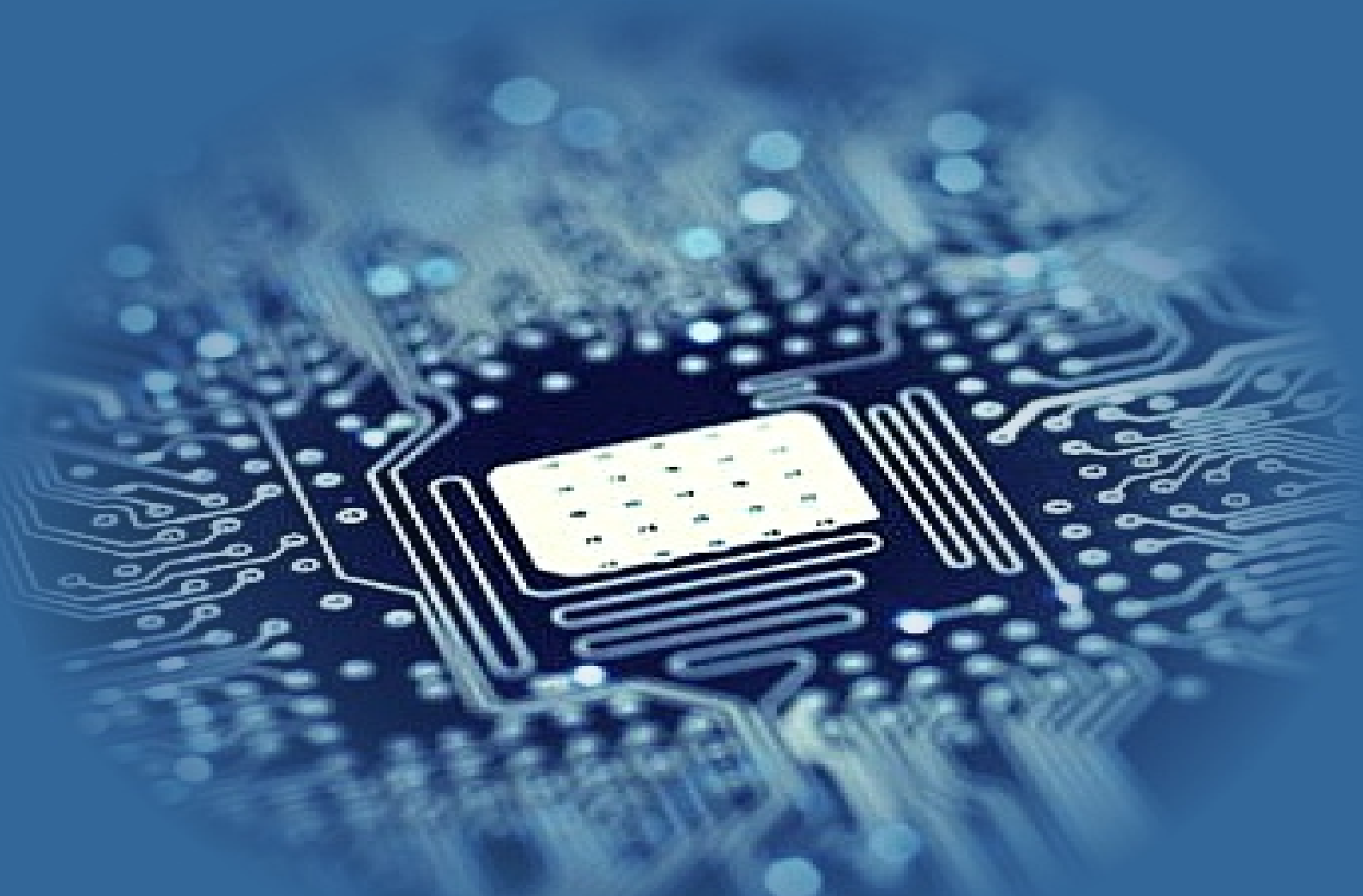


COMPUTER MODELLING AND NEW TECHNOLOGIES

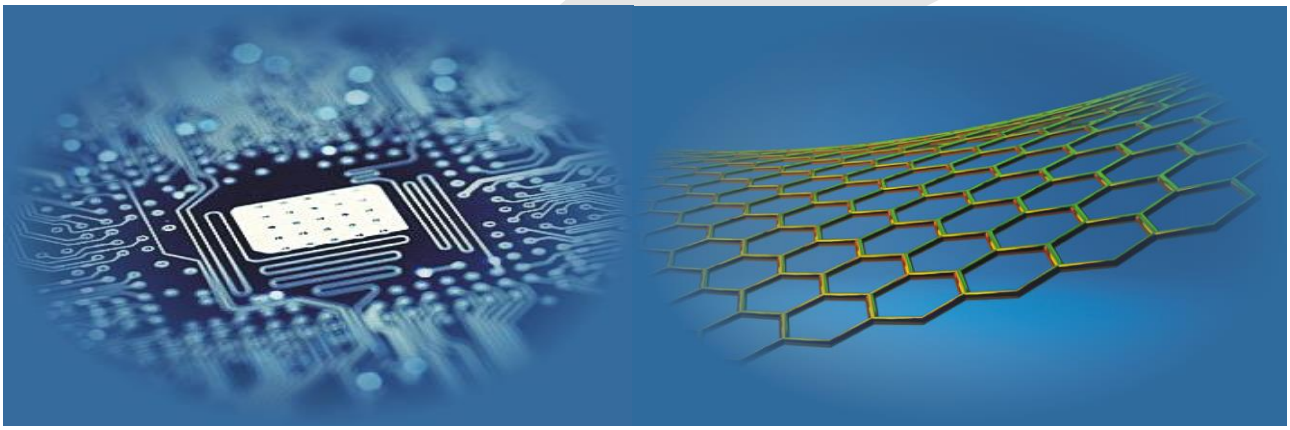
2017
VOLUME 21 NO 1
ISSN 1407-5806 ISSN 1407-5814 on-line

Latvian Transport Development and Education Association

Computer Modelling and New Technologies

2017 Volume 21 No 1

ISSN 1407-5806, ISSN 1407-5814 *(On-line: www.cmnt.lv)*



Riga – 2017

Journal topics:	Publisher	Supporting Organizations
• mathematical and computer modelling • computer and information technologies • natural and engineering sciences • operation research and decision making • nanoscience and nanotechnologies • innovative education	Latvian Transport Development and Education Association	Latvian Academy of Sciences Latvian Operations Research Society Fraunhofer Institute for Factory Operation and Automation IFF, Germany

Articles should be submitted in **English**. All articles are reviewed

EDITORIAL CORRESPONDENCE	COMPUTER MODELLING AND NEW TECHNOLOGIES, 2017, Vol. 21, No.1 ISSN 1407-5806, ISSN 1407-5814 (on-line: www.cmnt.lv)
Latvian Transport Development and Education Association	**Scientific and research journal** **The journal is being published since 1996**
68 Graudu, office C105, LV-1058 Riga, Latvia **Phone: +371 29411640** E-mail: yu_shunin@inbox.lv http://www.cmnt.lv	The papers published in Journal 'Computer Modelling and New Technologies' are included in: **INSPEC,** www.theiet.org/resources/inspec/ **VINITI,** http://www2.viniti.ru/ **CAS Database** http://www.cas.org/ **SCOPUS**

Editors' Remarks

****

Journey Home

by Rabindranath Tagore

The time that my journey takes is long and the way of it long.

I came out on the chariot of the first gleam of light, and pursued my voyage through the wildernesses of worlds leaving my track on many a star and planet.

It is the most distant course that comes nearest to thyself, and that training is the most intricate which leads to the utter simplicity of a tune.

The traveller has to knock at every alien door to come to his own, and one has to wander through all the outer worlds to reach the innermost shrine at the end.

My eyes strayed far and wide before I shut them and said `Here art thou!'

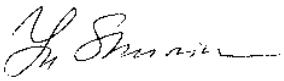
The question and the cry `Oh, where?' melt into tears of a thousand streams and deluge the world with the flood of the assurance `I am!'

Rabindranath Tagore (1861-1941)♠

****

This 21th volume No.1 includes research papers on **Information and Computer Technologies** and **Mathematical and Computer Modelling.**
Our journal policy is directed to fundamental and applied scientific researches, innovative technologies and industry, which is the fundamentals of the full-scale multi-disciplinary modelling and simulation. This edition is the continuation of our publishing activities. We hope our journal will be of interest for research community and professionals. We are open for collaboration both in the research field and publishing. We hope that the journal's contributors will consider collaboration with the Editorial Board as useful and constructive.

EDITORS

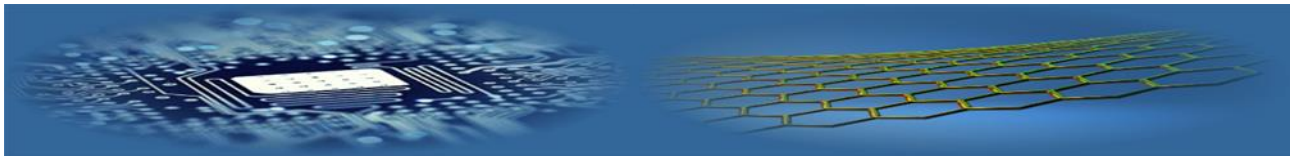


Yuri Shunin

Igor Kabashkin

♠ **Rabindranath Tagore (7 May 1861 – 7 August 1941),** was a Bengali poet, novelist, musician, painter and playwright who reshaped Bengali literature and music. As author of Gitanjali with its "profoundly sensitive, fresh and beautiful verse", he was the first non-European and the only Indian to be awarded the Nobel Prize for Literature in 1913. His poetry in translation was viewed as spiritual, and this together with his mesmerizing persona gave him a prophet-like aura in the west. His "elegant prose and magical poetry" still remain largely unknown outside the confines of Bengal.

CONTENT



Displaying formulas as embedded calculations in scientific literature, textbooks and educational web apps

Valery Ochkov[1], Evgeny Nikulchev[2*], Sasha Gurke[3]

[1]*National Research University "MPEI", Moscow, Russia 111250*
[2]*Moscow Technological Institute, Moscow, Russia 119334*
[3]*Elsevier, 230 Park Ave., New York, USA 10169*

**Corresponding author's e-mail: nikulchev@mail.ru*

Received 20 March 2017, www.cmnt.lv

Abstract

Mathematical expressions can be added or edited in electronic media as embedded calculations. Formulas in an electronic document are not just expressions written in a formal language, they are also an object that could be studied, helping to comprehend a text and avoid many typos and mistakes typical for mathematical formulas found in print. This article describes some features of formulas embedded in online scientific literature and educational web apps. The article is based on the experience of teaching of engineering subjects and math assisted by the visualization of calculations. The authors examine key features and practical application of calculations and formulas embedded into the scientific and educational texts.

Keywords

embedded computing,
electronic publications,
computational methods,
Mathcad,
online publications,
virtual laboratory,
Knovel Interactive Equations

1 Introduction

The introduction should briefly place the study in a broad context and highlight why it is important. It should define the purpose of the work and its significance. The current state of the research field should be reviewed carefully and key publications should be cited. Please highlight controversial and diverging hypotheses when necessary. Finally, briefly mention the main aim of the work and highlight the main conclusions. As far as possible, please keep the introduction comprehensible to scientists outside your particular field of research. References should be numbered in order of appearance and indicated by a numeral or numerals in square brackets, e.g., [1] or [2, 3], or [4–6]; see the end of the document for further details on references.

A lot of changes are taking place in educational and scientific content delivery. Traditional print is augmented and often supplanted by electronic materials with embedded features such as web links, bookmarks, comments, references to forums, and video clips or animations [1].

Embedded calculations play an important role among these features, enhancing comprehension of the math while enabling live computation [2].

There is a wide variety of software capable of supporting embedded calculations. For example, the symbolic math engine Maple has found a wide use in academia for differentiation and integration. The popular programming language Python is used for realization of algorithms in TeX documents [3]. Math formulas on HTML pages can be supported by Mathcad [4]. Cloud computing applications are also becoming more widespread [5].

Regardless of the technology is used, modern electronic textbook is not just a web-enabled text. It is also a self-contained learning tool [6]. It helps to comprehend math better, allows calculation and visualization of solutions in the form of graphs, and can be used for in-text exercises.

If there is an equation in a book, the reader should be able to use it for calculation right there. Additionally, readers should be able to change the variables in the equation. This also applies to tables and graphs containing functional dependencies [7].

The following are some observations from our pedagogical practice that illustrate the usefulness of embedded equations in education.

1. In a physics department, students can conduct lab experiments, and then use a computer to access a web page with live calculations, enter data points, and fit the curve using least squares method. This used to be the normal workflow. Now, an increasing number of students use smartphones and tablets. When teacher asked why they don't use the lab computer a student often replies: "I hardly know how to work with a standard keyboard and can input data faster on a tablet."

2. During math, physics, and chemistry classes students often use their mobile devices (smartphones, tablets, notebooks) to go to a math site and key in an algebraic or differential equation (see Fig. 2) to find an analytic or numeric solution that validates calculations written by the lecturer on the blackboard.

3. Finally, during a physics class, a student derived an equation describing string vibration and then created an animation for its visualization.

There is a new trend in the content and production of scientific papers, handbooks, monographs, and textbooks containing calculations. In the old days, authors submitted their work to publishers in the form of typed manuscripts with hard-written formulas and special characters. Now, publishers require electronic files with text in MS Word and

formulas created in a math editor or native software, scanned and inserted as images into text.

Formulas in scientific papers and textbooks are there for computing. In electronic formats, they can be used for computing right on the screen. Formulas that appear in print can be easily validated by using interactive electronic version. This is a revolutionary change!

Production of electronic handbooks with embedded interactive equations is described in the [8]. Making equations interactive exposes many errors and typos in the formulas. For example, multiple typos and errors were found during development of a series of web-based handbooks called "Heat Technology and Power Engineering" and corrected for the subsequent edition.

During peer review of scientific papers, we are often asked by editors to validate numeric solutions. While the damage caused by typos in regular text is negligible, those in formulas could lead to serious consequences. For example, while reviewing an article on the thermal conductivity of aqueous solution of NaCl, calculations were validated using Mathcad Calculation Server. This revealed that some equations and coefficients in the article had typos. When typos were corrected, the article was published error free and with a reference to "live" computing.

In this article, we will examine various aspects of computing with embedded formulas, and discuss ways of improving the quality and precision of solutions by using validated formulas.

2 Specifics of calculations using text-embedded formulas

Modern calculation apps such as Mathcad can handle the units of measurement. This is very convenient due to automatic conversion of input and output values with different units, and control over their dimensional compatibility. However, there are some problems.

There are two types of scientific equations – physical and empirical. The former deal with physical values and can be used with any appropriate units of measurement. Good examples of these formulas are $E = m \cdot c^2$ and $F = m \cdot a$. Calculation apps would solve these equations correctly regardless of the units as long as they have the right dimension and would throw an error when dimensions are not compatible. Empirical equations, created, as a rule, by statistical treatment of experimental (empirical) data, require numeric input in the stated units of measurement. Probably the simplest and best known empirical formula is a formula linking the height of an average man with his weight [4]: height of a person in centimeters is his weight in kilograms plus 100. This formula, written in Mathcad, and the units of measurement are shown in Fig. 1. The results in Fig. 1 are rounded, although variables in equations contain 15 digits in mantissa. This detail is an important aspect of the creation of "live" equations. In paper handbooks, monographs and textbooks, calculation of the systems of equations is done with the rounding of intermediate results and then using these results in other equations. In computer calculations, variables retain maximum precision. The result of these calculations would be slightly different from that in paper publications.

In handbooks, basic units of measurement are shown in the legends for physical formulas. Machine computing of these formulas is done without using the units. Here is a simple example: $F = a \cdot m$, where F is force in N (newton), a is acceleration in m/s^2, and m is mass in kg. For modern calculation applications input of the units is redundant because all appropriate conversions for the derived units are done automatically.

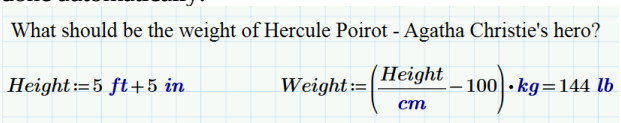


FIGURE 1 Input of a simple empirical equation

One shouldn't forget about physical problems with solutions that contain numbers beyond the range of computer calculations. They could have both, very large and very small parameters, resulting in wrong rounding due to simultaneous processing of large and small numbers. Also, solution algorithms for math models could become unsuitable if model equations contain variables that differ by the powers of magnitude. In this case, the errors for the large parameters could be on the same scale as parameters themselves, and could significantly distort the values for the small parameters. Before solving these equations, it is advisable to modify them to reduce the spread between the magnitudes of parameters and to reduce the number of arithmetic operations leading to accumulation of errors.

A solution is to replace the variables with dimensionless quantities prior to calculation, making all variables in a math model practically the same magnitude. The equations themselves become, as a rule, simpler due to elimination of "extraneous" constants and quantities and the real parameters of the process being modeled become exposed. The equations are made unit-independent by converting absolute values of some properties to relative, specific for a particular physical model. Usually, these are length, speed, time, and energy-related or mechanical properties.

Many physical constants are dimensionless by definition. These include, for example, the Mach number (compressibility factor), Reynolds number (viscosity factor), Poisson's ratio (a parameter of material elasticity), the Prandtl number (reflecting relationship between viscosity and thermal conductivity), etc. The angle measured in radians is a dimensionless quantity as well, and is calculated as a ratio between the lengths of an arc and a radius of a circle.

Let's use oscillation of a weight attached to a spring to illustrate how an equation can be made dimensionless. Suppose for the time t0 the initial extension of the spring is x_0 (m), k (N/m), v_0 (Hz) and m (kg). The characteristic time would have the dimension of $\sqrt{m/k}$ and, if the spring is stretched instantaneously, the interval of calculations would be so small it would be comparable to computational error, and representative values of x could be so large as to lead to a loss in precision. Let's make the quantities dimensionless by introducing new variables $\tilde{x} = x/x_0$, $\tilde{t} = t\sqrt{k/m}$, $\tilde{v} = \left(v\sqrt{m}\right)/\left(kx_0\right)$. The system of equations with new variables can be expressed as $\begin{cases} \dot{\tilde{x}} = \tilde{v} \\ \dot{\tilde{v}} = -\tilde{x} \end{cases}$.

Here, representative values of all the variables are comparable and, therefore, the system of equations can be solved numerically with required precision. In addition, by solving this system once, we can obtain a solution for the whole range of problems for different values of x_0, k and m. To get back the units, we need to reverse substitute the variables.

Getting rid of the units of measurement involves transitioning from the absolute physical quantities to the relative quantities specific for a given physical process. This transition is governed by the theory of similarity. For example, transition from the laminar flow to turbulent occurs at a certain value of Reynolds number, depending on the flow velocity, geometry and medium viscosity.

Making calculated quantities dimensionless opens a possibility of adding Reynolds and Prandtl numbers without getting a program error. Adding viscosity to thermal conductivity is not possible or, more precisely, it is not possible using software which can control units. This and other reasons lead to the return to dimensional physical quantities and rejection of their "shadows" – dimensionless quantities – the obvious drawbacks of which were felt by people who have provided them with pseudo-units such as degrees. Modern software has the means to remedy this situation.

Another obvious case of dimensionless quantity is the use of decibels. Bel is a common logarithm of a ratio of two unidimensional physical quantities, and a decibel is, of course, one tenth of a bel. By measuring a property in decibels we are actually converting a physical quantity into a dimensionless value on a logarithmic scale. Usually, decibels are used for measuring the volume of sound by comparing the intensity of sound waves I with the lowest level of sound perceivable by a human ear I_0 (typically $I_0 = 0.01$ W/m2). We can calculate directly the ratio I/I_0 but the range of this ratio is very wide making its use inconvenient. This is resolved by using logarithmic scale and the prefix deci $-0.1 \cdot \log(I/I_0)$. A very loud noise of a sledge hammer has a volume of 80 dB, whereas a conversation in a room –60 dB and a barely audible sound such as rustling of leaves is approximately 10 dB. Decibels, as well as other dimensionless quantities, would make calculation easier but, again, not if we use a computer.

Some time ago, during transition to the SI system of units, STM publishers and journal editors stopped accepting manuscripts that used other, still common units of measurement. This transition was painful for engineers and scientists who used foot, pound and other non-metric units. In spite of the prevalence of SI units, there are some customary, non-metric units that are still being used in many countries. For example, we will never learn to measure atmospheric pressure in pascals, or more precisely in hectopascals, preferring millimeters of mercury. Now, publishers are beginning to deviate from the rules requiring the use of SI units in articles and books. One of the reasons for that is unit-enabled calculation software.

Some equations in journals and books could be called pseudo-empirical. These are basically physical equations that in reality are empirical because they require units of measurement for computation (see Fig. 1). The following is an example from power engineering. At a given efficiency (η) of a power plant, the specific fuel consumption (b) can be calculated as $b = 123/\eta$. This means that a power plant with efficiency 34% consumes 362 (123/0.34) grams of fuel to generate one kilowatt-hour of electricity. This simple equation would result in an incomplete solution in Mathcad, i.e. a solution without the units (see the first line in calculations shown in Fig. 4). So, what is the problem? The reason is that, in the past, this equation was adopted for manual calculation and for calculation with software that could not handle the units (such as spreadsheet). However, if we remember what is equivalent fuel and its calorific value (CV), we can restore the original equation $b = 1/(CV \cdot \eta)$ and use it without any problems in Mathcad (see Fig. 4). The result will have the units and, in addition, will be more accurate. A power engineer using the original equation would find out that her plant consumes almost half a gram less fuel. On a large scale it represents a huge energy savings.

The case shown in Fig. 2 is quite common. The lack of units in the result shown in the first line in Fig. 4 points to a pseudo-empirical nature of this equation.

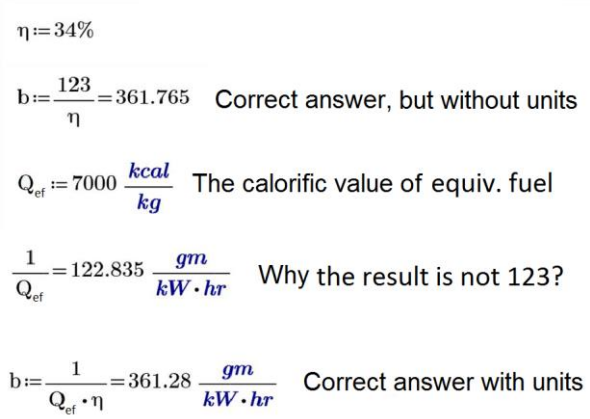


FIGURE 2 Working with pseudo-empirical equations

In the next example, the pseudo-empirical nature of the formula is not so apparent. This formula is used to convert concentration from one unit to another. When taken from a handbook and input into Mathcad, it produces an answer with correct units of measurement and, incredibly, with plausible results. The goal is to convert molarity (ratio of the solute to the volume of solution) of an aqueous NaCl solution to its molality (ratio of the solute to the mass of solution). The formula for this conversion can be easily found in a number of chemical handbooks and on the web (see Fig. 3).

http://erichware.info/sposob/ximia/ximikonc.htm

	0 < K < 100 % mass ratio	T titer	L molality	M molarity
M =	$\frac{10 * q * K}{wp}$	$\frac{1000 * T}{wp}$	$\frac{1000 * q * L}{1000 + wp * L}$	$\frac{1000 * yp}{v}$
L =	$\frac{1000 * K}{wp * (100 - K)}$	$\frac{1000 * T}{wp * (q - T)}$	$\frac{1000 * yp}{mb}$	$\frac{1000 * M}{1000 * q - wp * M}$
T =	$\frac{q * K}{100}$	$\frac{mp}{v}$	$\frac{q * wp * L}{1000 + wp * L}$	$\frac{wp * M}{1000}$
K =	$\frac{mp * 100\ \%}{mb + mp}$	$\frac{100 * T}{q}$	$\frac{100 * wp * L}{1000 + wp * L}$	$\frac{wp * M}{10 * q}$

wp – molar mass of solute (g/mole) mb – mass of solvent (g)
mp – mass of solute (g) yp – moles of solute (mole)
v – volume of solution (cm³) q – density of solution (g/cm³)

FIGURE 3 Equations for calculation of different types of concentration

We can copy an equation for calculation of molality (L) based on molarity (M) from the table shown in Fig. 3. To calculate, we need to input additional values: molecular mass of NaCl (mp) and solution density (q). Copy, input and

get … an incorrect answer (see the 2nd line of the calculations in Fig. 4).

In the table shown in Fig. 3, the variable K is defined as a mass percent without disclosing whether it is a mass of the solution or a mass of the solvent. An analysis of this table reveals that it is the mass of solution and not the solvent. However, other cases are not so apparent, leading to calculation errors. For example, in many analytical chemistry handbooks, solubility in water is given as a ratio of the mass of solute to the mass of solvent by default, without an explanation. On http://twt.mpei.ac.ru/MCS/Worksheets/Thermal/T90-T68.xmcd this quantity is provided in two units of measurement (ratio to the mass of solvent and to the mass of solution) to prevent calculation errors. There are lots of these "defaults". Take a temperature for example. Here is a typical problem: the temperature at the inlet of a heat exchanger is given as 25°C, find the temperature at the outlet if the temperature of the heat-transfer fluid in the heat exchanger increases by 5°C. The answer is 30°C. However, if this problem is input into Mathcad as is, the answer will be paradoxically 25°C + 5°C = 303.15°C. The answer can be explained by remembering that 5°C equal 5K but 25°C is equal 298.15K. Most users understand this default (Celsius scale and degree centigrade), found in many handbooks, but computers don't. Also, in these calculations, users should account for the year of the temperature scale, 1968 or 1990, in which the temperature is given.

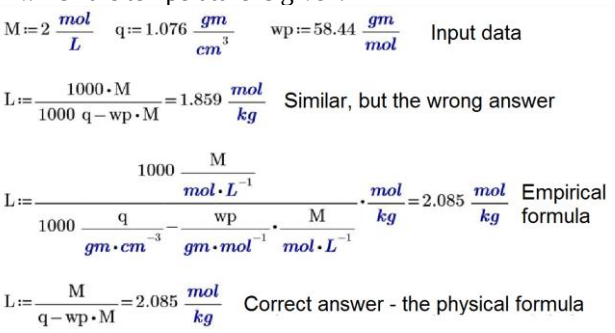


FIGURE 4 Working with a pseudo-empirical formula for conversion of different units of concentration

The problem here is that the equations in Fig. 3 have been adopted for the convenience of manual computation by using non-basic ("chemical") units of measurement: concentration, density (g/cm3 instead of kg/m3), molar mass (g/mol vs. kg/mol), mass (g vs. kg), volume (cm3 vs. dm3), etc. The formula for computation of molality from molarity can be, of course, used as empirical (see above) by adding required units of measurement (see 3rd line in Fig. 6) to obtain the correct result. However, it is better to go back to the initial physical formula by removing the coefficients (1000 – see the last line in Fig. 4), realizing that 1000 is the number of grams in a kilogram and the number of centimeters cube in a liter, etc.

As a result, most practicing chemists, when they need to convert concentration from one unit to another, try to avoid the existing formulas like those shown in the table in Fig. 3. Instead, they calculate using ratios. These ratios are not required, however, if you have a nearby computer with Mathcad or have an access to a website with appropriate application. There you can input and solve an algebraic equation linking, for example, the amount and mass of the solute for different units of concentration. Mathcad has a symbolic math engine that allows computing with the symbols of variables instead of their numerical values. Mathcad calculations with formulas for conversion of solution concentration from one unit to other using algebraic equations are shown in Fig. 5. This is a true improvement on handbook formulas. One can see both the formulas used for calculation and the corresponding physical law-based equations from which they are derived!

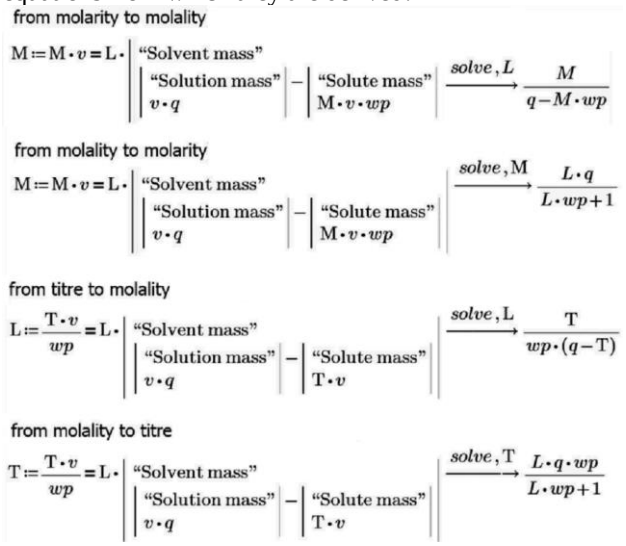


FIGURE 5 Generation of formulas for conversion of different types of concentration units

An example of solution for conversion of molality to molarity using formula, modified as shown in Fig. 5 on the basis of the physical law of conservation of mass, is provided in Fig. 6. Operator clearsym that is in the collapsible area is used to hash numeric values of variables during symbolic transformations.

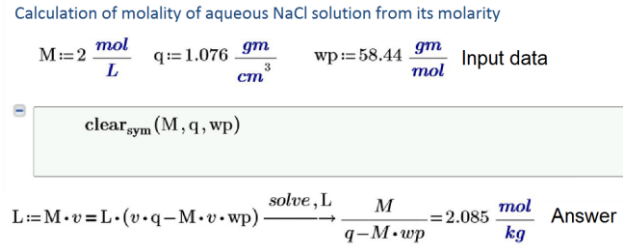


FIGURE 6 Computation of molality using modified formula

Fig. 7 shows a website for recalculation of concentrations using the above formula. This site was created some time ago and should be updated to show not just the formula itself but also the equation from which it was derived (see Fig. 6).

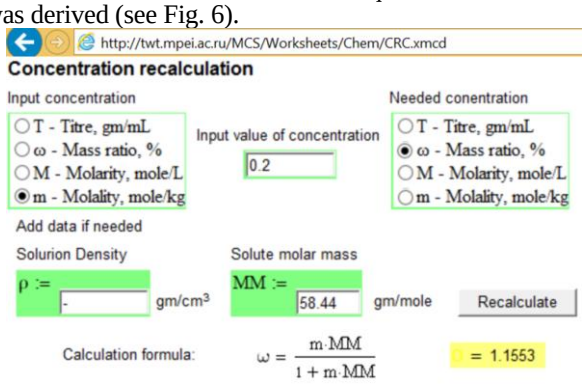


FIGURE 7 Website for recalculation of different types of concentrations

Similar results (Fig. 8) can be obtained by imputing relevant formulas online using Elsevier's Knovel Interactive Equations (https://app.knovel.com/ie/#welcome).

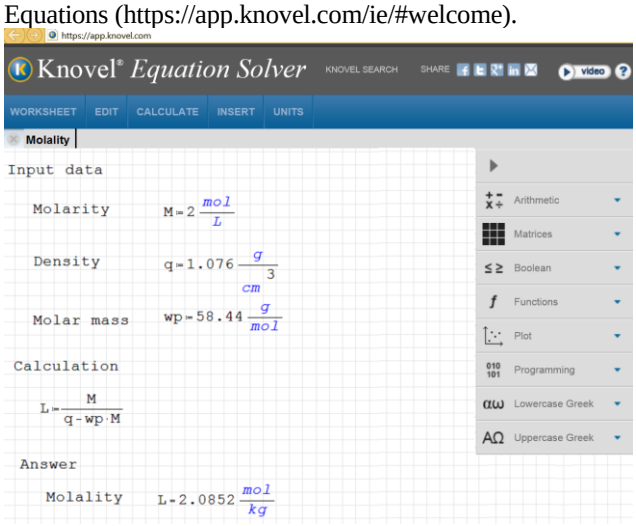


FIGURE 8 Working with Knovel Interactive Equations

Knovel Interactive Equations are based on a proprietary web-enabled math engine that supports a rapidly growing collection of several hundred validated equations and working examples in several subject areas, including Chemistry and Chemical Engineering, Electronics and Semiconductors, General Engineering, Mechanics and Mechanical Engineering, Metals and Metallurgy and Oil and Gas. Users can browse or search the collection, use the built-in Equation Solver to calculate, and export calculations for reports or sharing knowledge. They can also create a worksheet from scratch by combining text, math, images, and plots. The application has easy one-click access to a toolbox containing math functions, engineering units, programming structures, and math symbols used in engineering formulas.

Reference literature contains a huge number of pseudo-empirical formulas similar to those shown in the figures above. Using these formulas without proper adjustment can lead to computational errors but simplification of these (see Fig. 2) or their modification so that more convenient units can be used is no longer required in the computer age.

3 Conclusion

The website www.trie.ru, with interactive calculations developed and has been in operation for almost 10 years. During this time, >20,000 interactive calculations have been posted on this site, mostly related to power engineering, energy conservation, energy efficiency, structural materials, and properties of heat-transfer fluids. In addition, there are calculations in the areas of general science and engineering and educational subjects such as math, physics, chemistry (see above), heat exchange, thermodynamics, gas and hydrodynamics, electrical engineering, theoretical mechanics and other academic disciplines. Calculations are augmented with functions that can be referenced. These functions, when referenced, become visible in documents containing calculations. Many formulas are duplicated as text that can be easily copied and pasted in spreadsheets and programming code.

The technique of "live calculations" developed by the authors is currently used by the International Association for the Properties of Water and Steam (IAPWS) on its website www.iapws.org. This site has official formulations for water and steam in static PDF and, now, in "live" calculation formats.

We currently collaborate with Elsevier on Knovel Interactive Equations.

Calculation server that supports www.trie.ru is being developed at the Moscow Power Engineering Institute within the framework of the national research program.

Acknowledgments

Work performed partly funded by the Russian Foundation for Basic Research (grant number 16-08-01222).

References

[1] Zeidmane A 2013 Development of Mathematics Competences in Higher Education Institutions *International Journal of Engineering Pedagogy (iJEP)* **3** 11-7

[2] Booth S 2004 Learning and Teaching for Understanding Mathematics *Proc. of the 12 th SEFI MWG Seminar* Eds Demlova M and Lawson D 12–25

[3] Poore G M 2013 Reproducible Documents with PythonTeX *Proc. of the 12th Python in Science Conference* Ed. van der Walt S

[4] Ochkov V F, Orlov K A, Ko C K 2014 "Cloud" functions and templates of engineering calculations for nuclear power plants *Thermal Engineering* **6** 760–4

[5] Nikulchev E, Pluzhnik E, Biryukov D, Lukyanchikov O 2015 Designing applications in a hybrid cloud *Contemporary Engineering Sciences* **8** 963–70

[6] Flegg J, Mallet D, Lupton M 2012 Students' perceptions of the relevance of mathematics in engineering *International Journal of Mathematical Education in Science and Technology* **43** 717–32

[7] Ochkov V F, Ustyuzhanin E E, Ko C K, Shishakov V V 2015 Thermophysical databases: From tables to interactive internet resources and cloud templates *High Temperature* **53** 515–20

[8] Ochkov V, Orlov K, Voloshchuk V 2016 *Thermal Engineering Studies with Excel, Mathcad and Internet* Springer

AUTHORS

Valery Ochkov

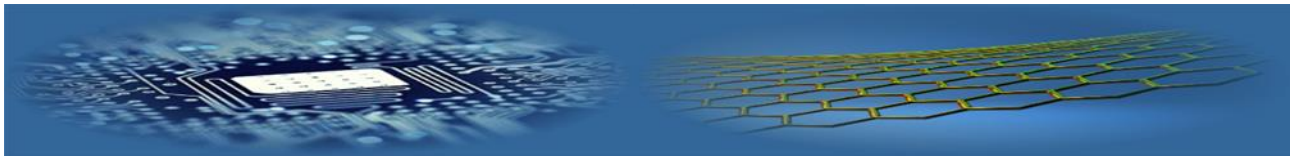
Current position, grades: Professor at the Department of Thermal Power Engineering of the National Research University "Moscow Power Engineering Institute" (Technical University)
University studies: Ph.D., Sc. D.
Publications: He is the author of over 200 regular and review papers in international scientific journals. He is the author of over 25 books including 3 in English (Springer Publishing House)
Experience: Valery F. Ochkov teaches courses at the "Moscow Power Engineering Institute" on information technology and optimization of power generation processes. Prof. Ochkov is also a founder and general manager of Trieru, an engineering consulting firm that develops simulators and analytical software for the power industry. He is a member of the working group on thermodynamics of the International Association for the Properties of Water and Steam (IAPWS). He is an author of more than 20 books and numerous journal articles on the use of math software for solving the problems in thermodynamics, heat and mass transfer and fluid dynamics. Prof. Ochkov is one of the creators of WaterSteamProT, an application for simulating thermophysical properties of water and steam. For more information about Prof. Ochkov, please see //twt.mpei.ac.ru/ochkov/v_ochkov.htm.

Evgeny Nikulchev

Current position, grades: vice-rector and professor of Moscow Technological Institute
University studies: Dr of Sci. (Nonlinear Dynamics, 2006), PhD (Computer Science, 2000)
Scientific interest: big information systems design and management, management theory, scientific research management
Publications: More than 40 publications in the computing services, cloud computing, artificial intelligence.
Experience: It is artwork in the development of computing services for software applications for different purposes. He was Chairman and member of the scientific committees of international conferences on information systems, e-learning; chief Editor of the journal Cloud of Science, guest editor of several journals.

Sasha Gurke

Current position, grades: Technical Engineering Fellow, Elsevier (USA)
University studies: Master degree in Chemical Technology (1972), St. Petersburg State University of Technology and Design, Russia
Scientific interests: engineering data normalization, calculations and taxonomies; publishing interactive content online
Publications: several papers and presentations
Experience: VP Product Development, SVP and Co-Founder of knovel (www.knovel.com); Editor at Chemical Abstracts Service (ACS)



Development of methods for determining the tracking software systems accuracy, application of a wireless communication device and self-contained power supply in heliostat units with centralized monitoring and control system

F A Satybaldiyeva[1*], D Beyer[2], A S Sarybaev[3]

[1]*Kazakh National Research Technical University named after K.I.Satpayev*
[2]*University of Applied Sciences Schmalkalden*
[3]*M. Auezov South Kazakhstan State University*

**Corresponding author's e-mail: feruza201200@gmail.com*

Received 20 March 2017, www.cmnt.lv

Abstract

This article deals with determination of accuracy in tracking software systems, and advantages of the heliostat automatic control system that utilizes measuring informational and control systems comprising wireless measuring instruments and information-processing equipment. The heliostat automatic control system covered in the article is one of the types of measuring informational and control systems. As distinct from other measuring informational systems, the described heliostat control system operates only when tracking parameters deviate towards the maximum permitted values.

Use of wireless communication between detectors, transducers and industrial logical controllers in modern optical SPS heliostat control systems is more advantageous than laying hundreds meters of cable.

To provide power supply, it is proposed to equip each heliostat with a self-contained power supply, since heliostat operates when concentrated solar radiation in the receiver is sufficient for steam generation, while the rest of the time it is in the standby mode. That is why use of a solar battery-powered self-contained power supply is more advantageous than use of centralized power supply from the industrial network.

Keywords

mathematical model,
automatic control systems,
heliostat,
wireless communication,
power supply

1 Introduction

The objective of accuracy determination in tracking software systems (TSS), just as in optical tracking systems (OTS), is to determine the misalignment angle between the Sun and the optical axis in concentrators or the set direction of reflection in heliostats.

In OTS, this angle is directly recorded by an optical detector (OD) that sends the signal for misalignment angle handling, with OTS also performing drive errors and rotational axes positions compensation due to available feedback. That is, when tracking using OTS, there are practically no requirements for drive errors and rotational axes positions. TSS operates as follows. The control program calculates the Sun's angular position in a coordinate system associated with a concentrator (or heliostat) — the Sun's angular coordinates are determined, based on which the software determines the angles of heliostat turning (the angles of concentrator turning are equal to the Sun's angular coordinates in one CS), the drives receive a signal for handling of these angles. That is, TSS cannot "see" the Sun, and thus there is no compensation of drive errors and rotational axes and, accordingly, control algorithm errors. But here we also obviously need control of concentrator turning angles, i.e. feedback is required as well. This task was considered and solved in the course of development of the Big Solar Furnace (BSF) heliostat software control system; the azimuthal and zenithal drive reducers were equipped with angular sensors at a pitch of about 36 arcsec. These sensors controlled the angles of heliostat turning in terms of azimuth and zenith. Field research of the sensors operation was performed. The experimental design was as follows: heliostat was put into the tracking mode using an optical sensor, while turning angles were determined by an angular sensor and further compared in time with the designed angles of turning. Assessment of actual angle deviations of heliostat turning angles in terms of azimuth and zenith against their designed values was carried out on the basis of the data obtained. It was established that difference between the angles does not exceed 1-2 arcmin, and the conclusion was made that it proves accuracy of the software tracking. It was also stated, that the BSF heliostat tracking systems operate practically in the on-off mode: "start-stop-start". It should be noted that characteristics of the misalignment angle sensor itself (angular value of dead spot) were not investigated.

2 Theory and calculations

As is known, solar cell concentrators (heliostats) must turn (keep track) of the Sun with a certain level of accuracy due to its apparent movement.

At that, along with concentrator's accuracy

characteristics, tracking accuracy is one of the factors influencing the solar radiation concentration, i.e. irradiance variation in the focus and mean concentrations at the receiver. Tracking accuracy, as was determined above, is clearly defined by a defocusing angle β or by an angular deviation of axial solar beams from either optical axis of the concentrator (the concentrator mode) or predetermined direction of solar beam reflection from the heliostat (the heliostat mode). At the initial approximation, the defocusing leads to a shift of a concentrated spot at the receiver. The methodical task setting (concepts of static and dynamic defocusing modes), as well as experimental and calculation studies of the influence of defocusing, virtually the only ones to the present day, were conducted in [2-5]. In these studies, allowed values for defocusing angles β were identified for energy concentrators. It was found that β must not exceed 16 *arcmin* at the acceptable flow reduction at the receiver by 10% for concentrators of solar power plants. The issue of permissible defocusing angles also becomes relevant due to problems of development of software systems for concentrators (heliostats) to keep track of the Sun. Acceptable values of defocusing angles obtained in [2-5] generally allow to make the conclusion on possible requirements for the tracking accuracy, although during the design of tracking systems it is still assumed that the tracking accuracy must not exceed 1 *arcmin*.

Flow densities at the receiver's surface elements and subsequently flows will be determined using the model suggested by Grilikhes V.A. [1]. This model assumes that even if beams are reflected from a non-precise concentrator, the angular dimension of the reflected solar cone will not change; and that concentrator inaccuracies are distributed according to some randomly set rule with the standard angular deviation σ of normals in relation to normals of a precise concentrator. In practice, this integral is defined numerically and actually is replaced by a sum of the following type. For example, when determining the flow on the area S_R of the receiver, it will be as follows:

$$E_A = \sum dS \sum_{S_R} \sum_{S_C} B(\mathbf{a}) * (\mathbf{n_M} * \mathbf{a}) * (\mathbf{n_A} * \mathbf{a}) * d\omega, \quad (1)$$

where $B(\boldsymbol{a})$ is the brightness of the reflected solar beam from the M point (the surface element of the concentrator in the direction of A point of the receiver (the unit vector $\boldsymbol{a}$); $\boldsymbol{n_M}$ and $\boldsymbol{n_A}$ are M and A area normals; $d\omega$ is the elementary solid angle with a vertex in the A point; S_R, S_C are surface areas of the receiver and the concentrator.

As stated in [2], the surface area of the concentrator S_C and its geometry, specular reflection factor (R_z), coordinates of M and A points, orientation of normal $\boldsymbol{n_A}$ in the center of the receiver's surface element dS_A, and incident irradiation parameters (for the Sun, it is usually a vector direction of the axial solar beam **c**, its angular radius φ_0 and the angular distribution of brightness $f(\varphi,\varphi_0)$ across the solar disk) shall be set in order to determine components of the irradiance integral. Based on these formulas, we developed an algorithm as well as software in C++ for calculation of both the irradiance and the flow from a paraboloid concentrator based on a flat receiver surface, taking into account the possibility to set the concentrator inaccuracies and the defocusing angle in the software.

The following sun spot radius r_P of the precise paraboloid will be used as the size scale, as well as for the purpose of results summarization [3]:

$$r_P = p*\varphi_0/[(1+\cos U_0)*\cos U_0], \quad (2)$$

where p is the focal paraboloid parameter ($p = 2f$, f is focal length); φ_0 is the angular radius of the solar disk; U_0 is concentrator's opening angle valid for a circular concentrator and effective for other shapes [4].

Overall influence of the defocusing angle β on irradiance distribution in the focal plane of a precise paraboloid concentrator ($\sigma = 0$) on a concentrated spot is shown in Fig. 1, where for comparison you may also see irradiance distribution for a non-precise concentrator (σ = 8 arcmin) for the case of $\beta = 0$. As we can see, assumptions in [1, 10] were confirmed, as sufficiently large defocusing angles mainly resulted in the shifting of spot leaving the irradiance curve's shape unchanged.

Fig. 1 shows that due to the uneven distribution of a concentrated spot, the influence of defocusing may vary for receivers of different radii (different average concentration).

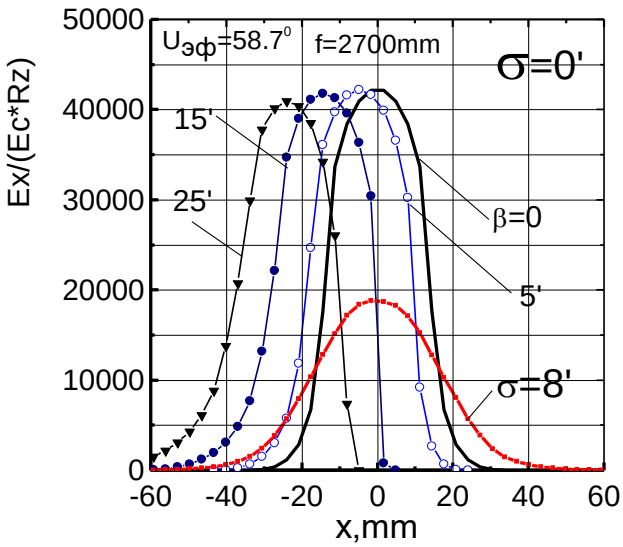


FIGURE 1 Irradiance distribution in the focal plane of a precise parabolic concentrator in the cross-section $y=0.05r_P$ (r_P = *31 mm*) at different defocusing angles β.

Thus, the following typical dimensions of the spot radii may be identified for the given concentrator with $U_0 = 58.7^0$ for the area radius r in fractions of the image spot radius of the precise concentrator r_P: *I* – the focal up to $r/r_P \approx 0.05$, where $C \approx 42{,}324$ (at $\sigma = 0$) and $C \approx 18{,}800$ (at $\sigma = 8$); II – the area of high mean concentration, up to $r/r_P \approx 0.25$ where $C \approx 40{,}440$ ($\sigma = 0$) and $C \approx 18{,}100$ ($\sigma = 8$); III – the border of sharp drop of the irradiance curve, up to $r/r_P \approx 0.35$ where $C \approx 38{,}600$ at $\sigma = 0$, while for the non-precise concentrator, it is up to $r/r_P \approx 0.7$ where $C \approx 12{,}200$ ($\sigma = 8$); IV – the border of a sloping part and the irradiance curve, the radius of which for

a precise concentrator is equal to $r/r_P \approx 0.65$ and C ≈ 20,500 ($\sigma = 0$), while for a non-precise concentrator it is $r/r_P \approx 1$ and C ≈ 8000 ($\sigma = 8$) . That is, even in the case of the non-precise paraboloid concentrator, the average concentrations are quite high within the spot. Moreover, which is important in practice, for the non-precise concentrator the main flow falls into the spot area of the precise concentrator, i.e. in the area of $r/r_P \approx 1$.

For the same concentrator Fig. 1 illustrates flow changes at these areas depending on the angle of defocusing β for both precise ($\sigma = 0$) and non-precise ($\sigma = 8'$) concentrators. To sum up the results, F_β flows at the receiver are shown in a relative form, as the fractions of flow incident on this area in the absence of defocusing $F_{\beta=0}$, i.e. when β=0. To allow determination of the $F_{\beta=0}$ flow fraction in the total flow, there are also $F_{\beta=0}$ values given in fractions of the total flow F_P [5] reflected from the concentrator.

Due to the fact that the flow change depending on β is quite minor, requirements for tracking inaccuracy α_C and its impact on reduction of flow $\Delta F/F_{\beta=0} = (1-F_\beta/F_{\beta=0})*100\%$ for the practical case — the non-precise concentrator with $\sigma = 8'$ — are presented in a tabular form (Table 1). Here we may clearly see the actual difference between the defocusing angle and tracking inaccuracies or the misalignment angle. Thus, in fact, to ensure the required defocusing angle, it is necessary that actuation angles α_a and α_h or projections of misalignment angle α at $\alpha_a = \alpha_h$ in the sensor planes were:

$$\alpha_a = \alpha_h = 0.7\alpha.. \quad (3)$$

TABLE 1 Flow reduction ($\Delta F/F_{\beta=0}$, %) at the receiver of r/r_P radius, depending on the misalignment angle α_C and tracking inaccuracy $\alpha_a = \alpha_h$ for a non-precise concentrator ($\sigma = 8'$).

$\alpha_C = \beta$, arcmin	$\alpha_a = \alpha_h$, arcmin	area radius, r/r_P			
		0.05	0.25	0.7	1.0
		$\Delta F/F_{\beta=0} = (1-F_\beta/F_{\beta=0})*100\%$			
0	0	0	0	0	0
1	0.7	0	-0.1	-0.1	-0.03
2	1.4	0.6	0.1	0.1	0.08
3	2.1	0.6	0.7	0.6	0.3
4	2.8	0.8	1.5	1.2	0.8
5	3.5	1.4	2.7	2.2	1.3
6	4.2	2.8	4.2	3.4	1.9
7	4.9	4.6	5.9	4.8	2.8
8	5.6	7	8.0	6.5	3.7
9	6.3	8.8	10.0	8.3	4.8
10	7	12	13	10,4	6.0

As Table 1 shows, in the case of tracking inaccuracy $\alpha_C = 2'$ ($\alpha_a = \alpha_h = 1.4'$), almost no reduction of flow and, consequently, concentration C occurs, and reduction of about 1% only occurs when $\alpha_C = 4'$ ($\alpha_a = \alpha_h = 2.8'$). Thus, tracking accuracy may be at the level of 4' ($\alpha_a = \alpha_h = 2.8'$) for solar furnaces, while for solar power plants tracking inaccuracy may be up to 7' ($\alpha_a = \alpha_h = 4.9'$) for the cases where the receiver radii (r/r_P) are at the level of 0.7-1 with acceptable flow reduction of 3%.

Having determined the software tracking accuracy in the Lab View virtual environment, we considered the model of application of a wireless communication device and self-contained power supply in heliostat units with centralized monitoring and control system [7, 11, 12].

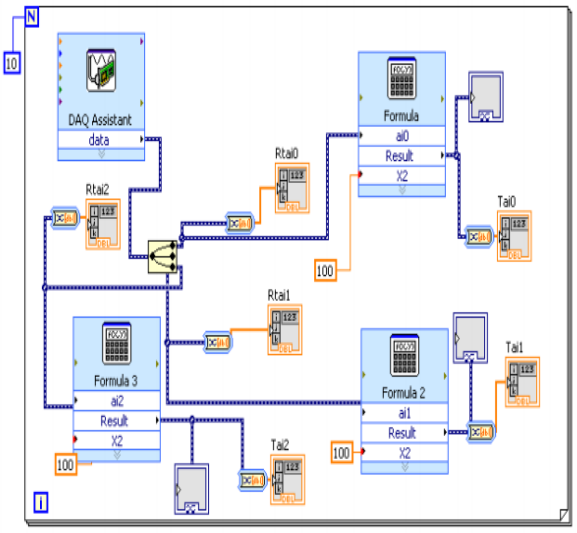


FIGURE 2 Creates a window of the diagram in Lab View

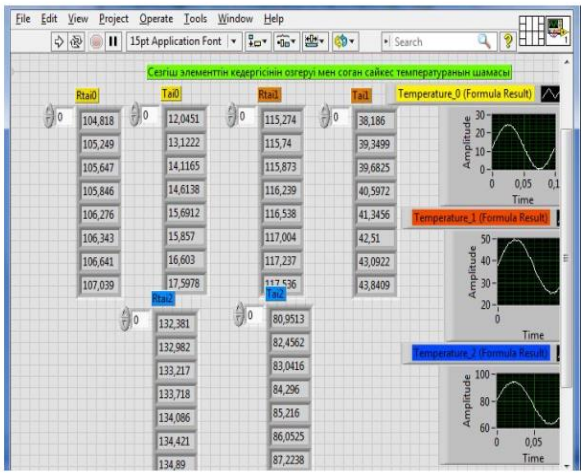


FIGURE 3 Meteorological parameters in LabView format opened with graphics editor National Instruments LabView

3 Experimental materials and methods

Currently, three types of heliostat automation systems are used, i.e.:

- the individual automatic control system;
- the centralized monitoring and control system;
- the combined monitoring and control system.

The operation concept of the individual automatic control system is that each heliostat operates individually, independently of each other, and each of them is equipped with tracking, positioning and orientation devices.

The centralized automation system monitors and controls the whole heliostat field from the operator's room.

The combined automation control system includes both of the above control systems [3, 17].

Let us consider the centralized control system for solar power station (SPS) automation. In addition to individual heliostat control system devices, the centralized control system includes centralized control devices with functions of feedback and control of each heliostat's orientation.

When controlling the SPS heliostat field, the control commands (certain codes of analog or digital signals) are sequentially transferred in a group or separately to each heliostat individually: start of operation, movement, stop, end of operation and reset. Control commands (electric signals) are transferred to heliostats via signal cables while power is transferred via power cables. The precise control of a corresponding heliostat is achieved by introduction of a specific system of signal coding and receiver addressing. Accuracy of heliostat control will be mainly determined by accuracy of automatic control system operation.

Figure 4 presents a block diagram of a centralized automatic monitoring and control system, which performs control from the primary system by means of continuous signals or impulses sent at specific time intervals according to the Sun's position on the dome of the sky. At that, the control commands must be generated (formed) in the primary control system.

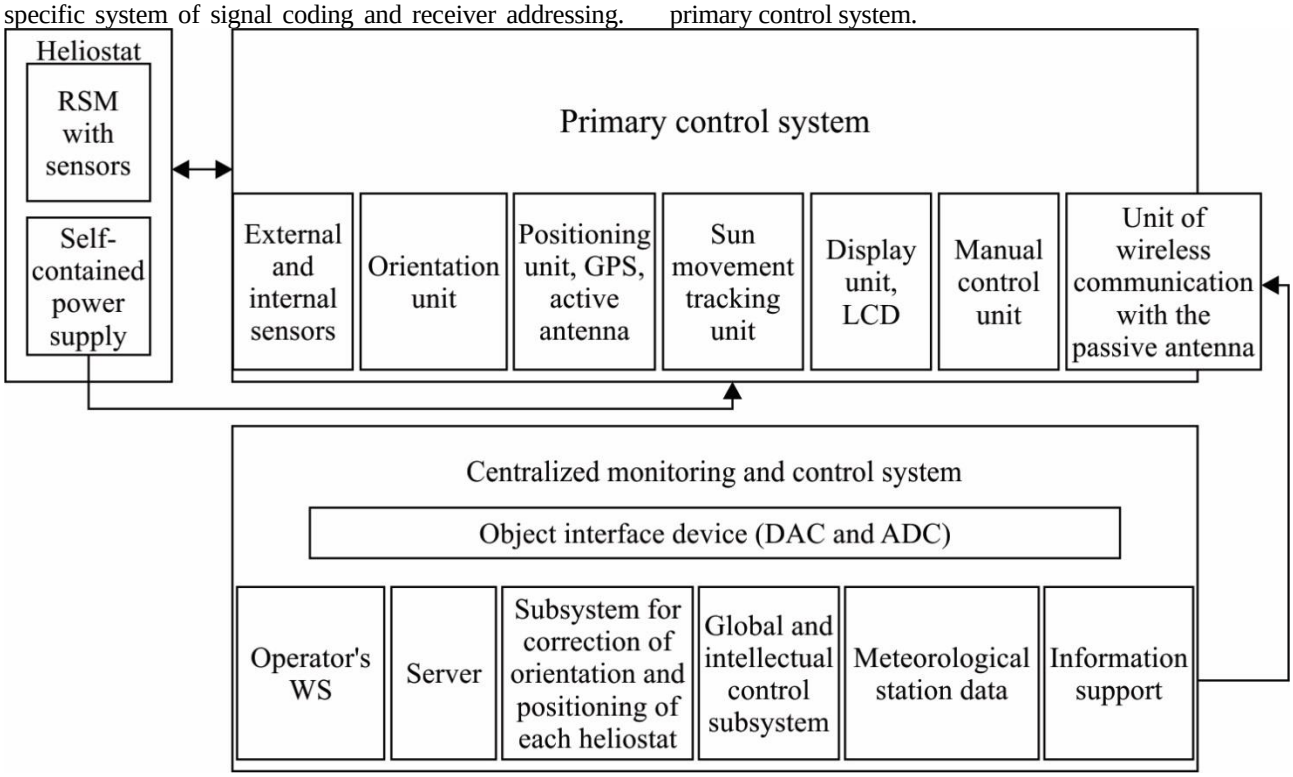


FIGURE 4 Centralized monitoring and control system

In case of use of the above automated control system (ACS) scheme with optical SPSs, each heliostat must be equipped with two electric drives with appropriate reduction gears to ensure zenithal and azimuthal turning and with four position sensors for precise pointing (on each heliostat) [1,8].

The considered automatic heliostat control system is one of types of measuring informational and control systems, which are complexes of measuring instruments and information-processing equipment. Their characteristic feature is that they are designed to obtain information on operating parameters values, which are characterized by the "pointed" or "misaligned" states, directly from the controlled (monitored) object. That is why such control systems must be functionally linked with the controlled object and must receive information directly from the object (from heliostat groups).

The matter of remote information transmission, i.e. communication channels, is of special importance for automated control systems on optical SPSs with a great number of heliostats.

The automatic control unit for all heliostats of an optical SPS system deals with signals transmitting information over a distance via wires (communication channels) [11, 14]. Their main, objectives are communication effectiveness and reliability, i.e. transmission of the greatest volume of information using the most economical way with the least distortion caused by various disturbances introduced by the communication channels themselves or due to other reasons.

For sequential control of turning of separate heliostat groups, the control signals must be transmitted via the communication channel in a specific sequence. This function may be performed by a multichannel system with time distribution of the channels. When building multichannel systems, the values of controlled variables are usually represented by a uniform parameter, such as DC voltage, resistance, etc. To transmit values via a communication channel, the uniform parameter is converted to an intermediate parameter, which ensures the least errors caused by instability of communication channel parameters or disturbances. At the receiving point, the intermediate parameter is converted to direct current or other signals allowing to register the values of controlled variables or reproduce them.

4 Results and analysis

Thus, analysis of design, technical parameters of heliostats on operating tower-type SPSs, as well as analysis of the composition, functions and operation of automation systems allow us to make the following conclusions [5, 9, 15]:

- land area utilization efficiency is not high, as heliostats are bulky and, accordingly, the area

occupied by a heliostat field is very large;

- the cost of a heliostat with the reflective surface area of 50 m^2, a tower with height of 80 m or more, equipped with a tracking system, position sensors, electronic positioning and orientation devices is relatively high;
- large overall dimensions and weight of heliostats complicate their control in terms of accuracy of sun rays pointing to a receiver (free play, deformation, etc.);
- due to the weight of mirrors, frame, rotary support mechanisms (RSM) and pillar of heliostat with reflective surface area of 50 m^2, the power consumption increases to 200 W/h;
- periodic switching on and off of motors of rotary support mechanisms is carried out by powerful, large, noncontact and contact electronic and electric power elements;
- electrical communication, signal transmission and exchange with heliostat's electronic devices and the automation system are carried out in the analog form, using a signal cable network;
- coordination of the industrial computer with controlled objects requires a great amount of DAC and ADC inputs and outputs;
- high power consumption by automation devices and the monitoring and control system.

In an optical system of tower-type SPS, the communication with the upper level of the automation system is carried out via a signal cable network. Heliostat power supply is arranged via a power cable network. Power and signal cables are laid in special ducts, trays and trenches. Currently, costs of cables and their laying are rather high, and they comprise 10% of the SPS's total cost.

With development of network and telecommunications technologies, wireless communication between detectors, transducers and industrial logical controllers is widely used in modern SCADAs. These devices are small and may be embedded directly into primary automation facilities. These devices are cheap and their application is more advantageous than laying hundreds meters of cable. That is why, to reduce costs and expenses on purchase and laying of signal cables, we propose to use cheap but effective wireless devices. Figure 5 below presents a schematic diagram of a receiver-transmitter developed by MikroElektronika on the basis of nRF24L01P microcircuit. As we can see, the diagram is simple, the additional elements connected to the microcircuit are required only for setting of its operating parameters.

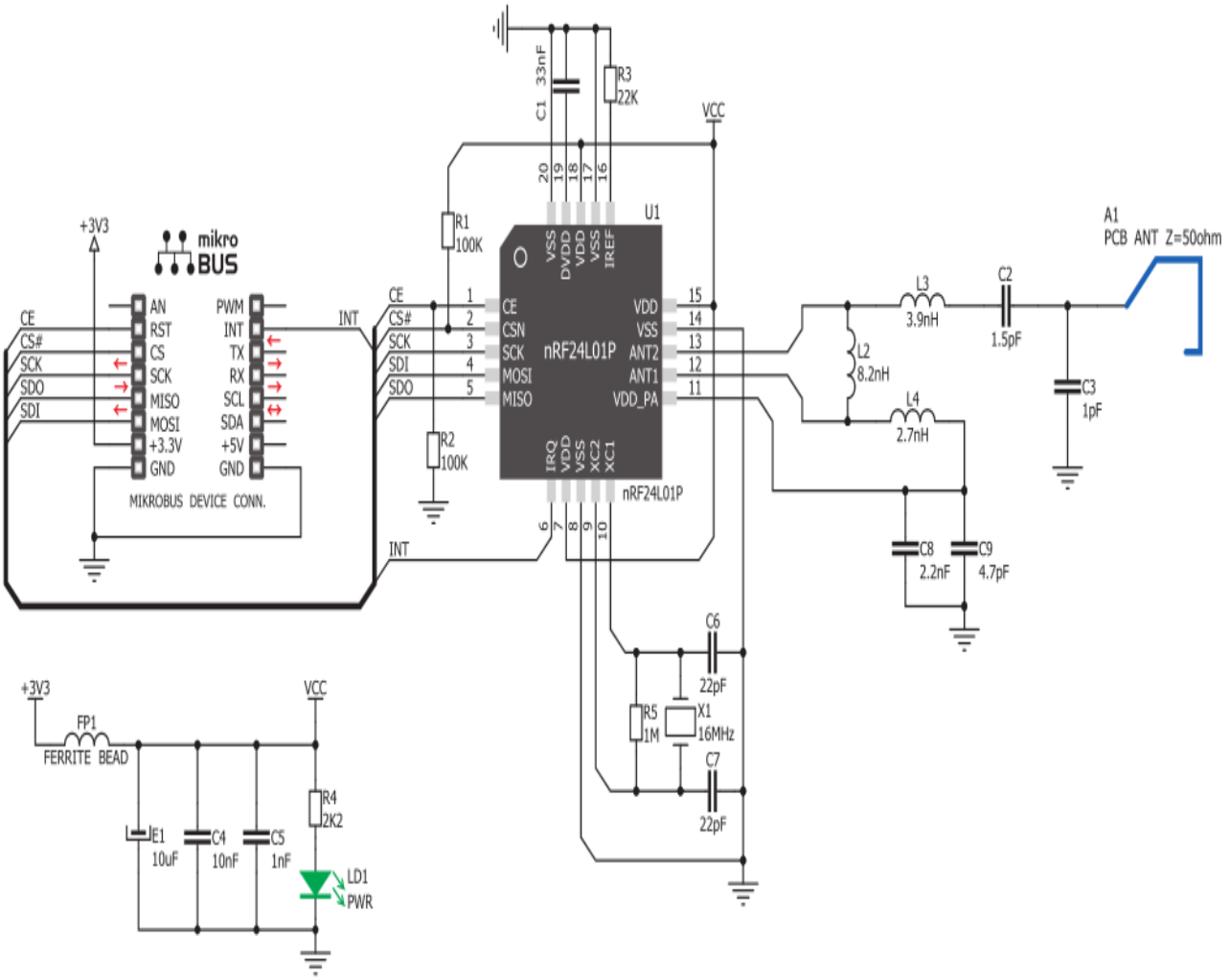


FIGURE 5 Schematic diagram of wireless device

This device can transmit both analog and discrete signals over a distance of up to two kilometers. The transmit and receive speed is 10 Mbit/s, the transmitter's operating frequency is 2.8 mHz, power supply is 5 V. The heliostat is equipped with RSM, electronic RSU motor control unit and wireless receiver and transmitter. To provide them with DC power, we propose to equip every heliostat with a self-contained power supply [11, 12, 15].

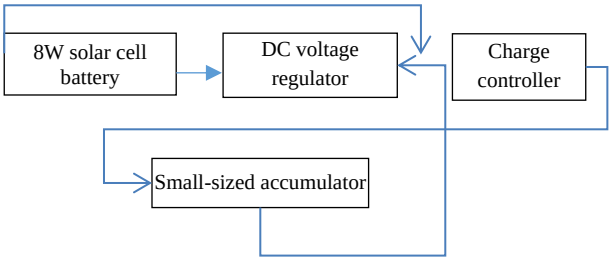


FIGURE 6 Block diagram of self-contained DC power supply composition

This power source (PS) consists of the following units presented in Figure 6.

This PS is cheap and small-sized. The solar panel is installed directly on the heliostat's reflecting surface. Heliostat operates only in weather when concentrated solar radiation in the receiver is sufficient for steam generation, while the rest of the time it is in the standby mode [15]. That is why use of a solar battery-powered self-contained power supply in our opinion is more advantageous than use of centralized power supply from the industrial network. It should be noted that this self-contained PS is designed to supply several adjacent small-sized heliostats, see Figure 7.

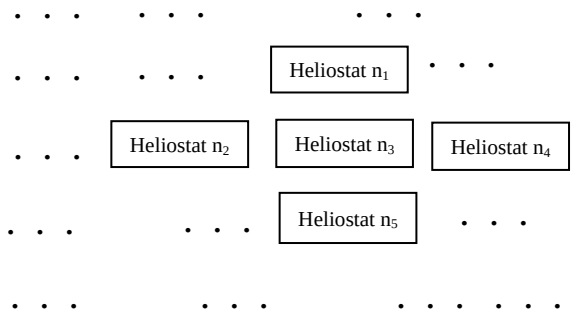


FIGURE 7 Heliostat field

For the purpose of saving the energy generated by the self-contained PS, control of n-group of heliostats and i-th heliostat in the n-group is carried out sequentially. For example, a self-contained PS provides power to five heliostats (Figure 7.). All the electronic control devices of the five heliostats are connected to the self-contained PS, but all RSM motors are disconnected from PS, and each RSM motor n_i sequentially connects to PS when performing functions of pointing, manual control, reset after maintenance or repair, setting to the next morning position and specific positioning in the period of extreme weather conditions. Thus, the monitoring and control methods fundamentally change when using a self-contained PS.

5 Conclusions

Tracking inaccuracy calculations and available conclusions given us the 15% efficiency factor based on the defined general relation between the defocusing angle β and misalignment angle α in measuring planes of the optical tracking sensor. [19] It was shown that for cases of both concentrator and heliostat modes of tracking, α is always equal to β, although the nature of their changing as well as their projections may vary, and that in practice the actual tracking inaccuracy in the sensor planes must not exceed $\alpha_a = \alpha_h \leq 0.7\beta$ [16].

Common link between the defocusing angle β and misalignment angle α in measuring planes of the tracking optic sensor has been determined, we demonstrated that $\alpha = \beta$ always, both in the concentrator and in the heliostat tracking mode, though the nature of their variation and their projections may differ, and in practice the actual tracking inaccuracy in sensor planes must not exceed $\alpha a = \alpha h \leq 0.7\beta$.

As distinct from other measuring informational systems, the described heliostat control system operates only when tracking parameters deviate towards the maximum permitted values. Thus, the considered diagram provides a much simpler way of comparing the obtained parameters against the preset parameters.

Acknowledgements

We would like to express a gratitude to reviewers for valuable comments.

References

[1] Abdurahmanov A A, Orlov S A, Bakhramov S A, Bourbeau A V, Klychev I, Fazil H K 2010 On the accuracy of tracking the sun concentrators *Solar technology* **4** 82-5
[2] U. Thein Lin 2009 *Stepper motor control system for moving the frame of the solar power plant. Natural and Technical Sciences.* M.: "The company Sputnik +" **1** 292-5
[3] Sarybaev S A, Satybaldyieva F A, Sarybay M A, Kultas A K 2014 Principles of controlling the orientation of the solar solar plants *Almaty Herald KBTU* **3**(30) 80-4
[4] Sarybaev A 2006 *Research and development of an automated control system of heliostats to extend the functionality of a large solar furnace capacity of 100kV* Abstract of Cand. Ph.D. T.: 25
[5] Klychev S I 2002 Simulation of optical-power characteristics of solar radiation concentrators *Solar technology* **3** 59-63
[6] Wolf R 2008 *Efficiency and Peak Capacity Gains through Solar-Thermal Feed Water Pre-Heating and Steam Generation using Compact Linear Fresnel Reflectors* Solar Heat & Power Europe GmbH. Mulheim/Ruhr. Germany
[7] Beysembekova R N, Sarybaev A S 2015 Schematic diagrams of the automated control systems, geolocation, and the equations of motion of solar installations for the position of the sun *II International Scientific and Practical Conference "Information and telecommunication technologies: education, science and practice"*
[8] Masafumi Yamaguchi A, Tatsuya Takamoto B, Kenji Araki C, Ekins-Daukes N 2005 A Multi-junction III-V solar cells: current status and future potential *Solar Energy* **79** 78-85
[9] Beysembekova R N, Sarybaev A N, Kocherova A N 2014 Software Development of an automated experimental solar control system

Bulletin of the Kazakh Academy of Sciences Almaty **5** 188-98
[10] Zahidov R A 1986 *Mirror concentration of radiant energy system* T.: "FAN" 173
[11] Ismailov S U, Satybaldiyeva F A, Beysembekova R N, Saribaev A S, Musabekov A A, Ismailov A S 2016 Electronic circuit based on pic microcontroller for heliostat departments with a support rotating mechanism for tracking the sun *Reports of the national academy of sciences of the republic of kazakhstan* **3**
[12] Archer B 1980 Comments on "Calculating the position of the sun" *Solar Energy* **25** 91
[13] Joseph J 1988 MichalskyThe Astronomical Almanac's Algorithm For Approximate Solar Position (1950-2050) *Solar Energy* **40**(3) 227-35
[14] Sarybaev A S, Beysembekova R N, Ismailov S U 2016 Development and research optimization system power output of thermal reactors by changing the orientation solar installations *The International Conference on Energy and Infrastructure Management (ICEIM-2016). School Of Petroleum Management Pandit Deendayal Petroleum University Gandhinagar- 382007, Gujarat, India 31-40*
[15] AlFaify S, Rammah Y S, Zahran H Y, Yakuphanoglu F, Mohd. Shkir 2016 Optical Properties of Nano-Rods PTCDA Thin Films: an Important Material for Optoelectronic Applications *Natural Sciences Publishing (NSP) OrganoOpto-Electronics (OOE)* **2**
[16] Yurchenko A V, Kozlov A V 2007 The long-term prediction of silicon solar batteries functioning for any geographical conditions *Proceedings of 22st European PV Solar Energy Conference and Exhibition. Milan 3-7 September 3019-22*
[17] Aipov P C, Yarmukhametov W R 2007 Improving the efficiency of solar power plants *Mechanization and electrification of SH* **9** 29-30
[18] Witzke A, Kaluza N J 2001 Test of Device Accelerated Ageing of Polymeric Material in High Concentrated Sunlight at the DLR Solar Furnace *5th Cologne Solar Symposium* 152-8

AUTHORS

Satybaldiyeva Feruza

Current position, grades: PhD student of Kazakh National Technical University.
University studies: Master degree in computer engineering and software from International University of Information Technologies.
Scientific interests: Optical tracking sensor
Publications:
1. Development and research optimization system power output of thermal reactors by changing the orientation solar installations. The International Conference on Energy and Infrastructure Management (ICEIM-2016). SCHOOL OF PETROLEUM MANAGEMENT PANDIT DEENDAYAL PETROLEUM UNIVERSITY GANDHINAGAR- 382007, GUJARAT, INDIA 2016, c. 31-40
2. Creation of control over the orientation of the solar systems helio concentration. INTERNATIONAL SCIENTIFIC AND PRACTICAL CONFERENCE Kyrgyz State University named I.Arabaeva 2015, pp 200-203
3. Software Development of an automated control system experimental solar№5.2014 Bulletin of the Kazakh Academy of Sciences Almaty 2014 pp.188-198

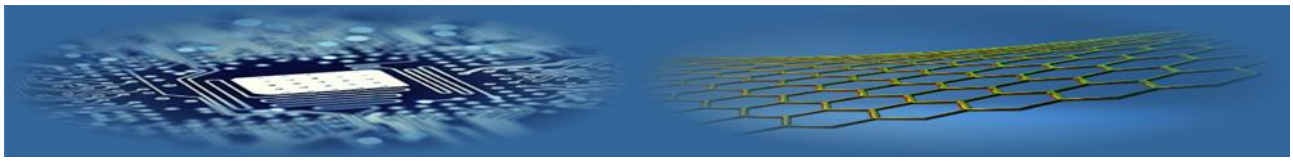
Experience: Experience in software development and process automation.

Ditmur Beyer

Current position, grades:professor University of Applied Sciences Schmalkalden

Sarybaev Abdushukur

Current position, grades: senior lecturer, Candidate of Technical Sciences South Kazakhstan National University after named M. Auezov



Combining web engineering methods to cover lifecycle

K Wakil[1, 2, 3*], D N A Jawawi[1]

[1]*Department of Software Engineering, Faculty of Computing, UTM, Malaysia*
[2]*Sulaimani Polytechnic University, Iraq*
[3]*University of Human Development, Iraq*

[*]*Corresponding author: karzanwakil@gmail.com*

Received 6 December 2016, www.cmnt.lv

Abstract

Web applications have rapidly evolved in the last decade, whilst web engineering methods have been lacking in the process development Web applications. One of the issues in web engineering methods is that no single web engineering method provides adequate coverage for the whole life cycle, because the web engineering methods are divided into three phases, which are; requirements, analysis/design, and implementation. Therefore, each method designed to special concern. It is obvious that we need to design a new method to cover the whole lifecycle to solve this issue. In this paper, we propose a framework for the new web engineering method through a combination of three methods comprising: Navigational Development Techniques (NDT) method for requirements phase; UML-Based Web Engineering (UWE) for analysis/design phase; and Interaction Flow Modeling Language (IFML) for the implementation phase. NDT and UWE are the most representative methods to develop web applications; while IFML is the newest method that focused on design and implementation. Our framework for the new method can support a whole lifecycle. Moreover, this method is more usable from developers.

Keywords

Web engineering,
methods,
combination,
lifecycle,
IFML

1 Introduction

Model Driven Web Engineering (MDWE) methods such as WebML [1], IFML [2], UWE [3] or OOHDM [4] during the last years have turn out to be established solutions for developing Web applications. These methods use Model Driven Development (MDD) ideas to attract high-level Web applications concepts into models and apply these models to derive applications automatically. The process of classic MDWE development consists of three phases [5]: (1) building a domain model, (2) defining a hypertext model and (3) defining the application's look and feel. A set of models is the outcome of the process that can create the last Web application using code generation. Moreover, the Model Driven Architecture (MDA) based development process establishes four phases of the development life cycle: analysis; platform independent design, where a Platform Independent Model (PIM) is built; Platform Specific Design, where a platform specific model (PSM) is built; and implementation [6].

Several issues within web engineering methods do exist. One of them is that there is no single method that covers the entire development life cycle in depth, and each method bears has its own particular strengths [7-8] as illustrated in Figure 1. As it is established in [9], the majority of the methods that are intended just for the hypermedia systems design partly cover the hypermedia systems life cycle and are further concentrated on the systems design. The web engineering community and several research groups are geared towards sustainable solutions to such variations, with some being solved by merging two methods like RUX-Method and UWE method to support Rich Internet Applications (RIA) [10], while the solution of others was obtained through enhanced methods like UWE metamodels in establishing novel modules of websites [11] although could never have all the issues completely solved.

To solve this issue, there is a necessity to design a new method. There are three ways to define a new web engineering method which include: extension existing methods; combine existing methods, and define new method. Nevertheless, in [9] the researcher recommended for new method to cover lifecycle with combination by employing common model, but in this paper we propose a framework for defining a new web engineering method through merging three approaches by metamodels with adopt strong models.

In order to design the framework for the new method, we selected the most representative methods [12] as follows: NDT method for the requirements phase; UWE for the analysis/design phase; and the newest method, IFML instead of WebML, for the implementation phase. For define a new method we propose a new framework, in this framework we use strong model in each method to cover lifecycle, here we use requirements model from NDT, design model from UWE, and Implementation model from IFML. The new method will be more usable and interoperable method to develop web applications.

The paper is organized as follows: Section 2 explains the background work undertaken for the web engineering methods during the lifecycle, and some combination between web engineering methods. In section 3, we analyze the most representative web engineering methods that used to support lifecycle. In Section 4, we propose a framework to design a new method to cover lifecycle. In the last section, we present some concluding remarks and suggestions for future research.

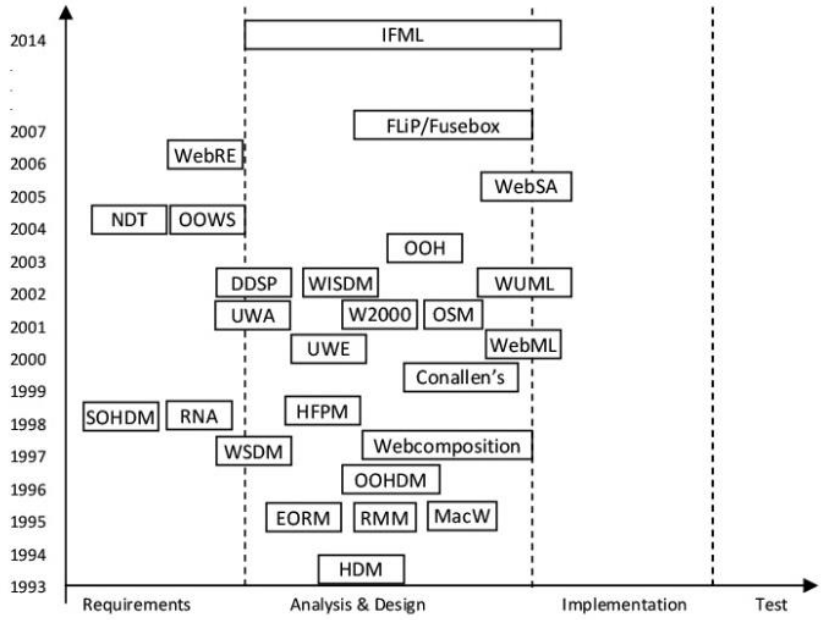


FIGURE 1 the evolution and coverage the best-known web development [7]

2 Background

Several MDWE methods were offered and they solved the complexities of methods to development web applications during the past years. However, they as well presented some restrictions and one of the limitations is that there is no single method to cover lifecycle deeply. Within this part, we evaluate several literature works to resolve lifecycle and combination within MDWE methods.

Numerous approaches, methods, and processes have been proposed in the educational and expert literature over the previous years to handle special features of Web development, and expert literature in MDWE [13]. Intricate interfaces, navigation, complex maintenance, safety features and unidentified remote users are among the serious difficulties pertinent to Web-based system improvement. Nevertheless, they resolved the challenges although they as well present some limits. One among the limits is lack in cover lifecycle [6, 14-15]. Lang and Fitzgerald [16] offer an all-inclusive record of overfly techniques and methods for Web hypermedia systems expansion. An explanation and comparative study of the better recognized of these Web development methods can be found in [17].

A significant perception in [6] as shown in Figure 1 is the assorted coverage by methods of the development phases. Within the Figure 1, every method is positioned in the phase where its major concentration lies. Therefore, even though the UWA Project [18] or WebML [19] offer some thought to necessities description and implementation, they largely highlight the examination and design phase. As can be viewed, most of Web development methods are focused on analyze and design phase, with perceptibly less concentration on the other life cycle phases.

Several union web engineering methods exist with every combination done for resolving a diverse difficulty such as; Preciado, J. C. et al. in [10] united RUX-Method and UWE to support RIA. They recommend a model-driven method to RIA development by uniting the UWE method for data and business logic sampling through the RUX-Method for the consumer interface sampling of RIAs.

Preciado et. al [20] employ RUX-Model [21], a MDWE Method for the systematic adaptation of RIAs UIs over existing HTML-based Web applications based on models in order to give them multimedia support, giving more efficient, interactive and instinctive user experiences. Amongst the phases of transformation proposed in RUX-Model, they have concentrated on the description of the connection procedure having the Web model being modified. This phase is decisive in the procedure because of it being the lone element of RUX-model that relies on the Web model chosen [22].

In [23] the researchers offered the Method Association Approach (MAA) that chooses and constructs appropriate methods from five model-driven web modeling methods. The MAA forms modeling approaches in definite web application domains for uses in diverse phases of their life sequence. The MAA places existing model-driven methods using metamodel ideas against key aspects of a particular web application continuum. Through the MAA, a design approach is built that flawlessly adjusts to web application aspects, and utilizes confirmed concepts of web design. The method has been confirmed using specialist corroboration and analysis of the two cases. In another study [24] the authors submitted a method and device support to construct web information networks that combine the employ of Scrum methods and Model-Driven Engineering (MDE). Such method and device permit performing fast design and corroboration of pre archetype models.

Daniel and Pozzi proposed a framework for the design and development of adaptive Web applications. The framework leverages on the integration of two well established methods: a conceptual model, complemented with a CASE tool for automatic code generation, and a language for expressing ECA rules, supported by an engine for rule execution. Such integration leads to a versatile and flexible adaptivity environment, whose advantage is twofold: on one hand,

conceptual modeling and automatic code generation support an efficient development process; on the other hand a detached rule engine allows us to widen the set of adaptivity requirements that can be handled and to overcome some limitations of current modeling approaches [25].

Huang et. al in the ref [26] presented an expanded lifecycle process model for the development of Web-based applications in SMEs. It consists of three sets of processes, meaning requirement processes, development processes, and evolution processes. Particularly, the post-delivery evolution processes are important to SMEs to develop and maintain quality Web applications with limited resources and time.

In the excellent work Rivero et. al, used User Interface prototypes (usually known as mockups) as a way to start the modeling process in the context of a mixed agile-MDWE process. To assist this process, we defined a lightweight metamodel that allows modeling features over mockups, interacting with end-users and generating MDWE models. Then, we conducted a statistical evaluation of both approaches (traditional vs. mockup-based modeling) [27].

The best combination to cover lifecycle presented in [8], in this paper suggested three web Engineering methods that are NDT,UWE, and WEML to cover lifecycle as shown in Figure 2, however this idea is best idea but for implementation very hard because used common model and need to new transformation model, moreover no tool support for implementation this idea.

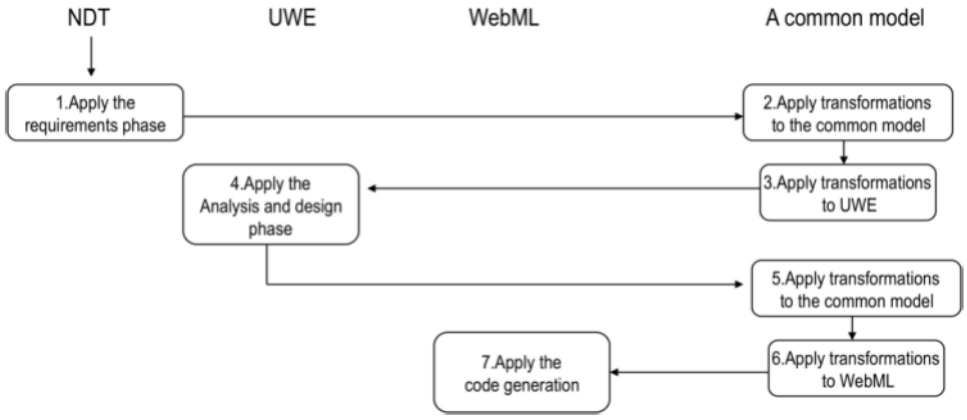


FIGURE 2 Use common metamodels to make approaches compatible[8]

One particular aspect of web engineering that remains problematic is the lack of integrated toolsets to support development methods and approaches, a long-standing difficulty alluded to some years ago in [28]. Because of the frequent changes in Web systems and the imperative to release fully functional upgrades quickly and often, Web development methods must be highly agile. The use of CASE tools that provide automated processes and enable rapid development/re-factoring is therefore necessary. In recent years, methods such as UWE, which offers a tool named MagicUWE [29], and WebML, which is supported by the WebRatio tool [30], have been greatly welcomed. Nevertheless, for CASE tools to be interoperable and interchangeable between and across Web development methods, it is essential that there must be a mechanism to facilitate the transformation and consistent integration of semantic metamodels. In this regard, MDWE offers much promise because it potentially enables Web developers to mix-and-match method fragments taken from different approaches and combine them into a tailored hybrid which is customized to the needs of a particular development project. This paper offers a critical view about this possibility by analyzing if approaches can be easily integrated or extended with new approaches.

3 Analyzing Web Engineering Methods

In this section we analyze the majority preventive methods that used to resolve lifecycle issue. The methods include NDT Method for Requirements phase; UWE for analyze and design phase and IFML for implementation phase, in the following we explained one by one.

3.1 NDT

NDT [31] is an MDWE methodological approach mainly focused on requirements and analysis. NDT describes a collection of PIM and CIM models and the set of revolutions by QVT to coin PIM from CIM. Similar to it happens in other methods, these metamodels are described by employing class diagrams. The necessities metamodel of NDT is an expansion of WebRE that comprises new ideas depending on the WebRE method. Moreover, it comprises two metamodels, the navigational and the content for the PIM level. The latter is the UML metamodel for class illustrations and the former is the metamodel for the UWE. One of the largely significant benefits of this method is its device support. A set of devices known as NDT-Suite, fabricated of four devices supports the MDE development process of NDT (this device-set can be got at [32]). Every NDT metamodel has a particular profile that is executed in Enterprise Architect [33]. The NDT method has modified the interface of this device through a set of device boxes having direct contact to every artifact of the method. This atmosphere is known as NDT-Profile. Besides, NDT-Suite comprises six other devices which are NDT-Driver, NDT-Report, NDT-Quality, NDT-Glossary, NDT-Prototypes, and NDT-Checker.

In Figure 3, the NDT development process can be defined as a bottom-up process. The process of development is concentrated on an extremely comprehensive necessities definition, directed by objectives that cover three sub-phases: necessities capture, requirements description, and

requirements corroboration. NDT simply covers the initial phases within the life cycle. Moreover, it is essential to highlight that workflows within NDT that shift from necessities to analysis are methodical. These workflows are described by means of the MDE paradigm. The need of offering a systematic procedure so as to create Web design models has been noticed by numerous investigation groups. These workflows might yet be mechanical if the development group utilizes its related device of NDT, the NDT-Tool.

When necessities are confirmed, the NDT procedure goes on by describing three models:

- The content model that is a class illustration. It articulates the static outlook of the system.
- The navigational model that reveals how consumers can navigate through the method.
- The abstract interface model that displays the theoretical interface of the structure.

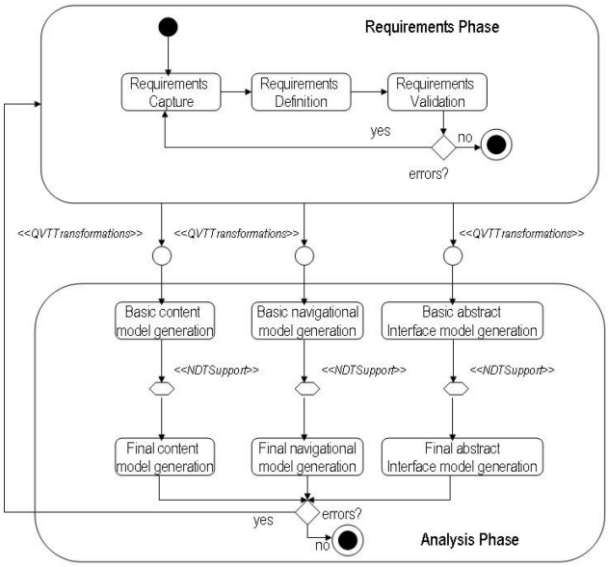


FIGURE 3 NDT development process [31]

3.2 UWE

UWE came up by 1998. The method was developed by the Web Engineering Group from the Ludwig-Maximilians-Universität München [34-35]. UWE is a software engineering approach based on UML [36] it uses the UML standard notation as much as possible and defines a UML profile to specify the peculiarities that introduce web applications. The major benefit of being UML compliance is that any CASE tool that supports the UML notation can be used to produce the UWE models [37].

UWE a method based on standards. UWE focused on visual modelling together with systematic design and automatic generation also UWE is a software engineering approach for the web applications whose objective is to cover the entire life cycle of Web application development.

UWE is said to be a small extension of UML, which provides UML profile for the web domain. UWE also provides some tolls which can which can be a lot helpful in model designing and consistency checks and automatic creation of web applications. There are two UWE plug-ins called ArgoUWE and MagicUWE which can support notation of UWE portfolio and design is also supported with the help of transformation [38].

UWE metamodel is a design considered as the conservative extension of the UML metamodel, in other words we can say that the modelling elements of a UML metamodel are inherited from the UML metamodel and they are not modified by adding new features or additions to the modelling elements class. Any additional features or relationship if we want to implement then they can be specified in different metamodel modelling element and then define OCL restrictions on additional static semantics and it is equivalent to well formatted rules in the UML specs. We can have benefit from the metamodelling tools that are depending upon equivalent XML Interchange (XMI) format by keeping them compatible with the Meta Object Facility (MOF) interchange metamodel [39].

The UWE Metamodel can be customized on the basis of a profile by mapping it to a UML profile.UWE metamodel for web applications can be created by using generic UML case tools and UML profiles or their extension i.e. typecast, objects those are tagged and OCL restrictions. CASE tools can be used to maintain UWE method but that actually depends on the technical feasibility. If we are bonded to use UML version then we also need to take the consideration of problems in its specification. UML metamodel is included in metamodel architecture for OMG for example, considering a metamodel arch with different levels in it, then a modelling element at level 2 is not an instance of exactly one element at level 3. This is also called as a 'loose metamodelling problem', which can be taken care of in new versions of UML [39-40] .

UWE approach proposes to build a set of CIMs, PIMs, and PSMs as results of the analysis, design and implementation phases of the model-driven process. The aim of the analysis phase is to gather a stable set of requirements. The functional requirements are captured by means of the requirements model. The requirements model comprises specialized use cases and a class model for the Web application. The design phase consists of constructing a series of models for the content, navigation, process, presentation and adaptivity aspects at a platform independent level. Transformations implement the systematic construction of dependent models by generating default models, which then can be refined by the designer,

as shown in Figure 4 [41].

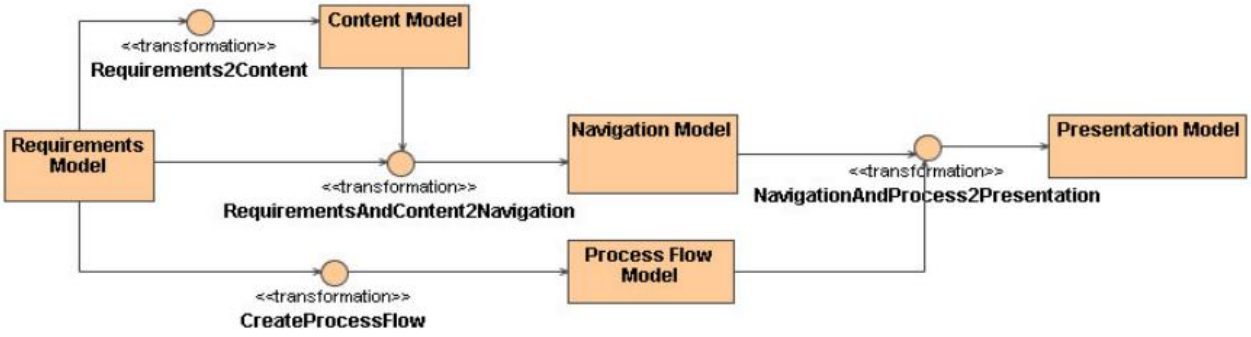


FIGURE 4 UWE core process [41]

3.3 IFML

IFML[2] supports the platform-independent description of graphical user interfaces for applications deployed or accessed on systems such as desktop computers, laptops, PDAs, mobile phones, and tablets. The main focus is on the structure and behavior of the application as perceived by the end user. The modeling language also incorporates references to the data and business logic that influence the user's experience. This is achieved by referencing the domain model objects that provide the content displayed in the interface and the actions that can be triggered by interacting with the interface.

The development of applications defined by interactivity is normally handled with agile techniques, which navigate diverse phases of "problem identification/design modification/implementation." The iteration of the creation method derives a partial version or a prototype of the system. Such an augmentable lifecycle is predominantly suitable for contemporary web and mobile uses, with the need of being installed swiftly and alter frequently throughout their lifetime to adjust to user prerequisites. Figure 5 offers a probable structural creation process hence positioning IFML within the activity flow.

Domain Modeling systematizes the key information objects established during conditions delineation into a broad and articulate setting model. Domain modeling delineates the key data sets established during conditions requirement into a domain model, normally a (characteristically visual) depiction of the necessary objects, their qualities and relationships.

Front-End Modeling plots the data manipulation and information conveyance functionality proposed by the requirements application conditions into front-end model. The operation of front-end modeling is at the conceptual angle, with IFML coming into play. The developer is at the liberty of utilizing IFML in the specification of front-end organization in a single or several top-level view containers, the internal formation of every view container regarding sub-containers, the constituents forming each view container's content, the events depicted by the components and vie containers, as well as how such events set off business events and revise the interface.

Business Logic Modeling delineates the business objects and the techniques needed to sustain the established use cases. UML dynamic and static figures are usually used in highlighting the objects interface as well as messages flow. Process-adjusted details (like UML functionality and sequence charts, BPMN process models, and BPEL service orchestrations) offer an efficient method of signifying the workflow across services and objects. The services highlighted in the business logic plan can be oriented in the front-end model to signify the operations to be set off through interface interaction. Being interdependent in nature, front-end, data, and business-logic structure events are performed in an iterative manner. The preference category of Figure 5 is simply indicative. Within some companies, the responsibility could commence at the structure of the front-end while the actions and data objects could be established at a later phase though analysis of the published information as well as the requested operations towards sustaining the interactions.

Architectural structure is the technique of delineating the network, hardware as well as the software elements that compose the architecture whereby the application offers its services to the users. The objective of the architectural structure is to establish the mixture of these components that adequately achieves the application needs as regards to scalability, efficiency, accessibility, security, and all together adhering to the economic and technical project limitations.

Implementation entails the approach of creating the software modules that convert the business logic, data as well as interface design into an application functioning on the opted design. Implementation of data situates the domain model onto a single or several data sources by merging the conceptual-level aspects with the formations of logical data (such as relationships and aspects to relational tables). The execution of business logic generates the software components required to sustain the identified use cases. The execution of individual entities may gain from the adoption of software designs, which systematize the manner in which fine-grain elements are devised and merged into a wider and highly reusable operational units and equally provide for nonfunctional needs like scalability, accessibility, security and competence. Translation of abstract-level ViewComponents and ViewContainers into the apposite aspects within the considered execution plan is done courtesy of interface accomplishment. It is possible for the ViewContainers and business objects to interoperate either in the server or client layer.

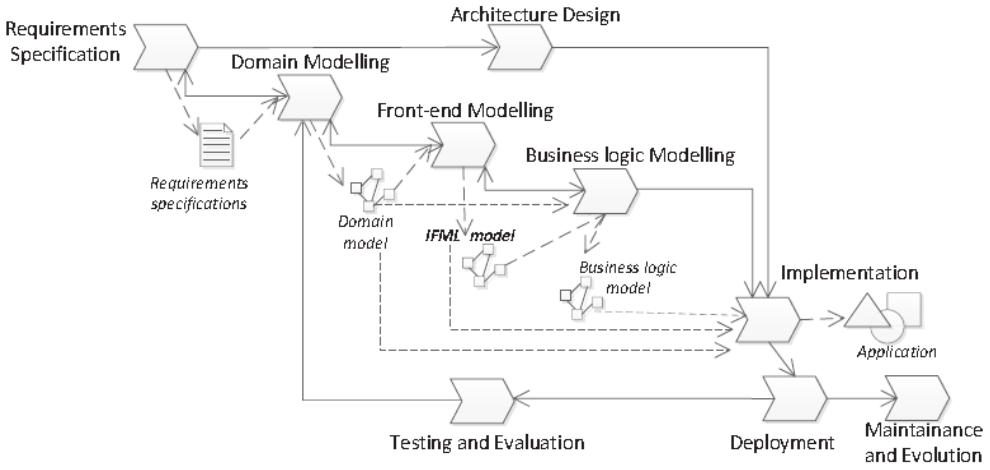


FIGURE 5 the role of IFML in the development process of an interactive application

Testing and Evaluation confirms the consistency of the installed application concerning the nonfunctional and functional conditions. The key important aspects for interactive model testing include:

- Functional Trialing: verification of the application behavior regarding the functional needs. Functional testing is disintegrated into classical events of module examination, system testing and integration testing.
- Usability Assessment: the nonfunctional prerequisites of accessibility, communication efficiency, and observance to merged usability values are confirmed against the generated front end.
- Performance Assessment: the application's response time and throughput ought to be examined in peak and average workload provisions. There is the need to monitor and examine the insufficient service levels, the usability design, so as to establish and get rid of bottlenecks.

4 Proposal of new web engineering method to cover lifecycle

In this section, we define a framework for a new web engineering method which can satisfactorily cover the whole life cycle. As mentioned in previous sections, we are using the most representative methods, namely, NDT, UWE, and IFML. Moreover, these methods are used to achieve better comparison and implementation. Each method has a particular strength in the process development of a lifecycle, as follows: NDT is a method focused on requirements; UWE is focused on analysis and design; and IFML is focused on Implementation. Already existing were some ideas covering life cycle, but these were difficult to implement. In this section, we define a new method through which to borrow models from one method to another. In Figure 6, we present answers to questions of "how, which, and where" a model can borrow attributes from another model.

- In Level A, we present the methods for particular strengths in a lifecycle. For the requirements phase, we have selected the NDT method; while for the analysis and design phase, we have selected UWE. For the implementation phase, we have selected IFML.
- In Level B, we present the important models which are used for the development of web applications. As mentioned in Section 3, NDT has three models which are: requirements capture; requirements definition; and requirements validation respectively. UWE has five models comprising: requirements model; content model; navigation model; process flow model; and presentation model. IFML has three models which are: Domain Model; IFML model; and Business Logic model. These also support code generation.
- Level C, presented the case tool by which to support methods, namely: NDT supported by NDT-Suite; UWE developed by ArgoUWE and MagicUWE; and IFML developed by WebRatiotool.
- Level E, presents a new web engineering method by combining strengths of each of the models NDT, UWE, and IFML, and implementing this inside IFML. In addition, we use WebRatio tool for the development of a new web engineering method.
- In Level D, we represented the strengths of each of the models, in particular, the model from NDT.

Furthermore, we recognized the need for transformation models for moving from one phase to another such as, CIM and PIM respectively.

As shown in Figure 6, we defined a new framework for a web engineering method; however, we could not implement this due to some challenges that were encountered. These challenges comprised: transformation models between levels became a significant challenge for implementation; a considerably long time was needed for implementation; the very complex work required was best suited for group work rather than an individual researcher; and there was a necessity to improve tool support. For these reasons, we could not apply a case study by our framework.

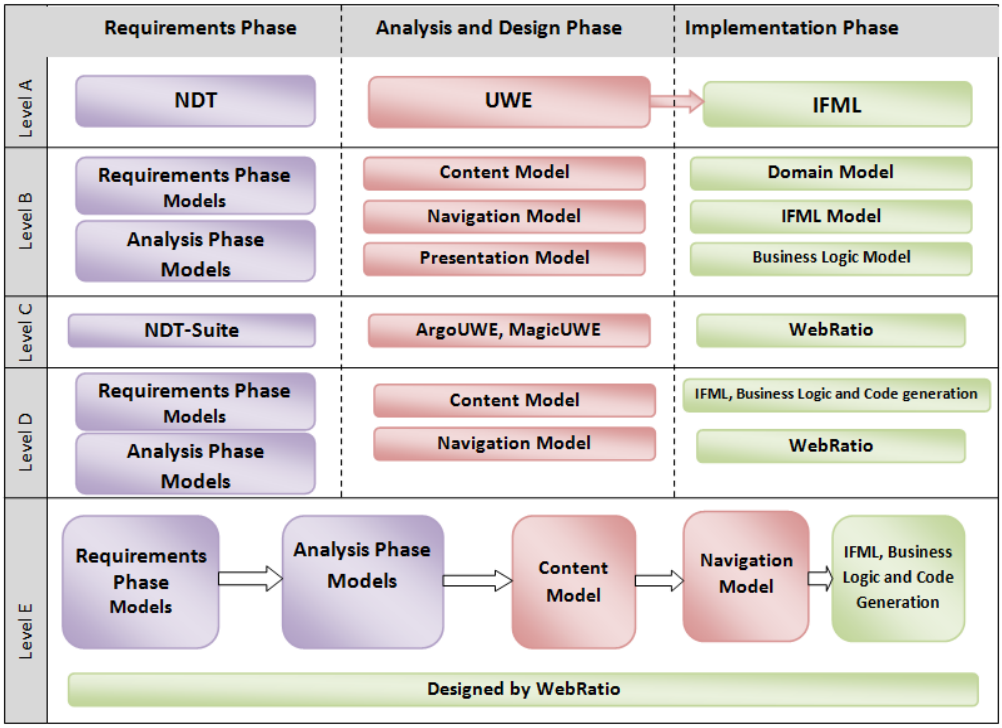


FIGURE 6 New Web Engineering method to Cover Lifecycle

5 Conclusion and future work

In this paper, we have defined a framework for a new web engineering method to cover lifecycle by using three methods that have particular strengths in the web engineering lifecycle. We propose a new web engineering method through a combination of three methods, namely: NDT method for the requirements phase; UWE for the analysis/design phase; and IFML for the implementation phase. Our method can support the whole lifecycle; it is also a compatible and interoperable method with which to support web development. Moreover, this method is more usable for implementation. Our recommendations for researchers are to implement this method by existing tools or extension of an existing tool. A new case tool can also be created; moreover, it can import more models to improve our method from other methods for the development of web applications. In addition, new models can be defined, including: adaptivity model; security model and so on. A suggestion for future work is the addition of model transformation and implementation case study by WebRatio Tool.

References

[1] Ceri S, et al. 2002 Conceptual modeling of data-intensive Web applications *Internet Computing, IEEE* **6** 20-30
[2] Brambilla M, Fraternali P 2014 Interaction flow modeling language: Model-driven UI engineering of web and mobile apps with IFML: Morgan Kaufmann
[3] Koch N, et al. 2008 UML-based web engineering *Web Engineering: Modelling and Implementing Web Applications, ed: Springer* 157-91
[4] Rossi G, Schwabe D 2008 Modeling and implementing web applications with OOHDM *Web engineering: Modelling and implementing Web applications, ed: Springer* 109-55
[5] Wimmer M, et al. 2007 On the integration of web modeling languages: Preliminary results and future challenges *Workshop on Model-driven Web Engineering (MDWE), held in conjunction with ICWE, Como, Italy*
[6] Hincapié Londoño J A, Duitama J F 2012 *Model-driven web engineering methods: a literature review* Revista Facultad de Ingeniería Universidad de Antioquia 69-81
[7] Jawawi K W a D N A 2017 Analyzing Interaction Flow Modeling Language in Web Development Lifecycle *International Journal of Advanced Computer Science and Applications (IJACSA)* **8**
[8] Aragón G, et al. 2013 An analysis of model-driven web engineering methodologies *International Journal of Innovative Computing, Information and Control* **9**
[9] de Koch N P 2001 *Software engineering for adaptive hypermedia systems* PhD Thesis, Verlag Uni-Druck, Munich
[10] Preciado J C, et al. 2008 Designing rich internet applications combining uwe and rux-method *Web Engineering ICWE'08. Eighth International Conference on* 148-54
[11] Wakil K, et al. 2014 Enhancement of UWE navigation model: Homepage development case study *International Journal of Software Engineering & Its Applications* **8**
[12] Aguilar J A, et al. 2015 An Analysis of Techniques and Tools for Requirements Elicitation in Model-Driven Web Engineering Methods *Computational Science and Its Applications--ICCSA ed: Springer* 518-27
[13] Wakil K, Jawawi D N 2015 Model driven web engineering: A systematic mapping study *e-Informatica Software Engineering Journal* **9** 107-42
[14] Lang M 2005 A critical review of challenges in hypermedia systems development *Information Systems Development, ed: Springer* 277-88
[15] Aragón G, et al. 2013 An analysis of model-driven web engineering methodologies
[16] Lang M, Fitzgerald B 2006 *New branches, old roots: a study of methods and techniques in web/hypermedia systems design*
[17] Escalona M, et al. 2007 The treatment of navigation in web engineering *Advances in Engineering Software* **38** 267-82
[18] Consortium U 2001 *Requirements Elicitation: Model, Notation, and Tool Architecture* ed: Ubiquitous Web Applications Consortium (Deliverable D6)
[19] Ceri S, et al. 2000 Web Modeling Language (WebML): a modeling

language for designing Web sites *Computer Networks* **33** 137-57
[20] Preciado J C, et al. 2007 An Approach to Support the Web User Interfaces Evolution *AEWSE*
[21] Rossi G, et al. 2007 *Web engineering: modelling and implementing web applications* Springer Science & Business Media
[22] Stearn B 2007 XULRunner: A new approach for developing rich internet applications *IEEE Internet Computing* 67-73
[23] Luinenburg L, et al. 2008 Designing web content management systems using the method association approach P*roceedings of the 4th International Workshop on Model-Driven Web Engineering (MDWE 2008)* 106-20
[24] Basso F P, et al. *Study on Combining Model-Driven Engineering and Scrum to Produce Web Information Systems*
[25] Daniel F, et al. 2006 Combining conceptual modeling and active rules for the design of adaptive web applications *Workshop proceedings of the sixth international conference on Web engineering* 10
[26] Huang W, et al. 2010 A novel lifecycle model for Web-based application development in small and medium enterprises *International Journal of Automation and Computing* **7** 389-98
[27] Rivero J M, et al. 2014 Mockup-Driven Development: Providing agile support for Model-Driven Web Engineering I*nformation and Software Technology* **56** 670-87
[28] Lang M 2002 Hypermedia systems development: do we really need new methods *Proceedings of the Informing Science+ IT Education Conference, Cork, Ireland* 883-91
[29] UWE 2015 MagicUWE - UWE Plugin for MagicDraw. Available: http://uwe.pst.ifi.lmu.de/toolMagicUWE.html
[30] WebRatio 2015 WebRatio. Available: http://webratio.com/
[31] Escalona M J, Aragón G 2008 NDT. A model-driven approach for web requirements *Software Engineering, IEEE Transactions on* **34** 377-90
[32] IWT2I 2015 Available: http://www.iwt2.org
[33] Enterprise Architect 2011 Available: http://www.sparx.org/
[34] Koch N P d 2001 *Software engineering for adaptive hypermedia systems* PhD Thesis, Verlag Uni-Druck, Munich
[35] Koch N, Kraus A 2002 The expressive power of uml-based web engineering *Second International Workshop on Web-oriented Software Technology (IWWOST02)*
[36] OMG O 2007 *Unified Modeling Language (OMG UML)* ed: Superstructure
[37] Torres Bosch M V 2008 A web engineering approach for the development of business process-driven web applications
[38] Colin A, Kühne T 2001 *The essence of multilevel metamodeling* «UML» 2001—The Unified Modeling Language. Modeling Languages, Concepts, and Tools, 19-33
[39] Atkinson C, Kühne T 2001 The essence of multilevel metamodeling *«UML» 2001-The Unified Modeling Language. Modeling Languages, Concepts, and Tools* **2185** 19-33
[40] Koch N, Kraus A 2003 Towards a common metamodel for the development of web applications *Web Engineering* 419-22
[41] Kraus A, et al. 2007 Model-Driven Generation of Web Applications in UWE *MDWE* 261

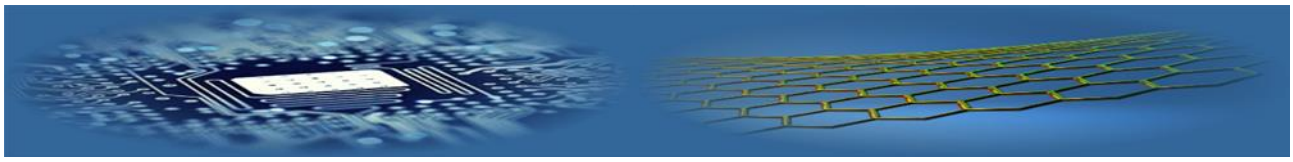
Authors

Karzan Wakil, Iraq

Current position, grades: lecturer at the University of Human Development-Iraq and Institute of Training and Educational Development in Sulaimaniyah-Iraq, received BSc.
University studies: Degree in Computer Science from Salahaddin University-Iraq-2006 and M.Sc. in Computer Science from University Technology Malaysia (UTM), Malaysia, 2013. Currently he is PhD student at UTM Malaysia His research areas are; Web Engineering, Software Engineering, Web Development, Model-Driven, Metamodel

Dayang N.A.Jawawi, Malaysia

Current position, grades: is an Associate Professor in Software Engineering Department, Faculty of Computing, Universiti Teknologi Malaysia (UTM).
University studies: B.Sc. in Software Engineering from Sheffield Hallam University, UK, and her M.Sc. and Ph.D. in the field of Software Engineering from Universiti Teknologi Malaysia. She has been an academic staff at Software Engineering Department for over ten years and she is currently serving as the Head of Department since October 2009. She is a member of the Software Engineering Research Group (SERG), K-Economy Research Alliance, UTM.



Model reservation resource of computer complexes system

Dmitry Kornev[1], Evgeny Nikulchev[2*]

[1]*Scientific Research Institute for Automation and Communication, Nizhny Tagansky str., 27, p. 1 Moscow, Russia*
[2]*Moscow Technology Institute, Leninsky Prospect, 38a, Moscow, Russia*

[*]*Corresponding author's e-mail: nikulchev@mail.ru*

Received 19 February 2017, www.cmnt.lv

Abstract

The article presents a method for calculating redundancy systems for platform management and organization of the movement of high-speed trains. The model is developed on the mathematical formalism of Petri nets. Calculated estimated time system failover. Proved the feasibility of using a majority redundancy.

Keywords:

information system,
Petri net,
information complex,
reservation of computer complex

1 Introduction

All high-tech systems require means of maintaining operability during sudden failures. This is particularly important for transport systems. Traffic safety is the main principle of the transport systems and provides high efficiency of the railways.

Currently, Russia's specialists creating an information platform, "High-speed rail". The platform is designed to traffic management of high-speed trains. The reliability of the platform should be provided with additional computing resources. These resources will be included with the sudden failure of platform elements.

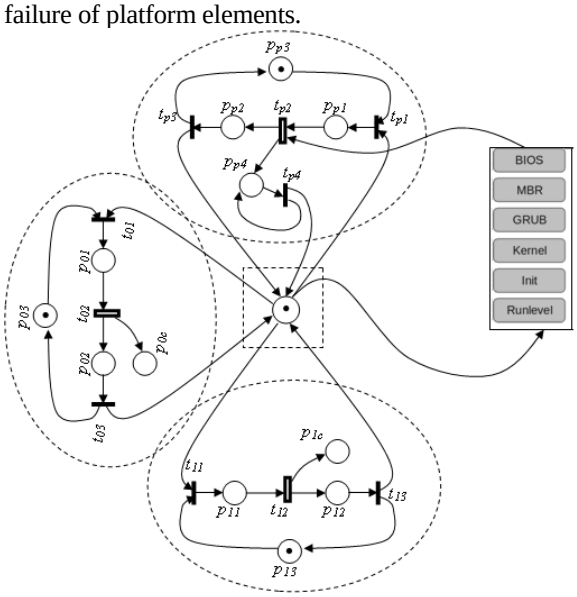


FIGURE 1 Model of interaction of the Platform and the VM in the event of failure of computing resources

Information resource characterized by the time ready to work. This time should match the characteristics of the control object. Platform deployed on virtual machines (VMs). These VMs enable communication platforms and each locomotive. Objects Management Platform is the system management and control locomotive. Information exchange interval "platform locomotive" is ΔT = 100 ms.

The timing of the inclusion of additional resources Platforms τp r carried out on the model of interaction between the host and VMs. Host constantly interacts with multiple VMs. The model used by the mathematical formalism of colored Petri nets (Figure 1).

2 Mathematical model

Object's model to the mathematical formalism of Petri net has the form [1, 2]:

$$\Pi = \{P, T, I, O, \mu\}, \quad (1)$$

$P = \{p_1, p_2, \dots, p_i, \dots, p_n\}$ - a plurality of positions;
$T = \{t_1, t_2, \dots, t_j, \dots, t_m\}$ - a plurality of transitions;
I - input function of transitions (this value determines the multiplicity of input arcs of transitions (t_j));
O - output function of transitions (this value determines the multiplicity of output arcs of transitions $O(t_j)$);
μ - marker.

Model of system "Platform-locomotive" is presented as a graph with two types of vertex - positions and transitions. Positions and transitions are connected by arcs.

Routes in the model represented by the equations:

$$|I(p_i)| = |\{t_j | p_i \in O(t_j)\}| = 1, \quad (2)$$

$$|O(p_i)| = |\{t_j | p_i \in I(t_j)\}| = 1, \quad (3)$$

$\{t_j | p_i \in O(t_j)\}$ - set of transitions, where p_i is the output
$\{t_j | p_i \in I(t_j)\}$ - set of transitions, where p_i is the input.
Transition $t_i \in T$ performed under the condition [3]

$$t_j: \mu(p_i) \geq \#(p_i, I(t_j)) \quad \forall\, p_i \in P, \quad (4)$$

$(p_i, I(t_j))$ - the multiplicity of the input position p_i of transition t_j.

After the transition $t_i \in T$ position p_i receives a new marker μ':

$$\mu'(p_i) = \mu(p_i) - \#(p_i, I(t_j)) + \#(p_i, O(t_j)). \quad (5)$$

Figure 1 shows:

- p_0 – resource Host Platforms;
- $p_{01}, \ldots, p_{k3}$ - the current state of the information environment k VM:
- $p_{01}, \ldots, p_{k1}$- resource of the Host; $p_{02}, \ldots, p_{k2}$ - it frees the host system resources; $p_{03}, \ldots, p_{k3}$- waiting host resource ; $p_{0c}, \ldots, p_{kc}$- counter operations;

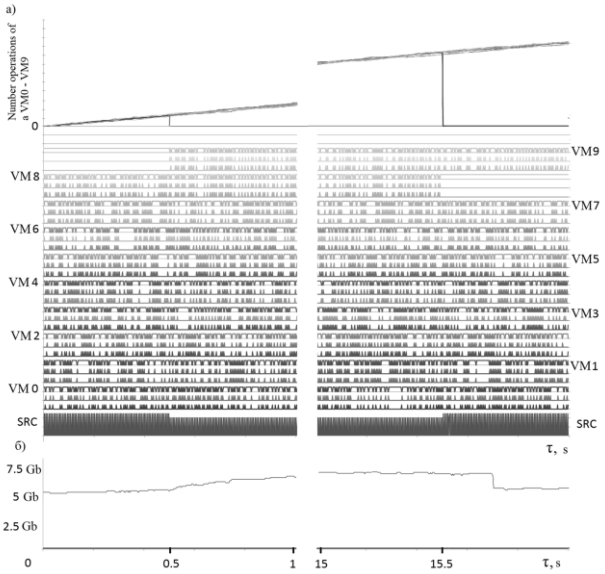


FIGURE 2 Deploying backup VM9: Simulation results (a); Experiment (b)

The mathematical model contains an additional structure [4, 5]:

$$P_{\mathrm{p}} = \{p_{\mathrm{p1}}, p_{\mathrm{p2}}, p_{\mathrm{p3}}, p_{\mathrm{p4}}\}, \quad (6)$$

$t_{01}, \ldots, t_{k3}$ - distribution of host resources and the generation of k VM requests: $t_{01}, \ldots, t_{k1}$ - providing a Host of resources ; $t_{02}, \ldots, t_{k2}$- working with Host resources; $t_{03}, \ldots, t_{k3}$ - host resources return.

$$T_{\mathrm{p}} = \{t_{\mathrm{p1}}, t_{\mathrm{p2}}, t_{\mathrm{p3}}, t_{\mathrm{p4}}\}, \quad (7)$$

p_{p1}- provide resource to reserve; p_{p2} - the release of the resource of reserve; p_{p3} - waiting for the resource for reserve; t_{p1} - providing resource to reserve; t_{p2} - work with the resource reserve host; t_{p3} - return resource of reserve.

3 Simulation

This structure is an informational reserve host. p_{p4} position provides job reserve. The memory of reserve sequentially loaded operating system modules: BIOS, MBR, GRUB, core, Init, Runlevel; this process reflects the transition t_{p2}. In the calculations of module loading time correspond to OS load time computer-based Core i3 Duo. Transition t_{p4} closes fully loaded software in the host memory reserve. At this moment the reserve receives information from the VM.

The calculation was carried out provided that: requests from locomotives ΔT, and one request service time τ_{max} had a uniform distribution with restrictions ΔT = 100 ms and ms τ_{max} = 99,5.

Figure 2a illustrates the functioning of the Platform in terms of Petri nets. The platform serves computers nine locomotives (VM0-VM8).

When τ = 0,5 s. is error on VM8 session on the Platform. At this point it begins to boot VM9 reserve. This is a time τ_p = 15s. Where τ = 15,5 to reserve loading over. Control is passed from VM8 to VM9. Similar experiments were conducted on actual host. Time unfolding real reserve was τ_p = 11 - 17. for different download host (Figure 2b, Table 1).

TABLE 1 Time deployment of VM backup

Software	time τ_p, s
Loading the virtual machine to the base operating system	7-12 s.
Loading virtual machine with snapshots	11-17 s.
Loading database	8-12 s.
Loading software with automatic driving	4-6 s.

4 Results and conclusions

The simulation results showed good agreement between calculated and experimental values τ_p time.

Modeling was carried out and on the Platform to deploy the reserve only with the basic software. In this case, the real time deployment of the reserve (VM9) is τ_p = 7 - 12. The calculated and experimental results unfolding VM9 with basic software are shown in Figure 3 and Table 1.

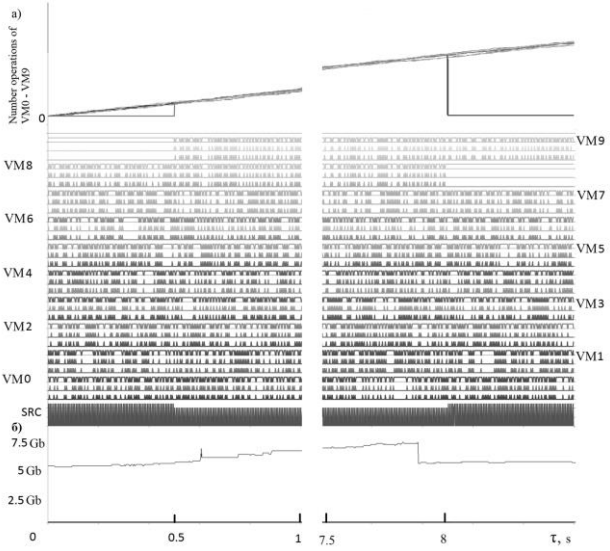


FIGURE 3 Deploying backup VM9 with basic software: simulation results (a); Experiment (b)

Modeling it possible to calculate the minimum reserve time of deployment on the platform. It is time $\tau_p \geq 5$s. For Platform train control system $\tau_p \geq 5$c is too large. The platform should deploy more VMs on the host, and give her control function of T2 = 100 ms.

This justifies the use for the Platform for the majority system redundancy.

Reference

[1] Kotov V 1984 *Petri nets* M.: Science: Home edition of Physical and Mathematical Literature 160 p.

[2] Leskin A 1989 *Petri nets in modeling and management* / A A Lesquin, P A Maltsev A M Spiridonov L.: Science 133 p.

[3] Zaitsev D 2013 Toward the Minimal Universal Petri Net *IEEE Transactions on Systems, Man, and Cybernetics: Systems* **144** 1–12

[4] Kornev D 2014 Study of algorithms of information systems using the mathematical formalism of Petri nets *Electronic Journal Cloud of*

Science **1**(1)
[5] Hee K van., Sidorova N M 2003 *Soundness and separability of workflow nets in the stepwise refinement approach* In van der Aalst, W. M. P.; Best, E. Voorhoeve Application and Theory of Petri Nets. Lect Notes in Comput Sci. 2678. Springer. pp. 337–56

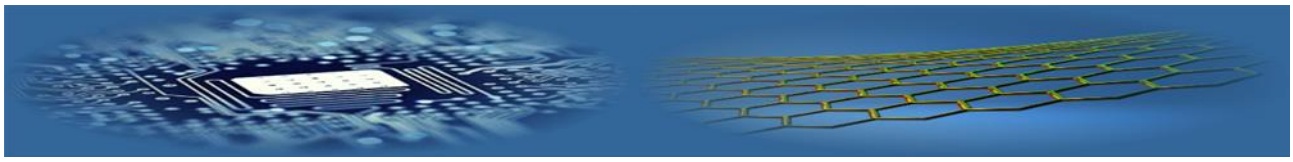
Authors

Dmitry Kornev 08.03.1989, Russia

Current position, grades: Candidate of Engineering Sciences. He graduated from the Moscow State University of Railway Engineering (MIIT) by specialty "Computers, Systems and Networks" in 2011. He got a degree in technical sciences in 2015.
Publications: 4 papers in peer-reviewed journals. Scientific interests: Information technologies.
Experience: Worked as engineer in Scientific Research Institute for Automation and Communication

Evgeny Nikulchev

Current position, grades: vice-rector and professor of Moscow Technological Institute
University studies: Dr of Sci. (Nonlinear Dynamics, 2006), PhD (Computer Science, 2000)
Scientific interest: big information systems design and management, management theory, scientific research management
Publications: More than 40 publications in the computing services, cloud computing, artificial intelligence.
Experience: It is artwork in the development of computing services for software applications for different purposes. He was Chairman and member of the scientific committees of international conferences on information systems, e-learning; chief Editor of the journal Cloud of Science, guest editor of several journals.



Application of artificial neural networks forhandwritten biometric images recognition

A Malygin[1], N Seilova[2], K Boskebeev[3], Zh Alimseitova[2]

[1]*Penza State University, Penza, Russia*
[2]*Kazakh National Technical Research University named after K.I. Satpaev, Almaty, Kazakhstan*
[3]*Kyrgyz State Technical University named after I. Razzakov, Bishkek, Kyrgyzstan*

Abstract

The development of information technology leads to new requirements for the development of security systems, identity authentication and other protection mechanisms. The article is devoted to the use of artificial neural networks for handwritten biometric images recognition that are used in high-authentication systems. There is given a general structure of the biometric-neural network authentication system, the structural scheme of information processing in biometric-neural network authentication systems, the structural scheme for learning the neural network converter of the biometric parameters vectors in the key code (password). There is formed and trained a network of neurons, are formed neural network containers on the basis of structures. The choice of the length of the biocode of neural network converters is substantiated. After graduation, testing is conducted and the probabilities of errors of the first and second kind are determined. There is given an example of a software implementation, where are given a learning mode, checking the results of training, testing the results of training.

Keywords:

artificial neural networks,
authentication,
biometric image,
first-kind errors,
second-kind errors

1 Introduction

Nowadays, the processes of informatization of modern society are actively underway. Internet space and digital mobile telephony are widely used. Almost all states declare their desire to create an e-government to provide services to their citizens. All these processes pose a number of problems before biometrics and cryptography. What was created in the last century, in terms of biometric technologies, does not work in the Internet space. Unfortunately, until now the Internet remains anonymous, impersonal environment with low confidence to it. This, on the one hand, is a breeding ground for all sorts of fraudsters, and on the other hand it undermines the confidence of ordinary citizens to new information technologies.

In the ordinary world, a government representative checks the citizen's credentials by his identity card or passport. In a virtual world, everything turns out to be more complicated, you can not use your identity card scan, even if it's a biometric identification card with a radio-readable RFID chip. You can not read the contents of the chip. Personal data contained in a bio-ID chip can not be used in the Internet space. For the Internet and other open information spaces, it is necessary to create special Internet passports or Internet identity cards, which on the one hand are biometric, and on the other hand are some cryptographic constructions that protect the confidentiality of personal data of their owners.

This problem is currently being given considerable attention. Researchers of the United States and NATO create and promote the relatively weak biometrics of open biometric images. Biometric authentication tools are built using fuzzy logic. Researchers from Russia, Belorussia and Kazakhstan suggest using neural network converters as biometric-code, while biometric authentication tools are built using large-scale artificial neural networks.

Researchers conducted in Kazakhstan, Russia and Belorussia in recent years have shown that existing biometric technologies can be significantly enhanced through the use of artificial neural networks [1-18]. They enrich the data in a continuous form, and usually for all input errors correction duplicate redundancy is enough, that is, 512 input biometric parameters the neural network converts into 256 bits of the output code with virtually no errors.

From the point of view of obtaining biometric properties, the neural network biometric converters are always better than "fuzzy extractors". This can easily be demonstrated by the example of poor biometric data, giving errors of 50% and more of biocode digits. Classical self-correcting codes are not capable to correct more than 50% of the errors, neural networks cope with this problem if their redundancy becomes threefold (inputs three times more than outputs).

2 The use of large neural networks

The use of large neural networks allows considering "good" biometric data along with "bad" and "very bad" biometric data. Moreover, the "worse" the biometric data is used, the more the network of artificial neurons should be and the more difficult it is to train.

The general structure of the biometric neural network authentication system is shown in Figure 1.

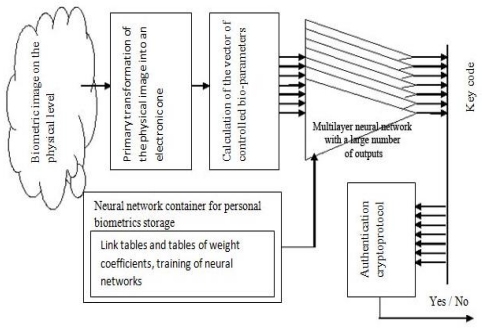


FIGURE 1 The general structure of the biometric neural network authentication system

It should be noted that for the solution of this problem artificial low-dimensional neural networks are unsuitable [12, 14, 17].

The process of input biometric image converting into an output long password (key) can be represented in the form of a diagram on Figure 2 [18].

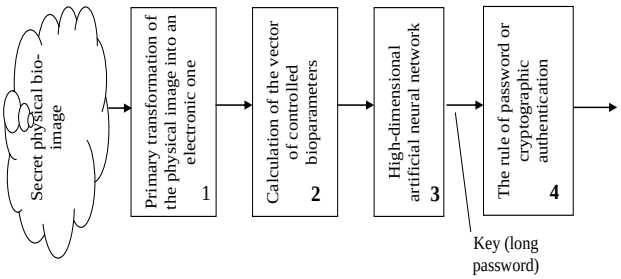


FIGURE 2 Structural diagram of information processing in biometric-neural network authentication systems

3 General training scheme of neural network converters biometry – code

The training scheme of the neural network biometric parameters vector converter into the key code (password) is represented by the scheme on Figure 3 [11]. For learning, you need N_1 of examples of the vectors "Own" and N_2 of examples of the vectors "Aliens".

Training of the artificial neural network should be carried out automatically (without human intervention in the process of artificial neural network parameters selection), the user must have a guarantee that his long password (key) involved in the training process will not be compromised.

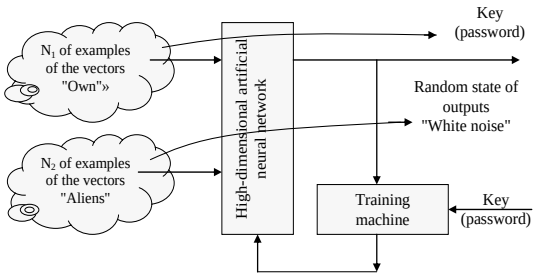


FIGURE 3 The structural scheme of training of the neural network biometric parameters vector converter into the key code (password)

During the training, the weight coefficients of the artificial neural network should be selected by the training machine in such a way that during the appearance of the elements of the vector "Own" on the inputs of the artificial neural network a long password (key) appears on the outputs of the artificial neural network. During the appearance on the inputs of the artificial neural network of the data vectors corresponding to the images "Alien", random states - "white noise" - should appear on the outputs of the artificial neural network. Training is carried out by alternately presenting of the images "Own" and "Aliens" with an intermediate selection of coefficients.

In order to use neural network bio-data enrichment it is necessary to be able to train single artificial neurons. [19-23].

4 Formation and training of the network of neurons. Neural Network Containers

In order to obtain the access biocode it is necessary to create a network of neurons with the number of outputs equal to the length of the key. The more inputs and outputs the neural network has the higher the quality of the decisions making. How strong this relationship is can be seen from the data in Table 1.

TABLE 1 The quality increase of solutions depending on the number of inputs and outputs of the artificial neural network

Number of inputs of the neural network	Number of outputs of the neural network	The probability of rejecting to "Own"	The probability of missing the "Alien"
5 inputs	1 output, 2 classes	$P_1 = 0,1$	$P_2 = 0,17$
48 inputs	1 output, 2 classes	$P_1 = 0,1$	$P_2 = 0,03$
480 inputs	1 output, 2 classes	$P_1 = 0,1$	$P_2 = 0,005$
480 inputs	256 outputs, 2^{256} classes	$P_1 = 0,1$	$P_2 = 0,000000001$
Training of the neural network with 480 inputs and 256 outputs was conducted on 20 examples of the image "Own" and on 128 examples of images "Alien" by the algorithm GOST R 52633.5-2011			

It can be seen from the Table 1 that a simple increase of the number of inputs of the biometric parameters is not very effective. It is much more important to increase in parallel the number of its outputs in parallel with the number of inputs. With the same number of inputs (480 inputs) an increase in the number of outputs from 1 to 256 gives a gain of about a billion times as the decisions made by the neural network (six zeros additionally appear). At the same time, other costs of computing resources increase approximately up to 100 times, an exponential relationship between the sizes of artificial intelligence and the quality of decisions made by it is seen.

One of the most important task is the choice of the structure used by the neural network. Typically, in the literature on artificial neural networks, networks are divided into single, double and triple layers, as well as networks with more than three neurons. Such a wide variety of neural network structures for biometrics is not relevant. GOST R 52633.5-2011 provides either single-layer or two-layer neural networks. For two-layer neural networks, the functions of the first and second layers are separated. Neurons of the first layer fulfill the function of biometric data enrichment and enriched data quantizing. If the quality of enrichment was not large enough, then the second layer of neurons corrects the biocode errors of the neurons of the first layer.

It should be noted that the second layer of neurons can always be replaced by the usual classical code that detects and corrects errors, but neural network error correction is more advantageous. The advantage of using neural network correctors – during the training of the second layer on the

examples of "Own" biocodes, the real indicator of stability of each of the biocode digits of is estimated.

Practice shows that the vast majority of biocode digits has a high stability, only isolated bits of code with an exactly known position are unstable.

The second layer of neurons train to correct unstable digits and simultaneously to hash (mix) both stable and unstable ones. All classical codes with errors detection and correction are constructed in the context of the hypothesis of equiprobable error distribution between code digits. Only because of this, classical self-correcting codes lose to neural network error correctors, which during the training take into account the real distribution of the stability indicators of the "Own" biocodes.

In addition to the number of network neurons layers, it is necessary to choose the number of inputs of each neuron and to specify the connection of inputs with the numbers of the network inputs. So, if the entire neural network has 480 inputs and the average information of the inputs is about 0.3 bits, then it is needed to use neurons with the number of inputs from 1 to 18 (depending on the quality of the biometric parameters used by the neuron and the correlation between them). The required number of inputs can be found only during the training of the neuron. Initially, a small number of inputs are randomly selected, if the quality of the solution does not reach the specified value, then the number of inputs of the neuron is increased. Ultimately, there is produced a single layer network of neurons, where each neuron has its own number of inputs connected randomly to the inputs of the entire network. After training, in addition, we obtain a table of weight coefficients for the input connections of each of the neurons.

A formally trained network is described by the tables of neuron connections and weight coefficients tables. If the network is two-layer, the tables of connection numbers and weight coefficients tables must be created for each of the layers of neurons. The layers of neurons are trained sequentially. After training the first layer of neurons, the examples of images "Own" and "All Aliens" are broadcast from the input of the neural network to the outputs of the neurons, there are received examples of the biocodes and neurons of the second layer are trained on them.

Tables of neural network connections and weight coefficients tables of the trained neurons form the so-called neural network containers. In the neural network container there is enough information to reproduce at the appropriate moment a software-trained neural network and to convert the biometric data of the person to the code of his cryptographic access key. The authentication procedure built using a neural network container is shown on Figure 4.

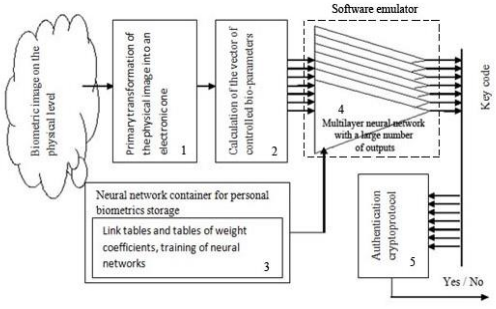


FIGURE 4 Block diagram of biometric authentication using a neural network container

Formation and the use of neural network containers for storage of human biometrics and interfaces of interaction with them are regulated by GOST R 52633.4-2012 [24].

The choice of GOST R 52633.5-2011 for a two-layer network of neurons (perceptrons) is conducted due to the fact that this neural network structure is universal and can solve any problems. M. Minsky and S. Papert [25] showed that a single perceptron is not universal.

In fact, the training of artificial neural networks in accordance with GOST R 52633.5-2011 at increasing the number of inputs for neurons is carried out more and more stable. Therefore, large neural networks after their training work more reliably than classical algorithms of multidimensional statistics and linear algebra. Constructing a quadratic decision rule in the form of a 480-dimensional hyperellipse is technically impossible, whereas it is quite possible to construct an approximation of this hyperellipse by 480-dimensional parallelepipeds (4 perceptrons with 480 inputs in the first layer of the network).

5 Rational choice of the biocode length of the neural network converters

If it is required 256 bits key code length for the subsequent cryptographic authentication, then from this condition it is unambiguous that the biometric-code neural network converter must have 256 outputs. If the neural network converter is single-layer, then the first layer must have 256 neurons.

As it was noted earlier, cryptographic protection is always much stronger than password and biometric protection. This fact is well observed when we change the number of neurons in the first layer of the network. In the first approximation, the probability of biometric errors $P_{2,Б}$ will decrease with the increasing of the number of used neurons, the corresponding decrease curves are shown in Fig. 5. However, the rapid growth of resistance (rapid decrease in the probability of $P_{2,Б}$) does not occur continuously. Usually a linear decrease in the probability $P_{2,Б}$ is observed only at the initial position of the growth of the number of neurons. Further, the growth indicator slows down and, starting from a certain moment, the probability of errors of the second kind generally ceases to decrease.

The moment of stop of the decrease in the probability of errors of the second kind depends on the informative nature of the biometric image of "Own". The Figure 5 presents two curves of the probability of errors reduce in the biometric component of authentication.

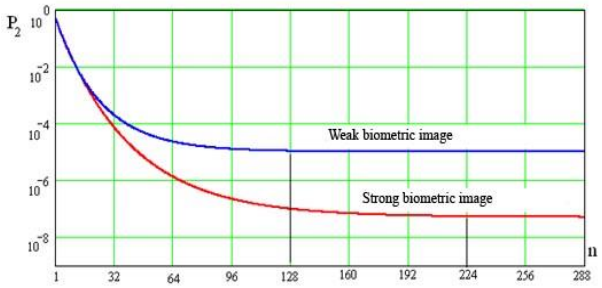


FIGURE 5 The influence of the number of outputs of a single-layer neural network on the probability of errors of the second kind

It can be seen from the Figure 5 that a "weak" small informative biometric image gives a higher error probability

($P_{2,Б} = 10^{-4,7}$), saturation for this image occurs with a key length of 128 bits, that is, with a 256-bit output length of 50% the redundancy of the code, which can be spent painlessly on detection and correction of errors.

A more "stronger" biometric image has a saturation area for codes longer than 224 bits ($P_{2,Б} = 10^{-7,1}$), that is, for this code the redundancy will be 12.5%. This redundancy can also be used to detect and correct errors by classical self-correcting codes. If this redundancy is not enough, then it is necessary to increase the output biocode length to the required.

In some cases (for example, using foreign cryptographic algorithms), the key length may be less than the length before the saturation start area of the biometric-neural network protection. This situation just corresponds to the second case of application of the relatively "strong" biometry (the lower curve of Figure 5) with the required key length of 128 bits. Formally, we can confine with 128 neurons in the first layer of a neural network, but this will lead to a certain losses of quality indicators.

In cases where it is technically advantageous to have a number of neurons greater than the length of the output cryptographic key (for example, 224 > 128), it is necessary to implement a hash of the biocode. After hashing, trim the hash function to the desired length and use it as a key.

In cases when keys are formed outside the biometric application, the value of the hash function can be brought to the specified one if it is added according to the module of two with the addition. If this is necessary, the hashing of the biocode "Own" is performed with "Salt". The “Salt” and the addition of the hash function are stored openly.

It should be noted that the above technique is suitable not only to shorten the length of the key, but also to increase it. This means that for a relatively "weak" biometric image of "Own" (the upper curve of Figure 5), we can dispense with a network of 128 artificial neurons, increasing the length of the biocode by hashing from 128 to 256 bits. This method is suitable for saving computing resources. Cryptographic hashing is usually performed about 1000 times faster than software emulation of an artificial neural network.

6 Neural network biocode correction by the second layer of neurons

In cases when the classic self-correcting codes give too much information loss (require too much redundancy) GOST R 52633.5-2011 recommends the use neurons of the second layer to correct the errors admitted by the neurons of the first layer. Before setting the neurons of the second layer, the standard recommends to estimate the stability of each of the digits of the biocode "Own". For this purpose, test examples of the image "Own" are submitted and the probabilities of errors of the first kind for each of the digits $P_{1,i}$ are calculated. Further, the stability indicators of each of the digits are calculated:

$$\gamma_i = 2 \cdot \left| P_{1,i} - 0.5 \right|, \quad (1)$$

where γ_i – stability indicator of i digit, taking values of 1.0 for absolutely stable digits and a value of 0.0 for unstable digits with equiprobable states "0" and "1".

Adjusting the neuron of the second layer, it is necessary to set its weight coefficients in proportion to the stability index (1). The sign of the weight coefficients is chosen randomly. The inputs of the correction neuron are connected to the digits of the biocode randomly, while converting the values of the digits of the biocode: the state "0" is converted to the state -1, the state "1" is assigned the value +1. As a result of multiplication of random signs of weight coefficients and random states of digits of the biocode ±1, a state close to zero appears at the output of the neuron adder.

The neuron adjusting goes to random cyclic permutations of the sign in the pairs of inputs. Obviously, in a neuron with 64 inputs, the maximum possible output state will be close to +64, and the minimum possible value will be close to -64. If the adjustable neuron must correct 1 error and give the state "1", then it is necessary to achieve the maximum value of the response to the examples of "Own" close to +2. If the adjustable neuron should correct 1 error and give a state of "0", then it is necessary to achieve the maximum response value to the examples of "Own" close to -2. The number of errors corrected by the neuron should always be about one unit less than the module of the maximum response to the examples "Own".

It should be noted that the second layer of neurons performs two functions. First, it corrects the biocode errors of the previous layer of neurons, and secondly, the adder of the second layer of neurons (mix) hashes the data of the codes "Alien".

7 Hashing of "Alien" data performed by neurons of the second layer

The ideal biometric converter should completely eliminate the uncertainty of the codes "Own" and maximize the entropy of the codes "Alien". The input entropies of continuous data of the examples of the image "Own" and of the examples of the image "Alien" are comparable:

$$H_{480}(\bar{v}) \approx \mathrm{H}_{480}(\bar{\xi}), \quad (2)$$

where $H_{480}(\bar{v})$ - input entropy of continuous data of the examples of the image "Own"; $\mathrm{H}_{480}(\bar{\xi})$ - input entropy of continuous data of the examples of the image "Alien".

After implementing of the neural network transformation, the situation changes:

$$\begin{cases} H_{256}(c) \approx 0; \\ H_{256}(x) \approx 256. \end{cases} \quad (3)$$

In fact, small entropy of the codes "Own" and the closeness of the entropy of the codes "Alien" to the limit value of 256 bits determines the closeness of the real converter to the ideal one.

A two-layer neural network improves its properties from layer to layer, designating the codes of the first layer by the index 1, and the codes of the second one – by the index 2, we can write

$$\begin{cases} H_{256}(c_1) \rangle H_{256}(c_2) \approx 0; \\ H_{256}(x_1) \langle H_{256}(x_2) \approx 256. \end{cases} \quad (4)$$

This situation is seen on the corresponding Hamming distance distribution of the codes "Own" and "Alien".

Figure 6 shows that at the output of the neurons of the first layer, the biocode has resistance to picking attacks of approximately 56 bits (the distance between the edges of the Hamming distance distributions of "Own" and "Alien").

After correcting the errors by the neurons of the second layer, the visible distance between the sets is more than 100 bits, but this is an imaginary (overestimated) resistance, due to the fact that the neurons of the second layer not only control the biocode "Own", but also hash the codes "Alien". The hashing properties of the second layer of neurons can be estimated as the ratio of the entropy of the codes "Alien" at the output of the neural network to the entropy of the biocodes at the output of the neurons of the first layer.

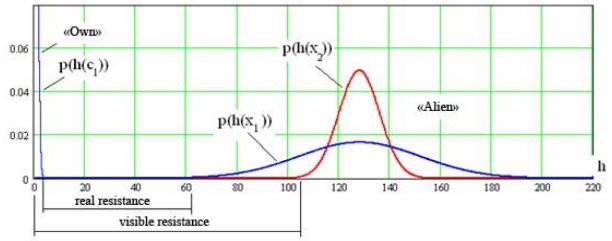


FIGURE 6 Compression of the Hamming distance distribution at the output of neurons of the second layer due to the hashing properties of neurons

The fact that the neurons of the second layer of the network have sufficiently strong hashing properties is explained by the random arrangement of weight coefficients and the rounding operation of the summation results performed by them. Any unidirectional operation leading to a decrease in the length of the code has some hashing properties, because the source data can not be restored from the short output codes. In our case, the summation results of the code correcting neuron can vary from -64 to +64, and its output code has only two values of "0" and "1". There is an operation of truncating the length of the 9-bit code to the 1st digit.

8 Probabilities of errors of the first and second kind

After training the system of biometric-neural network authentication it is necessary to evaluate the quality of training. The probability of an error of the first kind P_1 and the probability of an error of the second kind P_2 are estimated.

The user must know the real assessment of the resistance to the specific implementation selection attacks of biometric authentication after its training, built on the reproduction of a specific secret biometric image. Testing is carried out using N_1 - test examples of the vectors of the image "Own" and N_2-test examples of the vectors of the image "Aliens". The structural scheme of testing is shown on Figure 7.

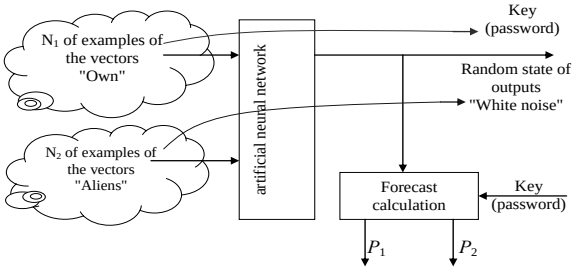


FIGURE 7 Structural scheme of testing the biometric-neural network authentication system after training

Any biometric protection is built on the fact that it is able to recognize the image "Own" and reliably to isolate many images "Alien" ("All Aliens"). Obviously, the means of biometric protection (biometric authentication) can be mistaken. The main task (task number 1) for biometrics is to provide the donor with a biometric image "Own". An error in the execution of this task is treated as an error of the first kind. The main characteristic of the efficiency of the biometric authentication tool is the probability of errors of the first kind P_1.

The second task of the biometric authentication tool is to prevent the donor from accessing the "Alien" image. The second most important characteristic of biometrics is the probability of errors of the second kind P_2 due to possible collisions of the images "Own" and "Alien" on the set of features (biometric parameters) under consideration.

Obviously, the probability of errors of the second kind P_2 will be less, if more biometric parameters take into account this or that biometric authentication tool. Only those biometric tools that analyze hundreds or even thousands of biometric parameters can be considered reliable. In this case, the attacker should not know the selected biometric image, only in this case the biometry can be considered as highly reliable.

9 Software implementation of handwriting image recognition

A necessary condition for training the skills of stable spelling of the recommended word and the correct formation of an impersonal base of biometric images is the familiarization by the donor with biometrics of the specialized software and hardware complex "Neuro-Test 1.2" [26].

The program uses an emulator of an artificial neural network, which has many outputs. The number of outputs of an artificial neural network is determined by the length of the biometric key generated by it. This excludes the hacking of the program through the detection and substitution of the last bit of the decisive rule. The program has a multi-bit decisive rule, the bit values combination of which is unique and unknown to the attacker. The program "NeuroTest 1.2" uses the algorithm of rapid automatic training of the artificial neural network with 256 outputs.

The main window of the program is shown on Figure 8.

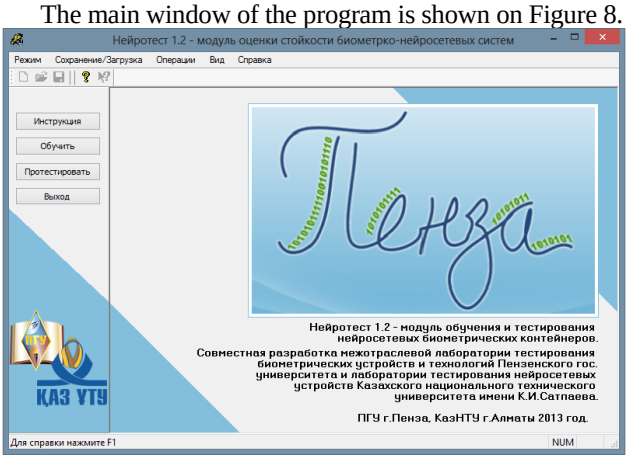


FIGURE 8 The main window of the program "NeuroTest 1.2"

Training mode. Training begins with the initialization of the training mode (the "Train" button or the "System training" mode).

Next, you need to reproduce a pre-defined handwritten word by your hand on the field of the “Genius” graphic tablet.

After entering the handwritten word, click the "Add" button, the lined field is cleared, and the number of the next

entered example appears in the right part of the window. If a hand trembles during entering a handwritten word-password, or the image is not typical, click the "Clear" button. In this case, the word is deleted without entering into the database of examples. You need to reproduce the word for 20 times. The input of handwritten images is controlled by the instructor.

You can view saved examples by clicking on the number in the right part of the window (Figure 9).

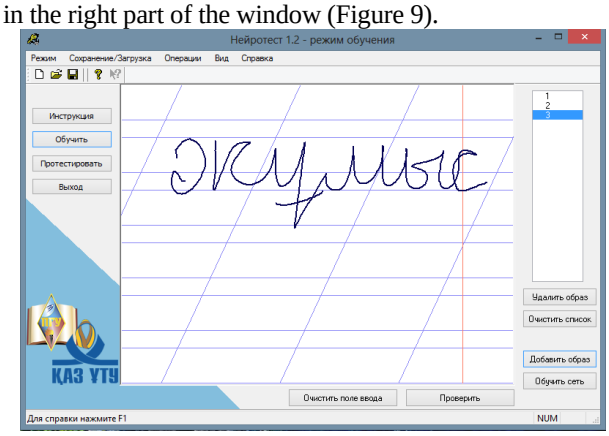


FIGURE 9 Handwritten images view

If any example does not match the handwriting, then it must be deleted by clicking on the "Delete image" button. If you do not like all the images or you need to change the word-password, then all the images in the list are deleted (the "Delete All" button).

Training of the system (neural network) is carried out by clicking the "Train" button. In this case, for a time up to 30 seconds a window appears with a forecast of the probabilities of system errors and with the number of the group to which you belong (Figure 10).

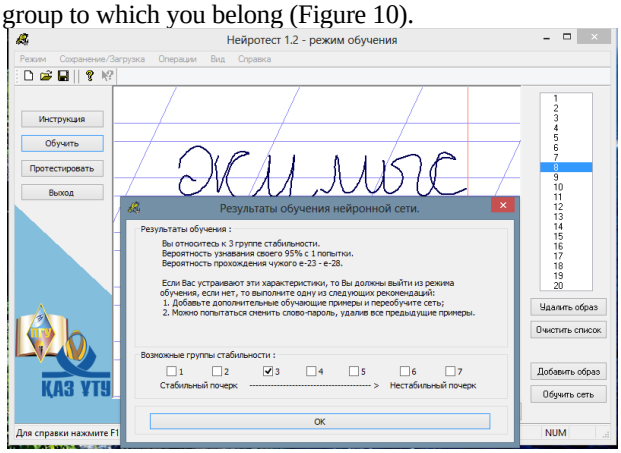


FIGURE 10 Completion of neural network training

If you do not like the group number and probabilistic characteristics, you can independently change the group number and retrain the neural network. It is important to remember that it is not recommended to change the group number by more than 1 to 2 positions, therefore, you can only switch to neighboring positions.

Perhaps the user was in the seventh group, for which the probabilistic characteristics of the system are too bad. You can change the training environment by deleting the most unlikely example. If there are difficulties with choosing the worst example, you can delete all the training examples and try to write them again, or add a few more training examples. After deleting the worst or after adding an additional example, click the "Train" button again. If you still can not get into the desired group, then the most unlikely example is deleted, or one more additional one is added and the network is trained again. If, after several attempts to re-train the network, you can not reach the desired strength group, you must try to change the word.

Controlling the recognition of "Own". After training the neural network it is necessary to check the quality of recognition by the system of "Own". The training test is performed by reproducing a handwritten word and by pressing the "Check" button. In this case, a window appears (Fig. 11) showing the key generated by the neural network in binary and hexadecimal encodings (the unmatched bits of the source key and of the generated one are marked with asterisks). The total number of unmatched binary key symbols is also displayed.

These data can be used for independent statistical testing of the system.

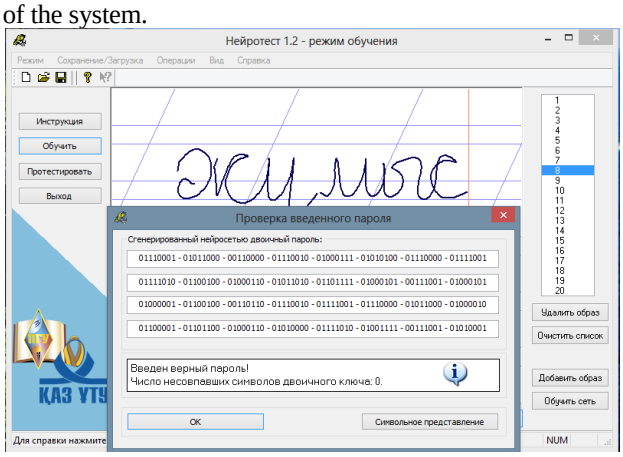


FIGURE 11 **Enter of an image** "Own"

If the system recognize the image "Own" badly, then it is necessary to add the unrecognized image to the base of the training examples and train the network again. After adding several new images, the network will be able to recognize the user inputting a handwriting image, but some characteristics of the system may deteriorate.

If the results of the training suit (GROUP 3 ... 4), then you need to save the data to disk. The folder number corresponds to the number of the biometric donor's personal identifier.

Test mode. The testing of the system begins with the initialization of the testing mode (the button "Test" or "Mode >> Network Testing").

Before you start testing, you need to make sure that the neural network is trained. In the test mode, you can check the input of handwritten images on the trained network, and you can also add a "good" image to the base of the training examples.

The system is tested by reproducing the handwritten word-password and by pressing the button "Check the entered password". The traffic light in the upper right corner lights up. The red light of the traffic light corresponds to a very large discrepancy between the dynamics of the reproduction of the handwritten word-password and the newly entered verification word (Figure 12).

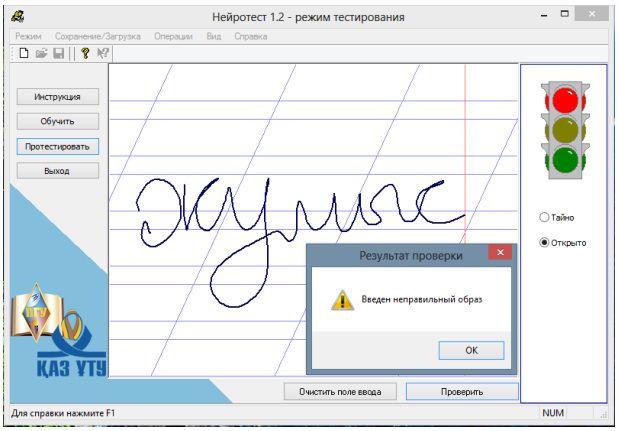


FIGURE 12 Example of entering an incorrect handwritten password

Frequent red light in the verification mode during the presentation of "Own" handwritten word indicates a poor recognition. The yellow light illuminates when several bits of the key do not match, the image is close to the reference one. The green light corresponds to the complete coincidence of the entered image with the reference one (the key is reproduced by the neural network without errors).

10 Conclusions

The two-layer network of neurons is universal in combination with the going next digital cryptographic authentication machine. This combination allows you to create any biometric information protection applications. An increase in the number of neurons in the layers of the network (in the first and in the second) is a very, very effective technique.

Nowadays, the best means of highly reliable biometric authentication provide the probability of errors of the second kind at a level of one billionth or less, that is, an attacker trying to overcome biometric protection must produce a billion different biometric images (for example, to reproduce with his hand a billion manuscript passwords). If the reproduction of one handwritten password takes 10 seconds, then the attacker will need 10 billion seconds, which is 321 years of continuous efforts. This is a lot more time for one person life. Alone, by the simply substituting of real biometric images it is impossible to overcome a highly reliable protection by the secret biometric image.

References

[1] GOST R 52633.5-2011 2012 *Information protection. Information protection technology. Automatic training of neural network converters of biometry access code* Standartinform: Moskow

[2] Volchikhin V I, Ivanov A I, Funtikov V A 2005 *Fast algorithms of training neural network mechanisms of biometric-cryptographic information protection* Monograph Penza State University: Penza p. 273

[3] Akhmetov B S, Ivanov A I etc 2013 Connection between high-dimensional entropy and high-dimensional correlation with mathematical expectation of coefficients modules of pair correlation *Bulletin of KBTU* 240-4

[4] Akhmetov B S, Ivanov A I etc 2013 Estimation of the probability of the errors of the neural network converters biometry-code on the basis of small samples. High technology - the key to sustainable development *Mat. II Int. Sci. Conf.* **1** 234-4

[5] Akhmetov B S, Ivanov A I etc 2011Synchronization of procedures for neural network training of handwriting features by the points of change in the direction of movement of the pen *Proceedings of the II Int. Scientific-prac. conf. "Information and Innovation Technologies: Integration of Science, Education and Business"* **2** 118-6

[6] GOST R 52633.3-2011 2012 *Information protection. Information protection technology. Testing the resistance of means of highly reliable biometric protection to selection attacks* Standartinform: Moskow

[7] Akhmetov B S, Volchikhin V I etc 2013 Algorithms for testing of biometric neural network information protection mechanisms *Monograph* LEM: Almaty p.152

[8] Akhmetov B S, Zakharov O S etc 2013 A method for estimating the probabilities of errors in neural network converters biometry-code using very small test samples. *Bul. of KazNTU* **3**

[9] Akhmetov B S, Ivanov A I etc 2013 An entropy-correlation approach to the calculation of the probability of joint occurrence of a large number of dependent events *Bulletin of KBTU* **2**(25) 55-5

[10] Akhmetov B S, Ivanov A I etc 2013 Biometric authentication of citizens in open information spaces. *Proceedings of the 1 st international scientific and practical conference "Intellectual information and communication technologies - the means of implementing of the third industrial revolution in the light of the" Kazakhstan-2050 "strategy"* 458-2

[11] Malygin A Yu, Akhmetov B S etc. 2012 Accounting of the correlations influence on the results of testing of the biometry-code converters *Information and telecommunication technologies: education, science, practice: Col. pr. of Int . Scientific-practical. Conf.* 34-4

[12] Ball Rud etc 2007 *A guide to biometrics Technosphere* Moskow p. 368

[13] Larin P Z, Rever E I 2004 Circle around the finger? Deception of Biometric Access Systems Using Dactyloscopic Identification of Personality *Information Security* **3**

[14] Arakala A, Jeffers J, Horadam K J - 2007 Fuzzy Extractors for Minutiae-Based Fingerprint Authentication *Advances in Biometrics (LNCS 4642). Springer* 760-10

[15] Ratha N K, Chikkerur S 2007 *Generating cancelable fingerprint templates IEEE Trans. PAMI* **29**(4) 561-12

[16] ISO / IEC 19784-1: 2006 *BioAPI - Biometric Application Programming Interface - Part 1: BioAPI Specification* http: //www.bioapi.org/

[17] GOST R ISO / IEC 19784-1-2007 2007 *Automatic identification. Biometric Identification. Biometric program interface. - Part 1. Specification of biometric software interface* Standartinform: Moscow

[18] GOST R ISO / IEC 19784-2-2010 2011 *Automatic identification. Biometric Identification. Biometric program interface. - Part 2. The interface of the supplier of the biometric archive function* Standartinform: Moscow

[19] Kruglov V V, Long M I, Golunov R. Yu 2001 *Fuzzy logic and artificial neural networks* Fizmatlit:Moscow p. 221

[20] Rutkovskaya D, Pilinsky M, Rutkowski L 2004 Neural networks, genetic algorithms and fuzzy systems *Trans. from Polish by Rudinsky I.D.* Hot line - Telecom

[21] Simon Khaikin 2006 *Neural networks: Full course* Williams:Moscow p. 1104

[22] Wesserman F 1992 *Neurocomputer technology: theory and practice* Mir:Moscow p. 240

[23] Ivanov A I 2000 *Biometric identification of the personality according to the dynamics of subconscious movements* Monograph Publishing House of the PGU:Penza

[24] GOST R 52633.4-2012 2012 *Information security. Information protection technology. Interfaces of interaction with neuronet converters biometry-code* Standartinform: Moscow

[25] Minsky M, Peipert S 1971 *Perceptrons* Mir:Moscow

[26] Akhmetov B S, Alimseitova Zh K, Malygin A Yu, Yubuzova H I 2015

Formation of the biometric base of handwritten images in the Kazakh language for biometric authentication programs of the individual *Proceedings of the II International Scientific and Practical Conference "Information and Telecommunication Technologies: Education, Science, Practice"* **2** 32-4

AUTHORS

Malygin A

Current position, grades: Professor of the military Department of the Penza state University
University studies: doctor of technical Sciences
Scientific interest: information security, neural networks, biometric systems
Publications (number or main)**:** More than 100 publications in the field of information security
Experience: teaches at Penza state University

Nurgul Seilova

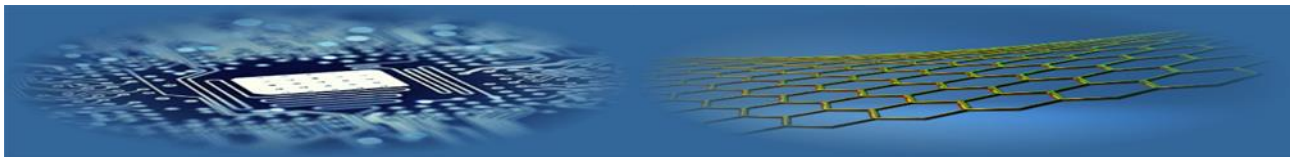
Current position, grades: Head of department Information Security
University studies: candidate of technical Sciences
Scientific interest: network technologies, Information Security
Publications (number or main)**:** More than 30 publications
Experience: 15 years of teaching experience and 3 years in managerial positions

Kylychbek Boskebeev

Current position, grades: Professor of Kyrgyz State Technical University after I.Razzakov
University studies: Ph.D. (Engineering)
Scientific interest: The development of Intellectual information systems
Publications (number or main)**:** 7 Russian Scientific Citation Index
Experience: experience of service is 20 years

Zhuldyz Alimseitova

Current position, grades: lecturer of the Kazakh national research technical University named after K. I. Satpayev
University studies: Kazakh national technical University (1999)
Scientific interest: information security, neural networks, biometric systems, cryptography
Publications (number or main)**:** More than 30 publications in the field of information security
Experience: teaches at Kazakh national research technical University named after K. I. Satpayev discipline on information security and information protection



Synthesis and simulation of digital pseudo-random impulse sequence generator based on PLIC FPGA Xilinx using CAD Vivado 2016.2 and development of acoustic noise generator scheme for the protection of information

A Zaurbek, N A Seilova, D Z Dzhuruntaev*

Kazakh National Technical Research University named after K.I. Satpayev, Almaty, Kazakhstan

**Corresponding author's e-mail: juruntaev@rambler.ru*

Received 27 February 2017, www.cmnt.lv

Abstract

In this work with the help of CAD Vivado 2016.2 system and Verilog hardware description language there were synthesized, simulated and built temporary digital pseudo-random impulse sequence generator diagrams based on CAD of FPGA families of the Xilinx company and eight-rate shift LFSR register, which can be used in cryptography to create a stream encryption algorithms. On the basis of a digital pseudo-random impulse sequence generator and active low-pass filter of the second order of Sallen - Key there was constructed an electric diagram of the acoustic noise generator that provides protection against wiretapping by using embedded devices, telephone conversations, laser wiretapping system and unauthorized dictaphone recording of confidential voice information by creating a masking vibration noise.

Keywords:

computer-aided design,
hardware description languages,
programmable logic integrated circuits,
synthesizing, simulation,
circuit simulation,
pseudo-random impulse sequence

1 Introduction

Programmable logic integrated circuits (PLIC) are one the fastest growing and promising elements of the digital circuitry. Today's widely used at present micro schemes of the programmable logic PLIC are the crystals on which there are hundreds of thousands or more of simple logic elements and triggers that allow to obtain the layout (prototype) of the digital device of almost any complexity.

Programmable logic integrated circuits provided to create quickly digital devices with an internal structure, defined by the user, that is, rapid conversion of one configuration of the digital circuit to another. In other words, the change of the principal electric scheme of the digital device in a PLIC crystal is performed by reprogramming. As a result, the cycle of creation of complex digital devices, including the development of multiprocessor systems and parallel processing systems with large amounts of data accumulation, is greatly reduced that reduces the cost of the whole project [1-4].

The most widespread in this area have PLIC with the FPGA type architecture (field-programmable gate array). PLIC FPGA advantage is their ability to provide not only high speed of processing, but also continuous processing and stable speed.

Integrated Circuits of a special purpose ASIC (abbreviation from English - Application-specific integrated circuit) run faster PLIC FPGA, but are not used to protect information systems. ASIC are programmed during manufacturing and have no possibility of reprogramming.

PLIC with FPGA architecture provides the broadest functional possibilities and the largest number of hardware resources, therefore, have the greatest interest in the development of digital devices. PLIC micro schemes are widely used in digital signal processing system, as well as in information protection issues, in particular, for generating a pseudo-random impulse (numbers) sequence.

The need of a digital pseudo-random number generator usage occurs in many problems. Digital hardware random number generators are used for static testing and in cryptography, where they are used to create cryptographic keys for the encrypted data [1, 2].

Considering the above, the task of digital devices developing, including digital circuit of the pseudo-random impulse sequence generator based on PLIC with FPGA architecture, is relevant and has practical interest for the information security.

For effective use of PLIC FPGA micro schemes there is a need to know and understand the different approaches and aspects of the synthesis and simulation of digital device circuits that is almost impossible without the use of computer-aided design. The most commonly used are special tools of CAD Vivado of the Xilinx company, PLIC FPGA manufacturer [5, 6]. The process of digital devices designing in CAD Vivado includes the following steps: the creation of modules of an initial description of the designed device, the project synthesis and implementation on the basis of PLIC of FPGA families of the Xilinx company and simulation of digital devices.

Nowadays, for the creation of digital devices within a reasonable terms and with a high quality based on the PLIC micro schemes, containing hundreds of thousands or millions of logic gates, there are used effectively Vivado 2016.2 CAD and hardware description language such as Verilog or VHDL [5-7]. High-level languages VHDL, Verilog and Verilog System in CAD Vivado are used for

modeling and for creation of the synthesizable descriptions. Vivado important property is the ability to control the entire development cycle using Tcl scripting language, which is the basis of the new format of the description of design constraints xdc (Xilinx Design Constraints). Xdc format has, in comparison with the ucf format, previously used, more flexible description of design constraints, which facilitate the construction of scalable projects.

The aim of this work is to create a project using CAD Vivado 2016.2, associated with the description of the designed device on a high-level Verilog language, synthesis, simulation, with the temporary work diagrams construction and with the diagram implementation of digital generator pseudo-random impulse sequence based on PLIC of FPGA families from Xilinx company and eight-rate linear feedback shift register, as well as the development of an electrical circuit of the noise generator, which creates an acoustic and vibro-acoustic noise to protect speech information.

2 Synthesis and simulation of digital pseudo-random impulse sequence circuits

In order to achieve the aims of the work there is launched CAD Vivado 2016.2 by the shortcut on your desktop. Opens the start screen of CAD Vivado 2016.2 Quick Start (Figure 1).

FIGURE 1 CAD Vivado 2016.2 Quick Start

As can be seen, on the Quick Start screen there can be created a new project, opened an existing one and you can familiarize with ready examples of projects or documentation CAD Vivado 2016.2.

In order to create a new project on the Quick Start screen select the command Create New Project and open the Project Name window (Figure 2).

In the Project Name field enter the name of the project, and in the Project location field specify the directory of location of the project, where it will be stored.

Thus, in the Project location field you can see the path to the previously created folder on the desktop.

Open the Project Type window and select the type of the RTL (Register Transfer Level) Project to create a project on the level of register exchanges using hardware description languages, for example, Verilog. Register transfer level (RTL) - a way of describing the operation (behavior) of synchronous digital circuits on the register level, logic signals and logic operations on the signals.

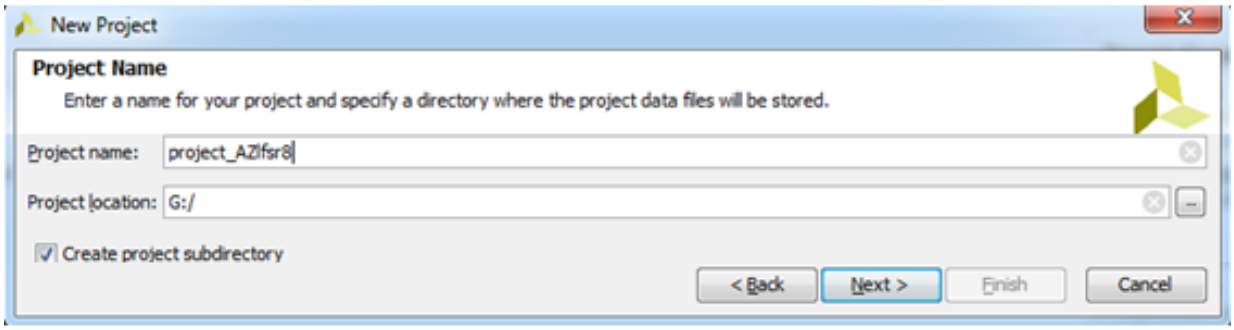


FIGURE 2 Project Name window for new project creation

Circuit or description entering of the designed device can be carried out by various methods, including circuit design.

In this work in order to describe the projected devices we use high-level Verilog language. Open the Add Sources window and then with the help of Create File button open a Create Source File panel (Figure 3).

In the File type field select Verilog, in the File Name field specify the name AZregoc8 file, and in the File location field specify the path where the file is stored.

Next, using a USB cable, connect BASYS 3 DIGILENT board to the computer.

In the FPGA folder select Basys3_Master.xdc file that allows you to add pre-defined rules and restrictions. Specify the location of the Basys3_Master.xdc file. Open the Default Part window (Figure 4).

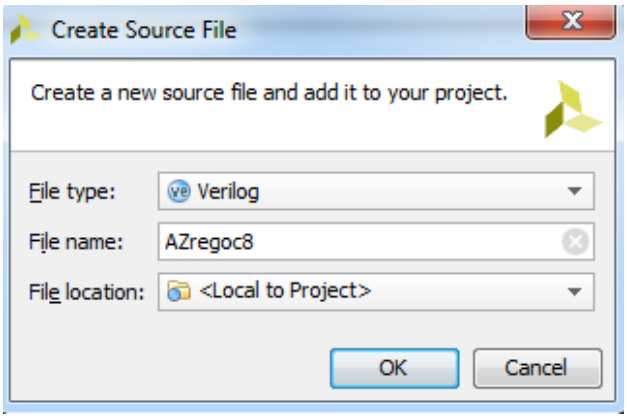


FIGURE 3 Create Source File panel

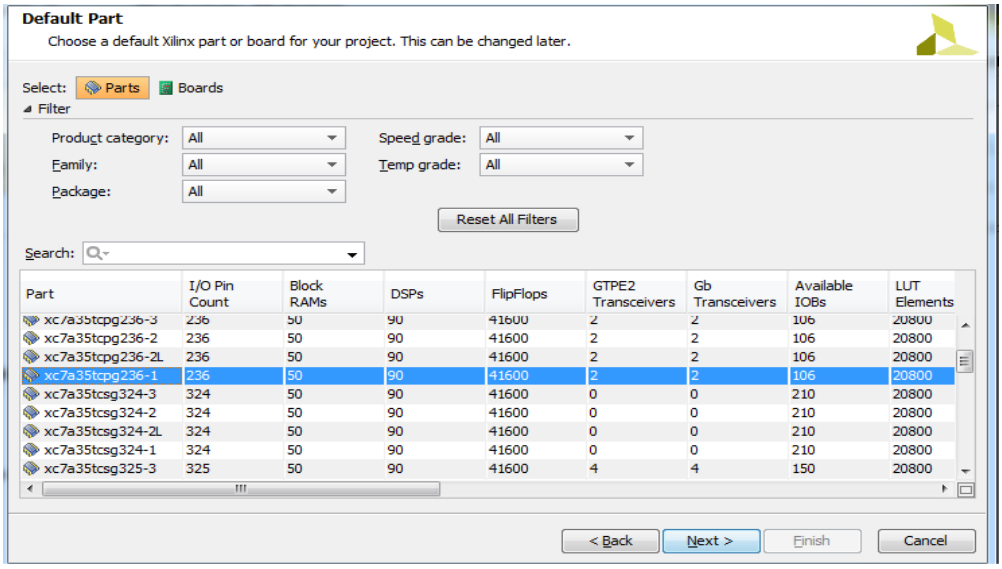


FIGURE 4 Default Part window

From the table Part, which lists the various micro schemes models of the BASYS DIGILENT board, choose xc7a35tcpg236-1 model and enter it in the Search field for the project implementation.

The window New Project Summary opens, click on the Finish button. In the Define Module window in the Port Name field specify the names of the input and output ports in accordance with the scheme or program on Verilog language of the designed digital pseudo-random impulse sequence generator on the basis of eight-rate shift register LFSR with synchronous pulse clk, rst reset signal, enable permission signal and output signals reg [7: 0] Q (Figure 5).

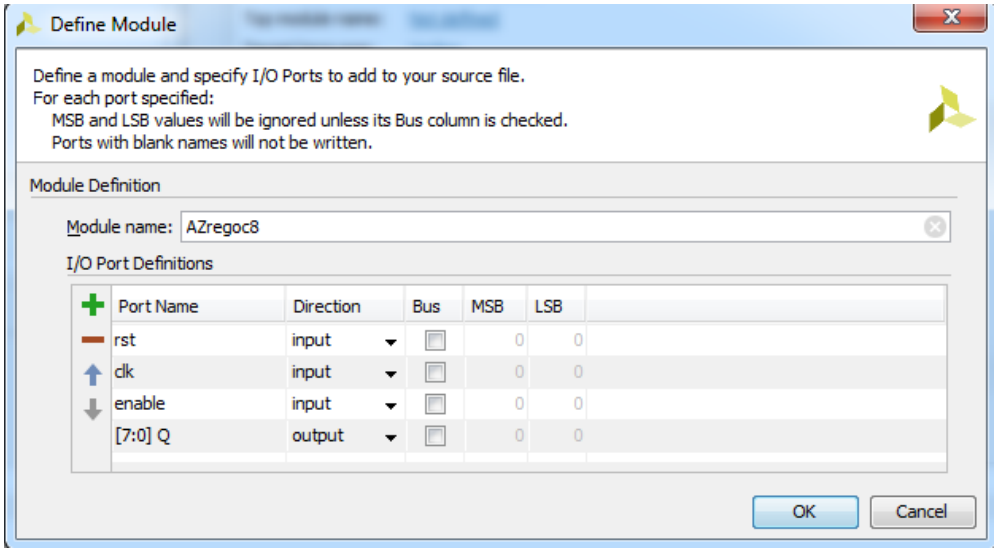


FIGURE 5 Define Modules window

Open the Project Manager window (Director - Project Manager). In this window open the Restrictions folder and select the file Basys3_Master.xdc in the folder costrs_1 (Figure 6). Open it with the Open Selected Source Files.

From the set of signals description Basys3_Master.xdc, which also shows the results of the BASYS 3 DIGILENT board, select and activate the input and output signals needed for our project. Basys3_Master.xdc file contains information for the development environment how the logical inputs and outputs AZregoc8 of the main module are connected with the location (LOC) of the physical legs of FPGA micro schemes, which are referred as PIN W5, PIN V17, PIN V14, etc. On the Figure 7 there is shown a stripped-down version of Basys3_Master.xdc file for this board.

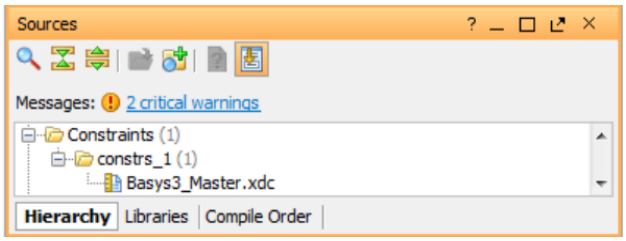


FIGURE 6 Project Manager window

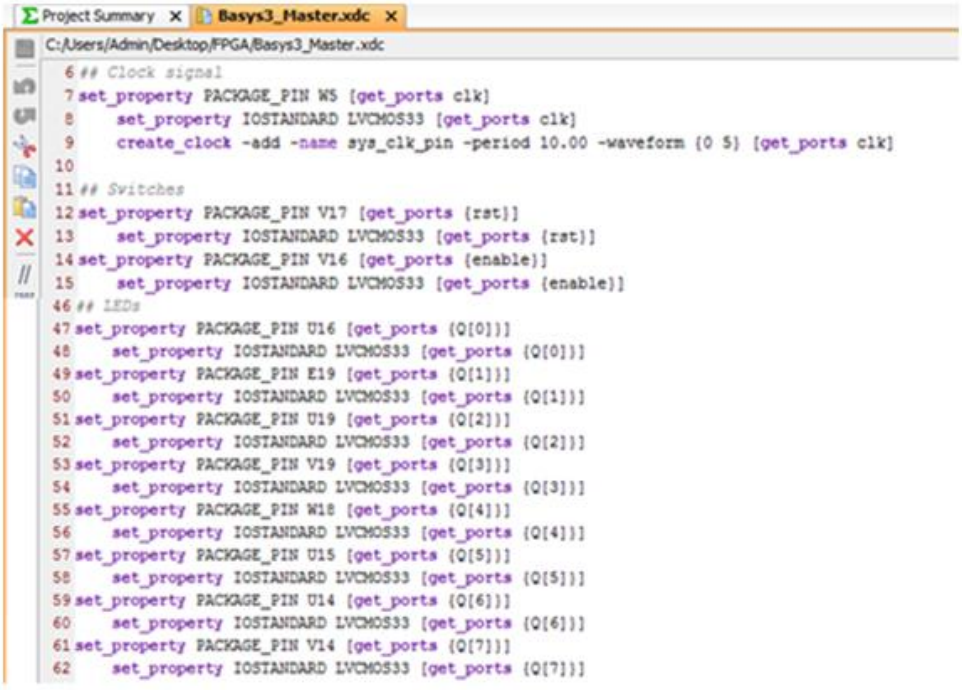


FIGURE 7 Basis3_Master.xdc file screenshot

Close Basys3_Master.xdc file. Next, in the Desing Sources folder open AZregoc8.v file and in the file add the pre-composed code text of the program on a high-level Verilog language in this file, which describes the operation of the digital generator pseudo-random impulse sequence diagram based on the eight-rate shift register LFSR. Close AZregoc8.v file. Below, on the Figure 8 there is a screenshots of AZregoc8.v file. Close AZregoc8.v file. So the project is created. Next, using the Project Settings and the bit sequence Bitstream - bin_file, carry out the project setting. In order to open the synthesized circuit in Synthesis folder click on the Run Synthesis command.

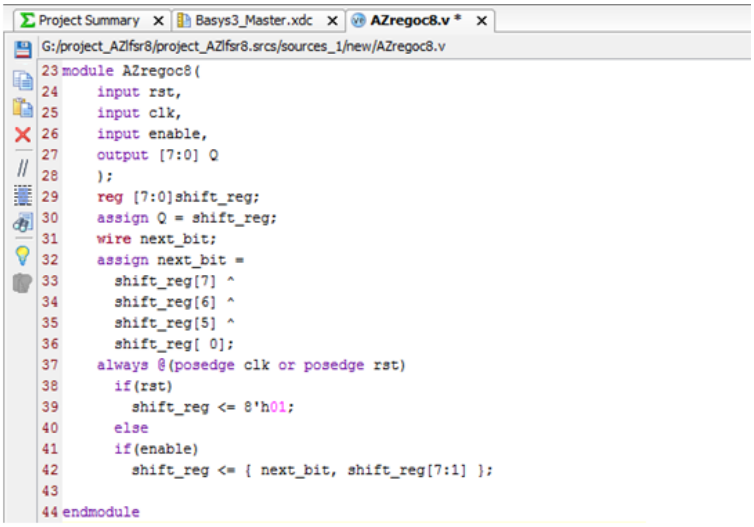


FIGURE 8 AZregoc8.v program screenshot

After the synthesis using the command Open Synthesized Design get a diagram of a digital pseudo-random impulse sequence generator based on the eight-rate shift register with the feedback LFSR, described on a high-level Verilog language (Figure 9).

In order to obtain a digital pseudo-random impulse sequence generator circuit based on RTL - register transfer level run the Open Elaborated Design command. In the Schematic window appears abstract diagram of a digital pseudo-random impulse sequence generator based on eight-rate LFSR register with inputs clk, rst, enable, and outputs Q [7: 0] (Figure 10).

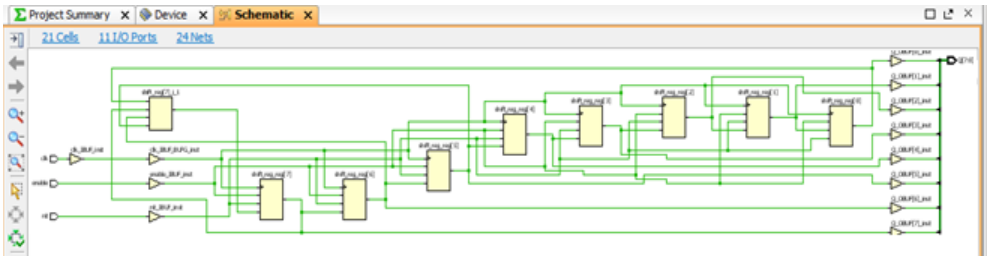


FIGURE 9 Screenshot of the digital pseudo-random impulse sequence generator based on the register LFSR

3 Construction of the temporary diagrams of the digital pseudo-random impulse sequence generator circuit operation

One of the major stages in the digital devices development based on PLIC is the simulation of the developed devices. In order to perform a behavioral simulation of the digital pseudo-random impulse sequence generator circuit based on the LFSR register run the Run Behavioral Simulation command. There will be opened Behavioral Simulation panel.

In the Objects or Untitled 1 window indicate the necessary parameters of the circuit simulation, the desired value of the clock impulses clk, rst reset signal and the enable signal of the digital pseudo-random impulse sequence generator circuit based on eight-rate shift register LFSR.

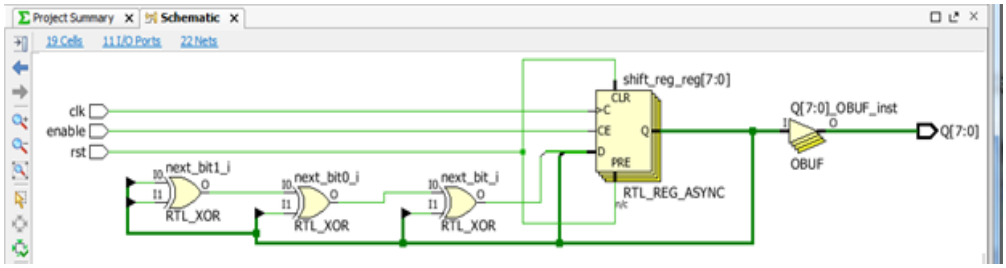


FIGURE 10 Screenshot of the the digital pseudo-random impulse sequence generator abstract circuit based on the LFSR register

At the beginning with the help of the Force Constant command make the shift register LFSR reset to "0". To construct a temporary operation diagrams of the digital pseudo-random impulse sequence generator based on the LFSR register select Force Clock command and indicate the threshold of the signal operation and the threshold of the trip signals.

Next, specify the time values for the beginning and completion of the register operation. Select the duration of the clock impulses clk. Specify the fill coefficient (the ratio of the impulse duration to the interval between impulses). Run the digital pseudo-random impulse sequence generator circuit based on the LFSR register whose content according to each clock impulse clk is shifted to the right, and the new calculated bit slides into the shift register LFSR to the left.

The calculation of the new bit is made by the operation OR (XOR). The calculation of the new value of the shift register LFSR occurs in the line shift_reg <= {next_bit, shift_reg[7: 1]}. The next shifted bit next_bit is calculated by the assignment operation: assign next_bit = ... (see Figure 8).

Figure 11 shows the temporary diagrams of the digital pseudo-random impulse sequence generator based on eight-rate shift register LFSR.

FIGURE 11 Screenshot of the digital pseudo-random impulse sequence generator based on eight-rate shift register LFSR temporary operation diagrams

For the implementation of the digital pseudo-random impulse sequence generator circuit based on the shift register LFSR (for the programming of PLIC) activate Run Implementation command on FPGA BASYS 3 DIGLENT board and run the Program and Debug Generate Bitstream. If the board is connected by the USB cable to a computer, the lights on the board start to blink in accordance with the impulses generated by the digital pseudo-random impulse generator based on the shift register LFSR (Figure 12). Therefore, the shift register LFSR with the feedback operates as the digital pseudo-random impulse generator with pseudo-random intervals between them.

With the help of CAD Vivado and Verilog language there have been created various digital devices projects (decoder, multiplexer, D-trigger, adder, register, digital counter, etc.), related to the creation of the xc7a35tcpg236-1 micro scheme describing modules of the Basys 3family, with the synthesis, simulation and project implementation on the basis of PLIC of the FPGA families from Xilinx company.

FIGURE 12 PLIC FPGA BASYS 3 DIGILENT board

The frequency of the generated impulse sequence of the shift register LFSR with the feedback allow to use it as a digital counter. A digital counter based on such generator has a simplified feedback circuit, unlike ordinary binary counters, and therefore, can operate at high clock frequencies. However, you must make sure that such a digital counter is never entered in the zero state. For this purpose there is an enable signal in the register. Unlike conventional digital counter, shift register LFSR with the feedback moves from one state to another not in a binary count, that allows to use it to generate a test signal for error detecting in logic circuits. Digital pseudo-random impulse sequence generator based on shift register LFSR is used very often for stream ciphers in cryptography. Large random numbers generated from the sequence bits of the digital generator based on shift register LFSR are strongly correlated and sometimes not even random at all. However, digital generators based on the shift register LFSR can be used to create basic cryptographic algorithms [9, 10]. It should be noted that the digital pseudo-random impulse sequence generator synthesized using CAD Vivado 2016.2, can be used to generate masking vibration noise.

4 Development of the acoustic noise generator scheme

Noise vibrations are created for protection from illegal removal of confidential acoustic (speech) information. For wiretapping and unauthorized recording of voice information attackers can use a variety of technical means: wiretapping devices, laser wiretapping system, stethoscopes, voice recorders, and others.

In order to mask speech signals there is proposed noise generator electrical scheme, which creates an acoustic and vibro-acoustic noise (Figure 13). The structure of the acoustic noise generator includes a digital pseudo-random impulse sequence generator, built on the basis of eight-rate shift register with the feedback on the triggers of the D-type, multivibrator which generates the clock impulses, logic elements (LE) "OR" (XOR) and "NOT" by which the feedback is carried out, active filter of the low frequencies of the second order at the operational amplifier (OA) and piezoceramic transducer ZQ.

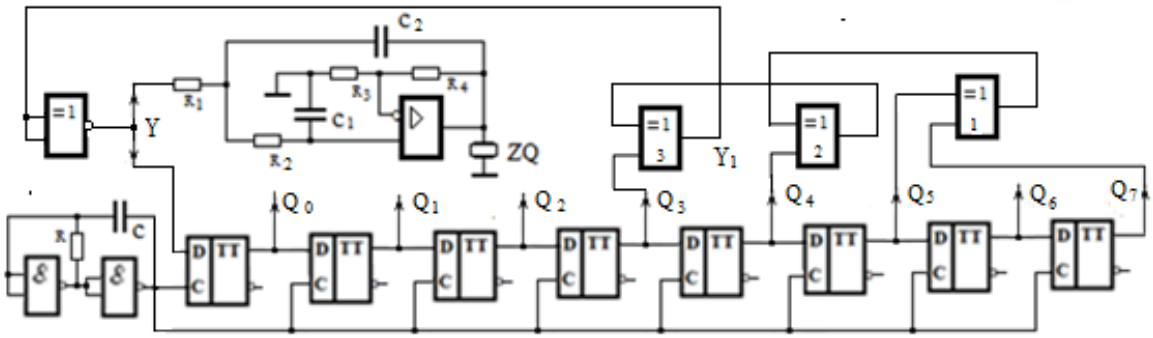


FIGURE 13 Digital pseudo-random impulse sequence generator and acoustic noise electrical scheme

Consider the operation principle of pseudo-random impulse sequence generator circuit on the basis of the shift register with the feedback. Zero state of the shift register with the feedback when the triggers of all bits are at logic 0 state (output trigger signals $Q_0 = Q_1 = \ldots = Q_7 = 0$) is not working. In other words, the output code 000 ... 0 is a disabled state, as it blocks the operation of the digital pseudo-random sequence generator. In the shift register there should be entered something other than zero, otherwise the shift register will remain at zero and the zeros will run around in the shift register.

Feedback is carried out from the trigger outputs of the 7th-, 5th-, 4th- and 3-rd register bit through the two-input logic elements "OR» (XOR – addition according to the module 2). To exclude the zero forbidden state of the shift register at the output of the last LE-3 XOR there is used an inverter (NOT logic element). Due to the use of the inverter the forbidden state of the digital generator will be a 1111 ... 1 code (instead of the code 000 ... 0), which in this case is eliminated by the initial resetting of the shift register to zero

at power-on signal R_a (shift register reset signal to zero is not indicated in the diagram). The generator produces a pseudo-random sequence of eight-bit codes from all register triggers, as well as pseudo-random sequence of zeros and ones from the output of any of the register triggers.

At zero register state at the logic element LE-3 "OR" output there will be a logic 0 signal ($У_1 = 0$), and the output of the LE "NOT" - logic 1 ($У = 1$) and the unit signal arrives to the trigger input of the zero bit of the shift register.

In order to obtain the first clock impulse of the multivibrator the feedback signal $У = 1$ from the output of the digital generator is recorded to the trigger of the zero register bit and at the same time the register content is shifted by one bit to the right. At the same time the register content is as follows: $Q_0 = 1$, $Q_1 = Q_2 = Q_3 = Q_4 = Q_5 = Q_6 = Q_7 = 0$ that corresponds to the number 1. The feedback signal is $У = 1$. Therefore, after the second clock impulse to the zero bit trigger the signal of a logic 1 is recorded, and the register content is shifted right again by one bit and the register will has a number 3 ($Q_0 = Q_1 = 1$, $Q_2 = Q_3 = Q_4 = Q_5 = Q_6 = Q_7 = 0$).

The feedback signal У will still be equal to 1 ($У = 1$). This 1- signal is recorded to the trigger of the zero bit after the third clock impulse and after register content shift will have number 7. After next multivibrator clock impulses the bit triggers state and the register content (register number) will vary pseudo-randomly, that means - generated sequence of numbers (impulses) will be pseudo-random before 2^8-1 impulse.

In general, at n-bit shift register there can be generated m - code sequences of pseudo-random impulses where $m = 2^n - 1$. The pseudo-random code numbers (impulses) sequence differs from a true random interval, but within a period has no differences from the truly random.

Pseudo-random code numbers sequence corresponding to $m = 2^n - 1$ can be removed from any trigger output of shift register any bit as the same sequence comes with temporary shift from the trigger output of each bit. At a relatively large n value a pseudo-random sequence is virtually identical to the random sequence.

It should be noted also that acoustic noise generated by a digital pseudo-random impulse sequence generator, also provides protection against wiretapping by using embedded devices and dictaphone recording in the office of the head of the organization or the negotiations conducted in specially designated rooms for this purpose.

The digital output of the shift register, generating a maximum length sequence, can be converted into white noise with limited band by using a low frequency filter whose cutoff frequency is considerably lower than register clock frequency. The useful range of the noise generated by a digital pseudo-random sequence generator extends from the low frequency border, reverse to the repetition period, to high frequency border equal to about 20% of the clock frequency (at this frequency the noise power per hertz falls by 0.6 dB) [5]. In order to use a part of the spectrum, much closer to the clock speed, it is advisable to apply filters with a steeper cut, such as Butterworth, Chebyshev or Sallen - Key filters.

In this work, in order to create acoustic noise to the output of a digital pseudo-random impulse sequence generator there is connected the active low-pass filter (LPF) of a Sallen - Key second-order on the basis of the operational amplifier, a load of which is a piezoceramic transducer ZQ (Figure 13). In the Sallen-Key scheme the capacitors C_1 and C_2 are selected with the same capacity. Resistors R_1 and R_2 are selected with the same resistance. Typically, the minimum capacity is selected. Such capacitors have a maximum stability in characteristics. Then, define the resistance value of R_1 and R_2:

$$R_1 = R_2 = 1/2\pi f_p C,$$

where $C = C_1 = C_2$, f_p – pole frequency.

The pole resonance frequency is determined according to the following formula:

$$f_p = 1/\sqrt{R_1 R_2 C_1 C_2}\ .$$

Resistors R3 and R4 in the Sallen-Key scheme determine the voltage gain in the same way as in a conventional inverting amplifier circuit.

In the scheme of the active RC filter the amplifier is covered by both negative and positive feedbacks. Positive feedback depth ratio is determined by resistors R1 and R2 or capacitors C1 and C2. The operational amplifier operates according to the non-inverting amplifier circuit.

Active LPF, cutoff frequency of which is small in comparison with the frequency of the multivibrator clock impulses, converts digital noise (pseudo-random impulse sequence) to the analog.

Digital noise is a temporary random process, similar by its properties to the physical noise process and therefore is called "pseudo-random process".

5 Conclusion

In conclusion, it should be noted, that in this work with the help of CAD Vivado 2016.2 system and Verilog hardware description language there was synthesized, simulated and built temporary digital pseudo-random impulse sequence generator operation circuits based on PLIC of the FPGA families of the Xilinx company, which can be used in cryptography to create stream encryption algorithms.

On the basis of the digital pseudo-random impulse generator and active low-pass filter of the Sallen - Key second order, the load which is a piezoceramic transducer ZQ, there was built a electrical circuit of the acoustic noise generator which by creating a masking vibration noise provides protection against wiretapping by using embedded devices, laser wiretapping system and unauthorized dictaphone recording of confidential voice information.

In future it is expected to improve the cryptographic strength of the generated sequences with relatively large periods, linear complexity and good statistical properties through improvement of the digital pseudo-random impulse sequence generator based on the use of shift registers LFSR with different clocking (with a complicated clocking scheme).

This work relates to the field of information security for the creation of cryptographic keys for encrypted data transfer, and can also be used in systems of confidential speech information protection (for example, for protection of negotiations in the office of the head or in a room specially designated for this purpose) by means of an acoustic noising at the audio signals frequencies.

References

[1] Harris D, Harris S 2013 *Digital circuit design and computer architecture* Second edition. Morgan Kaufman Publishing House, English Edition 1619 p.

[2] Tarasov I. E., Pevtsov E F 2012 *Programmable logic circuits and their application in circuit design: Textbook* M.: MSTU MIREA 184 p.

[3] Zotov V Y 2003 *Design of digital devices based on PLIC of the Xilinx company in CAD WebPACK ISE* M.: Hotline-Telecom 624 p.

[4] Amosov V V 2007 *Circuitry and design tools of digital devices* BHV-Petersburg 542 p.

[5] *VIVADO - a new development tool of Inline Group* plis.ru/docum/sredstvarazrabotki_i_ip/vivado_- novoe sredstvo razrabotki

[6] *The new version of CAD Vivado Design Suite 2016.1* www.komponenta.ru/about / news / novaya-versiya-sapr-vivado-design-suite

[7] Polyakov A K 2003 *VHDL and VERILOG languages in the design of digital equipment* M.: SOLON-Press 320 p.

[8] *The shift register with linear feedback* https: //ru.wikipedia. org / wiki

[9] Kuznetsov V M 2011 *Pseudo-random sequences generator on digital delay elements. Abstract of dissertation for the degree of Doctor of Technical Sciences* Kazan: KAI

[10] Sizonenko A B 2012 Multichannel digital sound source based on the recurrence shift register *Journal "Special equipment and communication"* **3**

AUTHORS

Zaurbek Aizhan

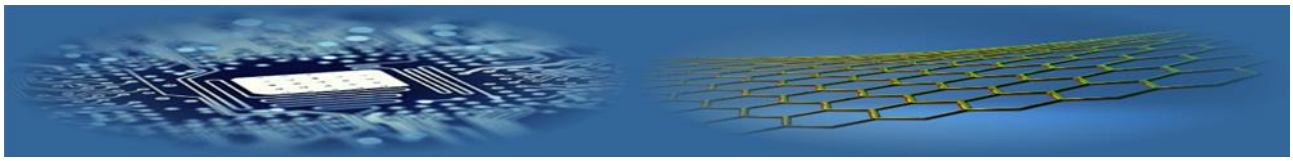
Current position: student of group ВТиПО-13 p (5b070400) – (Computing engineering and software -13 Russian) Kazakh National Research Technical University to the name K.I. Satpayev, 4-course.
Publications: More than 7 publications

Nurgul Seilova

Current position, grades: Head of department Information Security
University studies: candidate of technical Sciences
Scientific interest: network technologies, Information Security
Publications: More than 30 publications
Experience: 15 years of teaching experience and 3 years in managerial positions

Juruntaev Joldas

Current position, grades: Associated professor of department of Informative safety
University studies: doctor of technical Sciences
Scientific interest: Synthesis and design of digital devices on basis Programmable logic integrated circuits FPGA with the use automatic projection system Vivado 2016.2; technical equipments of defence of acoustic (speech) information; author of two textbooks circuit "Technology" on Russian and Kazakh the languages produced on permission of Department of Education and Science of Republics of Kazakhstan**.**
Publications: more than 60 scientific reasons, more than 50 Teaching developments.
Experience: 50 years of teaching experience.



A modified non-monotone method with 3-1 piecewise NCP function for nonlinear complementary problem

Ke Su, Dan Yang*

College of Mathematics and Information Science, University of Hebei, 071002, Baoding, China

**Corresponding author's e-mail: yangdan7458@163.com*

Received 21 December 2016, www.cmnt.lv

Abstract

In this paper, we proposed a modified nonmonotone method for nonlinear complementarity problem, different from the existed methods, we transform the original problem to a semi-smooth equation by using a piecewise NCP function, and combined with the nonmonotone line search. Only one nonlinear equations need to be solved per iteration so that the computational costs are reduced. Under some suitable assumptions, we give the convergence properties of the proposed method and the numerical results to show that our method is efficient.

Keywords:

nonlinear complementary problem,
piecewise NCP function,
nonmonotone,
global convergence

1 Introduction

We consider the following nonlinear complementarity problem NCP: find $x \in R^n$, to satisfy:

$$x \geq 0, F(x) \geq 0, x^T F(x) = 0, \tag{1}$$

where $F: R^n \to R^n$ is continuously differentiable P_0 -function, i.e., for all $x, y \in R^n$ with $x \neq y$, there exists an index i_0 such that

$$x_{i_0} \neq y_{i_0}, (x_{i_0} - y_{i_0})[F_{i_0}(x) - F_{i_0}(y)] \geq 0 \cdot \tag{2}$$

The nonlinear complementarity problem is one of the important types of variational inequalities, mainly comes from the actual problem in the field of economy (such as balance problem) and relevant issues in the field of physics, mechanics and engineering (such as the discretization of infinite dimensional problem [1-3]). Therefore, the problem NCP (1) has attracted great attention due to its various application. One way of solving the nonlinear complementarity problem (1) is to construct a Newton method. This method is to solve a system of nonlinear equations:

$$H(x,s) = \begin{pmatrix} s - F(x) \\ \varphi(x,s) \end{pmatrix} = 0,$$

which is equivalent to (1). Among them, $\phi(x,s)$ is a kind of NCP function satisfies:

$$\varphi(x,s) = \begin{cases} 3x - \dfrac{x^2}{s} & if\ s \geq x > 0, or\ 3s > -x \geq 0, \\ 3s - \dfrac{s^2}{x} & if\ x > s > 0, or\ 3x > -s \geq 0, \\ 9x + 9s & if\ x \leq 0\ and - x \geq 3s, or\ s \leq -3x \leq 0. \end{cases}$$

$F: R^n \to R^n$ is continuously differentiable P_0 -function.

Chen et al [4] investigate a semismooth Newton algorithm for P_0 -NCP. In order ro improve the numerical results, Zhang et al [5] replace the monotone line search with a non-monotone line search when they fulfill the algorithm. Qi H D and Qi L Q [6] propose a new QP-free method which ensures the strict feasibility of all iterates based on the Fischer-Burmister NCP function. They also prove that the method has global convergence without isolatedness of accumulation point and strict complementarity condition. D.G.Pu et al [7] minimizes a smooth function subject to smooth inequality constraints. This iterative method is to solve a nonsmooth equations that are obtained by the multiplier and the Fischer-Burmeister NCP function. Liu and Pu [8] present 3-1 piecewise NCP function for new nonmonotone QP-free infeasible method. This method proved globally convergent without a linear independence constraint qualification.

Motivated by the above ideas, we construct a Newton method based on the solution of nonlinear equations obtained by the 3-1 piecewise NCP function for the P_0 -NCP function. The acceptance of a trial step is more flexible by means of nonmonotone techniques. In this algorithm, we only need to solve one nonlinear equations per iteration so that the computational costs are reduced. The method has proved to be implementable and globally convergent without a strict complementarity.

This paper is organized as follows. In the next section, we introduce the 3-1 piecewise linear NCP function and the properties of it. The nonlinear complementary problem is transformed into equivalent system of nonlinear equations. In Sect.3 introduces the algorithm. In Sect.4 proves the algorithm to be implementable and presents the algorithm's convergence theory.

2 Preliminaries

Function $\phi: R^2 \to R$ is called an NCP function when $\phi(a,b)=0$ if and only if $a \geq 0, b \geq 0$ and $ab=0$. The 3-1 piecewise linear NCP function is defined as:

$$\varphi(a,b)=\begin{cases} 3a-\dfrac{a^2}{b} & if\ b \geq a > 0, or\ 3b > -a \geq 0, \\ 3b-\dfrac{b^2}{a} & if\ a > b > 0, or\ 3a > -b \geq 0, \\ 9a+9b & if\ a \leq 0\ and\ -a \geq 3b, or\ b \leq -3a \leq 0 \end{cases} s \cdot (3)$$

If $(a,b) \neq (0,0)$, then

$$\nabla\varphi(a,b)=\begin{cases} \begin{pmatrix} 3-\dfrac{2a}{b} \\ \dfrac{a^2}{b^2} \end{pmatrix} & if\ b \geq a > 0, or\ 3b > -a \geq 0, \\ \begin{pmatrix} \dfrac{b^2}{a} \\ 3-\dfrac{2b}{a} \end{pmatrix} & if\ a > b > 0, or\ 3a > -b \geq 0, \\ \begin{pmatrix} 9 \\ 9 \end{pmatrix} & if\ a \leq 0\ and\ -a \geq 3b, or\ b \leq -3a \leq 0 \end{cases} . (4)$$

Detailed Property and application of piecewise NCP function see [9].

It is easy to check the following Proposition.

Proposition 2.1 For the function $\phi(a,b)$ the following holds.

1. $\varphi(a,b)=0 s \Leftrightarrow a \geq 0, b \geq 0$ and $ab=0$;a
2. the square of φ is continuously differentiable;
3. φ is twice continuously differentiable everywhere except at the origin, but it is strongly semismooth at the origin and is a pseudo-smooth NCP function.

Constuct function: $H: R^{2n} \to R^{2n}$

$$H(x,s)=\begin{pmatrix} s-F(x) \\ \varphi(x,s) \end{pmatrix}. \quad (5)$$

For the update of s, we require it infinitely close to F(x), So we order s-F(x)=0. Therefore, contacting the first line of Proposition 3, we know that nonlinear complementarity problem (1) is equivalent to solving the minimization problem:

$$\begin{aligned} &\min \quad \psi(x,s) \\ &\psi(x,s)=\| H(x,s) \| \end{aligned}. \quad (6)$$

3 Algorithm

For solving (6), we need to introduce the following symbols

$$(\xi_i^k \eta_i^k)=\begin{cases} (1,\ 1) & (x,s)=(0,0) \\ \nabla\varphi(x,s) & otherwise \end{cases}, \quad (7)$$

$i=1,2\cdots n$, obviously, $\xi_i^k > 0$ and $\eta_i^k > 0$.

Compute the Jacobian matrix $V(x^k,s^k)$ of $H(x^k,s^k)$, we get

$$V(x^k,s^k)=\begin{pmatrix} -F'(x^k) & I \\ diag(\xi_i^k) & diag(\eta_i^k) \end{pmatrix},$$

where I is identity matrix of $n \times n$, $diag(\xi_i^k)$ or $diag(\eta_i^k)$ denotes the diagonal matrix whose i th diagonal element is ξ_i^k or η_i^k, respectively.

We now present the algorithm combining a Newton method with the nonmonotone line search, the following algorithm is obtained d and λ by calculating system of nonlinear equations, which from the Hessian of H. In order to solve: min $\psi(x,s)$, we adopt the nonmonotone line search based on [8], so that the trial step is more flexible.

Algorithm 3.1

Step 0. Initialization:

Given initial point:

$$x^0, s^0 \in R^n,\ \tau \in (0,1),\ 0 < \theta < 1,\ k=0.$$

Step1. If $\psi(x^k,s^k)=0$ then stop. Otherwise, calculation of the search direction:

Calculate d^k and λ^k by solving the following linear system in (d,λ) :

$$V_k \begin{pmatrix} d \\ \lambda \end{pmatrix}=\begin{pmatrix} F(x^k)-s^k \\ -\varphi(x^k,s^k) \end{pmatrix}, \quad (8)$$

Step 2. Nonmonotone line search.

Step 2.1. If

$$\psi(x^k+d^k, s^k+\lambda^k) \leq \theta\psi(x^k,s^k), \quad (9)$$

$$\phi(x^k+d^k, s^k+\lambda^k) \leq \theta \max_{0 \leq r \leq m(k)-1} \| \phi^{k-r} \|, \quad (10)$$

where $m(0)=0,\ 0 \leq m(k) \leq \min\{m(k-1)+1, M\}$, M is a positive constant.

Then let $x^{k+1}=x^k+d^k$, $s^{k+1}=s^k+\lambda^k$, go to step 3.

Otherwise go to step 2.2.

Step 2.2. Let $x^{k+1}=x^k+\alpha_k d^k$, $s^{k+1}=s^k+\alpha_k \lambda^k$.

Where $\alpha_k=\tau^j$ $(0<\tau<1)$ and j is the smallest non-negative integer and satisfied: (10).

Step 3. Update: let $k=k+1$ and go to step 1.

4 Convergence

In this section, we discuss the global convergence property of algorithm with the nonmonotone line search. In order to achieve the convergence of the algorithm, we give some Assumptions as follows:

Assumption 4.1

A. $F: R^n \to R^n$ is continuously differentiable P_0 -function, so that $\nabla F(x)$ is positive semidefinite.

B. F is Lipschitz continuously differentiable, namely, there exists a constant L such that for all $x_1, x_2 \in R^n$, $y_1, y_2 \in R^{2n}$

$$\| F(x_1)-F(x_2) \| \leq L \| x_1-x_2 \|,$$

$\| H(y_1)-H(y_2) \| \le L \| y_1-y_2 \|$.

Lemma 4.1 If $\phi^k \neq 0$ then given any $\varepsilon > 0$ there is a $\bar{t} > 0$, such that for any $0 < t \le \bar{t}$ and any k,

$$\left\|\varphi^k\right\|^2 - \left\|\varphi(x^k + td^k, s^k + t\lambda^k)\right\|^2 \ge (2-\varepsilon)t\left\|\varphi^k\right\|^2$$

Proof: If $\varphi^k \neq 0$ implies

$$diag(\xi^k)\cdot d^k + diag(\eta^k)\cdot\lambda^k = -\varphi(x^k, s^k). \quad (11)$$

We define that if $(x,s) \neq (0,0)$, then $(\bar{\xi}_i^k, \bar{\eta}_i^k) = (\xi_i^k, \eta_i^k)$.

Otherwise, $\bar{\xi}_i^k d^k + \bar{\eta}_i^k \lambda^k = \phi_i'((x^k, s^k),(d^k, \lambda^k))$.

Where $\varphi_i'((x^k, s^k),(d^k, \lambda^k))$ is the direction derivative of $\varphi_i(x,s)$ at (x^k, s^k) in the direction (d^k, λ^k). Let $diag(\bar{\xi}^k)$ or $diag(\bar{\eta}^k)$ denote the diagonal matrix whose ith diagonal element is $\bar{\xi}_i^k$ or $\bar{\eta}_i^k$, respectively.

Clearly, for all i,

$$\varphi_i(x^k + td^k, s^k + t\lambda^k) - \varphi_i^k - t(\bar{\xi}_i^k d^k + \bar{\eta}_i^k \lambda^k) = o(t). \quad (12)$$

It follows by the definition of above, we have

$$\left\|\varphi^k + t(diag(\bar{\xi}^k)d^k + diag(\bar{\eta}^k)\lambda^k)\right\|^2 = (1-2t)\left\|\varphi^k\right\|^2 + t\left\|diag(\bar{\xi}^k)d^k + diag(\bar{\eta}^k)\lambda^k\right\|^2. \quad (13)$$

It follows from (12) and (13) that, given any $\varepsilon > 0$, there is a $t > 0$, such that for any $0 < t \le t$,

$$\left\|\varphi^k\right\|^2 - \left\|\varphi(x^k + td^k, s^k + t\lambda^k)\right\|^2 \ge (2-\varepsilon)t\left\|\varphi^k\right\|^2.$$

Hence, this lemma holds.

Lemma 4.2 For all k, there is an $\alpha_{\min} > 0$ such that $\alpha_k \ge \alpha_{\min} > 0$.

Proof: Assume $\phi^k \neq 0$ for sufficiently large k, it follows by Lemma 4.1 that, for all k, $\phi^k \neq 0$ and any $\alpha \le \min\{\frac{1-\theta}{2-\varepsilon}, \bar{t}\}$.

$$\left\|\varphi(x^k + \alpha d^k, s^k + \alpha\lambda^k)\right\|^2 \le [1-(2-\varepsilon)\alpha]\left\|\varphi^k\right\|^2 \le \theta\left\|\varphi^k\right\|^2 \le \theta^2 \max_{0\le r\le m(k)-1}\left\|\varphi^{k-r}\right\|^2.$$

Lemma 4.3 If $H(x^k, s^k) \neq 0$ then V^k is nonsingular.

Proof: Assume $H(x^k, s^k) \neq 0$, If $V^k(u,v)^T = 0$ for some $(u,v)^T \in R^{2n}$, where $u = (u_1, u_2 \cdots u_n)$, $v = (v_1, v_2 \cdots v_n)$, then

$$-F'(x^k)u + Iv = 0, \quad (14)$$

$$diag(\xi^k)u + diag(\eta^k)v = 0. \quad (15)$$

From the definitions of ξ_i^k and η_i^k we know that $\xi_i^k > 0$ and $\eta_i^k > 0$ for all i.

So, $diag(\eta^k)$ is nonsingular. We have

$$v = -(diag(\eta^k))^{-1} diag(\xi^k)u. \quad (16)$$

Puting (16) into (14), and multiplying by u^T, we have

$$-u^T F'(x^k)u - u^T(diag(\eta^k))^{-1} diag(\xi^k)u = 0.$$

By the fact that $F(x)$ is the P_0-function, so all the principal minor determinant of $\nabla F(x)$ is non-negative, that is to say, $\nabla F(x)$ is positive semidefinite. And matrix $(diag(\eta^k))^{-1} diag(\xi^k)$ is positive definite. Therefore $u = 0$. It follows from (16) that $v = 0$. Hence, V^k is nonsingular.

Lemma 4.4 If V^* is an accumulation matrix of $\{V^k\}$, then V^* is nonsingular.

Proof: It is clear that V^k is nonsingular for all $k = 0,1,2....$ Since ξ_i^k and η_i^k are bounded without loss of generality, let $\xi_i^k \to \xi_i^*$, $\eta_i^k \to \eta_i^*$ and let $x^k \to x^*$, then

$$V^k \to V^* = \begin{pmatrix} -\nabla F(x^*) & I \\ diag(\xi_i^*) & diag(\eta_i^*) \end{pmatrix}$$

Let $(u, v) \in R^{2n}$ be the solution of $V^*(u,v)^T = 0$

$$-\nabla F(x^*)u + Iv = 0, \quad (17)$$

$$diag(\xi^*)u + diag(\eta^*)v = 0. \quad (18)$$

In the next section, V^* is proven to be nonsingular, which is equivalent to showing that $(u,v) = (0,0)$

First, consider such an $j \in J$ for which $\xi_j^* = 0$. From the definition of the 3-1 piecewise NCP function, it is only possible in the second area and $x > s > 0$ or $3x > -s \ge 0$,

$$\xi_j^k = \left(\frac{s}{x}\right)^2 \to 0.$$

Hence $\eta_j^k = 3 - \frac{2s}{x} \to 3 \neq 0$. Then for such an $j \in J$, we deduce that the matrix $diag(\eta_j^*)$ is nonsingular, and $v_j = 0,\ j \in J$ by (18).

For $j \notin J$ such that $\xi_j^* \neq 0$, substituting (18) into (17) and multiplyng (17) by v_j^T, then

$$v_j^T \cdot \nabla F(x^*) \cdot \sum_{j:\xi_j^* \neq 0} \frac{\eta_j^*}{\xi_j^*} \cdot v_j + v_j^T I v_j = 0.$$

$\nabla F(x^*)$ is positive semidefinite together with the $\xi^* > 0, \eta^* \ge 0$ implies $v_j = 0,\ j \notin J$.

This proves $(u,v)=(0,0)$ and hence V^* is nonsingular.

Lemma 4.5 Suppose the Assumption 4.1 holds, $\phi(x^k,s^k)\to 0$, as $k\to\infty$.

Proof: In view of convenience, if for all sufficiently large k (10) holds, define $\|\phi^{l(k)}\|=\max\limits_{0\le r\le m(k)-1}\|\phi^{k-r}\|$, where $k-m(k)+1\le l(k)\le k$.

Since $m(k+1)\le m(k)+1$, then

$$\begin{aligned}\|\varphi^{l(k+1)}\|&=\max_{0\le r\le m(k+1)-1}\|\varphi^{k+1-r}\|\\&\le\max_{0\le r\le m(k)}\|\varphi^{k+1-r}\|\\&=\max\{\|\varphi^{l(k)}\|,\|\varphi^{k+1}\|\}\\&=\|\varphi^{l(k)}\|\end{aligned}\qquad(1)$$

So, $\|\phi^{l(k)}\|$ is monotone decreasing, which implies that the $\{\|\phi^{l(k)}\|\}$ converges.

It follows from (10) that $\|\phi^{l(k)}\|\le\theta\|\phi^{l(l(k)-1)}\|$.

Since $\theta\in(0,1)$, therefore $\{\|\phi^{l(k)}\|\}\to 0(k\to\infty)$.

Therefore $\|\phi^{k+1}\|\le\theta\|\phi^{l(k)}\|\to 0$ holds by the Algorithm 3.1.

That is, $\lim\limits_{k\to\infty}\|\varphi^k\|=0$.

Lemma 4.6 Suppose the Assumption 4.1 holds, $d^k\to 0,\lambda^k\to 0,H^k\to 0$, as $k\to\infty$.

Proof: Suppose the contrary that exists $\varepsilon_1>0,\ \varepsilon_2>0$ for a subsequence (x^k,s^k) such that $\|d^k\|\ge\varepsilon_1>0,\|\lambda^k\|\ge\varepsilon_2>0$. If $\phi^k\ne 0$, then (d^k,λ^k) is the decreasing direction of $\|\phi^k\|$ by lemma 4.1, which contradict $\lim\limits_{k\to\infty}\|\phi^k\|=0$. Hence, $d^k\to 0,\lambda^k\to 0$.

V^* is nonsingular from lemma 4.4 together with $V^*\begin{pmatrix}0\\0\end{pmatrix}=\begin{pmatrix}F(x^*)-s^*\\0\end{pmatrix}$. It is seen that $F(x^*)-s^*=0$ and $\phi(x^*,s^*)=0$, namely, $\psi(x^*,s^*)=0$, so $(x^k,s^k)\to(x^*,s^*)$ is the solving of NCPs.

5 Numerical tests

In this section, we implemented Algorithm 3.1 for solving NCP. All experiments were performed on a personal computer with 2.0 GB memory and Intel(R) Core(TM)2 Duo CPU 2.93 GHz. The operating system was Windows 7 and the computer codes were all written in Matlab 7.1.

In the following tables, IT denotes the number of iterations. CPU denotes the CPU time in seconds. x^k is the final value of x, FV denotes the value of $\|H(x,s)\|$ when the algorithm terminates. We considered the following 3 examples.

Example 5.1: Consider (1), where $x\in R^3$ and $F(x):R^3\to R^3$ given by $F(x)=\begin{pmatrix}x_2\\x_3\\-x_2+x_3+1\end{pmatrix}$.

This problem is from Example 4.4 in [11], which has infinitely many solutions $(0,\lambda,0)$, where $\lambda\in[0,1]$. The initial point x^0,s^0 is randomly generated whose elements are in the interval (0,10). The termination criterion is $\|H(x,s)\|\le 10^{-6}$. Parameters are chosen as follows: $\theta=0.6,\ \tau=0.9$. The test results are listed in Table 1 by using different starting points.

TABLE 1

x^0	s^0	IT	CPU	x^k
(9.5013,2.3114,6.0684)	(6.582,3.782,2.478)	6	0.018834	(-0.0000,0.5000,-0.0000)
(6.8128,3.7948,8.3180)	(8.459,5.248,6.254)	6	0.017559	(-0.0000,0.1285,-0.0000)
(4.4470,6.1543,7.9194)	(5.791,3.896,8.412)	4	0.015375	(-0.0000,0.9998,-0.0000)
(8.4622,5.2515,2.0265)	(7.685,3.365,2.489)	5	0.017096	(-0.0000,1.0000,-0.0000)
(3.0462,1.8965,1.9343)	(4.235,1.226,2.742)	4	0.010422	(-0.0000,0.8585,-0.0000)

Example 5.2: Consider (1), where $x\in R^3$ and $F(x):R^3\to R^3$ given by $F(x)=\begin{pmatrix}x_1-5\\x_2^3+x_2-x_3-3\\x_2+2x_3^3+x_3-3\end{pmatrix}$

This problem has infinitely many solutions. The initial point x^0,s^0 is randomly generated. The termination criterion is $\|H(x,s)\|\le 10^{-6}$. Parameters are chosen as follows: $\theta=0.6,\ \tau=0.9$. The test results are listed in Table 2 by using different starting points.

TABLE 2

x^0	s^0	IT	CPU	FV	x^k
(2,3,9)	(1,1,2)	14	0.016404	2.4007×10^{-7}	(5.0000,1.3428,0.7643)
(8,13,9)	(3,4,2)	14	0.012882	1.7533×10^{-7}	(5.0000,1.3428,0.7643)
(9,14,18)	(4,17,12)	16	0.014060	1.2517×10^{-7}	(5.0000,1.2027,0.7944)
(11,7,8)	(6,9,13)	14	0.015661	2.8667×10^{-7}	(5.0000,1.3428,0.7643)
(5,7,3)	(4.9,3)	12	0.023225	4.6498×10^{-7}	(5.0000,1.2027,0.7944)

Example 5.3: Consider (1), where $x \in R^4$ and $F(x): R^4 \to R^4$ given by $F(x) = \begin{pmatrix} x_1^3 - 8 \\ x_2 + x_2^3 - x_3 + 3 \\ x_2 + x_3 + 2x_3^3 - 3 \\ x_4 + 2x_4^3 \end{pmatrix}$

This problem is from Example 1 in [12]. The initial point x^0, s^0 select the following. The termination criterion is $\|H(x,s)\| \le 10^{-4}$. Parameters are chosen as follows: $\theta = 0.8$, $\tau = 0.6$. The test results are listed in Table 3 by using different starting points.

TABLE 3

x^0	s^0	IT	CPU	FV	x^k
(3,1,1,1)	(1,2,2,5)	5	0.033747	2.4217×10^{-5}	(2.0000,-0.0000,1.0000,0.0000)
(3,1,2,1)	(1,2,6,2)	5	0.014731	5.8588×10^{-6}	(2.0000,-0.0000,1.0000,0.0000)
(1,1,2,1)	(1,2,5,1)	5	0.016660	6.8160×10^{-6}	(2.0000,-0.0000,1.0000,0.0000)
(2,1,1,1)	(1,1,4,2)	5	0.026659	2.7528×10^{-5}	(2.0000,-0.0000,1.0000,0.0000)

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No.11101115, 61572011), the Natural Science Foundation of Hebei Province (No.A2014201033), the Key Research Foundation of Education Bureau of Hebei Province (No.ZD2015069).

References

[1] Cottle R W, Giannessi F, Lions J L 1980 *Variational inequalities and complementarity problems. Theory and Applications.* Wiley, New York
[2] Harker P T, Pang J S 1990 Finite-dimensional variational inequality and nonlinear complementarity problems, a survey of theory, algorithms and applications *Mathematical Programming, Ser B* **48**(2) 161-220
[3] Isac G 1992 *Complementarity Problems* Springer-Verlay, Berlin Heidelberg
[4] Chen J S, Pan S H 2008 A regularization semismooth Newton method based on the generalized Fischer-Burmeister function for P_0 - NCPs *Journal of Computational and Applied Mathematics* **220**(1-2) 464-79
[5] Zhang H C, Hanger W W 2004 A nonmonotone line search technique and its application to unconstrained optimization *SIAM Journal on Optimization* **14**(4) 1043-56
[6] Qi H D, Qi L Q 2000 A New QP-free,globally V,locally superlinear convergent feasible method for the solution of inequality constrainted optimization problems *SIAM Journal on Optimization* **11** 113-32
[7] Pu D G, Li K D, Xue W J 2005 Convergence of QP-free infeasible methods for nonlinear inequality constrained optimization problem *Journal of Tongji University* **33**(4) 525-9
[8] Liu A L, Pu D G 2014 3-1 piecewise NCP function for new nonmonotone QP-free infeasible method *Journal ref: Journal of Robotics and Mechatronics* **26**(5) 566-72
[9] Pu D G, Zhou Y 2006 Piecewise linear NCP function for QP-free feasible method *Applied Mathematics: A Journal of Chinese Universities* **21**(3) 289-301
[10] Pu D G, Kong X Q, Wang X C 2009 Filter QP-free method with piecewise linear NCP function *OR Transactions* **13**(2) 48-58
[11] Tang J Y, Dong L, Zhou J C 2013 A smoothing Newton method for nonlinear complementarity problems *Computational and Applied Mathematics* **32**(1) 107-18
[12] Ji U I, Chen G Q 2005 New simple smooth merit function for box constrained variational inequalities and damped Newton type method *Applied Mathematics and Mechanics* **26**(8) 1083-92

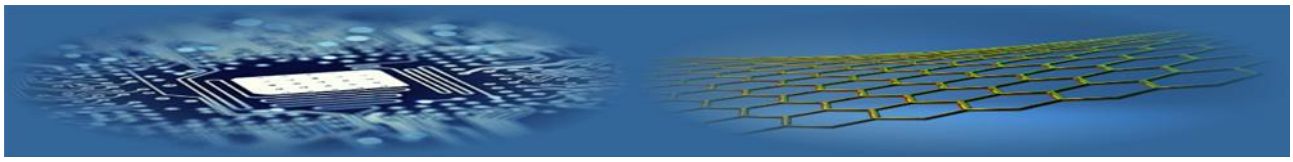
AUTHORS

Ke Su, born in February, 1978, Baoding City, Hebei Province, P.R.China

Current position, grades: the professor of College of Mathematics and Information Science, Hebei University, China
University studies: received her B.Sc. in Applied Mathematics from Hebei University in China, she received her M.Sc. from Hebei University in China
Scientific interest: mathematical programming, nonlinear optimization
Publications: more than 10 papers in various journals
Experience: She has teaching experience of 10 years, has completed four scientific research projects.

Dan Yang, born in October, 1991, Handan City, Hebei Province, P.R.China

Current position, grades: the graduate student of College of Mathematics and Information Science, Hebei University, China
University studies: received her B.Sc. in Applied Mathematics from Hebei Normal University in China
Scientific interest: mathematical programming, nonlinear optimization
Publications: 4 papers in various journals
Experience: She has teaching experience of 1 year, has completed two scientific research projects.



Improving the reliability of the managed of electric drives with a synchronous generator

Elena Loginova

Moscow Institute of Technology (MTI), Leninsky Prospect, 38a, Moscow, Russia

Corresponding author's e-mail: Ejy-loginova@mail.ru

Received 19 February 2017, www.cmnt.lv

Abstract

The article examines the dynamic properties of the excitation control system of synchronous generator of capacity 2200kW, which comprises a converter - bridge rectifier. It was investigated the influence of the type of bridge rectifier - of thyristor symmetric bridge and of asymmetric bridge - on transient processes of synchronous generator in emergency operation. This problem was solved by numerical methods.

A mathematical model of synchronous generator excitation control system. In order to make the model of the generator excitation system has been used plugin SIMULINK mathematical software MATLAB.

During simulation were calculated currents short circuit of the synchronous generator in the case of using in system excitation a thyristor symmetrical bridge and asymmetrical bridge. Calculations have shown that the use of symmetric thyristor bridge reduces times the short circuit fivefold. This allows us to recommend used a symmetrical bridge in the electric drive system.

Keywords:

excitation system of the synchronous generator, the short-circuit mode

1 Introduction

The main element of an autonomous power system is a synchronous generator with independent excitation. To supply the field winding synchronous generator is also used synchronous generator - Synchronous exciter CB. Synchronous pathogen runs on the load via a bridge rectifier (Fig. 1) [1, 2].

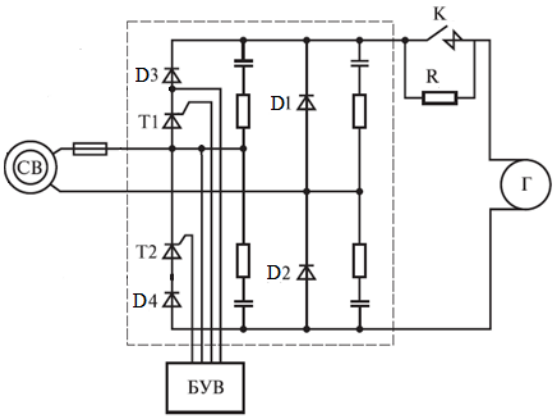


FIGURE 1 The system excitation synchronous generator. Г- is the rotor winding; СВ – is the synchronous exciter; БУВ -is thyristors control unit; T1, T2 – it thyristors; D1-D4 - it diodes; K- is contact of contactor

Bridge rectifiers are widely used in industrial drives because they are rugged, reliable and economical. The bridge is an circuit that provides the same polarity of output for either polarity of input. The most widely used scheme with asymmetrical bridge. The two arms of the bridge contains a diodes D1, D2, and the other two arms contains a thyristors T1, T2. The bridge rectifier provides rectification and change current an electrical load. Capacitors provides a low impedance path to the AC component of the output, reducing the AC voltage across, and AC current through, the resistive load. Therefore, the load current and the voltage variation is reduced as compared with what would be the case without the capacitor.

Block БУВ is designed to control thyristors.

1.1 THEORETICAL BACKGROUND

The figure 2 shows a diagram of an equivalent circuit of the bridge

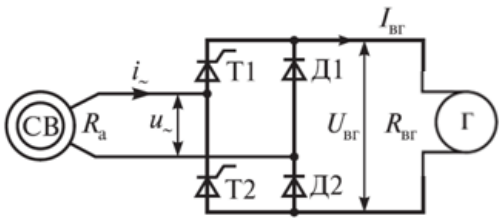


FIGURE 2 The equivalent circuit of the asymmetrical bridge: R_a - is the stator winding resistance CB; $u_{\sim}$ - is the is the stator voltage CB; $i_{\sim}$ - is the stator current CB; $U_{ВГ}$ – is the rectified voltage to the excitation winding; $I_{ВГ}$ – is the excitation current $Г$; $R_{ВГ}$ – Is the resistance of the field winding Г

The average output voltage of the bridge rectifier depends on the control angle of the thyristors α and is determined from the expression:

$$U_{\text{ВГ}} = U_{\text{ср}} \cdot \frac{1+cos\alpha}{2} - \Delta U_R - \Delta U_{\text{В}} - \Delta U_X, \quad (1)$$

where $U_{cp} = \frac{2\cdot\sqrt{2}}{\pi} \cdot U, B$ – is average rectified voltage (thyristor firing angle $\alpha = 0$);

$\Delta U_R, B$ - is voltage drop by resistance an rectification circuit;
$\Delta U_B, B$ - is the voltage drop across the diodes;
$\Delta U_X, B$ - this is the voltage drop of the switching thyristors.

Meaning ΔU_B very little; therefore the expression (1) can be written:

$$U_{\text{ВГ}} = U_{\text{ср}} \cdot \frac{1+cos\alpha}{2} - \Delta U_R - \Delta U_X. \quad (2)$$

The value ΔU_X for single-phase bridge rectifier:

$$\Delta U_X = \frac{x_s \cdot I_{\text{ВГ}}}{\pi}, \quad (3)$$

where x_s - is the resistance of the thyristor.

The value ΔU_R is a voltage drop across active resistance the armature winding for synchronous exciter

The value ΔU_R can be expressed as follows:

$$\Delta U_R = R_a \cdot I_{\text{ср}}. \quad (4)$$

The average current in the windings of the armature of the synchronous exciter I_{CP} is expressed:

$$I_{\text{ср}} = I_{\text{ВГ}} \cdot k_{\text{пр}}, \quad (5)$$

where k_{np} - is the conductivity factor rectifier.

Conductivity factor is a variable quantity; k_{np} depends on the ratio current at the input / output and switching angles valves of the rectifier. The coefficient is determined from the ratio of current in the stator winding CB and average current in the field winding Г:

$$\begin{cases} I_{\text{ср}} = I_{\text{ВГ}} \left(1 - \frac{\alpha + \frac{\gamma_2}{2}}{\pi} + \frac{2\gamma_1}{3\pi}\right); \\ I = I_{\text{вг}} \cdot \\ \cdot \sqrt{1 - \frac{\alpha + \frac{2}{3}\gamma_2}{\pi} + \frac{2\gamma_1(1+2cos^2\gamma_1) - 3sin2\gamma_1}{4\pi(1-2cos\gamma_1)^2}}, \end{cases} \quad (6)$$

where k_{np} – is γ_1 and γ_2 angles are switching thyristors T1 and T2.

Hence we obtain the formula for the coefficient of the conduction

$$k_{\text{пр}} = 1 - \frac{\alpha + \frac{\gamma_2}{2}}{\pi} + \frac{2\gamma_1}{3\pi}. \quad (7)$$

Thyristors control angles are determined from the following expressions:

$$cos\gamma_1 = 1 - \frac{I_{\text{ВГ}}}{I_{\text{к}}}, \quad (8)$$

$$cos(\alpha + \gamma_2) = cos\alpha - \frac{I_{\text{ВГ}}}{I_{\text{к}}}, \quad (9)$$

where I_K - is short-circuit current.

Its value is defined by the relationship:

$$I_{\text{к}} = \frac{\pi}{2} \cdot \frac{U_{\text{ср}}}{x_s} = \frac{\sqrt{2}U}{x_s} = \frac{(U)_m}{x_s}. \quad (10)$$

Thus, the voltage across winding excitation synchronous generator $U_{BГ}$ can be expressed as:

$$U_{\text{вг}} = U_{\text{ср}} \frac{1+cos\alpha}{2} - R_a k_{\text{пр}}(\gamma_1, \gamma_2, \alpha) I_{\text{ВГ}} - \frac{2x_s I_{\text{ВГ}}}{\pi}. \quad (11)$$

Also, the voltage $U_{BГ}$ can be determined through the resistance $R_{BГ}$:

$$U_{\text{ВГ}} = I_{\text{ВГ}} \cdot R_{\text{ВГ}}. \quad (12)$$

Equating expressions (11) and (12) and performing the conversion, we get the following:

$$U_{\text{ср}} \frac{1+cos\alpha}{2} - \frac{2x_s I_{\text{ВГ}}}{\pi} = I_{\text{ВГ}} R_{\text{ВГ}} + k_{\text{пр}} I_{\text{ВГ}} \cdot R_a. \quad (13)$$

Figure 3 shows the characteristic of the control rectifier $U_{BГ} = f(\alpha)$.

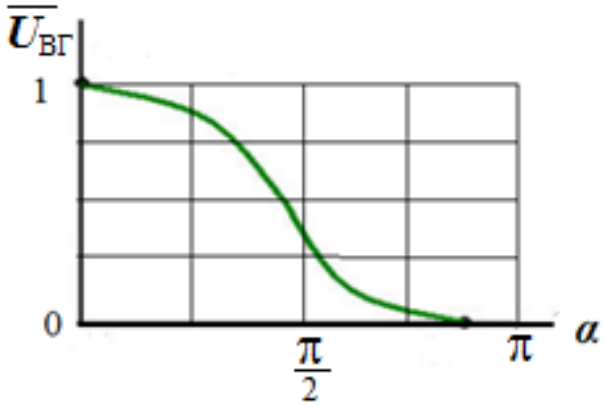


FIGURE 3 Control characteristic rectifier in relative units

$\overline{U_{BГ}} = f(\alpha)$ relationship shows that for any value of the angle of the voltage $\overline{U_{BГ}} \geq 0$.

If the generator excitation system uses a symmetric bridge, can get a negative voltage $\overline{U_{BГ}} < 0$ (Fig. 4).

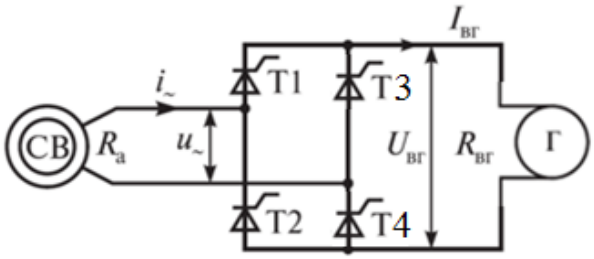


FIGURE 4 The equivalent circuit of the symmetrical bridge

This is very important in the short-circuit conditions.

The negative output voltage of the rectifier will quickly reduce the short-circuit current. This theory was tested using a simulation model of the thyristor excitation circuit of synchronous generator.

2 Simulation results and conclusions

Figure 5 presented a simulation model of a system of excitation synchronous generator with a symmetrical bridge. Modelling carried out in the annex Simulink Matlab package (Fig. 5) [3, 4].

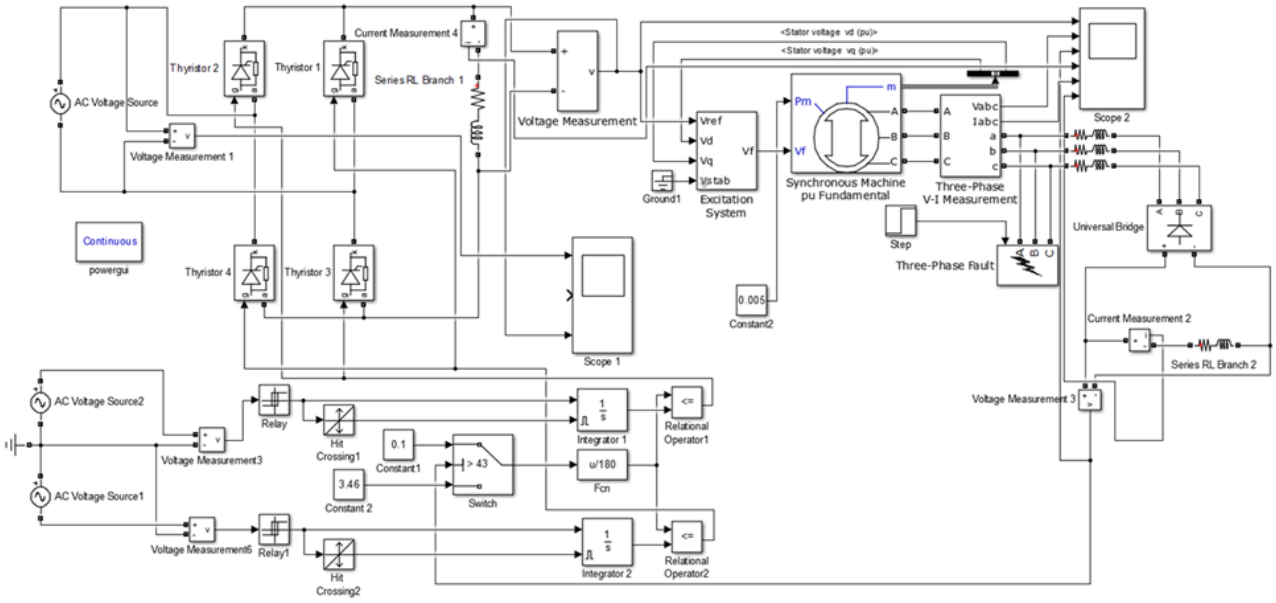


FIGURE 5 Simulation model of excitation system synchronous generator with a symmetrical bridge

Here is a self-contained energy system locomotive power 2200kWt. As the exciter of this system uses a single-phase synchronous generator. Exciter runs on the load via bridge rectifier. The load of exciter it is the winding exciter of traction synchronous generator. Traction synchronous generator runs on an electric motor (active-inductive load) through a three-phase bridge. The model simulates short-circuiting on output the three-phase bridge. Results are presented in Fig. 6.

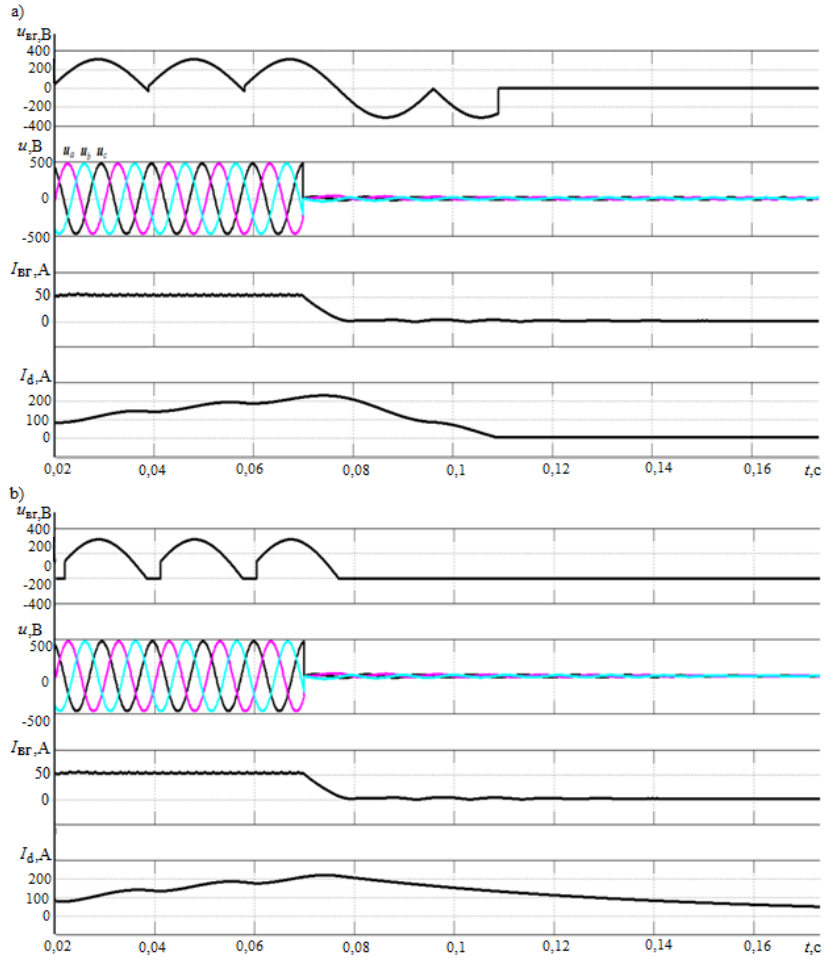


FIGURE 6 The results of the short-circuit mode simulation in circuit with a asymmetrical bridge (a); the results of the short-circuit mode simulation in circuit with a symmetrical bridge (b):u – is phase voltage of the generator; I_d – is generator load current

Calculations have shown that the use of symmetric bridge reduces times the short circuit fivefold. This allows us to recommend used a symmetrical bridge in the electric drive system.

References

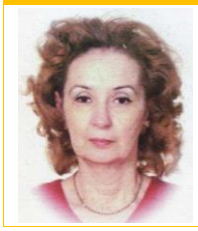
[1] Strzelecki R 2008 *Power Electronics in Smart Electrical Energy Networks* Springer p. 57

[2] Stutz M 2000 Conventional versus electron flow *All About Circuits* **1**, Chapter 1

[3] Chernyh I V 2007 *Simulation of electrical devices in MATLAB* SimPowerSystems and Simulink. - M: DMK Press 288 p.

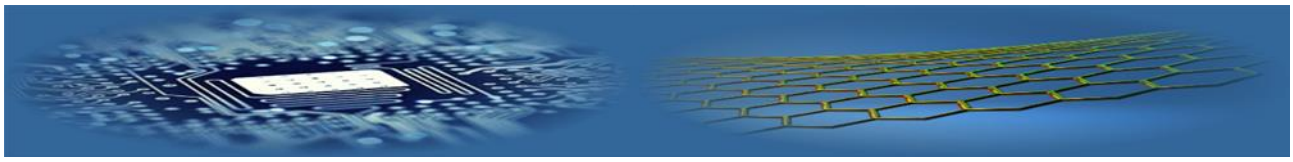
[4] Chernyh I V 2003 *Simulink: Tool simulation of dynamic systems* Internet edition 252 p.

AUTHORS



Loginova Elena, March 1955, Moscow, Russia

Current position, grades: Doctor of Technical Sciences, Professor.
Scientific interests: Traction electric drive of autonomous locomotive and control systems.
She graduated from the Moscow State University of Railway Engineering (MIIT) in 1977.
Responsible for research in the field of electric drive in the Moscow Institute of Technology. She is the author of two textbooks and over 100 scientific articles in the field of electric drive.
Experience: A professor at the Moscow State University of Railway Engineering (MIIT) and visiting professor at the Moscow Institute of Technology (MTI).
From 1977 to 1980 she worked in the Research Institute of Railway Transport.



Development of the SVM Classifier by means of the Hybrid Versions of the Particle Swarm Optimization Algorithm Based on the Grid Search

Liliya Demidova, Irina Klyueva*

Ryazan State Radio Engineering University Gagarin Str., 59/1, Ryazan, Russian Federation

Corresponding author's e-mail: i.aleschenko@yandex.ru

Received 3 April 2017, www.cmnt.lv

Abstract

In this article the approaches to the problem solving of searching of the parameters of the SVM classifier based on the hybridization of the particle swarm optimization algorithm (PSO algorithm) and the grid search algorithms with the aim of providing of high quality classification decisions have been considered. The paper presents two hybrid versions of the basic PSO algorithm, involving the use of the classical Grid Search (GS) algorithm and Design of Experiment (DOE) algorithm correspondingly. It is proposed to use the canonical PSO algorithm as the basic algorithm. The results of experimental studies confirm the application efficiency of the hybrid versions of the basic PSO algorithm with the aim of reducing of the time expenditures for searching the optimum parameters of the SVM classifier while maintaining of high quality of its classification decisions.

Keywords:

classification,
particle swarm optimization algorithm,
grid search algorithm,
SVM classifier,
radial basis kernel function

1 Introduction

Data classification is one of the most common problems of machine learning [1–7]. The solution of this problem requires creation of a classifier that assigns each input dataset the value of the label of one of the classes. Classification of new data is produced after passing through the stage of "learning", in which the input of the learning algorithm serves the data with already assigned labels of classes.

Currently, the SVM algorithm (Support Vector Machine, SVM) [1–7] is successfully applied for solving a wide spectrum classification problems in various applications. The SVM algorithm is a machine learning algorithm by precedents. The SVM algorithm implements the construction of binary SVM classifier.

The SVM algorithm implements creation of the separating hyperplane that divides objects with different class. Herewith, two parallel hyperplanes defining the boundaries of classes and locating at the greatest possible distance from each other are constructed on both sides of the separating hyperplane. It is assumed that the greater the distance between these parallel hyperplanes the smaller the average error of the SVM classifier. The vectors of features of the classified data nearest to the parallel hyperplanes are called the support vectors.

In most cases, the linear separability of objects of real datasets into classes is impossible. In this regard, the main feature of the SVM classifier in case of nonlinear separability of objects is the use of special function, called the kernel function. The kernel function is used to transfer the experimental dataset from the original space of features to the space of higher dimension in which the separating hyperplane is build.

In the process of learning of the SVM algorithm the one of the priority problems is to configure the parameters of the SVM classifier, the most important of which are the kernel function type, the values of the kernel parameters and the value of the regularization parameter.

The one of the following functions [6] is usually used as a kernel function that allows to separate the objects of different classes: linear function, polynomial function, radial basis function, sigmoid function.

The regularization parameter C allows finding a compromise between the maximizing of the gap separating the classes and the minimizing of the total error. In other words, the regularization parameter controls the ratio between the smooth boundary and the corrects data classification.

In case of radial basis kernel function (RBF) [6] it is necessary to determine the value of the coefficient σ of this function.

The simplest approach to settings of the SVM classifier parameters is based on a simple enumeration of the different combinations of the parameter values. For the purpose of setting parameters of the SVM classifier the grid search algorithms (in particular, the Grid Search algorithm) are most often applied [2]. Herewith, the cross-validation on the training dataset is used for each parameters combination corresponding to the specific grid node. As the result, the best combination of the parameters values is selected. This combination defines the certain grid node which is characterized by the best value of the cross-validation indicator.

Finding the optimal set of values for the parameters of the SVM classifier allows avoiding the problems of overfitting or the problems of underfitting of the SVM classifier. If the errors on the training and testing datasets are close to each other and small in value, such the SVM

classifier is recognized the sought for the solution of classification problems.

Since a complex, multi-extreme and multi-parameter objective function is used for the construction of SVM classifiers, it is advisable to use the search for its optimum from the whole space of possible solutions.

Currently, the optimization algorithms inspired by the natural biological systems are used widely. Such algorithms are the bioinspired algorithms for the stochastic optimization: genetic algorithm, particle swarm optimization algorithm, ant colony optimization algorithm, bee algorithm. These algorithms operate with sets of simple entities in the search space, simulating the intellectual behavior of a population in which each individual represents some alternative approximate solution.

In recent years, the particle swarm optimization algorithm (PSO algorithm) [4–9] is used in the solution of various applied optimization problems, based on the idea of possibility of the optimization problems solving by modeling of behavior of the animals groups.

The PSO algorithm is characterized by simplicity of implementation and, consequently, low algorithmic complexity. It is sufficient to determine only the value of the optimized function for the implementation of the PSO algorithm. In this regard, the PSO algorithm can be recommended for the search of the optimum parameters values of the SVM classifier.

Currently, there are various ways to improve the efficiency of the basic PSO algorithm, which can be divided into metoptimazine and combinational.

In this paper we propose to implement a combinational method of improving for the basic PSO algorithm by developing the hybrid versions with the use of the grid search algorithms. It is plan to use two grid search algorithms: the classic Grid Search algorithm (GS algorithm) and Design of Experiment algorithm (DOE algorithm) [2, 6, 7].

The aim of this paper is the development of the hybrid versions of the basic PSO algorithm based on the grid search algorithms and the comparison of their search characteristics. It is planned to test the developed hybrid versions of the PSO algorithm on real datasets in the framework of the problem solving of search of the optimum parameters values of the SVM classifier. The main indicators to measure the effectiveness of the developed algorithms are the search time of the optimum parameters values of the SVM classifier, the quality of data classification (overall accuracy, sensitivity, specificity, the number of support vectors). Herewith, the problem of binary classification has been considered.

2 Principles of the SVM algorithm implementation

As a result of SVM classifier learning, the separating hyperplane is defined (Figure 1) [6]. It can be represented by equation $< w, z > +b = 0$, where w is the vector-perpendicular to the separating hyperplane; b is the parameter which corresponds to the shortest distance from the origin of coordinates to the hyperplane; $< w, z >$ is the scalar product of vectors w and z . The condition $-1 << w, z > +b < 1$ specifies the strip that separates the classes. The wider the strip is, the more confidently we can classify objects.

The objects closest to the separating hyperplane and located on the bounders of the separating strip are called support vectors. They carry basic information about the separation of the classes.

One of the main problem in case of nonlinear separability of objects is to define the rectifiable type of the kernel function and select the optimal values for some set of parameters in order to build the effective SVM classifier.

The classification of the specific object can be performed using the following rule [6]:

$$F(z) = sign\ (\sum_{i=1}^{S} \lambda_i y_i \kappa(z_i, z)\ +\ b)\,, \qquad (1)$$

where λ_i is a dual variable of the Lagrange function; z_i is the object of the training dataset; $y_i \in Y = \{-1;+1\}$ is the number, which characterizes the class of the object z_i from the training dataset; $\kappa(z_i, z)$ is the kernel function; C is the regularization parameter ($C > 0$); S is the quantity of objects in the training dataset; $i = \overline{1,S}$.

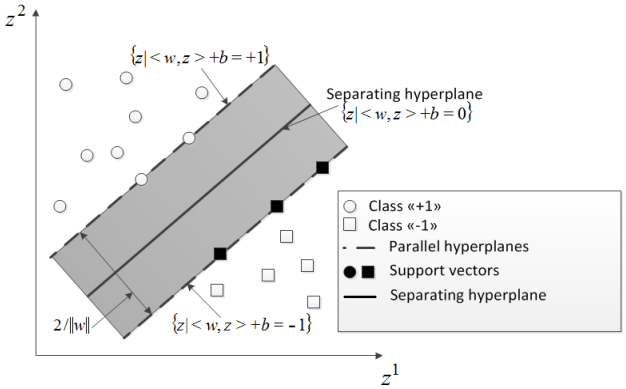


FIGURE 1 The separating hyperplane in the space D-2

The most complete mathematical description of the SVM algorithm is given in [5, 6].

The main problem of the SVM classifier learning is the absence of the recommendations for choice of the value of the regularization parameter C , the kernel function type $\kappa(z_i, z)$, and the kernel function parameters values, which provide the high data classification accuracy. This problem can be solved by means of the application of different optimization algorithms, in particular, with the use of the PSO algorithm.

The radial basis kernel (RBF) [6] is often applied for the SVM classifier development in case of nonlinear separation of objects into classes:

$$\kappa(z_i, z) = \exp(-\|z_i - z\|^2 / (2\sigma^2))\,, \qquad (2)$$

where $\sigma > 0$.

Herewith, it is necessary to determine the value of the parameter σ of the radial basis kernel function along with the value of the regularization parameter C .

3 The principles of implementation of the PSO algorithm and its hybrid versions

The search space in the PSO algorithm is filled with a population of particles each of which has some location and velocity in the space of the problem parameters at the concrete moment of time. In addition, each particle can remember its best location in the swarm and communicate

with other particles about the globally "best" location among all particles.

The value of the objective function is calculated for each particle. The particle location and the velocity are changed according to the certain rules [6, 7] after calculation of the new value of the objective function.

The basic principles of calculating of the new location and the new velocity of particles are given in [6, 7].

Currently, the different versions of the PSO algorithm are known. The canonical version received the traditional application and it's one of the most common versions of the PSO algorithm. In this version of the PSO algorithm it is proposed to perform the normalization of the speedup' coefficients in such way that the convergence of the algorithm not so much depends on the choice of their values [6, 7].

In recent years, the approaches implementing the hybridization of the PSO algorithm with other optimization algorithms in order to increase the efficiency of the classical PSO algorithm are widely used [6–8].

In this paper we present two hybrid versions of the PSO algorithm, involving the use of the classical "Grid Search" (GS) algorithm and the "Design of Experiment" (DOE) algorithm [6–8].

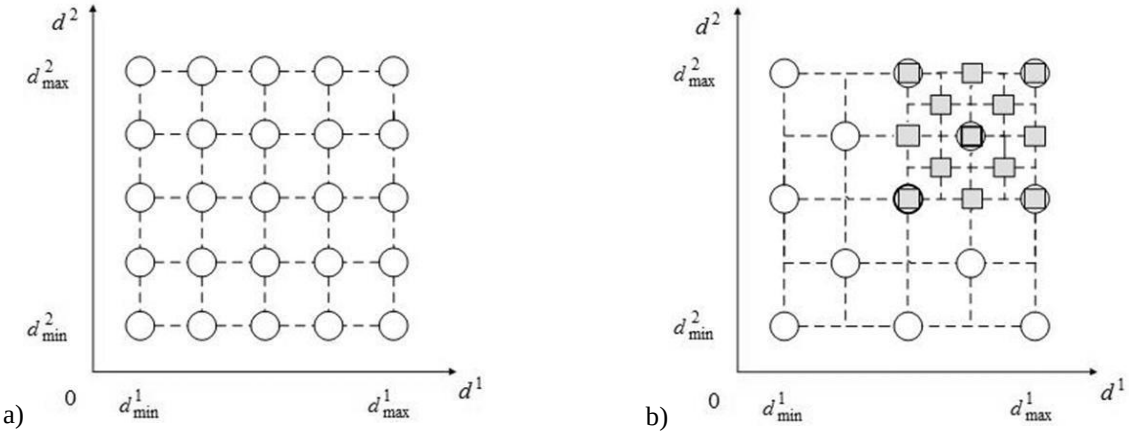


FIGURE 2 The grid formation: a) in the GS algorithm; b) in the DOE algorithm

In the GS algorithm the grid is created on the variation's ranges of the optimization parameters with a certain step for each parameter (Figure 2, a, a special case of the search space D-2) and the efficiency of all combinations of values of the optimization parameters on the grid is evaluated. A herewith all nodes of the grid are explored. The advantage of the search over all grid nodes is the thoroughness of the finding of the globally optimal solution.

The DOE algorithm is an alternative grid search algorithm. The advantages of the DOE algorithm (Figure 2, b) are the following.

- The search boundaries are iteratively improved until the conditions for stopping the search are not satisfied. After each iteration of the DOE algorithm, the search space is narrowed and refined so that "the best" founding node corresponding to the best value of the objective function will be the center of the search space.
- If the search process goes beyond the initially specified (acceptable) search ranges, the new boundaries of the search ranges will be defined in such way that the new search space in the DOE algorithm will be kept within the permissible boundaries of the search ranges.

The proposed hybrid versions of the PSO algorithm were developed, primarily, to solve the problem of search of the optimum parameters values of the SVM classifier based on the radial basis kernel function. These algorithms operating with the set of particles in the search space of D-2 can be applied for solving other optimization problems of the appropriate dimension. Also, these algorithms can also be adapted to the case of the search space with more higher dimensional.

At the creating of the hybrid version of the PSO algorithm it is proposed to execute the clarification of the position (coordinates) of the globally best particle in the swarm at the each iteration of the PSO algorithm with the use of the grid search algorithm and update the current swarm particle population. A herewith the "worst" particle should be removed from the swarm (the particle with the "worst" value of objective function), and the "best" particle founding by the grid search algorithm must be added instead.

In the hybrid PSO-GS algorithm, acceleration of search of the globally optimal solution in swarm is achieved by:

- additional grid search in the area of the potential globally "best" location in swarm;
- updating of the particle swarm population and removal of the "worst" particles.

The hybrid version of the PSO algorithm can be presented by the following sequence of steps.

Step 1. To determine the initial characteristics of particles in the swarm (coordinates and velocities). To determine the customizable parameters of the PSO algorithm (the number of particles in a swamp, the maximum iterations number of the PSO algorithm, the boundaries of the search ranges).

Step 2. To realize one step of the PSO algorithm. To correct the velocity $v_i \in R^n$ and the current coordinates $x_i \in R^n$ for each i-th particle ($i = 1, m$), to determine the coordinates of the globally "best" particle in the swarm.

Then the objective function value is calculated at each new point of the search space and check of each point is carried out in order to determine if its location is the best in swarm. For the problem to search a minimum of the function in the form $f(x) \to \min\limits_{x \in R^n}$ the best particle location will be considered as a point in the search space where the minimum value of this function is achieved at all of the algorithm iteration starting from the first iteration to the current.

Step 3. To determine the boundaries of the search ranges for the grid search algorithm (the classical GS algorithm or the DOE algorithm). A herewith to determine the grid size $[d^j_{\min}, d^j_{\max}]$ ($j=\overline{1,n}$) taking into account the maximum straggling $[r^j_{\min}, r^j_{\max}]$ of particles in the swarm at the current generation of the PSO algorithm. The values of coordinates of the globally "best" particle in swarm can be used as the values of coordinates χ^j ($j=\overline{1,n}$) of the "main" (central) grid node.

The minimum distance from the "best" particle in the swarm (the centroid of the grid) to the boundaries of the straggling range can be calculated as:

$$l^j = \min\{\chi^j - r^j_{\min}, r^j_{\max} - \chi^j\} \quad (3)$$

and the boundaries of the search ranges for the grid are defined as:

$$d^j_{\min} = \chi^j - l^j, \quad (4)$$

$$d^j_{\max} = \chi^j + l^j. \quad (5)$$

Step 4. To specify the coordinates of the globally "best" particle in the swarm using the grid search algorithm (the classical GS algorithm or the DOE algorithm). To check if the really clarification of coordinates of the globally "best" particle of the swarm is achieved. If the clarification is achieved (a new solution is obtained), then to transfer to step 5 or otherwise to transfer to step 6.

Step 5. To take as the new globally optimal solution at the current iteration of the PSO algorithm the solution obtained by implementing of the grid search algorithm at the step 4.To update the swarm particle population: to delete the "worst" particle and to add the "best" particle founding at the step 4.

Step 6. In the case of achievement of the algorithm breakpoint determined according to the maximum number of iterations or the finding of the global optimum with the given accuracy, to transfer to step 7 or otherwise to transfer to step 2.

Step 7. To accept the values of coordinates of the "best" particle in the swarm as the found value of the globally optimal solution and to complete the algorithm execution.

The features of implementation of the grid search algorithms which are used at the step 4 of the proposed hybrid version of the PSO algorithm are considered further.

The search ranges of the GS algorithm $[d^j_{\min}, d^j_{\max}]$ ($j=\overline{1,n}$) found on the base of formulas (4) and (5) at the step 3 of the hybrid version of the PSO algorithm are divided into the specified number of intervals, and the grid nodes are determined.

Then the value of the optimized (target) function in each grid node is calculated. As a result of implementation of the GS algorithm the "best" node with the "best" value of the objective function will be determined. The coordinates of this node can be used as the coordinates of the new globally best particle in the swarm.

The DOE algorithm is used to solve the optimization problems in the search space D-2 typically, but it can easily be adapted to perform the calculations in the space of the arbitrary dimension n. Since we plan to use the hybrid version of the PSO algorithm with the DOE algorithm to solve the optimization problems in the search space D-2 (i.e., when $n=2$), and, also, because of the good visibility of implementation of the DOE algorithm in this space, the further description of implementation of the DOE algorithm at the step 4 of the hybrid version of the PSO algorithm is given for the particular case in the search space D-2.

Step 1. To determine within the ranges boundaries $[d^j_{\min}, d^j_{\max}]$ ($j=1,2$) 13 grid nodes (Figure 2, b the nodes of the first iteration of the DOE algorithm are noted with markers of circular shape in white, and the nodes of the second iteration are noted with markers of square shape in grey color, herewith, the nodes which participate in multiple iterations are noted with double markers of circular and square shape). The central node (the centroid of the grid) with coordinates χ^j ($j=1,2$) (the example in Figure 2, b is the marker of circular shape with the selected contour) corresponds to the globally "best" particle of the swarm, and the width of the search ranges on the current iteration of the DOE algorithm is defined as $S^j = d^j_{\max} - d^j_{\min}$ ($j=1,2$).

The coordinates of the grid nodes are defined as the following (when moving along the grid from the lower left node from bottom to top, from left to right):

$$[\chi^1 - S^1/2, \chi^2 - S^2/2], \quad [\chi^1 - S^1/2, \chi^2 + S^2/2],$$
$$[\chi^1 + S^1/2, \chi^2 + S^2/2], \quad [\chi^1 + S^1/2, \chi^2 - S^2/2],$$
$$[\chi^1 - S^1/2, \chi^2], \quad [\chi^1, \chi^2 + S^2/2], \quad [\chi^1 + S^1/2, \chi^2],$$
$$[\chi^1, \chi^2 - S^2/2], \quad [\chi^1 - S^1/4, \chi^2 - S^2/4],$$
$$[\chi^1 - S^1/4, \chi^2 + S^2/4], \quad [\chi^1 + S^1/4, \chi^2 + S^2/4],$$
$$[\chi^1 + S^1/4, \chi^2 - S^2/4], [\chi^1, \chi^2].$$

Step 2. To calculate the value of the objective function at each node of the grid and to find the coordinates φ^j ($j=1,2$) of the node with the "best" value of the objective function.

Step 3. To override the width of the search ranges as $S^j/2$ ($j=1,2$) and to use the calculated values as the new values of S^j ($j=1,2$) for the next iteration of the DOE algorithm.

Herewith, the new boundaries of the search ranges are redefined for the next step as:

$$d^j_{\min} = \varphi^j - S^j/2, \quad 6)$$

$$d^j_{\max} = \varphi^j + S^j/2. \quad (7)$$

Step 4. Go to the step 1 if the number of iterations of the DOE algorithm is not exhausted, otherwise to complete the algorithm. Herewith, the values of coordinates of the "best" node φ^j ($j=1,2$) found at the current iteration of the DOE algorithm are applied as the new coordinates of the center node of the grid χ^j ($j=1,2$) (the example in Figure 2, b is marker of circular shape with the selected contour).

It should be noted that boundaries of the search ranges $[d^j_{\min}, d^j_{\max}]$ ($j=1,2$) for the first iteration of the DOE

algorithm are calculated based on formulas (4) and (5) at the step 3 of the hybrid version of the PSO algorithm, and for all other iterations of the DOE algorithm they are calculated on the base of formulas (6) and (7) at the step 3 of the DOE algorithm.

At implementation of the DOE algorithm the control for the acceptability of the new calculated boundaries of the search ranges is executed.

If at some of the current iteration of the DOE algorithm the coordinates of the "best" found node have been close to the current boundaries of the grid search ranges, then in case of building of the grid at the next iteration of the DOE algorithm the going beyond the originally defined (allowed) boundaries of the search ranges of the hybrid version of the PSO algorithm $[range^j_{\min}, range^j_{\max}]$ ($j=1,2$) is possible. If after the calculations according to the formulas (6) and (7) of the new boundaries of the grid search ranges $[d^j_{\min}, d^j_{\max}]$ ($j=1,2$) it is turned out that one of conditions $d^j_{\min} < range^j_{\min}$ for some $j = j^* \in \{1,2\}$ or $d^j_{\max} > range^j_{\max}$ for some $j = j^* \in \{1,2\}$ is produced, i.e., the going beyond the originally defined (allowed) boundaries of the search ranges of the hybrid version of the PSO algorithm takes place, the grid is narrowed to the new boundaries of the search ranges according to formulas:

if $d^j_{\min} < range^j_{\min}$ for some $j = j^* \in \{1,2\}$, then

$$d^{j^*}_{min} = \varphi^{j^*} - (\varphi^{j^*} - range^{j^*}_{min}), \quad (8)$$

$$d^{j^*}_{max} = \varphi^{j^*} + (\varphi^{j^*} - range^{j^*}_{min}). \quad (9)$$

If $d^j_{\max} > range^j_{\max}$ for some $j = j^* \in \{1,2\}$, then

$$d^{j^*}_{min} = \varphi^{j^*} - (range^{j^*}_{max} - \varphi^{j^*}), \quad (10)$$

$$d^{j^*}_{max} = \varphi^{j^*} + (range^{j^*}_{max} - \varphi^{j^*}). \quad (11)$$

As the result of implementation of this hybrid version of the PSO algorithm the search of the solution of one or another optimization problem can be carried out.

4 The results of experimental studies

The feasibility of application of the proposed hybrid algorithms is confirmed by the results of experimental studies. In particular, the problems of search of the optimum global solution of the several test functions and the problem of search of the optimum parameters values of the SVM classifier were considered.

The several versions of the PSO algorithm were used by performing experimental studies:

- the canonical PSO algorithm (hereinafter referred to as the basic PSO algorithm);
- the hybrid version of the basic PSO algorithm based on the classical GS algorithm (hereinafter referred to as the PSO-GS algorithm);
- the hybrid version of the basic PSO algorithm based on the DOE algorithm (hereinafter referred to as the PSO-DOE algorithm).

The software implementation of these algorithms was conducted by using a high level programming language Python (programming environment Python 3.5). Herewith, the SVM algorithm from the machine learning library Scikit-Learn was used.

The implementation of the optimization algorithms for the test functions. The comparative analysis of these three optimization algorithms was implemented within the framework of solving the problem of search of the global optimum of several test functions. In particular, the results of experimental studies for the Rastrigin, Rosenbrock and sphere objective functions are given in [7].

The obtained results [7] allow to say that the basic PSO algorithm is characterized by the worst values of the quality indicators, such as the average time of convergence, the average convergence rate, the average value of the objective function, the proportion of successful runs in comparison with the PSO-GS algorithm and the PSO-DOE algorithm.

Herewith, the PSO-DOE algorithm allows finding the global optimum of the test functions, on average, in less time than the PSO-GS algorithm does. A large proportion of successful launches is provided and a smaller error in calculating the values of the global optimum of the test functions is achieved by implementing of the PSO-DOE algorithm [7].

The implementation of the optimization algorithms for setting parameters values of the SVM classifier. The advisability of applying the PSO-GS algorithm and the PSO-DOE algorithm for problem solving of search of the optimum parameters values of the SVM classifier was confirmed experimentally.

The studies were conducted using data from Statlog project and from UCI machine learning library. The binary classification was performed for all datasets. The following datasets were used in the present work (Table 1):

- the dataset for medical diagnosis of heart disease – Heart (270 instances, 13 characteristics, the source is http://achive.ics.uci.edu/ml/machine-learning-databases/statlog/heart/);
- the dataset for credit scoring of applications for consumer credits – Australian (690 instances, 14 characteristics; the source is http://archive.ics.uci.edu/ml/machine-learning-databases/statlog/australian/);
- the testing dataset – МОТП12 (400 instances, 2 characteristics; the source is http://www.machinelearning.ru/wiki/index.php?title=Изображение:MOTP12_svm_example.rar).

The calculations using the hybrid versions of the PSO algorithm were performed with different total number of the grid nodes (i.e., in case of different total number of evaluations of the objective function in the grid nodes) calculating for the PSO-GS algorithm and the PSO-DOE algorithm correspondingly to formulas:

$$\gamma = (r+1)^2, \quad (12)$$

$$\gamma = 13 \cdot h, \quad (13)$$

where r is the number of splitting intervals on each j-th range of grid search $[d^j_{\min}, d^j_{\max}]$ ($j=1,2$); h is the number of iterations of the DOE algorithm.

The selection of the optimum parameters values of the SVM classifier was performed on the results of several experiments for different values of the parameters r and h (Figure 3, a special case for a dataset МОТП12). In the present work the following values were selected as optimum based on the criterion of minimum value of the time of the first finding of the optimum: $r=5$ and $h=5$.

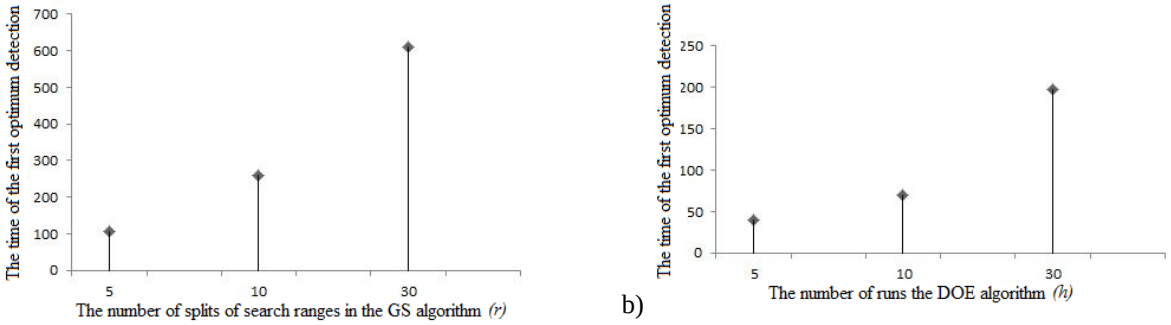


FIGURE 3 The determination of the optimal number of calculations on the grid with implementation of the hybrid versions of the PSO algorithm based on: a) the GS algorithm, b) the DOE algorithm

The radial basis kernel function (2) was used during the development of the SVM classifier. Therefore, the PSO algorithm and its hybrid versions were applied for searching the optimal values of two parameters of the SVM classifier: the regularization parameter C and the coefficient of the kernel function σ (i.e., the calculations were performed in the search space D-2). Herewith, it is supposed that the radial basic kernel function is a priori optimum in the context of solving classification problem.

The parameters values of the SVM classifier are relied as optimum if they provided the high classification accuracy and the minimum number of the support vectors in the training dataset.

The assessment of the classification quality can be performed using the different indicators of the classification quality: the overall accuracy (*Acc*), also called the total ratio of correct answers (the overall success rate, *OSR*); the sensitivity (*Se*), also called the indicator of completeness (the recall, *Re*); the specificity (*Sp*); the precision (*Pr*); the balanced F-measure (*F*1). These indicators are computed by the following formulas:

$$OSR = \frac{TP + TN}{TP + TN + FP + FN}, \quad (14)$$

$$Se = \frac{TP}{TP + FN}, \quad (15)$$

$$Sp = \frac{TN}{TN + FP}, \quad (16)$$

$$Pr = \frac{TP}{TP + FP}, \quad (17)$$

$$F1 = \frac{2 \cdot Pr \cdot Re}{Pr + Re}. \quad (18)$$

where TP is the number of true positive observations; TN is the number of true negative observations; FP is the number of false-positive observations (the error of type II); FN is the number of false negative observations (the error of type I); *Re*=*Se*.

The indicator of overall accuracy *OSR* presents the ratio of true predicted observations in relation to the total number of observations of the classifier.

The indicator of sensitivity *Se* presents which part of the total number of real positive observations is predicted as the positive, i.e. it shows how much the classifier is "pessimistic" in its assessments or how often it "throws off" the observations of the correct class (this occurs at low value of the indicator *Se*). This indicator is also called the indicator of completeness *Re*.

The indicator of specificity *Sp* presents which part of the total number of real negative observations is predicted as the negative.

The indicator of precision *Pr* presents how many of predicted positive observations are really positive, i.e. it shows how the classifier is optimistic in its assessments or how often it "prefers" (and this occurs at low value of the indicator *Pr*) to connect the observations of other classes to the specified.

The indicator of balanced F-measure (*F*1) calculates the harmonic mean between the indicator of precision *Pr* and the indicator of completeness *Re*. In formula (18) the same weight is assigned to both indicators.

To avoid the underfitting and the overfitting of the SVM classifier it was supposed that the high classification accuracy is achieved if the number of errors on the training and testing datasets is minimal, herewith, the number of errors of the SVM classifier on the training and testing datasets are virtually identical [7].

The same values of the parameters of the PSO algorithm and the same search ranges of the parameters values of the SVM classifier were defined for all runs of the optimization algorithms.

In order to ensure the objective comparison of the experiments results, the runs of the basic PSO algorithm and the proposed PSO-GS algorithm and the PSO-DOE algorithm for a particular dataset were initialized by the identical randomly generated initial population of particles.

Besides, the identical random partitions of the original dataset into the training and testing datasets are used. The size of the testing dataset was 20% of the original dataset during the process of the SVM classifier development.

The ROC analysis [6] was used for the quality assessment of the binary classification. The ROC curve, also known as the error curve, displays the ratio between the rate of correct positive classifications of the total number of

positive classifications (true positive rate – *TPR*) and the rate of incorrect positive classifications of the total number of negative classifications (false positive rate – *FPR*). The AUC (the area under the ROC curve) gives quantitative interpretation of the ROC curve. It is believed that the higher AUC is, the better the classifier is.

The ROC curves for the SVM classifiers built by using data of the testing datasets for three original datasets described above and the AUC indicator for each SVM classifier are presented in Figure 4.

The parameters setting of the SVM classifiers was performed by using the basic PSO algorithm and its hybrid versions.

At first glance the results of ROC analysis including the results of the comparative analysis of the values of AUC indicator present that the differences of the SVM classifiers is quite small and it is difficult to determine the quality classification. However, presenting the classification results in the form of Table 1, which shows the number of correctly and wrongly classified objects, the advantage of the classification quality should be given to the hybrid versions of the basic PSO algorithm.

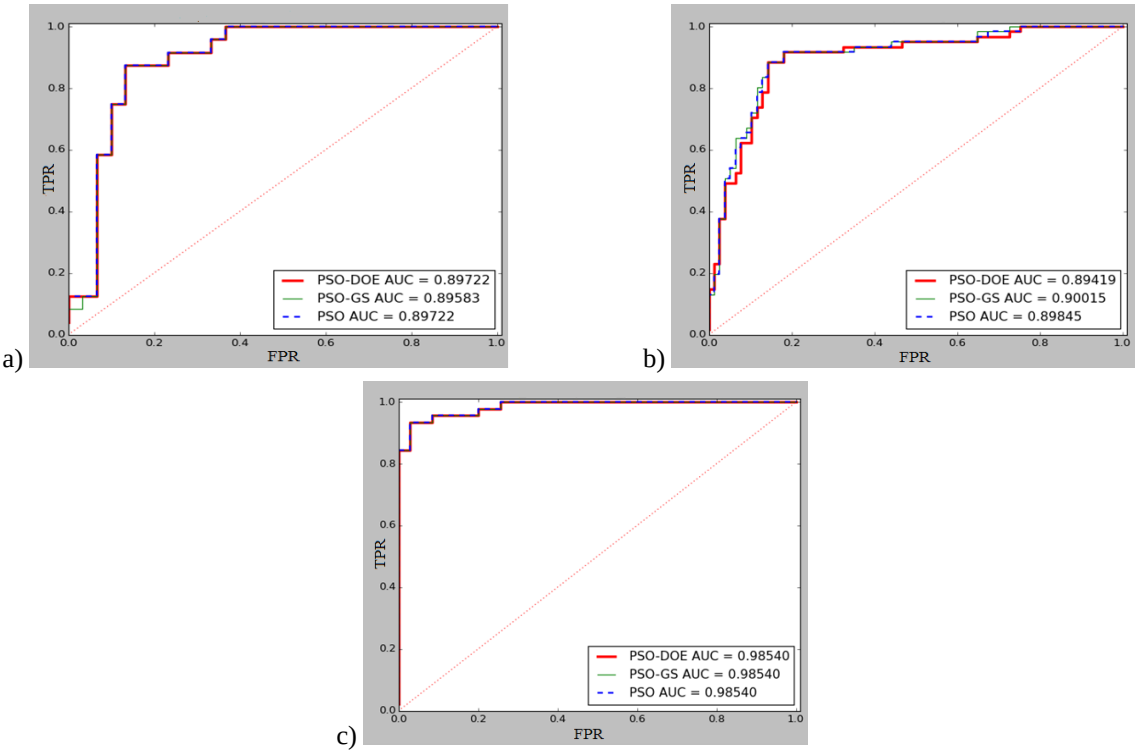


FIGURE 4 The ROC curves for the SVM classifiers built by using the basic PSO algorithm and its hybrid versions: a) for Heart dataset, b) for Australian dataset, c) for MOTII12 dataset

TABLE 1 The classification results

Dataset	Number of objects	Number of features	Version of PSO algorithm	Found parameters		Number of errors (class «1»/class «-1»)		Number of the support vectors	Accuracy (%)	Sensitivity (%)	Specificity (%)	F-measure	Iteration of the first detection	Search time (sec.)	
				C	x_1	At the training	At the testing							Time of the first detection	total
Heart	270	13	basic PSO	8.87	0.05	6 (2/4)	7 (3/4)	108	95.19	96.67	93.33	0.9571	17	523	642
			PSO-GS	9.82	0.05	5 (1/4)	7 (3/4)	107	95.56	97.33	93.33	0.9605	9	362	714
			PSO-DOE	**9.98**	**0.05**	**5 (1/4)**	**7 (3/4)**	**107**	**95.56**	**97.33**	**93.33**	**0.9605**	**6**	**243**	**712**
Australian	690	14	basic PSO	9.48	0.13	11 (5/6)	18 (7/11)	276	95.80	96.09	95.56	0.9532	12	1546	2872
			PSO-GS	9.73	0.13	12 (5/7)	18 (7/11)	273	95.65	96.09	95.30	0.9516	5	1031	3481
			PSO-DOE	**9.99**	**0.13**	**10 (5/5)**	**18 (7/11)**	**273**	**95.95**	**96.09**	**95.82**	**0.9547**	**4**	**789**	**3292**
MOTP12	400	2	basic PSO	9.89	9.45	12 (5/7)	4 (3/1)	122	96.00	96.10	95.90	0.9610	8	171	441
			PSO-GS	9.89	9.49	12 (5/7)	4 (3/1)	121	96.00	96.10	95.90	0.9610	4	107	653
			PSO-DOE	**10**	**9.47**	**12 (5/7)**	**4 (3/1)**	**121**	**96.00**	**96.10**	**95.90**	**0.9610**	**1**	**40**	**509**

Based on Table 1 we can conclude that the PSO-GS algorithm and the PSO-DOE algorithm solve the problem of searching optimum parameters of the SVM classifier more efficient than the basic PSO algorithm does. The hybrid versions of the PSO algorithm allow reducing the search time of the optimal solution by 3-5 times and the best values of the quality indicators of the SVM classifier are achieved. In particular, we received the highest values of the overall accuracy *OSR*, the sensitivity *Se* and the specificity *Sp*, and the smaller values of the number of the support vectors.

Herewith, the using of the PSO-DOE algorithm provides the best rate of convergence to the optimal solution in most cases (i.e., less time of the first detection of the optimal solution).

5 Conclusion

The results of experimental studies confirm the feasibility of application of the proposed hybrid versions of the PSO algorithm in the framework of the solving the problem of the effective SVM classifier development. The advantage of hybridization of the basic PSO algorithm with the grid search algorithms is the reducing of time for searching the optimum parameters values of the SVM classifier, while maintaining, and in some cases improving, the quality of classification decisions.

The obtained results were achieved by the union of capabilities of the PSO algorithm with the positive features of the grid search algorithms. In particular, the additional search on the grid in the area of the potential globally best position of the particles in the swarm was implemented for updating the population of the particle swarm and removing the "worst" particles.

The further research may be associated with the development of the recommendations for applying of the hybrid optimization algorithms in the framework of the solving of the problem of the SVM classifiers development for unbalanced datasets.

References

[1] Joachims T 2005 A support vector method for multivariate performance measures *In Proceedings of the International Conference on Machine Learning (ICML)* 377–84

[2] Lean Yu, Shouyang Wang, Kin Keung Lai, Ligang Zhou 2008 *Bio-Inspired Credit Risk Analysis: Computational Intelligence with Support Vector Machines* Springer-Verlag: Berlin Heidelberg p 247

[3] Vapnik V 1998 *Statistical Learning Theory* John Wiley & Sons: New York p 740

[4] Ren Y, Bai G 2010 Determination of optimal SVM parameters by using GA/PSO *Journal of Computers* **5**(8) 1160–8

[5] Demidova L, Nikulchev E, Sokolova Y 2016 Big data classification using the SVM classifiers with the modified particle swarm optimization and the SVM ensembles *International Journal of Advanced Computer Science and Applications* **7**(5) 294-312

[6] Demidova L, Klyueva I, Sokolova Y, Stepanov N, Tyart N 2017 Intellectual Approaches to Improvement of the Classification Decisions Quality On the Base Of the SVM Classifier *Procedia Computer Science* **103** 222-230

[7] Demidova L, Klyueva I 2016 Razrabotka i issledovanie gibridnyh versij algoritma roja chastic na osnove algoritmov poiska po setke *Vestnik Rjazanskogo gosudarstvennogo radiotehnicheskogo universiteta* **3**(57) 107–17 *(in Russian)*

[8] Demidova L, Klyueva I, Pylkin A 2016 The Study of Characteristics of the Hybrid Particle Swarm Algorithm in Solution of the Global Optimization Problem *5th Mediterranean Conference on Embedded Computing (MECO)* 322–5

[9] Jun Sun, Choi-Hong Lai, Xiao-Jun Wu 2011 *Particle Swarm Optimisation: Classical and Quantum Perspectives* CRC Press p 419

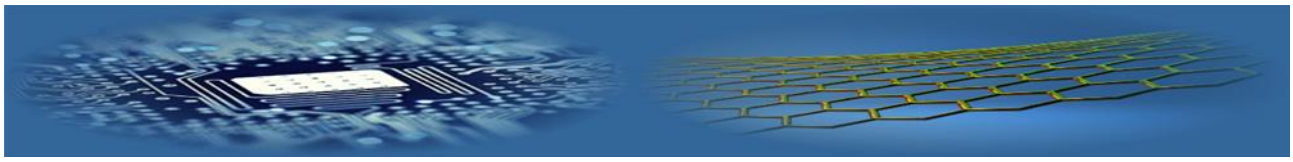
AUTHORS

Liliya Demidova, 1968, Russia

Current position, grades: professor of Ryazan State Radio Engineering University and Moscow Technological Institute.
University studies: Ph.D (technical sciences, 1994), Dr.Sc (technical sciences, 2009).
Scientific interest: fuzzy set theory, evolutionary algorithms, machine learning.
Publications: more than 250 publications in the artificial intelligence.
Experience: more then 27 years.

Irina Klyueva, 1989, Russia

Current position, grades: post-graduate student of Ryazan State Radio Engineering University.
University studies: Ph.D candidate.
Scientific interest: evolutionary algorithms, machine learning.
Publications: more than 15 publications in the artificial intelligence.
Experience: more then 3 years.



A literature review on algorithms for the load balancing in cloud computing environments and their future trends

Aanjey Mani Tripathi*, Sarvpal Singh

Madan Mohan Malviya University of Technology, Gorakhpur-273010 Uttar Pradesh, India

Corresponding author's e-mail: aanjeymanit09@gmail.com

Received 15 April 2017, www.cmnt.lv

Abstract

Cloud computing is a computing utility which provides basic service for computing. It is a high performance distributed computing which has the huge collection of virtual resources that can be easily accessed anytime using the internet similar to pay as you go, model. A cloud defines set of virtual computers connected to each other in a form of parallel and distributed system. It ensures the dynamic provision of resources based on service level agreement (SLA) to ameliorate one or more objectives. To attain this goal several research challenges have been faced in the area of cloud computing, And the Load balancing is one of them, which aim at equalizing the workload among all the obtainable nodes by minimizing execution time, minimizing communication delays, maximising resource utilization and maximising throughput. This paper disburses a literature review of existing load balancing algorithms suggested so for and categorized under different metrics enveloping the advantages and disadvantages of each. An overview of the important research challenges of these algorithms is presented at the end with some possible ideas for improvement.

Keywords:

Cloud computing,
Load balancing algorithm,
virtual machine

1 Introduction

Cloud computing [1] is a subscription based service like pay-as-you-go model [2] which delivers software, infrastructure and the platform kind of services [3]. These services are categorized as the Infrastructure as a service (IaaS), Platform as a service (PaaS), and Software as a service (SaaS) in the industry. Cloud computing is introduced to reduce the cost of the hardware and software. It also aims to make the next generation data center more powerful so that it can provide dynamic and flexible services to the consumer. Deployment of cloud computation makes the industry stronger and also gives the time to focus on innovation and creativity. This will lead the IT services [4] to the higher level and will help in developing the world [5].

Cloud computing is a darwinism of the parallel computing, grid computing, and distributed computing [6]). It deals with trading the resources in an efficient way according to the need of the user. Also, it is a large scale of heterogeneous resources that resides in the data centre [7]. The virtualization ability of the cloud computing hides the heterogeneity of the resources which makes it different from other computing technologies introduced previously. The other features include user-oriented approach which delivers the services as per user necessities and virtualization technology [8] that is used to pack the resources to make it scalable and flexible.

The working of the cloud computing is described as dispatching the tasks to the pool of resources which consists of several computers. It provides enormous services including storage, power, and several software services according to the need of the task [9]). The business and virtualization technology [3] used by the cloud computing have taken the technology to a new height, leaving the responsibility of resource allocation to the virtualization of virtual machine layer. Along with these advantages, Cloud computing faces a number of research challenges such as network level migration (10), ensuring appropriate access control [11], security [12], data availability [13], Official quagmire and transitive trust issues, Data lineage, data origin and unintended leak of sensitive information [14], besides this the most frequent problem in cloud computing is load balancing. By paying more attention to the load balancing [15], like, various new features are introduced in cloud computing. Balancing of Workloads among available nodes in cloud computing is an important facet. An efficient, effective load balancing scheme ensures an efficient resource utilization [16] by the provisioning of resources [17] to cloud users on demand basis by using pay-as-you-go-scheme. Load balancing equalizes the workload among the nodes by minimizing the execution time [18], minimizing communication delays, maximizing resource utilization and maximizing the throughput.

The motive of this paper is to survey of the maximum available algorithms that have been proposed for providing a contrast to these schemes on difference metrics that examine popular load balancing algorithms with the challenges that could be addressed. Here the load balancing algorithms have been partitioned in three main categories, static, dynamic and hybrid. To the best of our knowledge & efforts, this literature survey presents load balancing algorithms with a determined focus on cloud computing.

Briefly, the contribution of this paper is as follows.

- Giving an overview of existing cloud computing load balancing challenges.
- Providing a literature review of the existing load balancing algorithms and the way of their application.
- Advantages and disadvantages of existing load balancing algorithms.
- Future research challenges of load balancing in cloud computing.

This paper examines the related work and explore the load balancing algorithms that can be categorised all the static, dynamic and hybrid algorithms. Firstly the focus is on the cloud computing load balancing challenges, description of static, dynamic and hybrid algorithms, which further continues with the discussion over several parameters on which we determine the effectiveness of algorithms. At the end a comparative analysis of these algorithms on the discussed metrics is made which would help the future researchers in their work.

Many types of research have been done in the field of cloud computing and a number of challenges identified as that counts resource provisioning, job scheduling and load balancing. In this section, we analysed some papers of load balancing in cloud computing.

In [19] author contributed to this research area by providing survey and comparative analysis on five different meta-heuristic techniques of Cloud and Grid computing : Ant colony Optimization (ACO), Genetic Algorithm (GA), Particle Swarm optimization (PSO), League Championship Algorithm (LCA) and BAT algorithm [19]. They also dispensed the comparison of these algorithms. Although, this paper only restricted to Meta- heuristic techniques.

In [20] author discussed a number of existing load balancing algorithms and dispenses the comparison on certain metrics i.e. performance, scalability and overhead etc. continued by synthesis of algorithms on certain perspective such as energy consumption and carbon emission. However this paper mainly focuses on green computing based load balancing algorithm.

Another team of authors [21] in their paper provided an overview on distributed load balancing algorithm counting parameters i.e. fault tolerance, high availability and scalability. The paper investigated three algorithms Honeybee Foraging behaviour, active clustering and biased random sampling on parameters. Though, the paper mainly focused on distributed load balancing.

In [22] author targeted at two load balancing approaches static and dynamic scheme with computational synthesis on the performance of various load balancing algorithms. It also summarized advantages and disadvantages. However, main impetus of this paper is to analyse algorithms on the basis of time factor.

In [2] author have evaluated various load balancing policies. They focused their observations on criteria including average response time, datacentre service time and total cost. The simulation results & their work prove that the round robin algorithm performance was comparatively better than other methods. The scheme presented by their only covers limited parameters.

It is important to point out that none of the above discussed papers presents load balancing algorithms by including all three approaches static, dynamic and hybrid. Thus our work aims at including all the three approaches of load balancing algorithm with their comparative measures and covering the future challenges of each.

2 Load balancing strategies

2.1 CLOUD COMPUTING LOAD BALANCING CHALLENGES

Cloud comprises of massive resources and the management of these resources requires proper layout and high level planning. Before designing an algorithm, resource provision must be taken into consideration covering overall scenario and have to identify the main issues that could leave an impact upon the algorithm performance [17]. In this section we have discussed the challenges to be addressed while trying to propose an optimal algorithm to resolve the issues of the load balancing in cloud computing.

The Challenges to be taken into consideration are:

Spatial Distribution of the cloud nodes: Some algorithms are developed only for the intranet where nodes are closely located and where communication delays are avoidable. However the major challenges to develop a load balancing algorithm that could work well with the spatial distribution of the cloud nodes with the consideration of a number factors like [23]:

- Speed of network links between the nodes.
- The distance between the user and the task processing nodes.
- Distance between the nodes that involved providing the services.
- And the High Delay among spatial distributed nodes.

Environment: Cloud computing technology is an integration of both heterogeneous and homogeneous environment. Both environments have their own characteristics and their own differentiable criteria. For that purpose it is important to develop an efficient load balancing that works well for both environment (Mayanka Katyal 2013) (Al Nuaimi et al. 2012)

Storage/Replication: Any full replication algorithm could not provide an efficient utilization of the storage because, the same data gets stored at number of nodes. In case of full replication algorithm cost is the downside due to higher storage requirements. However with the partial replication algorithm we could save parts of the data space at each nodes (with a certain level of overlap) based on the capabilities like capacity and processing power of each nodes. With this capabilities it increases the resource utilization but resulted in raising the complexity of the load balancing algorithm in checking the availability of the data set parts across the different cloud nodes [23].

Algorithm Complexity: For an effective Load balancing algorithm Complexity should be low because layer complexity maximizes the complex operations, which may show negativity in performance issues of results or

degraded performance. Furthermore, when an algorithm requires more information and higher communication for monitoring and control, delays could prove to be troublesome and could result in efficiency drop. Therefore, load balancing algorithm must be at its simplest form (Al Nuaimi et al. 2012).

Point of Failure: Load balancing algorithm aims at controlling the load balance and collecting data from the different nodes. For this purpose the algorithm must be designed in such a way that it incurs no any single point of failure. Some centralized algorithms though provide an effective and efficient mechanism but have the issues of a single and central administrator for the entire system. The Distributed load balancing algorithms are more complex and require more coordination but they proved through simulation results that they are better approach and provided better solution. Hierarchical Load balancing algorithm are also a better solution as they work on master slave mode with the issues of threshold policies, information exchange criteria and failure intensity(Al Nuaimi et al. 2012)(Mayank Katyal 2013) (Al Nuaimi et al. 2012).

2.2 LOAD BALANCING METRICS

Load balancing distributes the local load among the resources and ensures resource utilization with higher user satisfaction. A suitable load balancing mechanism must have some properties that could make it distinguishable and useful. These properties should provide higher throughput, higher response time, must have fault tolerance, scalability, high performance, efficient resource utilization, and low overhead. Here we have discussed these important metrics as follows:

Throughput: Describes the sending and receiving rates of data of the total number of completed task on a given input at a given time unit. For better performance of cloud system high throughput rate is required. If the throughput is high then adoptability must be high [26].

Response time: Time taken by load balancing mechanism to respond for a submitted request [27, 28].

Fault Tolerance: Continue processing without stops if any node encounters a failure then the system redirect the work to another location of data. It is the capability of the mechanism [29].

Scalability: Scalability is a capability of the system to cope and perform under an increased or expanding workload. A system that scales well will be able to maintain or even increase its level of performance or efficiency when tested by larger operational demands [30].

Resource Utilization: Refers to the utilization of resources in system. An efficient load balancing algorithm must have higher resource utilization [31].

Overhead: Refers to the communication overhead caused by communication between the nodes during movement of tasks [29, 32].

Performance: Performance refers to effectiveness of the system after complete execution of load balancing algorithms. If all listed parameters perform well then it will maximize the performance of entire system [11, 33].

2.3 LOAD BALANCING ALGORITHMS

There are a number of load balancing algorithms which work to achieve their task on different layers of cloud with different level of complexities. For better load balancing researchers aim at developing more complex load balancing algorithms. But with prons it also increases its cons like processing load, overhead, and execution time (34). Load balancing algorithms can be categorized on the basis of spatial distribution of nodes (topology) and the environment shown in Figure 1. Table 1 summarizes the types of algorithms with their knowledge base information along with addressed issues and drawbacks. Table 2 classifies some existing load balancing algorithms on the basis of environment & topology.

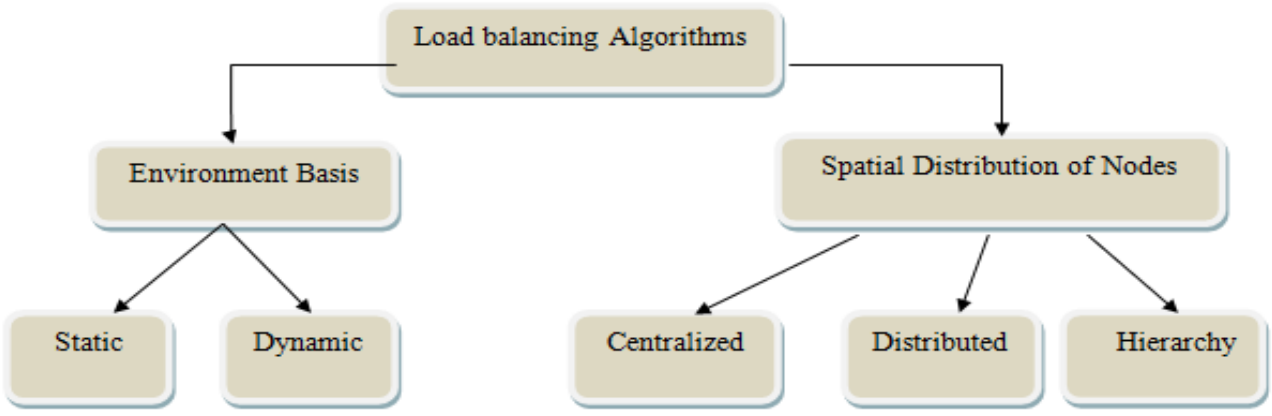


FIGURE 1 Categorizations of load balancing algorithm

TABLE 1 Types of algorithm with their knowledge base informationalong with addressed issues and drawbacks

Types of Algorithms	Knowledge Base	Addressed issues	Usage	Drawbacks
Static	Previous Knowledge is mandatory about each node statistics and their user requirements	Response time Resource Utilization Scalability Power Consumption and Energy Utilization Make span Throughput/Performance	Used in Homogenous Environment	Flexibility issues Scalability issues Is not compatible with changing user requirements as well as load
Dynamic	Run time statistics of every node are observed to embrace to changing load requirements	Under loaded processor location where load will be transfer by an overloaded processor. Task transfer to a remote machine. Load Estimation. Information Gathering. Limiting the number of migration. Throughput.	Used in Heterogeneous Environment	Complexity. Time Consuming
Centralized	Any single node or server is responsible for sustaining the statistics of whole network and updating it time to time	Threshold policies. Throughput. Communication between central server and processors in network. Failure Intensity. Associated Overhead.	Useful in small network which have low load	No fault tolerant. Overhead central decision making node
Distributed	Every processor which is the part of network responsible for load balancing and each maintain their own local databases (e.g. MIB) to make efficient load balancing decisions	Selection of processor that take part in load balancing. Migration time. Interprocessor communication. Information exchange criteria. Throughput. Fault tolerance.	Useful in large and heterogeneous environment	Complexity of Algorithm. Communication overhead
Hierarchy	Nodes at different levels of hierarchy communicate with the nodes below them to get information about the network performance	Threshold policies. Information exchange criteria. Selection of nodes at different levels of network. Failure intensity. Performance. Migration time.	Useful in medium or large size network with heterogeneous environment	Less fault tolerant. Complex

TABLE 2 Classification of some existing load balancing algorithm on the basis of Environment & Topology

Dynamic	Centralized	Distributed	Hierarchy
ESCEA Throttled Biased Random Token Rating Genetic algorithm Active Clustering INS (Index Name Server)	Round Robin Min-Min Genetic Algorithm	Biased Random Sampling Map Reduce Active Clustering INS (Index Name Server)	Map Reduce

Summary: Some of the load balancing algorithms is as follows.

Round robin algorithm

The round robin algorithm (Dave and Maheta (2014) is one of the most popular and simplest algorithm. It allocates the resources to task or requests on the basis of time quantum. In this, time is divided into multiple slices and is allocated to the requests. It utilizes the principle of time scheduling. The resources of the service provider are provided to the requesting client on the basis of time slice. The first node is selected randomly and then it allocates job to other node on time quantum in circular manner. In round robin, loads are equally distributed on all the nodes. The scheduler begin with a node and moves on following node after a VM is assigned to that node. The iteration continued until all the nodes have been assigned to at least one VM and this process continuously occurs and is restarted from the first node. Thus, in this case, the scheduler does not need to wait for the exhaustion of the resources of a node before moving to the next node. This deficiency has been controlled in the weighted round robin. Round robin maintains the allocation order of requests locally. It send the requests to that node which has the least number of connections, and because of this, for some periods of time, some nodes may be heavily loaded and some may remain idle [35]. This problem is solved by CLBDM (Central Load Balancing Decision Model) which is based on session switching at application layer. CLBDM calculates the connection time between user and the node and perform allocation on the basis of predefined threshold. Because of these features there has been lot of research carried out to improve performance of this algorithm. Round robin works efficiently when all the servers have the same or similar performance and are running with equal loads.

Genetic Algorithm

Dasgupta et al. (2013) proposed a Genetic Algorithm (GA) used as a soft computing approaches which uses the mechanism of natural selection strategy. The algorithm

balances the load of cloud infrastructure with an approach of minimizing the make span of tasks. From the simulation of this algorithm it is proven that it surpass the existing algorithms like First Come First Serve (FCFS), local search algorithm Stochastic Hill Climbing (SHC), and Round Robin (RR). In Genetic algorithm have three operations: selection, genetic operation, and replacement. All three operations are meant for the purpose of the spread-out search space, to apply complex objective function and to avoid being trapped into local optimal solution [19, 37, 38].

Index Name Server

To avoid the redundancy and storage replication of data Wu et al. (2012) have defined a novel architecture for data centre management mechanism Index Name Server (INS), which synthesize de-duplication with access point selection optimization techniques to upgrade the performance & efficiency of the cloud storage system. During deployment of network architecture Distributed hash table (DHT) used by Index Name Server to organize the distribution of all the data and nodes. To find an optimal path on a given weight it uses the concept of time and weight in ad-hoc network. According path preferences and to figure out the performance of each node and pick up the shortest path. To attain a flexible system performance and the best resource allocation there are several transmission matrices included in the environment and the records are table driven in INS. All the other schemes and method uses the backup strategy and it makes wastage of resources. But in INS, it excludes the scanning procedure of backup strategies and reduces the backup cost. [39] also defines future objective of INS to improve the accuracy of backup selection through considering data rates and formats, user habits, and on the basis of file formats and avoiding peak hours statistics. However, it is a centralized and complex algorithm suffers with single point failure issue (Al Nuaimi et al. 2012).

Ant Colony

Kalra et al. (2015) Joshi et al. (2014) Defined a novel Ant colony based algorithm to balance the load in cloud computing by locating the under loaded node, and experimentally proves this approach is to be more appropriate than the traditional approaches like First Come First Serve (FCFS), local search algorithm like Stochastic Hill Climbing (SHC), another soft computing approach Genetic Algorithm (GA) and some existing Ant Colony Based strategy [19]. ACO is a random search algorithm which works like ant colonies. Ants searches food and connect to each other through pheromone which is evaporative stuff on paths travelled. It also guarantees that QoS requirement of Ant colony Based load balancing policy in cloud computing customer job. All the jobs are predicted to be holding the same priority though Fault torrent issues are not taken into account. Here few suggestions and ideas for the future research work are proposed on the cloud scheduling technique too [41]. The pheromone value evaluation is conducted using fault tolerance and different function variation.

OLB

For better resource utilization and improvement in response time Aditya et al. (2015) Hans et al. (2015) defined Opportunistic load balancing (OLB) ignoring expected task execution time and thus could not achieve good scheduling performance in make span. It is static load balancing algorithm so there is no need to consider the current workload. Its aim is to keep each node involved in execution process of tasks [5]. Random execution of unexecuted task on currently available nodes is conducted. Processing of this algorithm is found to be slow because it does not calculate the current execution time.

Min-Min scheduling

Kokilavani et al. (2011) introduced Min Min Algorithm which take into account both, the minimum completion time and minimum execution time and selects nodes for executing tasks based on the Min-Min completion time. It is a static approach where the cloud manager firstly identifies the minimum execution time of unassigned task from the unassigned task set and minimum completion time of resources from all available resources. Being a static algorithm, it requires having prior knowledge of matrices related to the job. Then it assigns the task to the resource which has minimum execution time [44]. the job having maximum execution time has to wait for unspecified period of time to execute until all the tasks are assigned and updated. Results prove that this algorithm is a better scheme, that reduces the makespan than others and responds next to Genetic algorithm having rate of enhancement is also lesser in maximum scenarios [6]. Min Min algorithm also suffers with the starvation problem, and do not care about energy consumption [18].

CLBDM (Central load Balancing Decision Model)

Radojevic et al. (2011) discussed the Central load balancing decision model (CLBDM) that works as an automated administrator, compute the connection time between client and server on a given cloud resource by computing the overall execution time of task. If the connection time is over than a defined threshold then there an issue may occur (Al Nuaimi et al. 2012). And if the issue is endowed, then the task is terminated and assigned to another node using traditional round robin algorithm. CLBDM algorithm is refinement of Round Robin algorithm and is based on session switching at application layer. However, this algorithm suffered with single point failure and threshold might not be applicable in all cases.

WLC

Ren et al. (2011) introduced commonly used weighted least connection (WLC) concept that is one among the good dynamic algorithm. It does not consider parameters like distance between client and servers, service capability, processing speed, storage capacity and bandwidth. This algorithm, start with the predictions of weight of each service node and the number of connections on each service nodes. The WLC allocates the task to service node on the basis of[min [{C(Si)/W(Si)}], where C is number of allocated connections and W is the weight of service node [46]. Mean allocation is done on comparison of the sum of the connections with each service node and allocates the task having the least number of connections [47]. However, it also suffers with some issues like connections on service node cannot indicate the load well, and during long run constant weight cannot be corrected and the node is bound to divergence from the actual load condition due to which it faces load imbalance [48].

ESWLC

To handle log connectivity applications Ren et al. (2011) introduced Exponential smooth forecast based on weight least

connection (ESWLC) which allocates the resource with least weight to a task and take into account time series and tribulations. Based on the node and capabilities of the nodes, task is assigned to a node. ESWLC takes the decision to allocate a certain task to a node predicted on the basis of experience of node's cpu potency, number of connections, recollection, memory usage, the size of disk occupations. Exponential smoothing forecasting is a prediction based algorithm considering time series. This algorithm use historical data, and distinguish them using the smoothing factor. After smoothing, recent data has been makes a great impact on predictive value then long term data [48]. It establishes the training set using historical data, and then develop prediction model and the value is predicted for the next moment, which has minimum value next time continued by sending of next request connection (Al Nuaimi et al. 2012).

LBMM

Wang et al. (2010) proposed Load balancing Min Min (LBMM) using three levels of parameters for the allocation of resources in dynamic environments. This algorithm uses OLB (Opportunistic load balancing) as it base algorithm. It embraces Min Min scheduling and load balancing mechanism [18] which can avoid the non-essential assignment and utilizes the better executing efficiency. In Min Min algorithm workload of each node does not consider. It only recognizes the completion time of every task. Due to this some nodes may always get busy and some node may still remain idle. Therefore, load imbalance has been raised and the execution time of every node has been decreased. This algorithm has been processed in three layers from. In the first one the task assigned by the request manager to an appropriate service manager. The task has been divided in logical autonomous subtask by the service manager in the second phase and the execution of subtask finished in the last one. The selection of service node to execute the task has been done on the basis of the remaining CPU space (node availability), remaining memory and the transmission rate [42, 43]. However this algorithm reduces the makespan and increases the resource utilization.

Biased random sampling

[49] Defined Biased random sampling algorithm is a distributed load balancing algorithm. It creates the virtual graph which works as a knowledge base for this algorithm. And the virtual graph is a graph that represents the connection between every node and through this we know the appropriate load on the server. Every node is assumed as a vertex node and each node have a degree to represents unused resource. Each node must have one in-degree. It also uses the walk length parameter for processes, which is the traversal from one node to another. For allocating the task to a node, it begins with a random node and compares the walk length with the threshold and if it is equivalent or more than the threshold value formerly load balancer allocates the task to that node and decreases the degree of that node by one. If the degree of that node is less than one, then it is forwarded to next node which is the neighbour node of current node and walks length has been incremented by one (Randles et al. 2010). Biased random sampling algorithm performs very well with the equal or higher number of resource and provide high throughput with the utilization of increased system resources. However, it encounters performance degradation if the number of server increases due to additional overhead to calculate the walk length.

Three phase hierarchical scheduling

Wang et al. (2011) introduced to reduce the execution time of each node, three phase hierarchical scheduling has been proposed including multiple phases of scheduling. These phases comprises BTO (Best Task Order), EOLB (Enhanced opportunistic load balancing), and EMM (Enhanced Min Min). In three phases hierarchical scheduling algorithm request monitor performs as a head of the network and is liable for observing the service manager which in turn monitors service nodes. Task execution order which is based on demand task order scheduling and service priority defined by best task order scheduling algorithm in the first phase. It stores all the tasks, subsequent tasks and tasks which are in waiting queue, in a job queue on the basis of their execution order decided by BTO. With this it reduces the waiting time and execution time of tasks. In the next phase, it uses Enhanced opportunistic load balancing algorithm that consolidate traditional opportunistic load balancing and service manager threshold. And on the basis of the job characteristics the service manager threshold allocates job on suitable node using OLB. In third phase which is Enhanced Min Min combines Min Min scheduling and service node threshold that allocates the node with the guarantee of minimum execution time taken by that node to execute (Mayanka Katyal 2013). It may be possible that EMM chooses the best service node first and then use the service node threshold to execute the task in shortest time. Three phases hierarchical scheduling algorithm confirms that jobs are executed faster and in an effective way. However this algorithm is developed under static algorithm [49].

Honey Bee Behaviour

To maximize the throughput in cloud computing paradigm Dhinesh Babu et al. (2013) have developed Honey bee behavior load balancing algorithm. This algorithm is motivated by honey bee behaviour of food findings. Bees widely search for the food and upon finding the location of food, they broadcast through waggle dance and this dance provides an idea about quality ,quantity and location as well as distance of the food. Using this idea, other bees start to acquire the food. Then again they return and perform wagle dance which provide the same ideas useful for rest of others. Same aproach is applied in cloud computing for load balancing, in this when any Virtual machine has been overloaded then it migrate the task to underloaded VM, here tasks is considered as bees and food sources are VMs (Randles, Lamb, and Taleb-Bendiab 2010). After migrating the task it will update the details about load on that machine and available tasks with their priorities. This information is useful for other waiting tasks to choose VM based on their criteria as discussed. It also confirmed that a VM which has less number of high precedence task and if a high precedence task assigned to this then that task will be executed at its first. Sorting of VM will be in ascending order according to their load. This algorithm maximizes the throughput and reduces the waiting time in queue due to priority based techniques. Though here overhead is also low , but at the same time response time of VMs is found to be low.

DDFTP

Mohamed et al. (2013) defined dual direction downloading algorithm for FTP servers which provides fast

and reliable download of files. DDFTP, a dynamic algorithm which divides the file into two parts and servers start processing on the basis of a certain pattern i.e. a file m divided in to m/2, then one server start downloading form the 0 zero block in incremental order (left to right) and other one start downloading from m in reverse order (right to left). Less communication between servers increases the performance and decreases the network overhead. The task is considered as accomplished when two servers start to download file on their decided patterns and new task can be assigned to servers. DDFTP also guarantees that the full utilization of communication channels i.e. if a channel bandwidth is low and other one has high then centre of file m can be change, means with high bandwidth channel download more in comparison to the low one [35]. Means m rely on the load on servers and the bandwidth of communication channels. (Al Nuaimi et al. 2012) suggested some improvements for resource utilization using partial replication whereas sustaining the similar level of performance. [48]

Enhanced map reduce

Vakil et al. (2015) defined Map Reduce which is a programming model which was implemented for the processing large data sets. Map Reduce firstly breaks the input file of job into even sized chunks and then performs replication, for the fault tolerance objective. Every single chunk developed by a map task which generates a list of key value sets. Based On the key Output of Map is split and stored in buckets. After finishing all map tasks, reduce task phase gets started which apply reduce function on map outputs corresponding to each key. Map Reduce running on a particular cluster, consisting of a master node that holds information about the data chunks. Enhanced map reduce overcome the many shortcoming of Map reduce using some other factors.

HTV

Bhatia et al. (2013) defined a method to increase the performance of data center HTV dynamic load balancing algorithm. HTV algorithm incessant examines the available resources to know the status of the node and stores in queue. Node will be sorted in the queue according their weight factor and is updated each time when persistent monitoring is done. Weight factor derived from the parameter load on the server and the response time of nodes. To allocate the resources for a new job, it will refer the queue dynamically that provides high performance and efficiency. There are some steps involved is HTV, Node information queue which stores the information regarding nodes taking into account parameters namely available space (in respect of memory and processor) and the performance of the node. HTV performance algorithm determines the load of specific node, total available space, performance and stores this information in the queue. Now, the load balancer uses these details for the proper allocation and distribution of the resources. [54] also suggested addition of some other parameters for better improvement like load on specific server and priority of user task.

OLB+LBMM

Wang et al. (2010) combined the Opportunistic load balancing (OLB) and Load balancing Min-Min for the better executing efficiency. In OLB, Each node has opportunity to execute the task and its keep busy each node. Each task is split into subtask. LBMM considers the completion period for the job, the node having minimum completion period is executed first. However, it suffers with the load imbalance i.e. some node possess heavy load and some are idle. For this purpose OLB algorithm has been added to this algorithm. which provides better completion time along with the better resource utilization and response time [5]. Furthermore, cloud computing is not only static, it may be dynamic also. Here overhead is maximized with energy consumption also being its drawback. [20]

Stochastic hill climbing

[55] developed stochastic hill climbing which is a local search algorithm. There are two types of procedure to solve any optimization problem, first is the complete method, which provides a valid solution or prove that no such solution exists. Unfortunately, this type of algorithm requires exponential time in worst case. The other is the incomplete method that does not provide guarantee for valid solution rather than this method provides satisfying solution with high probability. However, these algorithms are most popular because of speed, effectiveness, and simplicity. Stochastic hill climbing is an incomplete method type algorithm which continuously moves to uphill and stops at 'peak' where not any neighbour have high value. Now, this algorithm chooses a random element from the uphill assignments. The probability to choose an element from that may vary with the steepness of uphill move. It starts mapping the assignment from the set of assignments and each assignment element is evaluated on some criteria which are closer to valid assignment. The best element from the set will be the next assignment [4]. This operation is repeated till the solution or to the stopping benchmark. Thus, in stochastic hill climbing algorithm have two components first one is the candidate producer used to draw one solution candidate to a set of possible successors and the second one is an evaluation measures which grades each valid solution (or invalid full assignments). Refining the evaluation leads to enhanced (or closer to valid) solutions. [55] suggested using other soft computing techniques for better improvement.

Compare & Balance

Sahu et al. (2013) defined a dynamic cost efficient compare and balance algorithm for better utilization of the resources which is based on probability to compare load of nodes/hosts. If the load of any randomly selected host has been low, then it transfer extra load on that host. It minimizes the migration time using live migration technique. A traditional load balancing algorithm only considers the memory, CPU and Bandwidth of host in any datacentre. But in dynamic compare and balance algorithm, two concepts have been used. First it is required to optimize at host level in cloud system i.e. CPU, memory and bandwidth and second, optimize the cloud system on basis of threshold which is decided on behaviour of user application. It is a green computing algorithm which tries to improve host machine efficiency by minimizing number of active host. In this, VM migration executed from the high cost to low cost physical host. It assumed that every physical host has sufficient memory. Its only disadvantage is its overhead [4, 42].

3 Open issues and future trends

In this section we have discussed considerable load

balancing algorithms concerned that have not been completely and comprehensively studied till now as a research prospective. We point out issues of some algorithms and exploring these in aspects of future scope.

- Round robin algorithm work efficiently when all the servers have the same or similar performance and are running with equal loads. Performance degrades with the different load on the servers because server with minimum resources receives the next job even it has not yet been able to process the current job. Need to develop an algorithm which solves this shortcoming by utilizing novel task distribution models.
- In genetic algorithm, we can apply variation of the crossover and selection strategies as a future work for getting more efficient and tuned results.
- INS algorithm is complicated to implement and to avoid such implementation complexity need to change in the structure which makes it less complex with same performance.
- In future work of Ant colony needs to study the triggering method of ant generation and the approach for pheromone update in order to considerably minimizing the searching time for candidate nodes.
- OLB algorithm for static environment with centralised balancing and the processing of this algorithm is found to be slow because it does not calculate the current execution time. For this in the future work we have to develop an algorithm which calculates the current execution time.
- Min-Min algorithm suffers with the starvation problem, and do not care about energy consumption. However, the biggest drawback is load imbalance and which one is the central issue for cloud providers. In future work of this, have to develop an algorithm which reduces the makespan and increase the resource utilization.
- CLBDM suffers with single point failure and the threshold might not be applicable in all cases. In the future scope of this is to develop an algorithm in distributed nature with good fault tolerance.
- Three Phase Hierarchical scheduling, however this algorithm is developed under static algorithm. So, for better performance need to develop an algorithm with dynamic environment.
- DDFTP suggested some improvements for resource utilization using partial replication although maintaining the similar level of performance.
- HTV also suggested addition of some other parameters for better improvement like load on specific server and priority of user task.
- Stochastic Hill Climbing suggested using other soft computing techniques for better improvement.

Another important concept for future research works which get low attention in current load balancing mechanisms are task migration and failure management features; hence they would have added to existing algorithm for ameliorating their efficiency.

4 Conclusions

Cloud computing is higher service able now a days. Thus, the load balancing turns into an enormous task that must require solving. There are many distinct mechanism suggested by the scientists and researchers to solve the threats of the load balancing and none of them any single algorithm has been addressed all the issues of load balancing. Each algorithm considers only limited issues i.e. some algorithms considers resource utilization and some high throughput. Some perform well with the static environment and some with dynamic. So, after studied a number of state of the art on load balancing algorithms it confirms that an algorithm which considers all the issues impossible to develop.

This paper widely analyzes the number of load balancing algorithms in cloud computing based on environment and spatial distribution of nodes. We also summarize the types of algorithms with their knowledge base information along with addressed issues and drawbacks. Also, we have contrived the relative scrutiny of dissimilar algorithms of load balancing with the positive factors. Meanwhile all the algorithms which we discussed are not completely sufficient; thus, need to develop a new algorithm with also consider the factors such as fault tolerance and scalability. In the further study, we can concludes the efficiency of the load balancing algorithm which affected by many parameters. Therefore, before developing any new load balancing algorithm we have to conclude many new parameters for the better performance.

Acknowledgments

The material of this paper is based on load balancing algorithm in cloud computing and partially financially supported by TEQIP-II, Madan Mohan Malviya University of Technology, Gorakhpur, India.

References

[1] Voorsluys W, Broberg J, Buyya R 2011 Introduction to Cloud Computing *Cloud Comput Princ Paradig* [Internet] (jan):1–41

[2] Mohapatra S, Smruti Rekha K, Mohanty S 2013 A Comparison of Four Popular Heuristics for Load Balancing of Virtual Machines in Cloud Computing *International Journal of Computer Applications* p. 975–8887

[3] Buyya R, Buyya R, Yeo CS, Yeo CS, Venugopal S, Venugopal S, et al. 2009 Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility *Futur Gener Comput Syst* [Internet] Elsevier B.V. **25**(6) 17

[4] Sahu Y, Pateriya R K, Gupta R K 2013 Cloud server optimization with load balancing and green computing techniques using dynamic compare and balance algorithm *In Proceedings: 5th International Conference on Computational Intelligence and Communication Networks, CICN 2013* p. 527–31

[5] Wang S C, Yan K Q, Liao W P, Wang S S 2010 Towards a load balancing in a three-level cloud computing network *In Proceedings: 2010 3rd IEEE International Conference on Computer Science and Information Technology, ICCSIT 2010* p. 108–13

[6] Etminani K, Naghibzadeh M 2007 *A Min-Min Max-Min selective algorihtm for grid task scheduling* [Internet]. 2007 3rd IEEEIFIP International Conference in Central Asia on Internet p. 1–7

[7] Armbrust M, Fox A, Griffith R, Joseph A, R H 2009 *Above the clouds: A Berkeley view of cloud computing* Univ California, Berkeley, Tech Rep UCB [Internet] 07–013

[8] Lombardi F, Di Pietro R 2011 Secure virtualization for cloud computing *J Netw Comput Appl* [Internet] Elsevier **34**(4) 1113–22 Available: http://dx.doi.org/10.1016/j.jnca.2010.06.008

[9] Randles M, Lamb D, Taleb-Bendiab A 2010 A comparative study into

distributed load balancing algorithms for cloud computing *24th IEEE Int Conf Adv Inf Netw Appl Work WAINA* [Internet] 551–6
[10] Zhang Q, Cheng L, Boutaba R 2010 Cloud computing: State-of-the-art and research challenges *J Internet Serv Appl.* **1**(1) 7–18
[11] Casola V, Cuomo A, Rak M, Villano U 2013 The CloudGrid approach: Security analysis and performance evaluation *Futur Gener Comput Syst.* **29**(1) 387–401
[12] Ryoo J, Rizvi S, Aiken W, Kissell J, State P 2014 *Cloud Security Auditing* (December)
[13] Carvalho M, Cirne W, Brasileiro F, Wilkes J 2014 Long-term SLOs for reclaimed cloud computing resources P*roc ACM Symp Cloud Comput - SOCC '14* [Internet] 1–13
[14] Haryani N, Jagli D, Sangita O, Dhanamma J, Jagli1 M D, Solanki R, et al. 2014 Dynamic Method for Load Balancing in Cloud Computing *Int Conf Circuits, Syst Commun Inf Technol Appl* [Internet] **5**(4) 336–40
[15] Milani A S, Navimipour N J 2016 LER URGENTE - Load balancing mechanisms and techniques in the cloud environments: Systematic literature review and future trends *J Netw Comput Appl* [Internet] Elsevier
[16] Zhao J, Yang K, Wei X, Ding Y, Hu L, Xu G 2016 A Heuristic Clustering-Based Task Deployment Approach for Load Balancing Using Bayes Theorem in Cloud Environment *IEEE Trans Parallel Distrib Syst* [Internet] **27**(2) 305–16
[17] Zhang J, Huang H, Wang X 2016 Resource provision algorithms in cloud computing: A survey *J Netw Comput Appl* [Internet] Elsevier **64** 23–42
[18] Gopinath P P G, Vasudevan S K 2015 An In-depth analysis and study of load balancing techniques in the cloud computing environment. *Procedia Comput Sci* [Internet] Elsevier Masson SAS **50** 427–32
[19] Kalra M, Singh S 2015 A review of metaheuristic scheduling techniques in cloud computing *Egypt Informatics J* [Internet] Ministry of Higher Education and Scientific Research **16**(3) 275–95
[20] Kansal N J, Chana I 2012 Cloud Load Balancing Techniques : A Step Towards Green Computing *IJCSI Int J Comput Sci Issues* [Internet] **9**(1) 238–46
[21] Randles M, Lamb D, Taleb-Bendiab A 2010 A comparative study into distributed load balancing algorithms for cloud computing *In: 24th IEEE International Conference on Advanced Information Networking and Applications Workshops, WAINA 2010* p. 551–6
[22] Aditya A, Chatterjee U, Gupta S 2015 A comparative study of different static and dynamic load balancing algorithm in cloud computing with special emphasis on time factor *International Journal of Current Engineering and Technology* p. 1898–907
[23] Al Nuaimi K, Mohamed N, Al Nuaimi M, Al-Jaroodi J 2012 A survey of load balancing in cloud computing: challenges and algorithm *Proc - IEEE 2nd Symp Netw Cloud Comput Appl NCCA 2012* 137–42
[24] Mayanka Katyal A 2013 A Comparative Study of Load Balancing Algorithms in Cloud Computing Environment *Int J Distrib Cloud Comput* [Internet] **1**(2) 14
[25] Al Nuaimi K, Mohamed N, Al Nuaimi M, Al-Jaroodi J 2012 A survey of load balancing in cloud computing: challenges and algorithms *In Proceedings - IEEE 2nd Symposium on Network Cloud Computing and Applications, NCCA 2012* p. 137–42
[26] Patel N, Chauhan S 2015 A survey on load balancing and scheduling in cloud computing *Int J Innov Res Sci Technol.* **1**(7) 185–9
[27] Sharma A 2014 Response time based load balancing in cloud computing p. 1287–93
[28] Fahim Y, Ben Lahmar E, Labriji E H, Eddaoui A, Ouahabi S 2015 The load balancing improvement of a data center by a hybrid algorithm in cloud computing *Colloquium in Information Science and Technology, CIST* p. 141–4
[29] Voorsluys W, Broberg J, Venugopal S, Buyya R 2009 *Cost of virtual machine live migration in clouds: A performance evaluation* Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) **5931** LNCS:254–65
[30] Nguyen V H, Khaddaj S, Hoppe A, Oppong E 2011 A QoS based load balancing framework for large scale elastic distributed systems. *Proceedings - 10th International Symposium on Distributed Computing and Applications to Business, Engineering and Science, DCABES 2011* p. 146–50
[31] Sun H, Zhao T, Tang Y, Liu X 2014 A QoS-aware load balancing policy in multi-tenancy environment *Proceedings - IEEE 8th International Symposium on Service Oriented System Engineering, SOSE* 2014 p. 140–7
[32] Shao X, Jibiki M, Teranishi Y, Nishinaga N 2015 Effective Load Balancing Mechanism for Heterogeneous Range Queriable Cloud Storage *IEEE 7th Int Conf Cloud Comput Technol Sci. 2015* 405–12
[33] Grid C 2014 Performance evaluation of load sharing policies on. *Science*
[34] Dave S, Maheta P 2014 Utilizing round robin concept for load balancing algorithm at virtual machine level in cloud environment *Int J Comput Appl* [Internet] **94**(4) 23–9
[35] Desai T, Prajapati J 2013 A survey of various load balancing techniques and challenges in cloud computing *Int J Sci Technol Res* [Internet] **2**(11) 158–61
[36] Dasgupta K, Mandal B, Dutta P, Mandal J K, Dam S 2013 A Genetic Algorithm (GA) based load balancing strategy for cloud computing *Int Conf Comput Intell Model Tech Appl* [Internet] Elsevier B.V. **10** 340–7
[37] Zhao C, Zhang S, Liu Q, Xie J, Hu J 2009 Independent tasks scheduling based on genetic algorithm in cloud computing *5th Int Conf Wirel Commun Netw Mob Comput* [Internet] 1–4
[38] Zha J, Wang C-D, Chen Q-L, Lu X-Y, Lai J-H 2015 Server consolidation based on hybrid genetic algorithm *Ninth Int Conf Front Comput Sci Technol* [Internet] 370–5
[39] Wu T-Y, Lee W-T, Lin Y-S, Lin Y-S, Chan H-L, Huang J-S 2012 Dynamic load balancing mechanism based on cloud storage *Comput Commun Appl Conf (ComComAp)* 102–6
[40] Joshi A, Goudar R H 2014 Advanced computing, networking and informatics *Smart Innov Syst Technol* [Internet] **28**(2) 233–40
[41] Shoja H, Nahid H, Azizi R 2014 A comparative survey on load balancing algorithms in cloud computing *5th Int Conf Comput Commun Netw Technol ICCCNT 2014*
[42] Hans A, Kalra S 2015 A comprehensive study of various load balancing techniques used in cloud based biomedical services **8**(2) 127–32
[43] Kokilavani T, George Amalarethinam D I 2011 Load balanced minmin algorithm for static metatask scheduling in grid computing. *Int J Comput Appl.* **20**(2) 43–9
[44] Chen H, Wang F 2013 User-priority guided min-min scheduling algorithm for load balancing in cloud computing *Parallel Comput Technol (PARCOMPTECH) Natl Conf. 2013* 1–8
[45] Radojevic B, Zagar M 2011 Analysis of issues with load balancing algorithms in hosted (cloud) environments *Proc 34th Int Conv MIPRO* 416–20
[46] Ren X, Lin R, Zou H 2011 A dynamic load balancing strategy for cloud computing platform based on exponential smoothing forecast *IEEE Int Conf Cloud Comput Intell Syst.* 220–4
[47] Lee R, Jeng B 2011 *Load-Balancing Tactics in Cloud*
[48] Samanta P, Mondal R K 2016 Load balancing through arranging task with completion time **9**(5) 273–82
[49] Rahmeh O A, Johnson P, Taleb-Bendiab A 2008 A dynamic biased random sampling scheme for scalable and reliable grid networks. *INFOCOMP J Comput Sci* [Internet] **7**(4) 1–10
[50] Wang S C, Yan K Q, Wang S S, Chen C W 2011 A three-phases scheduling in a hierarchical cloud computing network *Proc - 3rd Int Conf Commun Mob Comput C 2011* 114–7
[51] Dhinesh Babu L D, Venkata Krishna P 2013 Honey bee behavior inspired load balancing of tasks in cloud computing environments *Appl Soft Comput J.* **13**(5) 2292–303
[52] Mohamed N, Al-Jaroodi J, Eid A 2013 A dual-direction technique for fast file downloads with dynamic load balancing in the cloud *J Netw Comput Appl* [Internet]. Elsevier **36**(4) 1116–30 Available: http://dx.doi.org/10.1016/j.jnca.2013.01.006
[53] Systems D 2015 Load balancing in MapReduce on homogeneous and heterogeneous clusters: an in-depth review *Mohammad Javad Kargar and Meysam Vakili* **15** 149–68
[54] Bhatia J, Patel T, Trivedi H, Majmudar V 2013 HTV dynamic load balancing algorithm for virtual machine instances in cloud *Proc - Int Symp Cloud Serv Comput ISCOS 2012* 15–20
[55] Mondal B, Dasgupta K, Dutta P 2012 Load balancing in cloud computing using stochastic hill climbing - a soft computing approach *Procedia Technol* [Internet] **4** 783–9

[56] Dave S, Maheta P 2014 Utilizing round robin concept for load balancing algorithm at virtual machine level in cloud environment *Int J Comput Appl*. **94**(4) 23–9
[57] Yu Q, Chen L, Li B 2015 Ant colony optimization applied to web service compositions in cloud computing *Comput Electr Eng* [Internet] Elsevier Ltd **41** 18–27
[58] Gao R, Wu J 2015 Dynamic load balancing strategy for cloud computing with ant colony optimization *Futur Internet* [Internet] **7**(4) 465–83
[59] Nishant K, Sharma P, Krishna V, Gupta C, Singh K P, Rastogi R 2012 Load balancing of nodes in cloud using ant colony optimization *UKSim 14th Int Conf Comput Model Simul* [Internet] 3–8
[60] Zhang Z, Zhang X 2010 A load balancing mechanism based on ant colony and complex network theory in open cloud computing federation *Ind Mechatronics Autom (ICIMA) 2nd Int Conf. 2010* 2 240–3

AUTHORS

Aanjey Mani Tripathi, Gorakhpur, India

University studies: Madan Mohan Malviya University of technology, Gorakhpur, U.P., India
Scientific interest: Cloud Computing, MANET, Sensor Network
Publications: 8
Experience: 4 years
He received the B.Tech degree in Information Technology from Dr. R. M. L. Awadh University, U.P. India in 2010, and the M.Tech degree in Computer Science &Engineering from Madan Mohan Malviya Engg. College, U.P., India in 2013. Now he is doing Ph.D. From Madan Mohan Malviya University of Technology, U.P., India from 2014 in the area of cloud computing.

Sarvpal Singh

Current position, grades: Dr. Sarvpal Singh is an associate professor in the Department of Computer Science & Engg. Madan Mohan Malviya University of Technology, U.P., India.
Scientific interest: wired/wireless networks, mobile computing, cloud computing, and wireless systems
Publications: 18
Experience: 18

AUTHORS' INDEX

INFORMATION AND COMPUTER TECHNOLOGIES

Displaying formulas as embedded calculations in scientific literature, textbooks and educational web apps

Valery Ochkov, Evgeny Nikulchev, Sasha Gurke

Computer Modelling & New Technologies 2017 ***21****(1) 7-12*

Mathematical expressions can be added or edited in electronic media as embedded calculations. Formulas in an electronic document are not just expressions written in a formal language, they are also an object that could be studied, helping to comprehend a text and avoid many typos and mistakes typical for mathematical formulas found in print. This article describes some features of formulas embedded in online scientific literature and educational web apps. The article is based on the experience of teaching of engineering subjects and math assisted by the visualization of calculations. The authors examine key features and practical application of calculations and formulas embedded into the scientific and educational texts.

Keywords: embedded computing, electronic publications, computational methods, Mathcad, online publications, virtual laboratory, Knovel Interactive Equations

Development of methods for determining the tracking software systems accuracy, application of a wireless communication device and self-contained power supply in heliostat units with centralized monitoring and control system

F A Satybaldiyeva, D Beyer, A S Sarybaev

Computer Modelling & New Technologies 2017 ***21****(1) 13-19*

This article deals with determination of accuracy in tracking software systems, and advantages of the heliostat automatic control system that utilizes measuring informational and control systems comprising wireless measuring instruments and information-processing equipment. The heliostat automatic control system covered in the article is one of the types of measuring informational and control systems. As distinct from other measuring informational systems, the described heliostat control system operates only when tracking parameters deviate towards the maximum permitted values.

Use of wireless communication between detectors, transducers and industrial logical controllers in modern optical SPS heliostat control systems is more advantageous than laying hundreds meters of cable.

To provide power supply, it is proposed to equip each heliostat with a self-contained power supply, since heliostat operates when concentrated solar radiation in the receiver is sufficient for steam generation, while the rest of the time it is in the standby mode. That is why use of a solar battery-powered self-contained power supply is more advantageous than use of centralized power supply from the industrial network.

Keywords: mathematical model, automatic control systems, heliostat, wireless communication, power supply

Combining web engineering methods to cover lifecycle

K Wakil, D N A Jawawi

Computer Modelling & New Technologies 2017 ***21****(1) 20-27*

Web applications have rapidly evolved in the last decade, whilst web engineering methods have been lacking in the process development Web applications. One of the issues in web engineering methods is that no single web engineering method provides adequate coverage for the whole life cycle, because the web engineering methods are divided into three phases, which are; requirements, analysis/design, and implementation. Therefore, each method designed to special concern. It is obvious that we need to design a new method to cover the whole lifecycle to solve this issue. In this paper, we propose a framework for the new web engineering method through a combination of three methods comprising: Navigational Development Techniques (NDT) method for requirements phase; UML-Based Web Engineering (UWE) for analysis/design phase; and Interaction Flow Modeling Language (IFML) for the implementation phase. NDT and UWE are the most representative methods to develop web applications; while IFML is the newest method that focused on design and implementation. Our framework for the new method can support a whole lifecycle. Moreover, this method is more usable from developers.

Keywords: Web engineering, methods, combination, lifecycle, IFML

Model reservation resource of computer complexes system

D Kornev, E Nikulchev

Computer Modelling & New Technologies 2017 ***21****(1) 28-30*

The article presents a method for calculating redundancy systems for platform management and organization of the movement of high-speed trains. The model is developed on the mathematical formalism of Petri nets. Calculated estimated time system failover. Proved the feasibility of using a majority redundancy.

Keywords: information system, Petri net, information complex, reservation of computer complex

Application of artificial neural networks for handwritten biometric images recognition

A Malygin, N Seilova, K Boskebeev, Zh Alimseitova

Computer Modelling & New Technologies 2017 ***21****(1) 31-38*

The development of information technology leads to new requirements for the development of security systems, identity authentication and other protection mechanisms. The article is devoted to the use of artificial neural networks for handwritten biometric images recognition that are used in high-authentication systems. There is given a general structure of the biometric-neural

network authentication system, the structural scheme of information processing in biometric-neural network authentication systems, the structural scheme for learning the neural network converter of the biometric parameters vectors in the key code (password). There is formed and trained a network of neurons, are formed neural network containers on the basis of structures. The choice of the length of the biocode of neural network converters is substantiated. After graduation, testing is conducted and the probabilities of errors of the first and second kind are determined. There is given an example of a software implementation, where are given a learning mode, checking the results of training, testing the results of training.

Keywords: artificial neural networks, authentication, biometric image, first-kind errors, second-kind errors

MATHEMATICAL AND COMPUTER MODELLING

Synthesis and simulation of digital pseudo-random impulse sequence generator based on PLIC FPGA Xilinx using CAD Vivado 2016.2 and development of acoustic noise generator scheme for the protection of information

A Zaurbek, N A Seilova, D Z Dzhuruntaev

Computer Modelling & New Technologies 2017 ***21****(1) 39-46*

In this work with the help of CAD Vivado 2016.2 system and Verilog hardware description language there were synthesized, simulated and built temporary digital pseudo-random impulse sequence generator diagrams based on CAD of FPGA families of the Xilinx company and eight-rate shift LFSR register, which can be used in cryptography to create a stream encryption algorithms. On the basis of a digital pseudo-random impulse sequence generator and active low-pass filter of the second order of Sallen - Key there was constructed an electric diagram of the acoustic noise generator that provides protection against wiretapping by using embedded devices, telephone conversations, laser wiretapping system and unauthorized dictaphone recording of confidential voice information by creating a masking vibration noise.

Keywords: computer-aided design, hardware description languages, programmable logic integrated circuits, synthesizing, simulation, circuit simulation, pseudo-random impulse sequence

A modified non-monotone method with 3-1 piecewise NCP function for nonlinear complementary problem

Ke Su, Dan Yang

Computer Modelling & New Technologies 2017 ***21****(1) 47-51*

In this paper, we proposed a modified nonmonotone method for nonlinear complementarity problem, different from the existed methods, we transform the original problem to a semi-smooth equation by using a piecewise NCP function, and combined with the nonmonotone line search. Only one nonlinear equations need to be solved per iteration so that the computational costs are reduced. Under some suitable assumptions, we give the convergence properties of the proposed method and the numerical results to show that our method is efficient.

Keywords: nonlinear complementary problem, piecewise NCP function, nonmonotone , global convergence

Improving the reliability of the managed of electric drives with a synchronous generator

Elena Loginova

Computer Modelling & New Technologies 2017 ***21****(1) 52-55*

The article examines the dynamic properties of the excitation control system of synchronous generator of capacity 2200kW, which comprises a converter - bridge rectifier. It was investigated the influence of the type of bridge rectifier - of thyristor symmetric bridge and of asymmetric bridge - on transient processes of synchronous generator in emergency operation. This problem was solved by numerical methods.

A mathematical model of synchronous generator excitation control system. In order to make the model of the generator excitation system has been used plugin SIMULINK mathematical software MATLAB.

During simulation were calculated currents short circuit of the synchronous generator in the case of using in system excitation a thyristor symmetrical bridge and asymmetrical bridge. Calculations have shown that the use of symmetric thyristor bridge reduces times the short circuit fivefold. This allows us to recommend used a symmetrical bridge in the electric drive system.

Keywords: excitation system of the synchronous generator, the short-circuit mode

Development of the SVM classifier by means of the hybrid versions of the particle swarm optimization algorithm based on the grid search

Liliya Demidova, Irina Klyueva

Computer Modelling & New Technologies 2017 ***21****(1) 56-63*

In this article the approaches to the problem solving of searching of the parameters of the SVM classifier based on the hybridization of the particle swarm optimization algorithm (PSO algorithm) and the grid search algorithms with the aim of providing of high quality classification decisions have been considered. The paper presents two hybrid versions of the basic PSO algorithm, involving the use of the classical Grid Search (GS) algorithm and Design of Experiment (DOE) algorithm correspondingly. It is proposed to use the canonical PSO algorithm as the basic algorithm. The results of experimental studies confirm the application efficiency of the hybrid versions of the basic PSO algorithm with the aim of reducing of the time expenditures for searching the optimum parameters of the SVM classifier while maintaining of high quality of its classification decisions.

Keywords: classification, particle swarm optimization algorithm, grid search algorithm, SVM classifier, radial basis kernel function

A literature review on algorithms for the load balancing in cloud computing environments and their future trends

Aanjey Mani Tripathi, Sarvpal Singh

*Computer Modelling & New Technologies 2017 **21**(1) 64-73*

Cloud computing is a computing utility which provides basic service for computing. It is a high performance distributed computing which has the huge collection of virtual resources that can be easily accessed anytime using the internet similar to pay as you go, model. A cloud defines set of virtual computers connected to each other in a form of parallel and distributed system. It ensures the dynamic provision of resources based on service level agreement (SLA) to ameliorate one or more objectives. To attain this goal several research challenges have been faced in the area of cloud computing, And the Load balancing is one of them, which aim at equalizing the workload among all the obtainable nodes by minimizing execution time, minimizing communication delays, maximising resource utilization and maximising throughput. This paper disburses a literature review of existing load balancing algorithms suggested so for and categorized under different metrics enveloping the advantages and disadvantages of each. An overview of the important research challenges of these algorithms is presented at the end with some possible ideas for improvement.

Keywords: Cloud computing, Load balancing algorithm, virtual machine

