# PROBABILISTIC DATA ANALYSIS FOR PREDICTING MEAN TIME BEFORE CRITICAL INTEGRITY LOSSES OF COMPLEX SYSTEM WHEN EXPLICIT QUANTITATIVE REQUIREMENTS TO INTEGRITY ARE NOT SPECIFIED

A. Kostogryzov[1], A. Nistratov[2], G. Nistratov[3]
[1]*Federal Research Center "Computer Science and Control", Russian Academy of Sciences*
[2]*Russian Power Agency of Ministry for the Power Generating Industry*
[3]*Research Institute for Applied Mathematics and Certification*
*Moscow, RUSSIA*
e-mail: [1]Akostogr@gmail.com

**Abstract**

When the requirements to system integrity are not specified (including dangerous manufacture, "smart"equipment, robotic systems, artificial intelligence systems etc.) the questions are appeared: What time can pass before crossing of conditional border of a possible critical loss of integrity? How to define this conditional time border quantitatively (to use preventive countermeasures in real time)? The mathematical approach to solve a problem of predicting mean time before critical integrity losses for complex system, when explicit quantitative requirements to integrity are not specified, is proposed.
***Keywords:*** data science, integrity loss, prediction

## 1 Introduction

For modern complex systems there are often problems of predicting mean time before critical integrity losses - it is understood as virtual crossing a conditional border of integrity. System (element) integrity is defined as such system (element) state when system (element) purposes are achieved with the required quality and safety. Reservoirs with water on coal mines may serve an example of system when quantitative requirements to integrity are defined obviously. Filling reservoir above defined upper border is inadmissible because of possible overflow of the reservoir and mine flooding. And filling of the reservoir below defined bottom border is inadmissible from the point of view of readiness to fire-prevention actions. However at systems operation in the conditions of uncertainty (for example, in dangerous manufacture) there are cases when integrity borders are fuzzy or are not specified. So, for enterprise equipment the monitored values parameters are always to be kept within working ranges. But really for some parameters there are deviations not only from working, but also from normative borders. And thus critical integrity losses do not appear (emergencies, failure, quality or safety losses), though long deviations from normative requirements potentially conduct to integrity losses. The questions are: What time can pass before crossing of such conditional border of a possible critical loss of integrity? How to define this conditional time border quantitatively (to use preventive countermeasures in real time)? Similar questions are not only for dangerous manufacture, but also for "smart"equipment, robotic systems, artificial intelligence systems. Integrity losses of requirements to quality, reliability or safety at systems operation are critical independent on borders which are specified obviously or not specified [1]. It is proposed the mathematical approach to solve such problem.

# 2 Description of idea

The mathematical model, allowing to calculate the probability of integrity during given prognostic period ($T_{given}$) should be selected. Analytical impression should reveal dependability on frequency of dangerous influences (defining the beginning of influencing), mean activation time, mean recovery time, time between the end of diagnostics and the beginning of next diagnostics, diagnostics time. Considering that the prediction is useful for such time $T_{given}$, for which it is possible to undertake preventive actions, and this time practically equals to activation time (when "integrity"may be lost after beginning of influencing), we define values of these parameters as unknown and designate these by one unknown "x".
Further, setting confidence probability of integrity during given prognostic period we solve the analytical equation to find unknown "x". The maximum value of all revealed solutions "x"gives more wide opportunities for a choice and use of adequate preventive countermeasures against possible losses of integrity. This is the found solution.

# 3 Selected probabilistic model

For every element and whole system the next limited set of two elementary events is proposed: "integrity is provided"(when from integrity point of view no additional actions are needed) and alternatively "integrity is lost"(when some actions are needed for recovering lost integrity).
For calculation in point of given prognostic period $T_{given}$ the probabilistic model "Protection against dangerous influences"[1] is selected. It allows to estimate technology of periodical system diagnostics. During diagnostics the recovery of lost integrity is initiated (if needed). The next metrics are used for probabilistic prediction: the probability $P(T_{infl}, T_{activ}, T_{betw.}, T_{diag}, T_{given})$ of integrity during given prognostic period (if all time during this given period element or system will be in elementary event "integrity is provided") and the probability to lose integrity (if at least once during this given period element or system will be in event "integrity is lost") - as addition to 1 the probability of integrity.
If element or whole system is presented as "black box"the input data for probabilistic prediction are the next: given prognostic period $T_{given}$; frequency of dangerous influences on "black box"$\frac{1}{T_{infl}}$ (defining the beginning of influencing); mean activation time $T_{activ}$ (when "integrity"may be lost after beginning of influencing); time between the end of diagnostics and the beginning of next diagnostics $T_{betw.}$; diagnostics time $T_{diag}$.

The next assumption is used: diagnostics time includes recovery time.
There are possible the next variants for technologies 1 and 2: variant 1 - the given prognostic period $T_{given}$ is less than established period between neighboring diagnostics ($T_{given} < T_{betw.} + T_{diag}$) - see Figure 1; variant 2 - the prognostic period $T_{given}$ is more than or equals to established period between ne ighboring diagnostics ($T_{given} \geq T_{betw.} + T_{diag}$).

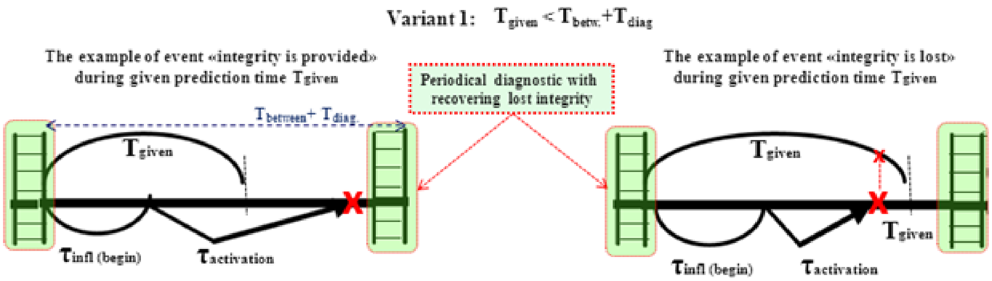


Figure 1: Some elementary events (left - "integrity is provided", right - "integrity is lost" during $T_{given}$)

The next formulas for probability distribution function (PDF) of time between the losses of element or system integrity are proposed under the condition of independence.
PDF for the model, variant 1: the probability of integrity is equal to

$$P_{(1)}(t) = 1 - \Omega_{infl}(t) * \Omega_{activ}(t), \tag{1}$$

where $\Omega_{infl}(t)$ - is the PDF of time between neighboring dangerous influences; $\Omega_{activ}(t)$ - is the PDF of activation time. These PDF $\Omega_{infl}(t)$ and $\Omega_{activ}(t)$ may be exponential PDF, $\Omega_{infl}(t) = 1 - e^{\frac{-t}{T_{infl}}}$, $\Omega_{activ}(t) = 1 - e^{\frac{-t}{T_{activ}}}$ - see rationale in [1]. For different threats a frequency of dangerous influences for these PDF is the sum of frequencies of every kind of influences.
PDF for the model, variant 2: the probability of integrity is equal to

$$P_{(2)}(T_{given}) = N(\frac{T_{betw} + T_{diag}}{T_{given}})P_{(1)}N(T_{betw} + T_{diag}) + (\frac{T_{rmn}}{T_{given}})P_{(1)}(T_{rmn}), \tag{2}$$

where $N = \frac{T_{given}}{T_{betw}+T_{diag}}, T_{rmn} = T_{given} - N(T_{betw} + T_{diag})$. The probability of integrity $P_{(1)}(t)$ is defined by (1).

# 4 Solution for "Black box"

The proposed method allows to estimate mean time before critical integrity losses of element or system, presented as "black box", for the given confidence probability to lose integrity $P_{conf}(T_{given})$. The found mean time before critical integrity losses is a solution $x_0$ of equation:

$$P(T_{infl}, x, T_{betw}, T_{diag}, x) = P_{conf}(x) \tag{3}$$

concerning of unknown parameter x. For calculations the formulas (1) - (2) are used. $T_{infl}$ is the mathematical expectation of PDF $\Omega_{infl}(t)$, it is defined by parameter statistics. The others parameters $T_{betw.}, T_{diag}$ in (3) are known. According to properties of function $P(T_{infl}, T_{activ}, T_{betw.}, T_{diag}, T_{given})$ the maximal x exists when $0 < P(T_{infl}, x, T_{betw}, T_{diag}, x) < 1$. I.e. the mean time before critical integrity losses is equal to maximum $X_0$ from all defined $x_0$ as solution of (3).

# 5 Solution for complex system

The model above is applicable to the system presented as one element "Black box". The main result of such system modelling is the mean time before critical integrity losses $X_0$ depending on threats, periodic diagnostics and recovery time [1]. For a complex system with parallel or serial structure integrated mean time before critical integrity losses is estimated from $X_{0i}$ , defined for i-th element by using method for "Black box". Let's consider the elementary structure from two independent parallel elements that means logic connection "OR"or series elements that means logic connection "AND". The mean time before critical integrity losses $X_0$ for system combined from two independent elements is equal to $X_0 = \frac{1}{\frac{1}{X_{01}} + \frac{1}{X_{02}}}$ for series connection and $X_0 = X_{01} + X_{02} - \frac{1}{\frac{1}{X_{01}} + \frac{1}{X_{02}}}$ for parallel connection.

It is correct for assumption that random values of time before critical integrity losses are distributed exponentially with mean $X_{0i}$.

**Example.** Let the frequency of dangerous influences is 1time a month ($T_{infl} = 1 month$), time between the end of diagnostics and the beginning of next diagnostics is equal to 8 hours ($T_{betw} = 8 hours$). Diagnostics and recovery time is neglected ($T_{diag} = 0$). If confidence probability is about 0.95, prognostic mean time before critical integrity losses is equal to 222 hours for obligatory adequate reaction every 8 hours. For rare events 1time a year the level of probability to lose integrity is about 0.9996.

For confidence probability of integrity about 0.99 estimated mean time before critical integrity losses may be very hard. Therefore confidence probability is recommended to set on level from 0.8 to 0.99 in dependence on system or element importance and possibilities for counteractions or from practice statistics. The approach is implemented by the Joint-Stock Company "Siberian Coal Energy Company"which is the leading coal producer in Russia and one of the world's largest coal companies [1].

# References

[1] V. Artemyev, A. Kostogryzov, Ju. Rudenko, O. Kurpatov, G. Nistratov, A. Nistratov (2017). *Probabilistic methods of estimating the mean residual time before the next parameters abnormalities for monitored critical systems.*. Proceedings of the 2nd International Conference on System Reliability and Safety (ICSRS- 2017), December 20-22, 2017, Milan, Italy, pp. 368-373.