysis of multi-state systems and system consisting of multi-state elements.

REFERENCES (in chronological order)

(1986) Ushakov, I.A. Universal Generating Function (in Russian). Engineering Cybernetics, No. 5.
(1987) Ushakov, I.A. Universal generating function. Soviet Journal Computer Systems Science, No.3.
(1987) Ushakov, I.A. Optimal standby problem and a universal generating function. Soviet Journal Computer and System Science, No.4.
(1987) Ushakov, I.A. Solution of multi-criteria discrete optimization problems using a universal generating function. Soviet Journal of Computer and System Sciences , No. 5.
(1988) Ushakov, I.A. Solving of optimal redundancy problem by means of a generalized generating function. Elektronische Informationsverarbeitung und Kybernetik, No.4-5
(1995) Gnedenko, B.V., and Ushakov I.A. Probabilistic Reliability Engineering. John Wiley & Sons.
(1998) Levitin, G., Lisnianski, A., Ben Haim, H., and Elmakis, D. Redundancy optimization for series-parallel multi-state systems, IEEE Transactions on Reliability, , No. 2.
(1999) Levitin, G., and Lisnianski, A. Importance and sensitivity analysis of multi-state systems using universal generating functions method, Reliability Engineering & System Safety, No. 65.
(2000) Ushakov, I.A. The method of generating sequences. European Journal of Operational Research, No. 2.
(2000) Levitin, G., and Lisnianski, A. Optimal Replacement Scheduling in Multi-state Series-parallel Systems. Quality and Reliability Engineering International, , No. 16.
(2001) Levitin, G. Redundancy optimization for multi-state system with fixed resource-requirements and unreliable sources. IEEE Transactions on Reliability, No. 50.

(2001) Levitin, G., and Lisnianski, A. A new approach to solving problems of multi-state system reliability optimization, Quality and Reliability Engineering International, No. 47.
(2002) Levitin, G. Optimal allocation of multi-state elements in linear consecutively-connected systems with delays. International Journal of Reliability Quality and Safety Engineering, No. 9.
(2003) Levitin, G. Optimal allocation of multi-state elements in linear consecutively-connected systems. IEEE Transactions on Reliability, No. 2.
(2005) Levitin, G. The Universal Generating Function in Reliability Analysis and Optimization. Springer.

EMPIRICAL EVALUATION OF ACCURACY OF MATHEMATICAL SOFTWARE USED FOR AVAILABILITY ASSESSMENT OF FAULT-TOLERANT COMPUTER SYSTEMS

Vyacheslav Kharchenko

National Airspace University "KhAI", Kharkov, Ukraine

e-mail: v_s_kharchenko@ukr.net

Peter Popov

Centre for Software Reliability, City University London, United Kingdom

e-mail: ptp@csr.city.ac.uk

Oleg Odarushchenko, Valentina Zhadan

Poltava National Technical University named after Yuri Kondratiuk, Poltava, Ukraine

e-mail: skifs2005@mail.ru, valentinaodarus@gmail.com

ABSTRACT

Dependability assessment is typically based on complex probabilistic models. Markov and semi-Markov models are widely used to model dependability of complex hardware/software architectures. Solving such models, especially when they are stiff, is not trivial and is usually done using sophisticated mathematical software packages.

We report a practical experience of comparing the accuracy of solutions stiff Markov models obtained using well known commercial and research software packages. The study is conducted on a contrived but realistic cases study of computer system with hardware redundancy and diverse software under the assumptions that the rate of failure of software may vary over time, a realistic assumption. We observe that the disagreement between the solutions obtained with the different packages may be very significant. We discuss these findings and directions for future research.

1. INTRODUCTION

Dependability of computer systems is evaluated using probabilistic models in which the measure of interest is typically reliability, availability, etc. Often Markov chains are used in this process [1, 2, 3].

System modelers are often interested in transient measures, which provide more useful information than the steady-state measures. As models grow in size, closed-form solutions of transient measures become infeasible and in practice the models typically are solved using numerical methods.

Accurate dependability assessment of complex computer systems is an important issue. In many cases the accuracy of the assessment, e.g. in safety critical applications or in applications when poor dependability may lead to huge financial losses, is an essential part of the development process. The assessment methods and tools must provide high confidence in the assessment results and in many cases various regulation bodies would require the tools used in development to be certified to meet stringent quality requirements. To the best of our knowledge no such requirements (for software quality) are *not* in place for the software packages used in assessment. The modelers/assessors of complex computer systems are left with the choice – either to use the most accurate assessment algorithms, typically developed by researchers, and use an implementation of those in specific assessment or instead use the solutions available out of the box in the best known off-the-shelf mathematical packages available on the market. The first option – own implementation

of research results is not ideal because of the poor quality of the software code that one should expect from a prototype implementation. The second option for achieving accurate assessment – using off-the-shelf math software – is the focus of this paper.

Among the best known off-the-shelf math packages are Maple (Maplesoft), Mathematica (Wolfram Research) and Mathcad (PTC). These math packages enjoy high reputation among the respective customers earned over several decades by providing a wide range of solutions, and good support with regular updates.

The difficulties with transient numerical analysis of Markov chains have been studied extensively in the past – we survey the important related research. The main difficulty in analysis is the stiffness of the models. Given the extensive work on the issue in 80s and 90s would expect that the availably software packages used by modelers would provide accurate solution to stiff Markov chains. Surprisingly, this does not seem to be the case as this paper illustrates.

The paper is organized as follows: in section 2 we state the problem formally and describe a contrived example of fault-tolerant computer system which we use to compare several well known mathematical packages. In section 3 we present the results obtained. In section 4 we discuss the findings and their implication. Section 5 provides a survey of the important results on solving stiff Markov chains reported by others. Finally, in section 6 we conclude the paper and suggest ways forward.

2. PROBLEM STATEMENT

In this paper we study the accuracy of popular mathematical packages likely to be used in the dependability assessment of complex fault tolerant computer systems. The method of study used is as follows:

- Define a model of the system to be used in comparison;
- Develop solutions for system dependability, e.g. transient system's availability in the interval [0,t], using the instruments available by the chosen packages. As a benchmark solution to compare the solutions obtained with off-the-shelf software packages we used a highly specialized software utility, EXPMETH, which is an implementation of a method for solving stiff Kolmogorov equations [4];
- Compare the obtained solutions.

EXPMETH was developed more than 15 years ago and validated extensively on a range of models [5]. It was developed in Pascal program language and used to assess availability of a chosen system model. We have chosen a system described by a stiff Markov chain: the ratio between the rates of failure and repair ranges between 4 and 8.

The system we chose in the study is a fault-tolerant computer system with two hardware channels each executing software control as shown in Figure 1.

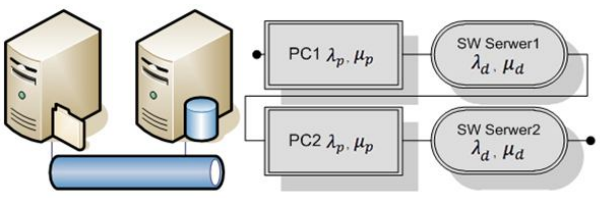


Figure 1. Reliability block diagram of the chosen fault-tolerant system

We assume that software run of the two hardware channels is diverse [5], i.e. non identical but functionally equivalent software copies are deployed, which offers protection against software design faults. In addition, we assume that the rates of failure and repair of software will vary over time, e.g. as a result of executing the software in partitions as discussed in [6].

The model to be used in the study is shown in Figure 2. We omit a more detailed justification of the chosen model as this is outside the scope of this paper as the focus of the study is the accuracy of the solutions provided by off-the-shelf math packages. Yet, we would like to stress that the model is plausible. Two channel configurations is very widely used in many safety-critical application, e.g. in instrumentation of nuclear plants. The variation of failure rates and repair is also a plausible concept – software may well perform different tasks with different importance, which would justify different degree of testing, hence different rates of failure and repair in the respective partitions.

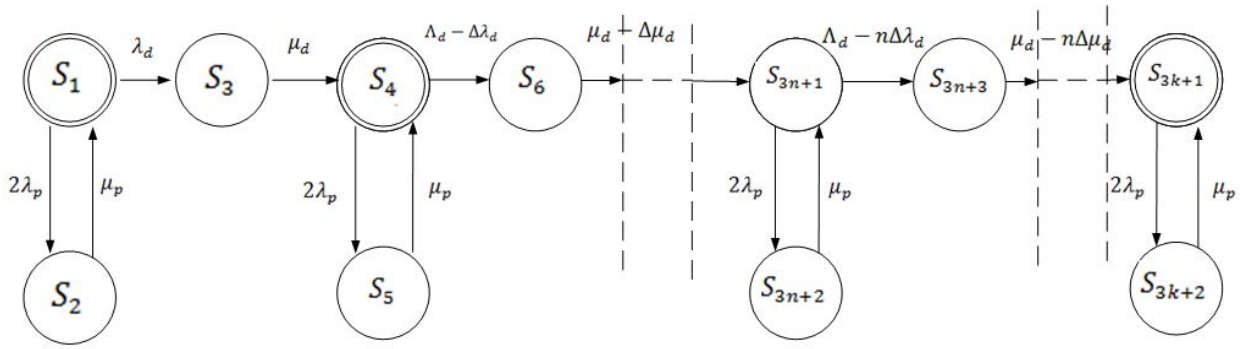


Figure 2. Model of the system to be studied. A 2-hardware channels, 2-software versions fault tolerant computer system with a variation of the rates of software failure and repair.

Informally, the operation of the system is as follows. Initially the system is working correctly – both hardware and software channels deliver the service as expected. If during the operation one of the hardware channels has failed the system operation will be failed over to the second channel until the first channel is "repaired". Similarly, a software component may fail, in which case a failover will take place, etc.

An important feature of the model is that as a result of software repair (e.g. restart of the failed channel) we assume that the rate of software failure of both channels will deteriorate by a small constant $\Delta\lambda_d$. Clearly this is just an assumption which might be unrealistic in many cases and its justification might require a detailed analysis of the application that software implements. However, the assumption captures a plausible phenomenon – variation of software failure rates which is well accepted in practice: various software 'aging effects' are indeed modeled by an increased rate of software failure. We conclude therefore, that the model is adequate for our purposes in this paper – a study of the accuracy of the solutions to stiff Markov chains offered by off-the-shelf math packages.

3. MODEL PARAMETERS

The model parameters are defined next:

- λ_p and μ_p – hardware failure and repair rates;
- λ_d and $\Delta\lambda_d$ – software initial failure rate and a step of failure rate decrease after software

recovers from failure;

- μ_d and $\Delta\mu_d$ – initial software repair rate and a step of software repair rate decrease after software recovers from failure.

System's behavior is modelled as a Markov process (Markov chain) that has a number of states, and the transition probabilities between these, $P_k^{(i)}$, depends only on the current state, i.e.

$$P\left\{x^{(i+1)} = x_k^{(i+1)}\Big|_{x^{(0)}\ldots x^{(i)}}\right\} = P\left\{x^{(i+1)} = x_k^{(i+1)}\Big|_{x^{(i)}}\right\} \quad (1)$$

The reader can notice that the model presented in Figure 2 is built with a set of similar fragments – the model fragments are topologically identical and only differ in terms of the values of model parameters.

From the model (Fig. 2) we can derive the following system of Kolmogorov equations:

For the initial fragment these are:

$$\frac{dP_1(t)}{dt} = -(2\lambda_p + \Lambda_d)P_1(t) + \mu_p P_2(t) \quad (2)$$

$$\frac{dP_2(t)}{dt} = -\mu_p P_2(t) + 2\lambda_p P_1(t) \quad (3)$$

$$\frac{dP_3(t)}{dt} = -\mu_d P_3(t) + \Lambda_d P_1(t) \quad (4)$$

For the internal fragments these are:

$$\frac{dP_{3i+1}}{dt} = -(\Lambda_d - i\Delta\lambda_d + 2\lambda_p)P_{3i+1}(t) + (\mu_d - (i-1)\Delta\mu_d)P_{3i}(t) + \mu_p P_{3i+2}(t) \quad (5)$$

$$\frac{dP_{3i+2}}{dt} = -\mu_p P_{3i+2}(t) + 2\lambda_p P_{3i+1}(t) \quad (6)$$

$$\frac{dP_{3i+2}}{dt} = -(\mu_d - i\Delta\mu_d)P_{3i+3}(t) + (\Lambda_d - i\Delta\lambda_d)P_{3i+1}(t) \quad (7)$$

And for the final fragment these are:

$$\frac{dP_{3k+1}}{dt} = -2\lambda_p P_{3k+1}(t) + \Delta\mu_d P_{3k}(t) + \mu_p P_{3k+2}(t) \quad (8)$$

$$\frac{dP_{3k+2}}{dt} = -\mu_p P_{3k+2}(t) + 2\lambda_p P_{3k+1}(t) \quad (9)$$

With the following initial conditions:

$$P_1(0) = 1, P_2(0) = 0, P_3(0) = 0, \ldots, P_{3i+1}(0) = 0, P_{3i+2}(0) = 0$$
$$P_{3i+3}(0) = 0, \ldots, P_{3k+1}(0) = 0, P_{3k+2}(0) = 0. \quad (10)$$

The availability function is defined as the sum of probabilities that system is in one of the states for which at least one of the channels is working, which is defined by the following sum:

$$P_a(\mathrm{t}) = \sum_{i:S_i \in MS_p} P_{S_i}(t) \tag{11}$$

4. RESULTS

Figure 2 represents a stiff Markov chain, characterized by the fact that some of the eigenvalues of Jacobean matrix $\frac{\partial f}{\partial y}$ are large in absolute value with negative real part, while some other eigenvalues are with small positive real part. Getting an accurate solution requires selecting a very small integrating step, which limits the distribution of the error. In such circumstances the classical implicit Runge-Kutta methods provide incorrect result even if a small step of integration is used. The error is caused by the rounding errors which accumulate over the large number of small steps. We consider only the implicit Runge-Kutta methods, because the explicit Runge-Kutta method can provide correct solution even if the system is stiff. In case of solving stiff ordinary differential equations (ODE) the numerical method have to satisfy the following condition:

1. convergence (the method has to converge to ODE);
2. special requirements for stability;
3. the method must pass successfully certain calculation tests.

The software utility EXPMETH implements a special numerical method of solving stiff differential equations – the "exponential" method, studied by O. Arushanyan and S. Zaliotkin [4]. It is based on an accurate representation and calculation of the matrix exponent. The results obtained with EXPMETH were used in the study as a reference solution (to compare the results obtained with the tools/methods). The exponential method was implemented step by step in each of the math packages included in the comparison: Mathcad 15, Maple 15 and Mathematica 8.0.1. We also used the standard solution built in the respective math packages for solving ordinary differential equations as detailed below:

1. In Mathcad 15 we used the built-in function *Stiffr(P,0,10000,D,J)*, with arguments:
   - *P* – the initial state vector of differential equations system;
   - *0, 10000* – time interval on which system availability was computed;
   - *D* – the system of differential equations (defined above);
   - *J* – eigenvalues of the respective Jacobean matrix.
2. In Maple 15 we used the built-in function *DSolve (odesys, numeric, vars, options)*, where:
   - *odesys* – is the set of ODE(s) and the initial/boundary conditions;
   - *numeric* – a parameter used to instruct dsolve to find a numerical solution;
   - *vars* - (optional) can be any indeterminate function of one argument, or a list of them such functions, representing the unknowns of the ODE problem;
   - *options* - equations of the form "keyword = value". In our case this parameter was used to select specify the method of integration, e.g. (stiff=true,method=rosenbrock) were used to set the stiff property to true and select the Rosenbrock method of solving the system of differential equations.
3. In Mathematica 8.0.1 we used the built-in function - *NDSolve{ODE},{t,1,10000}, Method->{"ExplicitRungeKutta}* , where:
   - *{ODE}* – defines the set of ODE(s) and the initial/boundary conditions;

- *{t,1,10000}* – indicates that the solution in time domain is sought on the interval [1,10000];
- *Method->{"ExplicitRungeKutta"}* – the explicit Runge-Kutta method was used.

Figure 3, 4 and 5 show the results obtained for system availability using each of the 3 math packages with the methods described above together with the results obtained using the exponential method computed both in the respective package and using EXPMETH utility.

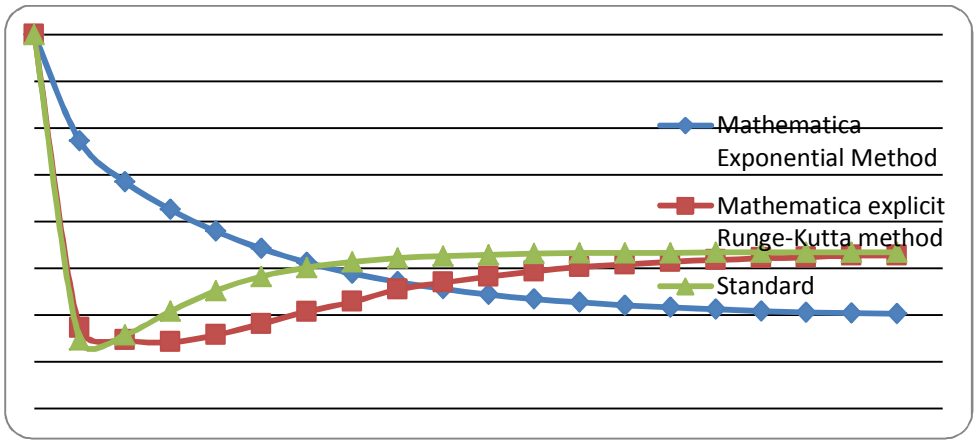


Figure 3. Availability function plots the results obtained with Mathematica 8.0.1 vs. EXPMETH.

The explicit Runge-Kutta method generally follows the solution provided by EXPMETH, while the exponential method is hopelessly inaccurate – starts with gross overestimation of system availability and gradually declines to a significant underestimation of system availability.

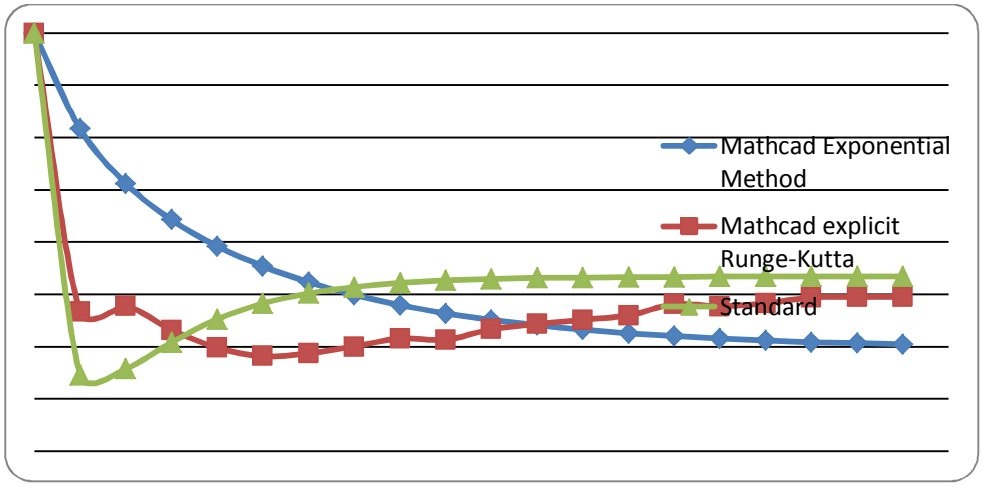


Figure 4. Availability function plots calculated with Mathcad 15 vs. EXPMETH.

The explicit Runge-Kutta method performs worst than in Mathematica although it also generally follows the solution provided by EXPMATH. The exponential method again is very poor – starts with gross overestimation of system availability and gradually declines to a significant underestimation of system availability.

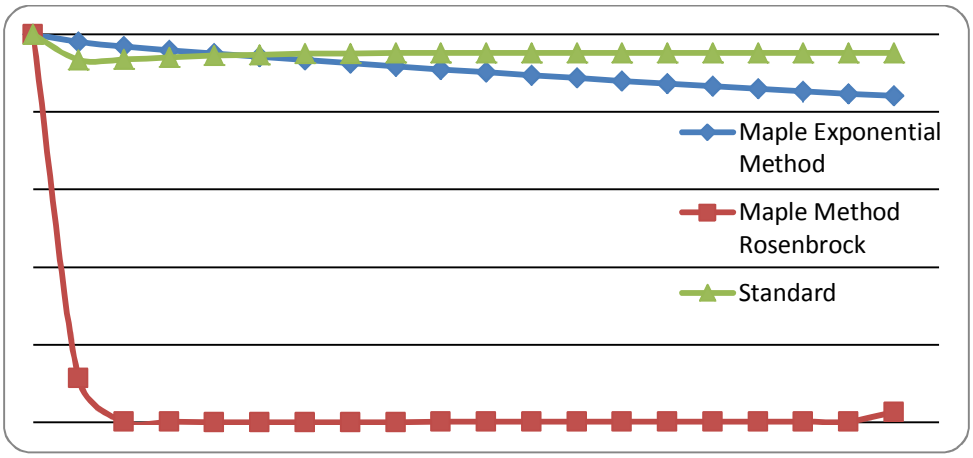


Figure 5. Availability function plots calculated with Maple 15 vs. EXPMETH.

It is really striking how inaccurate the Rosenbrock method is despite being explicitly said to target stiff Markov models. The exponential method shows the same pattern of poor accuracy – starts with gross overestimation of system availability which is gradually replaced by a significant underestimation of system availability.

We also included in the study the simulation solver of the well known tool Mobius, developed and maintained by the Performability group at the University of Illinois at Urbana Champaign (UIUC). Since this solver differs in nature from the other three, we will provide a more extensive description of how the system model shown in Figure 2 was developed using the SAN (stochastic activity networks) formalism of Mobius.

The fragments of the system model discussed above were explicitly specified as *atomic models*: the system model consists of 7 fragments (the initial fragment, 5 internal fragments and the final final fragment). Figure 6, 7, 8 and 9 illustrate of initial, two of the internal fragments and the final fragment build using the Mobius SAN formalism.

Tables 1 and 2 map the fragments to the atomic models: Table 1 shows the mapping between the states. Table 2 – shows the transitions between fragments.

Table 1. Fragments to the atomic models

fragment / state	Initial	Internal 1	Internal 2-5	Final
Both software components (SC) working	sw_working	sw_working	sw_working	sw_working
Both hardware components (HC) working	hw_working	hw_working	hw_working	hw_working
First SC failed	-	sw_fail	sw_failure1	sw_failure1
Second SC failed	-	-	sw_failure2	sw_failure2
First HC failed	hw_w1	hw_w1_int1	hw_w1_int2	hw_w1
Second HC failed	hw_w2	hw_w2_int1	hw_w2_int2	hw_w2
Both SC failed	-	syst_failed	syst_failed	syst_failed
Both HC failed	system_failed	system_failed	system_failed	system_failed

Table 2. Transitions between fragments

fragment / transition	Initial	Internal 1	Internal 2-5	Final
First HC failed	fail1	fail1_int1	fail1_int2	fail1
Second HC failed	fail2	fail2_int1	fail2_int2	fail2
First SC failed	-	sw_fail	sw_fail1	sw_fail1
Second SC failed	-	-	sw_fail2	sw_fail2
Recovery after first HC elimination	recov1	recov1_int2	recov1_int2	recov1
Recovery after second HC elimination	recov2	recov2_int2	recov2_int2	recov2
Recovery after first SC elimination	-	sw_recovery1	sw_recov1	sw_recov1
Recovery after second SC elimination	-	-	sw_recov2	sw_recov2

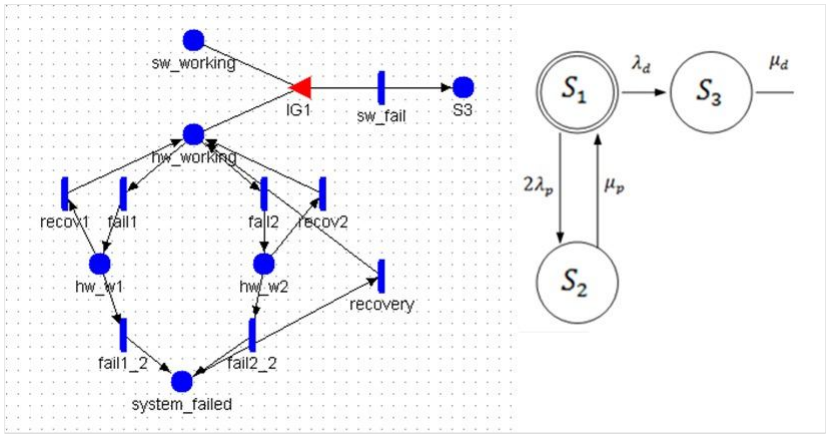


Figure 6. The atomic model of the initial fragment using Mobius SAN.

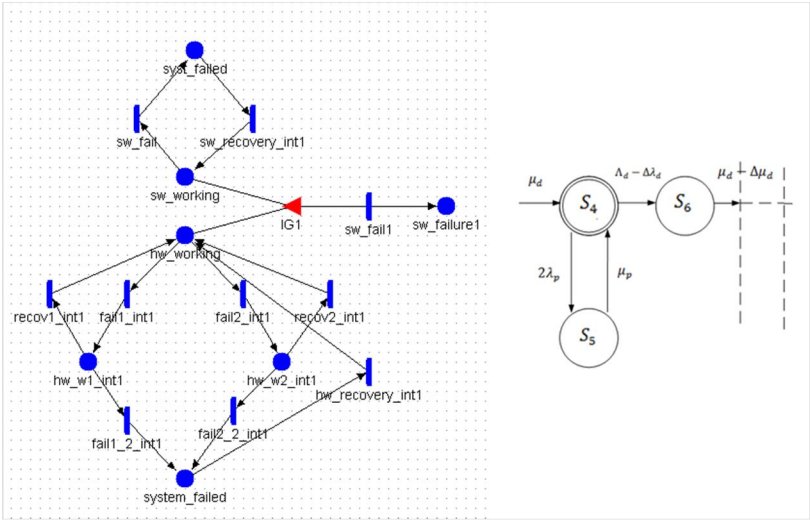


Figure 7. The atomic model showing the transition from the initial fragment to the first internal fragment.

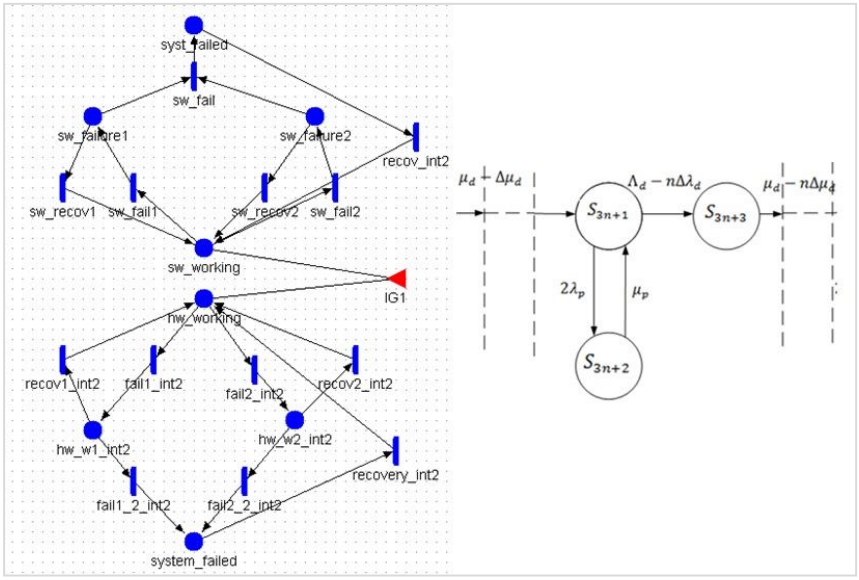


Figure 8. The atomic model with transitions between the internal fragments.

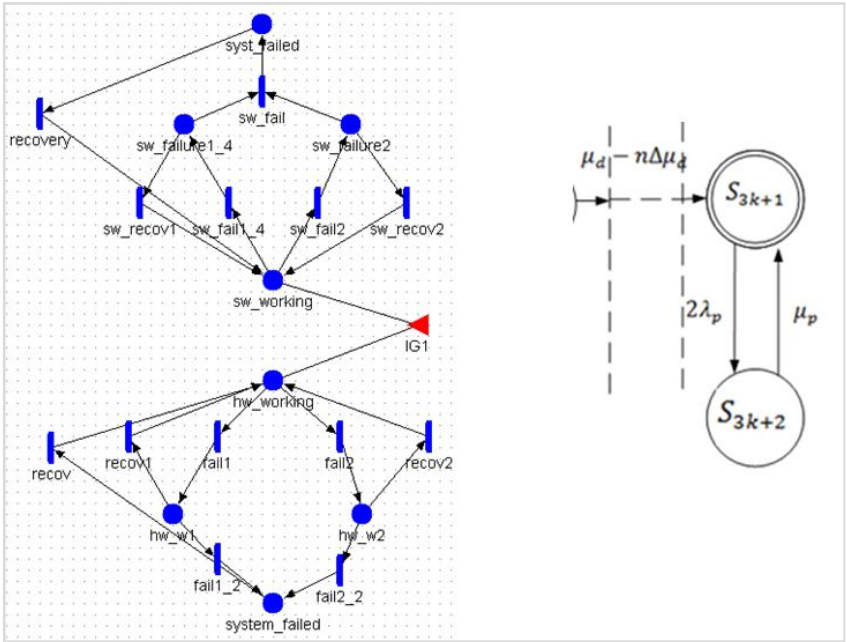


Figure 9. The atomic model with transitions between the last internal and the final fragment.

Figure 10 shows the system model using the SAN replication and joins. For further detail on the SAN syntax, the reader is encouraged to consult the SAN documentation.

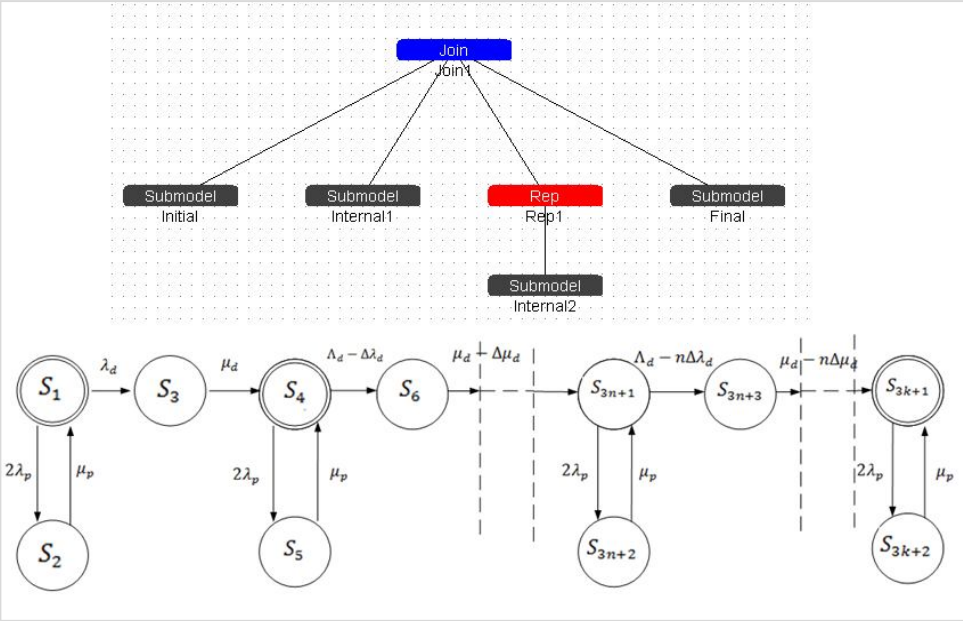


Figure 10. The system model, shown as a SAN compound model using REP and JOIN formalisms built in SAN.

System availability was computed via Monte Carlo simulation (simulation solver) with predefined confidence intervals. Figure 11 shows the results from the simulation solver with the respective confidence intervals against the reference availability obtained with the EXPETMETH utility.

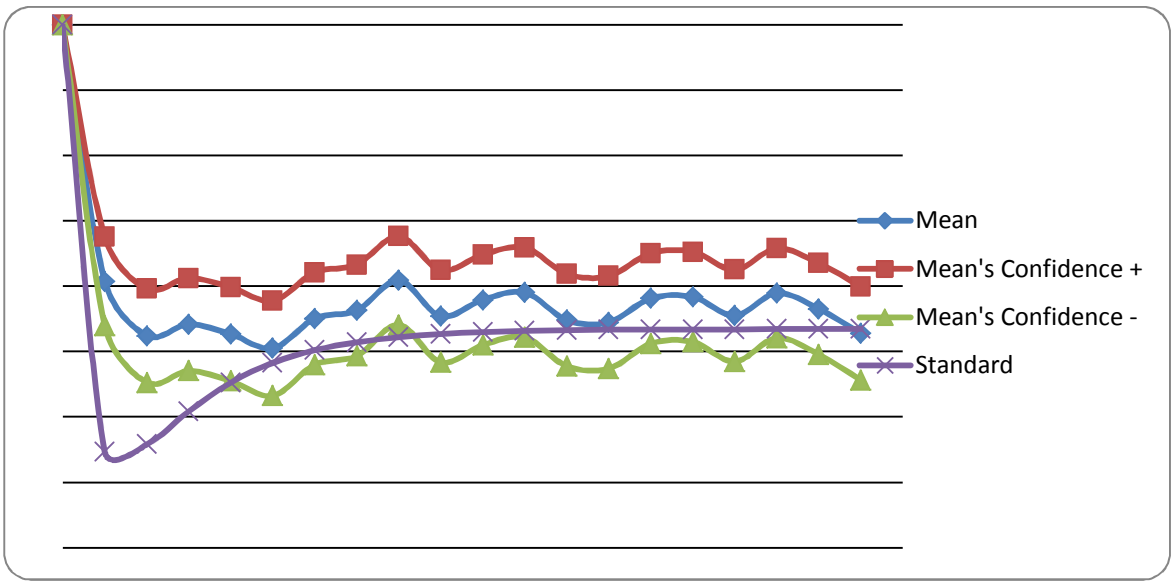


Figure 11. System availability calculated using SAN Mobius vs. EXPMETH.

The results obtained with all packages included in the comparison are shown in Figure 12.

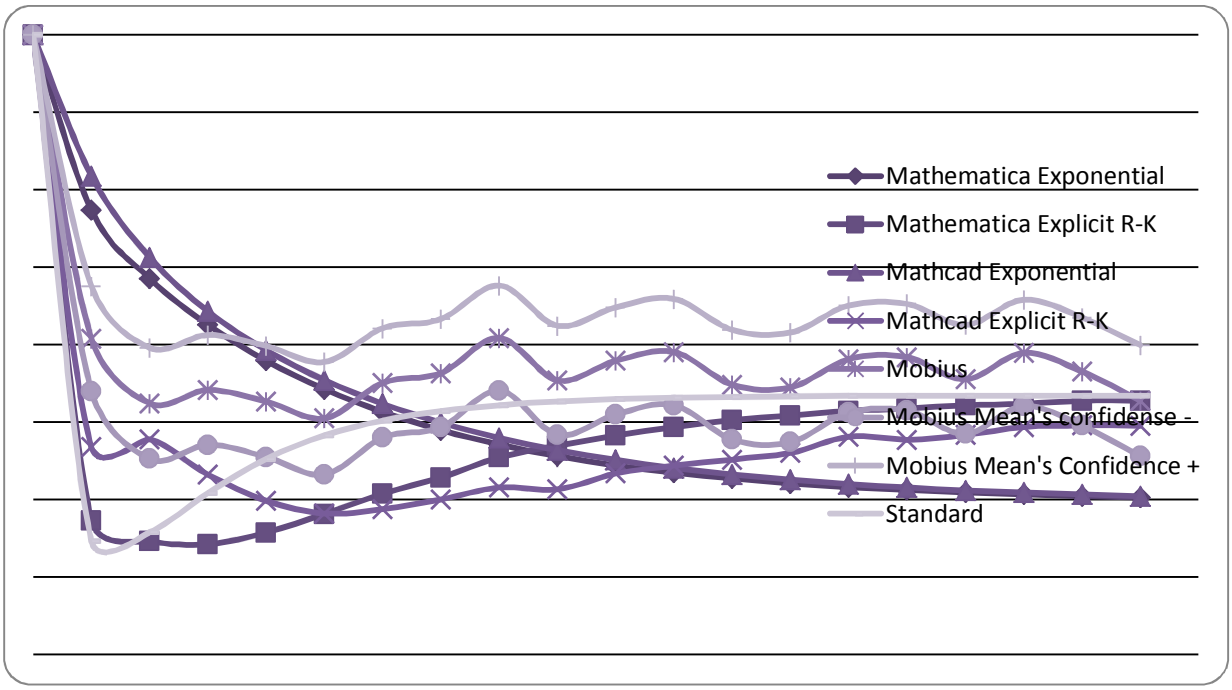


Figure 12. System availability function obtained with different tools/methods.

5. DISCUSSION

It is really striking how different the results obtained with different tools and methods are. The standard implementations built in the math packages for solving ordinary differential equations perform poorly: while the explicit Runge-Kutta method generally produce similar results (and generally follow those obtained with EXPMETH), the differences are non-negligible between Mathematica and Mathcad. The Rosenbrock method built in Maple to deal with stiff models seems to be an outlier – the availability it offers is too pessimistic.

Somewhat surprisingly, the exponential method implemented with the standard math packages offer results which are very different from those obtained with EXPMATH despite the fact that the algorithm is the same. One wonders what cause such a significant discrepancy.

The simulation solver seems to be close to the reference result obtained with EXPMETH – most of the time the EXPMETH computed availability is within the confidence intervals produced by the simulation solver.

The conclusions from this empirical comparison are alarming – the accuracy of the solutions offered by the popular math packages do not inspire high confidence! If one is to make a decision as to which of the results to trust one would really have very little to base their decision on. Clearly, further investigation is needed into which of the results should really be trusted.

6. RELATED RESEARCH

An extensive discussion of several problems related to obtaining accurate transient solution for Markov chains is presented in [7]. These authors note that stiffness is due to a large ratio of the model parameters but also, an observation made earlier by others, may be caused by the “mission time” if there exists a solution component whose variation is greater than 1/t, where t is the mission

time. They propose an extension to the standard methods for solving Kolmogorov equations such as Runge-Kutta and TR BDF2 method which offer *stability* despite the stiffness of the model to be solved.

Other relevant studies, [8] and [9], looked at the error of approximate methods of solving stiff Markov models using as a benchmark a M/M/1/k system, for which an exact solution is known.

In [10] and [11] the authors present an approximate method for solving stiff Markov models in which *aggregation* of the states is used: the states in the original model are divided into fast (i.e. those that have at least one fast outgoing transition), slow and fast recurrent (i.e. with fast outgoing and incoming transitions). The authors suggest that a good approximation of cumulative measures can be obtained when the original model is reduced to a model with slow states only and this derived model is solved. The authors experimentally evaluated the accuracy of the proposed method and report acceptable results for the cases when a non-stiff model is derived as a result of the proposed aggregation.

[12, 13] offer an empirical comparison of *uniformisation* methods of solving Markov chains. The authors point out that the standard uniformisation method proposed by Jensen performs poorly on stiff Markov models and concentrate on *Adaptive Uniformisation*. The problem is studied very extensively on complex contrived examples.

An interesting empirical study of the accuracy of scientific software (i.e. developed to process complex dataset from deep shelf oil exploration) was conducted by Less Hatton and Andy Roberts [14], which in nature is very similar to the study presented here. The authors conducted a thoroughly controlled experiment and compared the results from 15 independently developed very complex software packages, developed to the same specification. They report on the results obtained with 9 of the packages included in the study. All packages were subjected to the same large datasets collected from complex array of sensors. The results observed from the packages disagreed dramatically. The authors discussed why scientific software is of so poor quality.

7. CONCLUSIONS

The analysis of the results on system availability obtained with different off-the-shelf math packages allows us to draw the following conclusions:

- Each of the packages included in the study allows a modeler/assessor to evaluate system availability: each package either has a built in method which is appropriate for the task or allows one to build a routine (e.g. exponential method) to do so. The results obtained with the different packages, however, differ very significantly.
- The greatest discrepancy between the reference solution (obtained with EXPMATH) is observed with Maple 15 package when Rosenbrock's method for stiff Markov chains is used. The usage of the built-in function "dsolve" with 20 differential equations, leads to a large number of commands that are very similar in syntax, which greatly reduces the usability and seems to limit the scope for detecting and fixing faults. We scrutinized further the impact of model complexity on the accuracy of the solutions obtained with this method and observed that the accuracy deteriorates quickly with the increase of model complexity.
- The results obtained with Mathematical 8.0.1 and Mathcad 15 with the explicit Rung-Kutta method are closest to the reference solutions obtained with EXPMETH, which

is not surprising. We observed, however, that Mathematica 8.0.1 is sensitive to the initial period of operation: we observed a "drop" in availability at the beginning of the interval for which system's availability is computed.

- In terms of performance the packages performed as follows: Mathematica 8.0.1 took 01:16.9 seconds to compute the solution, Maple 15 – 07:44.3 seconds and Mathcad 15 – 00:25.34 seconds.
- The satisfactory solution was obtained using the simulation package Mobius.

In summary, if we are to rank the math packages included in the comparison, we would rank highest Mathematica 8.0.1 and Mathcad 15 (using the explicit Runge-Kutta solver) and the simulation solver of Mobius.

REFERENCES

[1] Volkov, L. (1981): Managing the operation of the aircraft systems: Tutorial. Moscow: Vyshaya Shkola. 368 p.

[2] Ventsel', E., Ovcharov, L. (2000): Probability theory and its applications in engineering. Moscow: Nauka. 480 p .

[3] Ventsel', E., Ovcharov, L. (1988): The theory of stochastic processes and its engineering applications, Moscow: Nauka. 384 p .

[4] Arushanyan, O., Zaletkin, S. (1990) : Numerical solution of ordinary differential equations using FORTRAN, Moscow: Moscow State University, 336 p.

[5] Littlewood, B., Popov, P. & Strigini, L. (2001): Modelling software design diversity - a review //ACM Computing Surveys. Vol. 33, №1. pp. 177 - 208.

[6] Popov, P. & Manno, G. (2011): The Effect of Correlated Failure Rates on Reliability of Continuous Time 1-Out-of-2 Software. in Computer Safety, Reliability, and Security (SAFECOMP 2011). Naples, Italy: Springer.

[7] Malthora, M., J.K. Muppala, & Trivedi K.S. (1994): Stiffness-Tolerant Methods for Transient Analysis of Stiff Markov Chains. Microelectronics Reliability. Vol. 34, № 11, pp. 1825-1841.

[8] Reibman, A., et al. (1989): Analysis of stiff Markov Chains. Operations Research Society of America. Vol. 1, № 2, pp. 126 - 133.

[9] Reibman, A. & Trivedi K.S. (1988): Numerical Transient Analysis of Markov models. Comput. Opns. Res. Vol. 15, № 1. pp. 19-36.

[10] Bobio, A. & Trivedi K.S. (1986): An Aggregation Technique for the Transient Analysis of Stiff Markov Chains // IEEE Transactions on Computers. Vol. C-35, № 9. pp. 803 - 814.

[11] Bobio, A. & Trivedi K.S. (1990): Computing Cumulative Measures of Stiff Markov Chains Using Aggregation // IEEE Transactions on Computers. Vol. 39, № 10. pp. 1291-1298.

[12] Diener, J.D. (1994): Empirical Comparison of Uniformisation Methods for Continuous Time Markov Chains, in Electrical and Computer Engineering // The University of Arizona. 97 p.

[13] Diener, J.D. & Sanders W.H. (1995): Empirical Comparison of Uniformisation Methods for Continuous-Time Markov Chains, in Computations with Markov Chains // W.J. Stewart, Editor. Kluwer Academic. pp. 547 - 570.

[14] Hatton, L. & Roberts A. (1994): How Accurate is scientific Software? // IEEE Transactions on Software Engineering. pp. 785 - 797.

ISSN 1932-2321