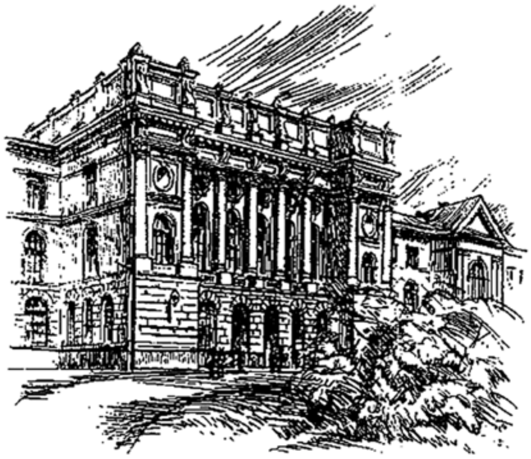
THE MINISTRY OF EDUCATION AND SCIENCE OF THE RUSSIAN FEDERATION



ST. PETERSBURG STATE POLYTECHNICAL UNIVERSITY JOURNAL

Computer Science. Telecommunications and Control Systems

1(188) 2014

Polytechnical University Publishing House
Saint Petersburg
2014

ST. PETERSBURG STATE POLYTECHNICAL UNIVERSITY JOURNAL
COMPUTER SCIENCE. TELECOMMUNICATIONS AND CONTROL SYSTEMS

The journal is published under scientific and methodical guidance of the Russian Academy of Sciences since 1995. The journal is published since 2008 as part of the periodical edition «Nauchno-tekhnicheskie vedomosti SPbGPU» (ISSN 1994-2354).

The journal is included in the List of Leading Peer-Reviewed Scientific Journals and other editions to publish major findings of PhD theses for the research degrees of Doctor of Sciences and Candidate of Sciences.

The publications are presented in the VINITI RAS Abstract Journal and Ulrich's Periodical Directory International Database.

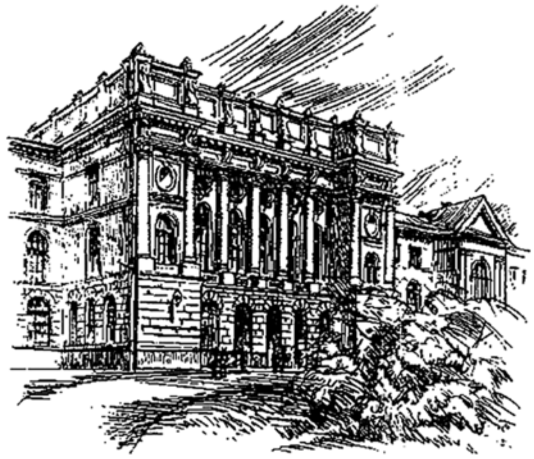
The journal is registered with the Federal Service for Supervision in the Sphere of Telecom, Information Technologies and Mass Communications (ROSKOMNADZOR). Certificate ПИ № ФС77-51457 issued Oct. 19, 2012.

Subscription index **47517** in the «Journals and Magazines» catalogue, Rospechat agency.

The journal is on the Russian Science Citation Index (RSCI) database
© Scientific Electronic Library (http://elibrary.ru/).

The address: 195251 Polytekhnicheskaya Str. 29, St. Petersburg, Russia.

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ



НАУЧНО-ТЕХНИЧЕСКИЕ ВЕДОМОСТИ

САНКТ-ПЕТЕРБУРГСКОГО ГОСУДАРСТВЕННОГО ПОЛИТЕХНИЧЕСКОГО УНИВЕРСИТЕТА

Информатика. Телекоммуникации. Управление

1(188) 2014

Издательство Политехнического университета
Санкт-Петербург
2014

НАУЧНО-ТЕХНИЧЕСКИЕ ВЕДОМОСТИ САНКТ-ПЕТЕРБУРГСКОГО ГОСУДАРСТВЕННОГО ПОЛИТЕХНИЧЕСКОГО УНИВЕРСИТЕТА
ИНФОРМАТИКА. ТЕЛЕКОММУНИКАЦИИ. УПРАВЛЕНИЕ

Журнал с 1995 года издается под научно-методическим руководством Российской академии наук. С 2008 года выпускается в составе сериального периодического издания «Научно-технические ведомости СПбГПУ» ISSN 1994-2354.

Журнал с 2002 года входит в Перечень ведущих рецензируемых научных журналов и изданий, в которых должны быть опубликованы основные результаты диссертаций на соискание ученой степени доктора и кандидата наук.

Сведения о публикациях представлены в Реферативном журнале ВИНИТИ РАН, в международной справочной системе «Ulrich`s Periodical Directory».

Журнал зарегистрирован Федеральной службой по надзору в сфере информационных технологий и массовых коммуникаций (Роскомнадзор). Свидетельство о регистрации ПИ № ФС77-51457 от 19.10.2012 г.

Подписной индекс **47517** в каталоге «Газеты. Журналы» Агентства «Роспечать».

Журнал включен в базу данных «Российский индекс научного цитирования» (РИНЦ), размещенную на платформе Научной электронной библиотеки на сайте http://www.elibrary.ru

При перепечатке материалов ссылка на журнал обязательна.

Точка зрения редакции может не совпадать с мнением авторов статей.

Адрес редакции и издательства: Россия, 195251, Санкт-Петербург, ул. Политехническая, д. 29.
Тел. редакции (812) 552-62-16.

Message from the incoming Editor-in-Chief

It is my great pleasure to introduce this February issue of the «St. Petersburg State Polytechnical University Journal. Computer Science. Telecommunication and Control Systems», which has been prepared and published totally in English. More than 5 years history of the Journal demonstrates its ability, as well as a potential of the Editorial Board and Editorial Council, to solve all spectrum of the questions starting from the peer review process and scientific content selection, and finalizing by the technical details and a paper work. Thus, the publication of the first time English version is a quite natural result of the Journal development. Taking into account a very ambitious program «5-100-2020», in which the University is an active player, it is especially important because of the globalization and internationalization of the science and education.

Traditionally the issue consists of 5 parts devoted to the problems of ICT, electronics, hardware and software theory and design, as well as computer simulation methods both in technology and human disciplines. The special section consists of materials from Zabbix International Conference 2013. The feature of Zabbix Conference is that it does not have its own printed proceedings, but only short abstracts available on the Conference web site. Publications of the Conference papers are a kind of an extension of the specialist audience having interest on the topic.

On behalf of the Journal staff I would like to thank all authors who have responded to our proposal to publish English issue of the «St. Petersburg State Polytechnical University Journal. Computer Science. Telecommunication and Control Systems». I would like to express my gratitude to the reviewers for their valuable comments and suggestions that improved the quality of the manuscripts. Special thanks are to our Colleagues from international team of the Editorial Board who has joined us in a final stage of the issue preparation.

Prof. Dr. *Alexander Korotkov*

THE 6TH INTERNATIONAL CONGRESS ON ULTRA MODERN TELECOMMUNICATIONS AND CONTROL SYSTEMS ICUMT 2014

http://www.icumt.info/2014

October 6 – 8, 2014
St. Petersburg, Russia

HIGHLIGHTS

- 6th ICUMT 2014 and Workshops proceedings planned to be published in IEEE Xplore (application pending) and indexed by relevant databases (previous ICUMT editions have been indexed in IEEE Xplore and Scopus)
- Submission deadline June 17, 2014. Workshops have own deadlines.
- Technically co-sponsored by IEEE R8

Important dates

Paper submission (via EDAS conference management system): June 17, 2014
Notification of acceptance: August 18, 2014
Camera ready version: September 03, 2014

CALL FOR PAPERS

About the Conference

ICUMT is an annual international congress providing an open forum for researchers, engineers, network planners and service providers targeted on newly emerging algorithms, systems, standards, services, and applications, bringing together leading international players in telecommunications, control systems, automation and robotics. The event is positioned as a major international annual congress for the presentation of original and fundamental research and engineering results.

Tracks

ICUMT 2014 will consist of two tracks, Telecommunications Track (ICUMT-T), and Control Systems, Automation and Robotics Track (ICUMT-CS). ICUMT-T is an annual international congress providing an open forum for researchers, engineers, network planners and service providers in telecommunications targeted on newly emerging algorithms, systems, standards, services, and applications. ICUMT-CS is an annual international congress providing an open forum for researchers, engineers and practitioners interested in control, automation and robotics targeted on newly emerging algorithms, systems, standards and applications.

Topics of interest include but are not limited to:
Telecommunications track (ICUMT-T) and
Control Systems, Automation and Robotics track (ICUMT-CS).
For details, please, see the Conference site

http://www.icumt.info/2014

Contents

Содержание

Проблемы передачи и обработки информации

Радиотехника, антенны, СВЧ-устройства

Математическое моделирование: методы, алгоритмы, технологии

Анализ загруженности и производительности облачных вычислений

UDC 004.383.8,004.85=111

E.N. Benderskaya

SOFT COMPUTING BASED ON NONLINEAR DYNAMIC SYSTEMS: POSSIBLE FOUNDATION OF NEW DATA MINING SYSTEMS

The article describes how the structure of AI systems is formed through an incoming image in nonlinear dynamic systems. The main steps of a new approach to solve the problem of image recognition, as well as its strengths and limitation are presented. An example of the use of a chaotic dynamic system to solve a clustering problem is shown.

SOFT COMPUTING; NONLINEAR DYNAMIC SYSTEM; AI; CHAOTIC DYNAMICS; IMAGE RECOGNITION; PATTERN RECOGNITION; TURING MACHINE.

Е.Н. Бендерская

МЯГКИЕ ВЫЧИСЛЕНИЯ НА БАЗЕ НЕЛИНЕЙНЫХ ДИНАМИЧЕСКИХ СИСТЕМ: ВОЗМОЖНАЯ ОСНОВА ДЛЯ НОВЫХ СИСТЕМ ИЗВЛЕЧЕНИЯ ЗНАНИЙ

Рассмотрены вопросы формирования структуры интеллектуальной системы с помощью входного образа на основе нелинейных динамических систем. Представлены основные составляющие нового подхода к решению задач распознавания образов, его достоинства и ограничения. Приведен пример использования хаотической динамики системы для решения задач кластеризации.

МЯГКИЕ ВЫЧИСЛЕНИЯ; НЕЛИНЕЙНЫЕ ДИНАМИЧЕСКИЕ СИСТЕМЫ; ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ; ХАОТИЧЕСКАЯ ДИНАМИКА; РАСПОЗНАВАНИЕ ИЗОБРАЖЕНИЙ; РАСПОЗНАВАНИЕ ОБРАЗОВ; МАШИНА ТЬЮРИНГА.

As today there is a large number of problem solving methods in various fields. Many methods are heuristic-based and it is not always clear which type of problem is best solved by a method, and what parameters will yield the best result for a given problem.

When solving any problem, the question arises of the best method to use in order to find a solution which satisfies the initial requirements. The classic approach is decomposition (usually functional decomposition) of the initial problem into sub problems and finding the best methods for each sub problem. Often, the researchers use only well-known methods or previously tested methods. In that case, there is a random component in the choice of methods from a large number of many possible ones.

Because of this the second way to solve this task is the development of AI systems which pick the methods and the parameters which are best for solving a certain problem. Searching and application take place instead of developing new AI system in those cases when required AI systems exist for a subject area. These systems fall into three main categories:

- multi-level automatic system with an expert at the highest level (advisory systems);
- a group of methods which solve problems by a majority rule;
- universal methods (the results will be less accurate than specialized methods).

One of the main problem is when developing AI systems is ensuring that the system is sophisticated enough for what needs to be solved.

Such systems should be able to recognize both simple and sophisticated images, coming closer to the abilities of the biological counterparts.

One of the possible ways of solving this dilemma is using multi-level systems for making decisions which would include a preliminary assessment of the problem. Usually the higher level is automated, but not completely automatic and requires an expert. Advisory systems allow increasing in the number of possible methods to be considered in addition to those which the developer already knows when trying to find a solution. Also such systems allow us to take into account of the accumulated knowledge about the features of each method, and based on task input data and the requirements for the solution. It can give recommendations about the best method and optimum settings. However, the final decision depends on the expert developer.

Due to the complexity of formal representation of the selecting process the optimal method for solving a problem, we should offer somewhat redundant but fairly effective approach. It can help to create a system which would include most suitable methods, their use for solving the problem, and subsequent selection of the best solution by some quality criteria or majority rule [14]. For example, the systems of decision rule committees in the theory of pattern recognition and the formal algebra of events used on the set of these rules, designed by Yi. Zhuravlev [14].

The other design approach which scientists from different fields carried out is the creation of a fairly universal method of solving the problem. Such a method would be suitable for a large number of conditions for solving the problem. Also the method would be insensitive to the deviation of the actual data from the data embedded a priori in the method. Such an approach generates methods that are universal and they give the results close to the optimal solution on average. In this case, the quality of the solution may be much lower than potentially achievable. This is applied to the main indicators of quality, such as the probability of the correct solution, the accuracy of the solution, as well as to the secondary indicators – complexity, cost of memory and time consumed. The difficulty of finding a suitable method, as well as developing a general method lies in the fact that the complexity of the method (and therefore the structure used) has to be adequate for solving the problem given. It would seem that the more complex the method, the wider the range of problems it can solve. However, simple problems, when solved by a complex method, often produce unsatisfactory results. Figuratively speaking, the additional degrees of freedom in the method, being unaffected by the input data, generate errors. This can be most clearly demonstrated on a neural network with an excessive number of elements for solving a simple problem. Instead of learning, with subsequent generalization the network does not produce patterns. It simply stores the input examples, and completely repeats the features of the training examples, which may be related not to the features of the input space, but to the peculiarities of measurement and acquisition of data. As a result, incorrect results are obtained from the test data (or even worse, when the network is already in use).

Consider ways to ensure adequate structural complexity:

1) principle of adjustment (development) of the structure (method) for a particular problem;

2) synergic governance principles;

3) principle of minimum description length.

In the first case the development of the neural network for a specific application is assumed, and thus the adequacy of the structure complexity and problem complexity is ensured [7]. To implement this method, developed by A. Galushkin, one must pass the priori information to the primary and secondary optimization functions, which are then used to determine the adequacy of the structure.

A. Kolesnikov has developed a whole theory of synergistic control [10, 11], with maximum use of the dynamics features of the object being controlled during development for «nonviolent» control and maximum use of the object's own dynamics to achieve a certain goal (subspace, trajectory or point) [10, 11]. By using this approach, it is assured that the control system and control object are of adequate complexity. Similar to the first principle of adjusting to the problem is the principle of minimum description length, (proposed by A. Potapov) illustrated in detail for image recognition tasks [13]. To compare methods of problem-solving, a metric is created which corresponds to the length of the description of the method which can be used to solve the problem and thus the method with a minimum value of the metric is selected.

Analysis of the main existing approaches leads to the idea of a new approach that would combine all three principles to ensure the adequacy of the system structure [7] for the complexity of the problem. An approach which will be proposed is based on the assumption that the complexity of the system can be controlled by its response to dynamic changes in the input image directly. For this purpose, one can use a dynamic self-organizing system, which is sensitive to changes in input data and interpretation of the structure of the system and, accordingly, its complexity. It must involve the concepts of the dynamic system complexity and the complexity of the attractor.

By structural complexity in this case we mean not only the complexity of connections in the system itself, but also the complexity of the generated image in the phase space (attractor), which reflects the dynamics of the system and hence the complexity of the task.

Chaotic Dynamics – New Opportunities to Solve Complex Problems

Trend analysis of the mathematical apparatus of the static and dynamic point of view also leads to the idea of using a highly sensitive-to-changes-in-the-input space nonlinear dynamic systems for the development of intelligent systems. It contains in its dynamics all the possible problem solutions, simple as well as complex. And unlike the artificial construction of a universal approach, there is a universal system that organizes itself, adjusts to the solution.

Mathematical methods of nonlinear dynamics and chaos can be regarded as the next stage in the development of mathematical methods. There is a tendency to shift from deterministic to statistical models with more complexity, to chaotic, which can be deterministic but due to nonlinearity. A large number of elements lead to complex and often unpredictable behavior.

Fig. 1 is a schematic representation of development stages of the mathematical apparatus in terms of the complexity of methods, models and objects which can be described based on them. The convention of these steps is that it does not take into account the time for the models to come into existence. Many of them were offered long ago, but due to the lack of suitable computational tools for modelling at the time, they could not be applied, but now these models are quite popular.

The development of mathematical methods and models from the point of view of a logic device (focus on static, the left column of Fig. 1) can be represented as follows. In the beginning

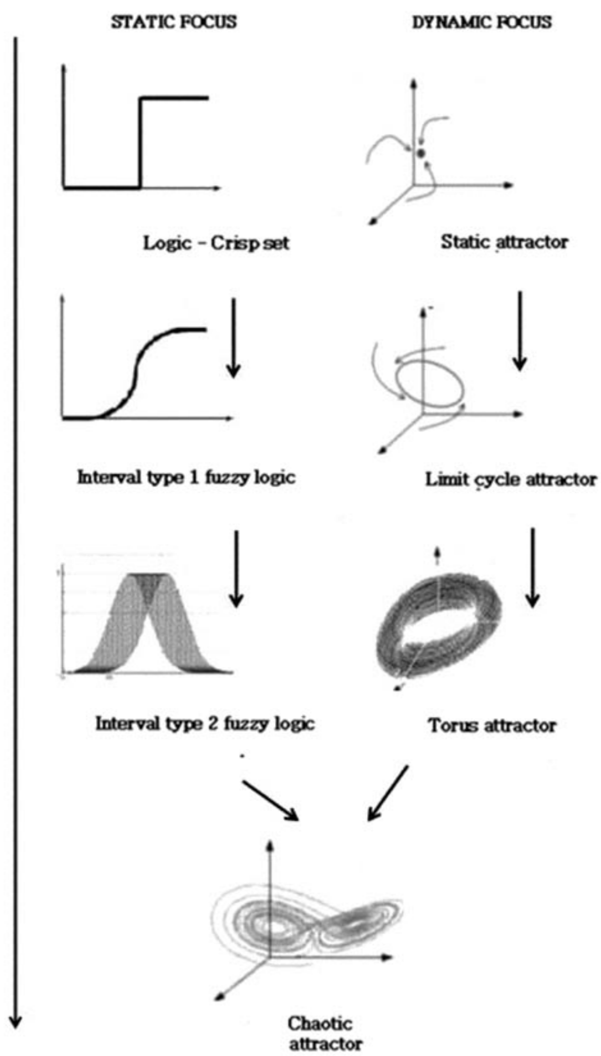


Fig. 1. Evolution of formal methods: dealing with uncertainty

there was classical logic which operated with clear numbers and precise sets. Largely this is why classical computational architectures require exact and specific input of the source data when performing calculations. It is impossible to do where some complex and hard to formalize problems exist.

A significant breakthrough in the field of information processing and overcoming linguistic uncertainty was the introduction of the concept of «fuzzy sets» and development of the theory of fuzzy logic. Now it is possible to perform operations simultaneously at a certain interval. The element on which the operations are performed is now an interval instead of a single point.

Further development of the theory of fuzzy sets and fuzzy logic is in some sense going via the extensive path: finding fuzzy sets of the second type, which are in reality «interval on an interval», increasing the dimension, etc. This, of course, enhances the capabilities of devices which deal with complexly organized and uncertain data, but, nevertheless, it is not as effective as the transition from a number of intervals.

One can observe the mathematical apparatus becoming more and more complex from the point of view of dynamic models when looking at the example of attractors attracting sets of dynamic systems as they become more complex (right column of Fig. 1). Firstly, models of systems the dynamics of which converge to the set of individual points of attraction in the phase space (point attractor), then to the set of closed trajectories (attractor type: limit cycle, torus), and finally to the set of trajectories that define a location in the phase space in the form of an infinite number of changing states (chaotic attractors).

For static models, the next level of generalization, in order to extend the ability of making calculations simultaneously on a whole set of possible solutions, is also modelling with a chaotic attractor.

When looking at the trends in neural networks, we realize the necessity of using the capacity of chaotic dynamic systems for solving problems of AI and accomplishing related tasks (e. g. coding and information transfer). The functioning of the dynamic neural network with an irregular structure makes it possible to form a solution on the boundary of order-chaos, which corresponds to a variety of different structures of the output space, extremes of which are ordered dynamics (cycle) and turbulent dynamics (lack of structure in general).

This is the next step in the development of the neural network structure, as in this case, not only the weights of the network are adjusted, but a collective solution is found by a set of nonlinear elements of the same type, each one having unstable dynamics, but as a whole, under the influence of the input data, they form a stable dynamic system.

Control of the Structural Complexity of the System

Control of chaos is often associated with the task of suppressing chaotic oscillations – the shift of the system to a stable periodic motion, or to a state of equilibrium. In a broad sense, it is the transformation of the chaotic behavior of the system into regular behavior or chaotic, but with different properties.

The challenges arising from the chaos control problem are much different from the traditional problems of automatic control [1, 2]. Instead of classic control goals, such as bringing the trajectory of the system to a set point or to a given movement, soft goals are set to chaos control: creating modes with partially specified properties, qualitative change in the phase portrait of the system, synchronization of chaotic oscillations and others. Unlike traditional control operations, in physical application of chaos theory the focus is not on finding the most effective way of achieving goals, but on researching the fundamental possibility of achieving it, on determining a class of possible movements by the controlled physical system [1, 2].

Study of the dynamics of ensembles consisting of a large number of nonlinear elements, is one of the main trends in the theory of nonlinear oscillations and waves. The main factor in the dynamics of ensembles of oscillating systems, which leads to an ordered-time behavior, is the synchronization of the ensemble elements. Numerous studies show that space-distributed random vibrating systems have many beneficial properties. In some of them self-synchronization occurs with specific parameters of the system. By self-synchronization we mean the process which contains identical elements of the system, each of which is characterized by chaotic dynamics. It can be initialized in various ways, over time, and starts to oscillate synchronously without outside influence.

In the presence of external influence on the nonlinear dynamic system, we get a response that reflects both the external conditions of the problem and the input signals which characterize the problem being solved. With this approach instead of creating a model for solving the problem, the target setting is given – a required outcome of solving the problem and it is believed that the solution is not unique. In any case in the form of presentation it forms a variety, which can be interpreted as the only solution, or as a set of basic solutions.

Instead of the usual representation of the original problem to be solved as a set of functions for subsequent use or for splitting the system into separate parts, in the synergetic approach the synthesis and study are performed on the system as a whole. Changing the state of a particular element system may not affect the state of the system as a whole, however, the joint dynamics of all the elements defines a unique state of a macroscopic system. This state of the system will be the solution of the problem [3, 4].

Namely this, the occurrence of synchronization (collective behavior), allows living systems to adapt, learn, and extract information in real time to solve computationally complex problems (due to distributed information processing). Many elements with complex dynamics produce efficient computing [8, 9].

A computing device that implements the proposed approach can be a set of asynchronous models of dynamic systems that interact with each other and combine properties such as being hybrid and asynchronous, having clusters (no rigid centralization and dynamic clustering of related models), and being stochastic [8, 9]. O. Granichin developed a computational model for such a device that is based on the following set of basic parameters [8]:

set of computational primitives (dynamic models H_i with parameters from the set Q);

memory X – total space of states of all models;

feed S – dynamic graph with a finite bit string s of whether to include the models at certain nodes;

program G – the rules given by graph S are the rules (or goals) for «switches» of the tape and model parameters when the pair (x, q) appears at one of the «active» nodes in the switching set J;

cycle – the time interval between successive switches;

breakpoint set T.

One can speak of a generalization of a Turing machine [8] which can be represented as a chain of interrelated components $<A, H, Q, q, q0, X, x, x0, S, s, s0, J, G, T>$ where A is the set of models (computational primitives); H – the evolution operator; Q – the set of states (parameter values); X – the memory; S – the generic tape (graph); J – the set of switching; G – the program (goals); T – the breakpoint set. The main stages in the use of complex modes of operation of chaotic systems to solve practical problems can be represented by the following sequence.

The initial state is given and the goal is defined – to reach a certain state. It is assumed that the goal can be achieved by navigating through a trajectory that passes near one of the attractors. Then the system is started and the input signals corresponding to the task are given to it. After a transition process the system goes into an attractor. Searching takes place for a trajectory which is accessible using a small perturbation of the system and is close enough to pass next to the desired point or sequence of points which corresponds to the desired state of the system. If such a path is not found, random input is fed into the system in order to jump to another attractor until the goal is achieved.

The Chaotic Neural Network – Example of a Structure and Dynamics Determined by Input Images

In chaotic dynamics under the influence of external perturbation structures are produced, and it may initially include the entire set of possible options. Chaotic systems allow us to go to the next level of aggregation in the concept of computing process and perform the calculations simultaneously on a whole set of possibilities. This set will be shaped by external signals, thus providing an adequate complexity. In many ways, this is similar to the principles used in quantum computing, which contains the entire set of solutions until the answer is found.

We want to consider a relatively simple and clear example of the use of external images to form the structure of a system. We present the use of various metrics based on the input data for the calculation of the connection matrix in the chaotic neural network (CNN) [5, 6]. It is capable of solving the clustering problem only on the basis of input data without any additional and prior information about task.

A feature of this oscillatory neural network is chaotic dynamics of individual neurons outputs, and mutual, independent on the initial conditions, self-clusterization. For the use of CNN it allows to solve problems with minimal prior clustering information concerning the objects to be sorted into clusters. One can draw an analogy between the formation of functional and logical structures on CNN with self-generated functional clusters of activity in the brain to solve different problems.

CNN is a one-layer recurrent network in which the elements are connected to «each other» without having any connection back to «themselves»:

$$y_i(t+1) = \frac{1}{C_i}\sum_{i \neq j}^{N} w_{ij} f(y_i(t)), \quad t = 1...T, \qquad (1)$$

$$f(y(t)) = 1 - 2y^2(t), \qquad (2)$$

$$w_{ij} = \exp(-\mid x_i - x_j \mid^2 / 2a^2), \qquad (3)$$

where $C_i = \sum_{i \neq j} w_{ij}$, $i, j = \overline{1, N}$ is the scaling constant, computed by the algorithm presented in [5, 6]; w_{ij} is the connection strength (weight vector) between neurons i and j; N is the number of neurons, which is equal to the number of points in the input image, represented in the form of $X = (x_1, x_2, ..., x_m)$; m is the dimension of the image space; T is the simulation time. As shown in [5], for nonlinear transformation $f(y(t))$ one can use any mapping that generates chaotic oscillations, however, a logistic mapping (2) is preferred.

The training of CNN consists of assigning weight vectors, which are based on the ratio of the input image (3) and uniquely determine the field, which acts on all the neural networks. As this field is not uniform, the analysis and resolution of the difference equations system (1) are much more difficult.

Study of the dynamics of ensembles of systems consisting of a large number of nonlinear elements is one of the main directions of development of nonlinear oscillations and waves theory. The main factor in the dynamics of ensembles of oscillating systems, which leads to ordered space-time behavior of the ensemble, is the synchronization of the elements [12].

Analysis of the different images dynamics for CNN (input structures, reflecting the impact of external environment on the system) with the same system parameters allows one to see the varying «music» of vibrations at each of the clusters formed in the system. In Fig. 2, you can clearly distinguish ensembles of elements, the character of the output oscillations is very different, and allows one to talk about the existence of self-generated clusters system and the availability of fragmentary synchronization [5]. With this synchronization the instantaneous outputs of neurons belonging to the same cluster do not match either in amplitude or phase and do not have a fixed phase shift between any two sequences. By cluster fragmentation synchronization we mean synchronization in the sense that each cluster is characterized by a unique «melody» of vibrations, encoded in the temporal sequence of output values of neurons. The proposed method for detection of cluster synchronization is described in detail in [5, 6] and it is based on an analysis of the relative remoteness of the instantaneous output values of each neurons pairs in a varying time interval.

The difficulty of using chaos and developing chaos logic also reflects in the fact that the term «chaos» defines several fundamentally different modes of the system. To separate the useful chaos from the rest, the expression «determinate chaos» has come about. The word «determinate» was introduced to highlight the repeatability of the experiments, and therefore we should make the calculations with it, and we need to have the possibility of its application.

The need to address increasingly complex problems, and the opportunities that are

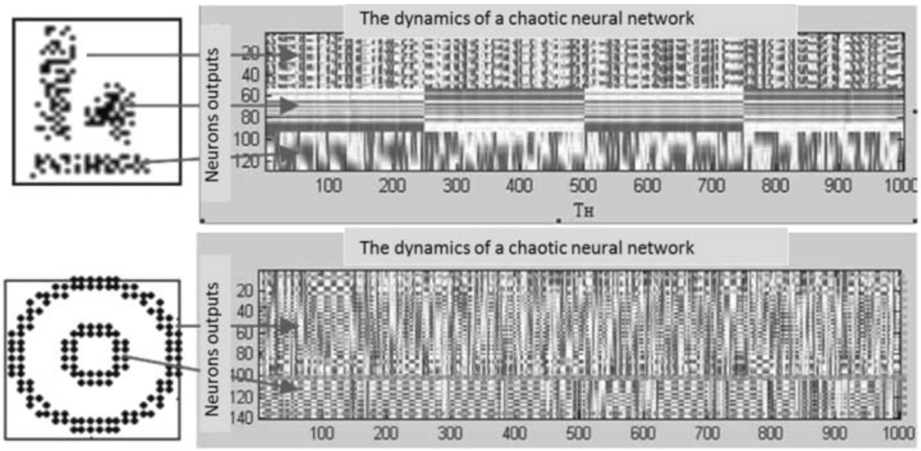


Fig. 2. Fragmentary synchronization for two different input images (one can see the «music» of oscillations of each cluster)

provided when using synergistic principles of analysis and synthesis, leads to the idea that for the complex challenges that have manifested emergent properties, the more effective approach is the holistic analysis as a whole, without division. This is not a departure from functional decomposition, but a significant addition to it, since during fragmentation of the system we often lose the uniqueness associated with system patterns.

Thus, we propose a general approach to solving different tasks – by reducing the original problem to a control problem, an optimization problem, or a problem of pattern recognition. This approach is similar to the neural network approach in the part, where problems of different types are reduced to the same type of problem and solvable by homogeneous network structures. In this approach, the complexity of the method (and the system to implement it) will be adequate to the complexity of the problem being solved just as it is in the formal synthesis theory of neural network structure through functions of primary and secondary optimization [7]. On the other hand when using a single approach it is possible to combine operations easily. For example, for information systems – it is the perception and storage, and actual processing of information. From association to storage and subsequent recognition, this is consistent with current ideas on how living systems solve problems.

REFERENCES / СПИСОК ЛИТЕРАТУРЫ

1. **Andrievskii B., Fradkov A.** Control of chaos: methods and applications. I. Methods. *Automation and Remote Control.* 2003, Vol. 64, Iss. 5, pp. 673–713.

2. **Andrievskii B.R., Fradkov A.L.** Control of chaos: method and applications. II Applications. *Automation and Remote Control.* 2004, Vol. 65, Iss. 4, pp. 505–533.

3. **Benderskaya E.N.** Nonlinear Trends in Modern Artificial Intelligence: A New Perspective. *Beyond AI: Interdisciplinary Aspects of Artificial Intelligence. Topics in Intelligent Engineering and Informatics.* Springer, 2013, Vol. 4, pp. 113–124.

4. **Benderskaya E.N., Zhukova S.V.** Multidisciplinary Trends in Modern Artificial Intelligence: Turing's Way. *AIECM – Turing 2012, Book Chapters: Artificial Intelligence, Evolutionary Computation and Metaheuristics.* Springer, 2013, pp. 320–343.

5. **Benderskaya E.N., Zhukova S.V.** Fragmentary Synchronization in Chaotic Neural Network and Data Mining. *HAIS 2009. LNCS.* Springer, Heidelberg, 2009, Vol. 5572, pp. 319–326.

6. **Benderskaya E.N., Zhukova S.V.** Clustering by Chaotic Neural Networks with Mean Field Calculated Via Delaunay Triangulation. *HAIS 2008. LNCS (LNAI).* Springer, Heidelberg, 2008, Vol. 5271, pp. 408–416.

7. **Galushkin A.I.** Neural Networks Theory. Springer, 2007, 396 p.

8. **Granichin O.N., Vasil'ev V.I.** Computational model based on evolutionary primitives. Turing machine generalization. *Internat. Journal of Nanotechnology and Molecular Computation.* 2010, Vol. 2, No. 2, pp. 30–43.

9. **Granichin O.N., Izmakova O.A.** A Randomized Stochastic Approximation Algorithm for Self-Learning. *Automation and Remote Control.* 2005, Vol. 66, Iss. 8, pp 1239–1248.

10. **Kolesnikov A., Veselov G., Monti A., Ponci F., Santi E., Dougal R.** Synergetic synthesis of dc-dc boost converter controllers: theory and experimental analysis. *Proceedings IEEE Applied Power Electronics Conference and Exposition APEC 17th.* Dalas, TX, 2002, pp. 409–415.

11. **Kolesnikov A., Veselov G., Popov A., Kolesnikov A., Kuzmenko A., Dougal R.A., Kondratiev I.** Synergetic approach to the modelling of power electronic systems. *IEEE 7th Workshop on Computers in Power Electronics.* Blacksburg, VA, USA, 2000, pp. 259–262

12. **Pikovsky A., Rosenblum M., Kurths J.** Synchronization: A Universal Concept in Nonlinear Sciences. *Cambridge Nonlinear Science Series.* Cambridge University Press, 2003, 432 p.

13. **Potapov A.S.** Principle of representational minimum description length in image analysis and pattern recognition. *Pattern Recognition and Image Analysis. (Advances in Mathematical Theory and Applications).* 2012, Vol. 22, No. 1, pp. 82–91.

14. **Zhuravlev Yu.I.** An algebraic approach to recognition or classification problems. *Pattern Recognition and Image Analysis. (Advances in Mathematical Theory and Applications).* 1988, Vol. 8, p. 59.

BENDERSKAYA, Elena N. *St. Petersburg State Polytechnical University.*
195251, Politekhnicheskaya Str. 29, St. Petersburg, Russia.
E-mail: helen.bend@gmail.com

БЕНДЕРСКАЯ Елена Николаевна – *доцент кафедры компьютерных систем и программных технологий Санкт-Петербургского государственного университета, кандидат технических наук.*
195251, Россия, Санкт-Петербург, ул. Политехническая, д. 29.
E-mail: helen.bend@gmail.com

UDC 004.4, 004.6=111

A.V. Fedorov, V.V. Potekhin

INTELLIGENT PROCESSING ALGORITHM OF ULTRASOUND IMAGES IN BIOMEDICINE

A software package for ultrasound imaging in the form of an elastogram has been developed to measure stiffness of human tissue. A technique for obtaining elastograms is based on the least squares method and identification of feature points in images. The method can be used for clinical purposes when ultrasound machines have no elastogram display.

ELASTOGRAPHY; ULTRASOUND IMAGE; IMAGE PROCESSING; MACHINE LEARNING.

А.В. Федоров, В.В. Потехин

ИНТЕЛЛЕКТУАЛЬНЫЙ АЛГОРИТМ ОБРАБОТКИ УЛЬТРАЗВУКОВЫХ ИЗОБРАЖЕНИЙ В БИОМЕДИЦИНЕ

Разработан программный комплекс для расчета жесткости органов человека на основе ультразвукового изображения для отображения эластограммы. Представлена методика получения эластограммы на основе метода наименьших квадратов и определения особых точек в изображении. Метод применим для проведения диагностики на аппаратах УЗИ, в которых отсутствует функция отображения эластограммы.

ЭЛАСТОГРАФИЯ; УЛЬТРАЗВУКОВОЕ ИЗОБРАЖЕНИЕ; ОБРАБОТКА ИЗОБРАЖЕНИЙ; МАШИННОЕ ОБУЧЕНИЕ.

Ultrasound elastography is a method that measures the deformation of tissue in response to an applied force to determine and display its stiffness. Tissue strain analysis gives a new dimension in ultrasound imaging, which includes such properties as quality and quantitative assessment of the tissue elasticity. Numerous applications in this field of diagnostic tools can provide more accurate diagnoses than one imaging modality alone [1]. In this paper, we describe an intelligent algorithm to acquire an elastography image, making use of conventional ultrasound equipment. We have developed a video streaming application that can help physicians to solve diagnostic problems. A multicore approach was employed to get near-real time execution of the algorithm.

With ultrasound elastography, diagnosticians may no longer need to use finger palpation to detect hard tissues. An elastogram will show if the tissue being examined is stiffer than the embedding tissue, which might suggest that this tissue can be abnormal, e. g., malignant nodules are stiffer than benign nodules. There are many systems that use built-in functions of elastography, exploiting the property of the ultrasound wave – shear-wave propagate faster in stiff tissue than in soft tissue, the tissue stiffness can be inferred by measuring shear-wave velocity [2, 3]. Unfortunately, such systems are very costly. The approach described in this paper aims to provide ultrasound elastography imaging with a standard non-modified ultrasound device not equipped with an elastography unit. All calculations were performed on a PC connected to an ultrasound machine.

All existing methods for providing additional information from an ultrasound image can be divided into three types [4].

Strain Imaging Methods:

Quasi Static Methods (Ophir et al., 1991; O'Donnell et al., 1991);

Dynamic Methods (Parker et al., 1990; Krouskop et al., 1987; Sandrin et al., 1999);

Radiation Force (Walker, 1999; Fatemi & Greenleaf, 1999; Nightingale et al., 2002; Lizzi et al., 2003).

Stress Imaging Methods:

Mechanical or Tactile Imaging (Sarvazyan et al., 1998; Wellman et al., 2001):

Computational Models;

Finite Element Modelling;

Using Surface Pressure Information.

Modulus Imaging Methods:

Iterative Modulus Reconstruction (Kallel et al., 1995);

Direct Methods (Solving PDE's) (Emelianov et al., 2000; Sumi et al., 1995);

Finite Element Inversion (Zhu et al., 2003).

In this paper, we discuss the implementation of the intelligent image processing algorithm for ultrasound elastography supported with multi-core calculations. An approach that does not allow quasi-static, shear wave or acoustic radiation force impulse (ARFI) imaging lacks precision and requires clinicians to put in some additional effort [5]. Our algorithm makes use of video streaming only and does not account for the velocity of a wave. These binding constraints imply that the system is to be used dynamically – the operator will have to apply some pressure with an ultrasound device on a periodic basis.

To start with, a video-stream from an ultrasound device is transmitted to a PC via the VGA-port. The program captures frames from this stream and displays them in the main window. The user selects a region of interest (ROI), and the program accumulates the images in the image storage. The images uploaded to the image storage are grayscale. To create a visual map, or elastogram, the program renders frames in real-time and, applying a special algorithm, calculates the relative stiffness of one tissue section in comparison with the other tissues in ROI. The algorithm is akin to the schema in neural networks. First, we compute the weights for each pixel or zone of pixels of the image stored in the image storage. Then, having matched up the actual image to the estimated weights and having made use of a threshold function, we can see the difference in density, as it is indicated by the color of every dissimilar pixel in the zone of pixels. A threshold function supports four color spaces – blue (soft tissues), green (soft tissue of medium density), yellow (stiff tissue of medium density), and red (the most rigid tissue). Before getting to the image storage, images are put through a filter. It is aimed at ensuring accuracy and helps to avoid saving artifacts. Another algorithm allows the image storage to get updated to the latest frame, which means getting the target image right after pressure was applied to the tissue.

The program can extract visual features from ultrasound images. When the user selects a region of interest, the program collects black and white images $I_1, I_2, I_3, \ldots, I_m$ compressed to an optimal size (to accelerate the program and image delivery) and saves them in an intermediate image stack of a given size:

$$S = \{I_1, I_2, I_3, \ldots, I_m\},$$

where n is a whole number.

The following actions can be performed with the selected area:

inverting colours;

equalizing the histogram;

displaying the elastogram;

identifying feature points and showing their displacement vectors.

When the user chooses an overlay of the elastogram, each current image C is compared with a stack of images, using the following formula:

$$E_{xy} = \frac{\sum_{n=0}^{m} (C_{xy} - I_{xy_n})^2}{n},$$

where E_{xy} is a part of the selection (may consist of one pixel or of the square area with sides of 10 pixels). This, with the addition of some RGB schemes, determines the color of the pixels on the elastogram. There is a choice of four colors: red (if $6 \geq E_{xy} \geq 0$), yellow (if $10 \geq E_{xy} > 6$), green (when $20 \geq E_{xy} > 10$), and blue (if $E_{xy} > 20$). These colors were selected in accordance with established practice. The thresholds were chosen on the basis of the study of the boundary correspondence between certain types of tissues that have different stiffness. After every fifth iteration (frame updating or a certain period of time), the oldest image is deleted and a new one is added to the stack. This methodology relies on image dynamics – after the tissue was subjected to external pressure, its deformed areas are highlighted blue, and its rather rigid areas are highlighted red.

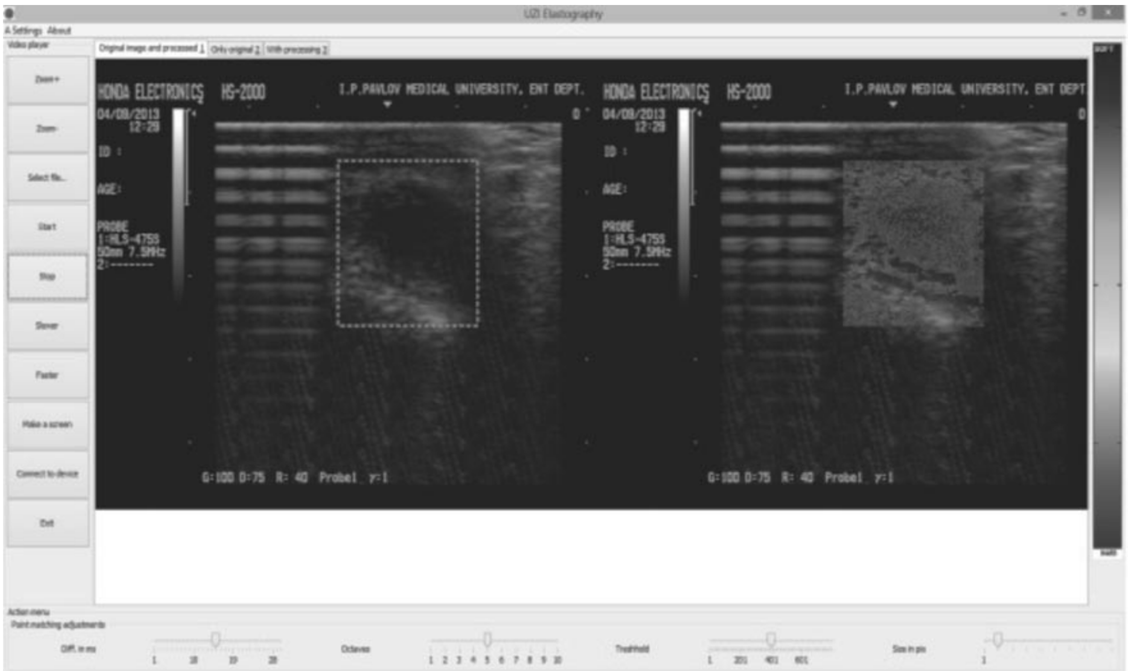


Fig. 1. Ultrasound and elastography images used to identify tissue stiffness

The elastography image gets displayed next to the corresponding original image. To remove noise, the user can increase the size of the E_{xy} by utilizing a special slider, and each pixel will be replaced with a cluster of them.

The image stack holds up to 30 images. If, after 5 iterations, the number of the images stored in the stack equals or exceeds its maximum capacity, the first 15 images are removed, and the program retains only the latest 15 images, followed by another 30 ones to fill the stack. These limitations are empirical in nature and stem from the acquired level of accuracy and the obtained results.

To further increase accuracy, there is another stack which is updated alongside with the original one and holds 50 large-scale images. These sample images are periodically com-

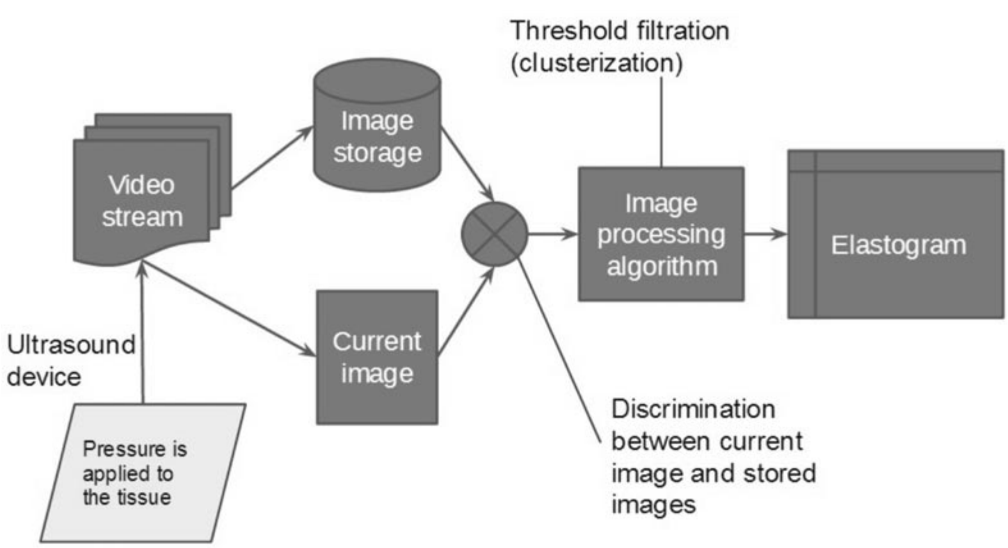


Fig. 2. Schematic algorithm for computing an elastogram

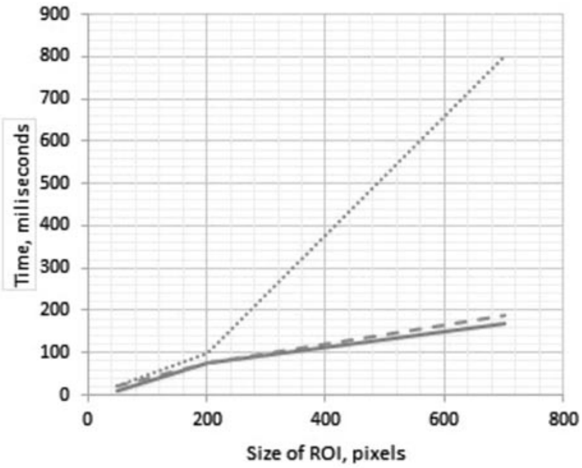


Fig. 3. Correspondence between the computation rate and the window size in combination with various parameters of parallelization
(••••••) without parallel computing;
(———) with parallel computing X axis;
(– – –) with parallel computing Y axis

pared with those from a smaller stack, so that severe errors in the calculations for an elastogram would not be missed. If any discrepancy is detected, both stacks are utilized to form the next frame to compute the elastogram.

The results of the computed analysis that involved examining a human lymphoma are shown in Fig. 1. The lymphoma can be seen within the selected region of interest in the left image. The post-processed image on the right side is an elastogram. As expected, the program indicates the rigidity of the lymphoma tissue with a red highlight, while the area around it is tinted green and blue, which implies that it is composed of fat or muscle.

Fig. 2 presents the scheme for computing an elastogram.

The multi-core approach was executed by doing parallel weight computation of the image storage in each frame. Parallel computing in the ROI was down by rows [6]. Such an approach allows for almost real-time elastography mapping. When the program is used on a PC with a Core 2 processor, the FPS is only 15, but with a CPU Core i7 processor, the FPS amounts to 45, which makes it possible for the radiologist to view a real-time video feed of the ultrasound exam and conduct an immediate analysis of the images. The correspondence between the computation rate and the window size, coupled with various parameters of parallelization, is shown in Fig. 3. With no parallelization, the computation time grows exponentially with the increasing size of the ROI. The computational procedure of creating an elastogram was tested not only without parallelization, but also with parallelization in the X and Y axes. The program produces similar results for parallelization both in the X axis and in the Y axis.

Having studied different approaches, we decided on such methods for detecting feature points as SIFT and SURF. Feature points are used to spot some contrast differences between images and determine how far these points are from each other. As soon as feature points have been detected, the program makes use of the K-means clustering algorithm to find the nearest points in the previous image and defines the displacement vector for each point [7]. But the application of these methods is complicated in large part due to the specific ultrasound imagining characteristics – ultrasound images feature unclear contours and there is no way to find any certain point. If we lower the threshold for finding feature points, computation time will increase exponentially [8]. The results of computing the SURF algorithm with different thresholds are presented in Fig. 4.

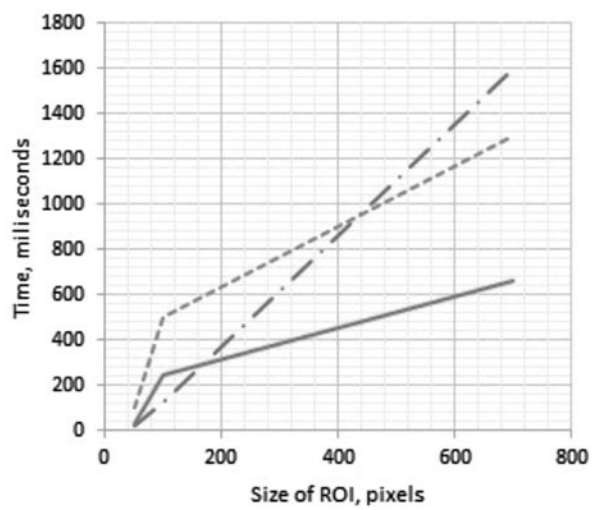


Fig. 4. Correspondence between computation time and the size of a ROI in relation to different values of the threshold in the SURF algorithm
(— · —) treshold equals 400;
(———) treshold equals 100;
(– – –) treshold equals 50

Compared with the SURF algorithm, the SIFT method turned out to take much longer to yield results and produced fewer feature points than the application requires in order to work properly. After applying one iteration, the user can see the displacement vectors of the feature points detected earlier.

The direct comparison method has proven to be the most effective of all.

We are planning to add functionality to the program by implementing the following features.

- Categorization of tissues according to their stiffness to enable anomaly detection.
- Machine learning aspect [9]. The learning aspect will involve learning on test cases featuring malignant and benign nodules [10].
- The function for selecting two ROIs and defining their correlation coefficient.
- The function for determining the Young's modulus for any point in the image.
- Utilizing finite element analysis of images.
- Transducing the contact force with which an ultrasound device is pressed against the skin, correlating the resulting data with the registered strain, and thereby determining more accurate measurements of tissue stiffness [11].
- Clustering selected points or pixels of any color to visually identify the contour lines.
- In the future, to classify diagnostic findings, the operator will have to save research results in a database.

In our laboratory, we used a General Electric 200 PRO ultrasound machine and a PC with a Core-i7 CPU that has 4 cores. The application has proven successful in laboratory testing, and now it is being tested by medical experts who are to estimate its precision and evaluate the utilization of the statistical parameter.

The major findings of the current study can be summarized as follows.

The given method suffers from several limitations: some outer tissues are of the same color as the tissues inside of them, and as the program works only with images, it simply cannot distinguish this hidden layer.

In 17 out of 20 video streams showing different malignant nodules, the program determined that the tissue surrounding these lumps were much stiffer than the other tissues and highlighted them red, which establishes the precision of the algorithm at 0.85. In contrast, the accuracy level of the elastography method based on the quasi-static approach is 0.95 [12].

Feature points algorithms are applicable only in cases with small-sized ROIs.

When used on a multicore PC, the application provides near real-time video stream processing.

The method is noise-resistant and not subject to perspective distortion.

The method is conditional on pressure being applied periodically to the area meant to be examined with elastography.

The program was written in the Java and Scala programming languages, thus it is possible to run it on any OS where the JDK v.7 is installed.

To obtain more accurate results, such features as the histogram equation, color inversion and binary comparisons between images were utilized in the program [13]. The user can choose which function to make use of and compare results. In addition, the user of the program can stop a video stream to examine the image. Alternatively, the user can slow a video down or speed it up when needed.

REFERENCES / СПИСОК ЛИТЕРАТУРЫ

1. **Benson J., Fan L.** A Complete Ultrasound Solution for Elastography, Tissue Strain Analytics, Siemens Medical Solutions USA, Inc., 2012.

2. **Si Luo, Cheoljin Lee, Kerem Karadayi, Yongmin Kim** Real-Time Ultrasound Elastography on a Multi-core DSP, Texas Instruments Incorporated, 2011.

3. **Kolesnikov M.P., Potekhin V.V., Tchernykh T.Y.** Active vibration control using modal analysis and rejection of resonance. *15th Internat. Congress on Sound and Vibration,* 2008.

4. **Saftoiu A., Vilmann P., Gorunescu F., Gheonea D.-I., Gorunescu M., Belciug S.** Endoscopic Ultrasound Elastography in the Diagnosis of Pancreatic Cancer. *Annals of gastroenterology*, 2010, Vol. 23(3), pp. 200-201.

5. **Varghese Tomy.** Elastography: Imaging the elastic properties of soft tissues with ultrasound. *Journal of Medical Ultrasonics*, 2002, Vol. 29, Iss. 4, pp. 155–171.

6. **Carlsen J.F., Ewertsen C., Lцnn L., Nielsen M.B.** Strain Elastography Ultrasound: An Overview with Emphasis on Breast Cancer Diagnosis. *Diagnostics,* 2013, No. 3.

7. **Bishop C.M.** Pattern recognition and machine learning, 2006.

8. **Luo Juan, Oubong Gwun** A Comparison of SIFT, PCA-SIFT and SURF. *Internat. Journal of Image Processing*, 2009, Vol. 3, Iss. 4.

9. **Varghese T., Zagzebski J.A., Frank G., Madsen E.L.** Elastographic imaging using a handheld compressor, *Ultrasonic Imaging*, 2002, Vol. 24, pp. 25—35.

10. **Garra B.S.** Tissue elasticity imaging using ultrasound, *Applied Radiology*, Apr. 2011.

11. **Chen L., Housden R.J., Treece G.M., Gee A.H., Prager R.W.** A hybrid displacement estimation method for ultrasonic elasticity imaging, CUED/F-INFENG/TR 615, Cambridge University Department of Engineering, Nov. 2008.

12. **Kapoor Atul, Kapoor Aprajita, Mahajan Goldaa** Real-time Elastography in Acute Appendicitis. *Ultrasound Med*, 2010, No. 29, pp. 871—877.

13. **Rangayyan Rangaraj M.** Biomedical Image Analysis, 2008.

FEDOROV, Aleksandr V. *St. Petersburg State Polytechnical University.*
195251, Polytekhnicheskaya Str. 21, St. Petersburg, Russia.
E-mail: aleksandr.v.fedorov@gmail.com

ФЕДОРОВ Александр Викторович — *студент кафедры систем и технологий управления Санкт-Петербургского государственного политехнического университета.*
195251, Россия, Санкт-Петербург, ул. Политехническая, д. 21.
E-mail: aleksandr.v.fedorov@gmail.com

POTEKHIN, Vyacheslav V. *St. Petersburg State Polytechnical University.*
195251, Polytekhnicheskaya Str. 21, St. Petersburg, Russia.
E-mail: slava.potekhin@gmail.com

ПОТЕХИН Вячеслав Витальевич — *доцент кафедры систем и технологий управления Санкт-Петербургского государственного политехнического университета, кандидат технических наук.*
195251, Россия, Санкт-Петербург, ул. Политехническая, д. 21.
E-mail: slava.potekhin@gmail.com

UDC 004.82:004.62=111

S.S. Zobnin, V.B. Polyakov

INTELLIGENT TIME SERIES STORING AND INFERENCE ENGINE IMPLEMENTATION WITH FOCUS ON PERFORMANCE AND HIGH LEVELS OF ABSTRACTION

This paper covers several aspects of intelligent time series database implementation. It also includes the description and analysis of a symbolic time series representation scheme. The paper focuses on various indexing and parallelization approaches in conjunction with actual backend storage engines. Special emphasis is made on identifying the problem of combining simple queries with time series pattern search and retrieval requests and finding a solution to this problem. The paper also considers query definition and provides the general architecture of a time series database with data mining capabilities.

TIME SERIES; DATABASE; DISTRIBUTED SYSTEM; INDEXING.

С.С. Зобнин, В.Б. Поляков

РЕАЛИЗАЦИЯ ВЫСОКОПРОИЗВОДИТЕЛЬНЫХ СИСТЕМ ХРАНЕНИЯ И ОБРАБОТКИ ВРЕМЕННЫХ РЯДОВ, ПРЕДСТАВЛЕННЫХ В ВИДЕ ВЫСОКОУРОВНЕВЫХ АБСТРАКЦИЙ

Рассмотрены аспекты реализации высокопроизводительных систем хранения и обработки временных рядов, представленных в виде высокоуровневых абстракций. Особое внимание уделено производительности конечного решения и способам увеличения производительности. Проведено сравнение различных высокоуровневых представлений временных рядов по разным критериям. Изучены различные распределенные низкоуровневые хранилища данных. Установлена и рассмотрена проблема эффективного выполнения комбинированных запросов к хранилищу временных рядов. Рассмотрена задача создания запроса для рассматриваемой базы данных временных рядов. Приведена конечная архитектура рассматриваемой системы.

БАЗА ДАННЫХ ВРЕМЕННЫХ РЯДОВ; РАСПРЕДЕЛЕННАЯ СИСТЕМА; ИНДЕКСИРОВАНИЕ.

Nowadays big volumes of time series data are generated in various fields of science and engineering. Most of the time such data is just stored and forgotten (or analyzed using standard visualization / retrieval tools «by hand»), but it is very tempting to have an opportunity to extract certain amounts of information that might be useful for immediate and strategical problem solving and control tasks, such as events, patterns, motifs, anomalies etc, in an automated or even automatic manner. The implementation of such system (which may also be called as intelligent database) is always accompanied by various problems. The main part of this paper contains a set of such problems and ways of solving them, referenced by related work and several contributions of this paper's author.

The main contribution of this paper is the integration of various techniques (which originated in signal processing, data mining, big data and distributed systems, engineering fields) connected with efficient and production-ready intelligent time series database implementation into a single reusable framework. Several secondary contributions are related to existing algorithm combinations and optimization. A novel time series approximation scheme is proposed. The purpose of this paper is to provide a general method (framework) of solving a range of tasks related to intelligent time series management and mining, as well as a general overview of the problem with some concrete ideas and practices.

Time Series Data, Representation Scheme

First of all, data itself, or, to be specific, the amount of time series and data quality is a problem. Hundreds of gigabytes of time series data to be processed in such system require high performance of storage backends. The quality or (and) specifics in data streams is also an issue: intelligent similarity search algorithms tend to depend on internal parameters of time series and have problems with generalization. Usually one should consider at least problems of noise (SNR) and data misses and provide ways of intelligent algorithms parametrization.

The general intelligent algorithm that works with time series data usually uses some intermediate time series representation. Many time series representations have been introduced (e. g. FFT, Wavelets, Piecewise-Constant Approximation, curve-based approximation etc). Recently it has been shown that symbolic representations of time series outperform most of other representations in terms of approximation performance and execution performance. Moreover, they allow using some frameworks which require symbolic input (e. g. Markov models, Kolmogorov complexity based methods, suffix trees), give good compression rate and give a lower bound on the real value (useful in data mining applications), reduce sensitivity to noise and greatly improve computational performance [1]. There are two types of time series symbolization: the first one is based on quantization and the second is based on temporal segmentation. It was found that quantization-based symbolization performs generally better on signals without temporal structure (the criteria are: information loss, accuracy, alphabet size) and vice-versa [2]. Several extensions for basic representation scheme have been developed, such as: symbolization based on kNN segmentation [3] where a relative frequency is used to determine the best approximation parameters, symbolization based on k-Means [4] which highly outperforms the standard scheme on highly Gaussian distributed streams. One may also employ a more intelligent approach: in order to optimize the algorithm for a particular dataset, a neural network may be tested and used either for an efficient rank-based parameter pick as in [3], or for an efficient segmentation scheme as in [4]. There is no doubt that symbolic representation of time series is most appropriate for similarity search in intelligent database (most arising) and it is the best solution for fast retrieval due to availability of various data structures from bioinformatics and computer science communities.

During the investigation of SAX time series representation the following specifics were found. SAX is very dependent on parameters (window size, word size, alphabet size). The window size parameter is responsible for capturing the dynamics of time series (data is normalized in each window and as a consequence). The word size and alphabet size parameters determine performance/space requirements of the transformation and operations over transformed time series. The word size and alphabet size correspond to discretization and quantization. It comes true that determining optimal parameters of SAX is an important and actual problem. Moreover, domain knowledge should be integrated in order to organize the most efficient storage and retrieval scheme. An example of such domain knowledge integration is manually defined stream groupings provided by an expert, the other example is unsupervised classification of time series, which determines parameters. It should be noted that it is also not optimal to generate an infinite number of parameter triples, because this requires a lot of additional computations. The optimization task should be defined to perform an on-line unsupervised classification that also minimizes the number of class reallocation in time.

Indexing Symbolic Time Series. Simple Queries. Approximations and Storage. Parallel Execution

Indexing is the core of a database, because there always are response and performance requirements of a database in production systems. One always faces a space-time dilemma when some indexing scheme is implemented, and uses an underlying representation (symbolic in our case) for the best performance. The chosen indexing scheme directly depends on queries to be executed against a database. In case of intelligent time series database, the main query is a «pattern query» or a similarity query. The other possible queries are event queries like «X more than 10 and Y less than 20» and more simple ones.

Here two combinations of indexing schemes are proposed. Each of these combinations allows fulfilling all query requirements.

Let's consider the first combination. In order to allow fast «direct» pattern match against symbols, a suffix tree is built over symbolic representations of time series. This can be called the first level of index. The second level of index is an index for similarity-based queries. A KD-tree is used for this purpose. The same KD-tree is used to execute event queries. Performance is improved via using the following methods. At first, one probabilistic suffix tree is used to perform faster first level index operations. Secondly, KD-tree performs partial queries and uses heuristics. Last, but not the least, a separate «reduces space» index could be built in the reduced feature space of time series, where the reduction is done via domain knowledge or general PCA or SOM scheme. These approaches were evaluated by the author of this paper. Alternative approaches to KD-tree index are other structures optimized for similarity search: general B-tree based structures [5], R-trees [6]. It should be noted, that the second index level techniques listed here are basic ones: trees need special measures to handle balancing issues, more parallelization and distribution considerations should be made. Even though a tree index can be equally distributed, processing time for each node needs prediction, which itself represents a forecasting problem that is not easily solved; nodes distribution mechanisms with random effects also have some difficulties.

The second combination of methods and techniques that allow fulfilling advanced querying requirements as well as performance requirements gives more focus on distributional and fault-tolerance capabilities of the system. Instead of using a suffix tree as a primary indexing structure, one can use a distributed hashing structure. There has been large research on distributed hashing structures recently, especially Distributed Hash Tables (DHT): CAN [7], Chord [8] and others. Generally, the problem of indexing time series was thoroughly studied in [9].

The advantage of the suffix tree as a first level indexing structure is its simplicity and general ability to perform suffix (substring) queries. DHT-like structures require much more additional space to allow such expressivity. On the other hand, DHT-based structures are purely distributed, thus naturally allowing distributed computations and storing.

A storage and cluster computation mechanism is required in order to fulfill strict requirements of productions environment. Although there is some research into standard backends usage for time series mining tasks [10], such DBMSs lack replication and fault-tolerance due to highly structured data predisposition. The integration of intelligent capabilities requires modifying the source code like in [11], which is error prone and not appropriate in many cases. A modern distributed computation framework is much more preferable, given the fact that parallelism can always be exploited and is implemented naturally in such frameworks. One more requirement is the ability of manual RAM caching of computations in such framework, the indexes are more preferable in RAM rather than on hard drive. As a result, considerable systems are [12–14].

It should be noted, that such distributed systems are executed on commodity hardware (i. e. standard hardware with no special requirements), comparing to specialized massively parallel solutions, for example [15] meaning that they are more universal and have bigger latencies comparing to specialized hardware/software massively parallel solutions at the same time. One implication is that the time series database system implemented on top of commodity hardware is not capable of handling hard real time tasks arising, for example, in military domains.

The General Architecture of an Intelligent System

The general architecture of an intelligent time series database can be represented in Fig. 1.

The central component is the intelligent database itself, where indexes are stored in memory, while raw data is stored in symbolic form on the disk. This may require a cluster of machines due to large size of indexes.

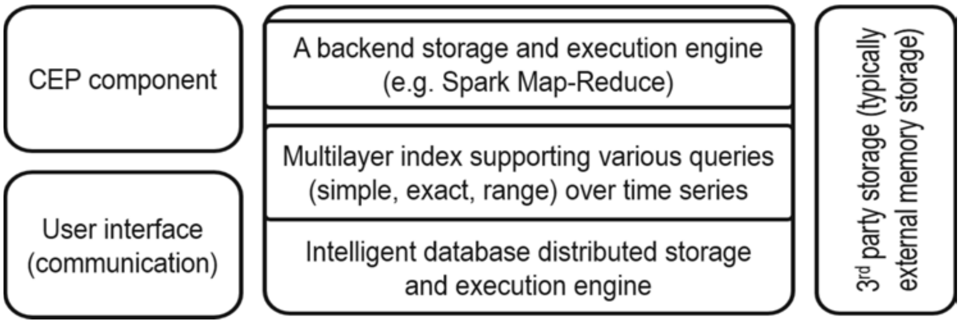


Fig. 1. The overview of architecture of an intelligent time series database

The system is highly coupled with some 3rd-party storage, because raw data is always stored somewhere. A good architectural solution is to create an «Adapter» for each 3rd party, the database is working with, so that the intelligent database itself and adapters are highly reusable.

The CEP component of the system is responsible for preprocessing data for detection of some event types. This component executes event processing queries against data, a good solution is to reuse some CEP engine while building indexes and use some declarative language to find events rather than to encode requests by hand. CEP component also provides additional stream processing capabilities for on-line analysis, i. e. stream manipulation, aggregation, joining etc.

It is very important to have a subsystem responsible for user communication so that an operator would have a way to present patterns for queries in convenient forms. An example of such assistance is a visual pattern constructor, where it is easy to define the shape of a curve (pattern) to be found in the database.

The concept of multilevel indexing is represented in Fig. 2.

As discussed in the previous session, the most optimal storing and indexing scheme for time series is the scheme containing both computer science techniques and bioinformatics techniques. This allows fulfilling performance requirements as well as provides capabilities to perform simple queries, exact time series match queries, range queries, N-nearest time series queries and others. Not only does such scheme allow such querying capabilities, but it also allows using distributed systems and networks concepts and experience: standard DHT or CAN implementations may be used as the second level of index.

It is convenient to have a distributed optimization component in common space to allow its reusability: the need of solving optimization tasks arises on multiple levels of the system. Currently a version of distributed

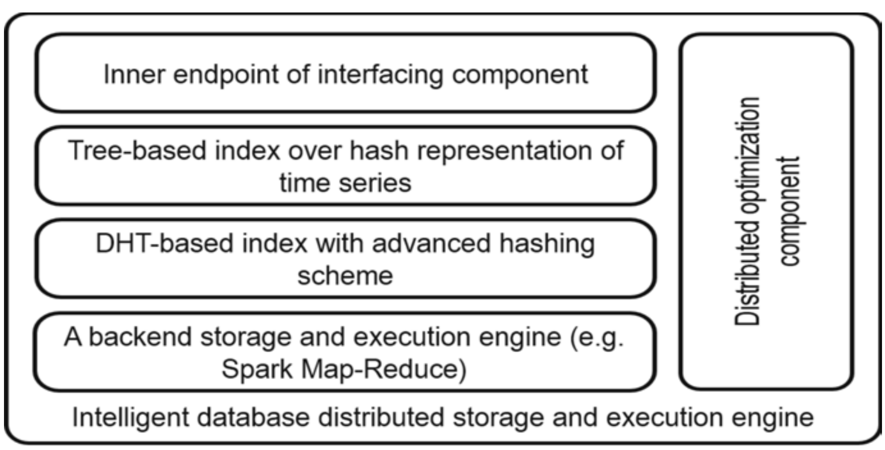


Fig. 2. The overview of architecture of an intelligent time series database

simulated annealing is used as an optimization framework. A special protocol is implemented for optimization of agents' communication. This protocol allows asynchronous optimization, incremental updates of «current optimum» and dynamic task reallocation.

The following figure (Fig. 3) shows in detail the process of parsing incoming and generating a distributed index for later querying as well as the process of query generation, processing and execution, one can also consider this figure as a general algorithm, i. e. how the system works.

A system constantly accepts incoming data from various origins. CEP/Stream processing (as well as any simple processing like stream multiplication) queries are constantly executed on incoming data in order to receive alternated streams.

These data streams arrive to the module responsible for generating symbolic representation. This very module contains additional domain knowledge, possibly automatically inferred from data, that allows more efficient storage and representation. For example, there can be additional logic for automatic unsupervised classification of incoming data streams. A simpler approach is to integrate domain knowledge of an expert in various forms. For example, an expert can set up proper groups of similar signals from all. Different parameters based on domain knowledge are then used for symbolic representation generation algorithm.

The output of the module described in the previous paragraph goes to a specially designed hash function that gives hash values to time series dividing them into equivalence groups (dimensionality reduction). Generated hash values are then mapped into DHT / CAN nodes and distributed around the network. This process is executed continuously because amounts of incoming data can be too large. The other continuously running indexing process is the process that performs indexing of hashes into tree structures. These structures are also ideally distributed as it was discussed in the previous chapters.

User defined queries, possibly from a special graphical interface for query construction, as well as constraints on window size and additional parameters are supplied to the other side of the system. This data is then translated into patterns suitable for search, a search query is executed and some post processing of search results are performed.

At least two general ways of the described system integration can be defined. The first one puts time series database in the front line (data acquisition layer). The second way of integration considers time series database as an

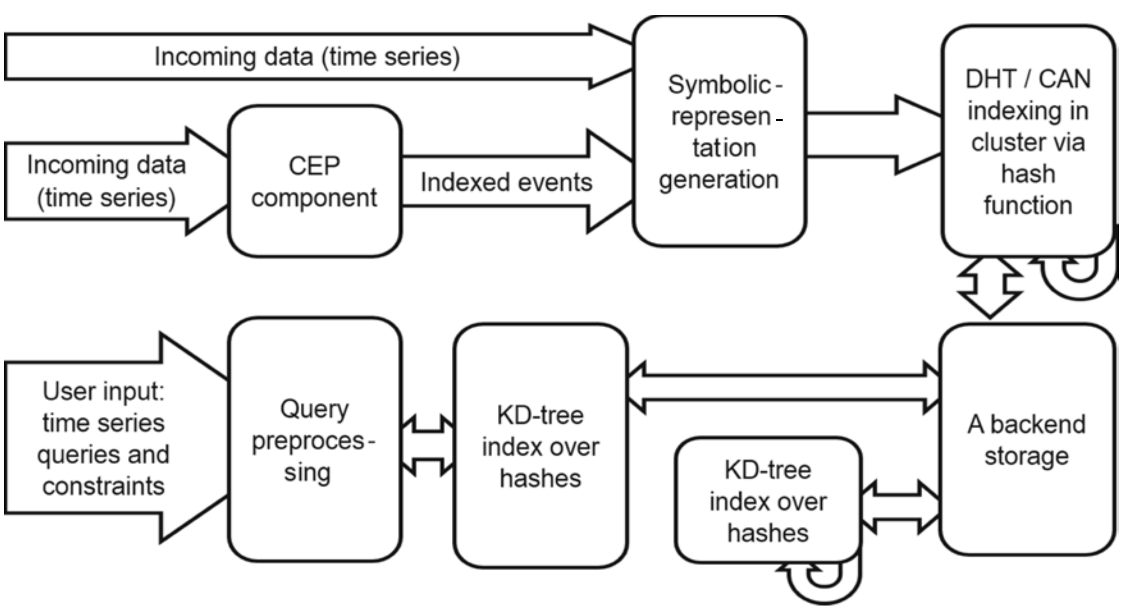


Fig. 3. Internal data flows

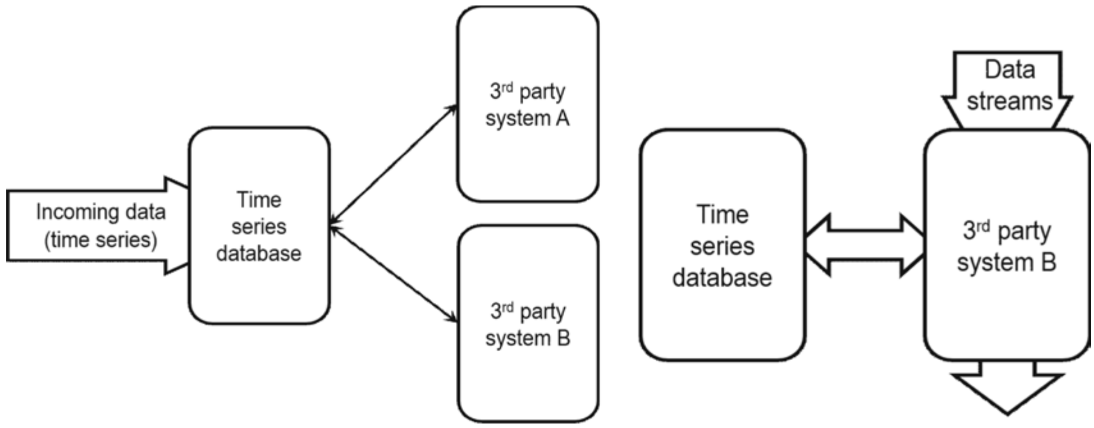


Fig. 4. Ways of integrating time series database with other systems: left, where time series database is considered as a part of data acquisition layer and right, where considered system is used as an auxiliary system for other products

auxiliary part of another bigger system. Both these cases are represented in Fig. 4.

The left case (Fig. 4, left) inserts time series database into the system circuit, thus reducing total system fault-tolerance. On the other side, in such way time series database receives on-line data and has this data indexed allowing other systems to work with this data at once. Thus, such way is the best for 3rd party systems that require intelligent time series manipulations with low latency. It is also possible to stream data in parallel to third party components having all pros at the same moment.

The right case (Fig. 4, right) puts time series database aside from real time series. In this way an intelligent database does not have data in real time, but it is not located in the main data circuit. Such way of integration is more suitable for cases when off-line processing is required for some existent systems as a separate module, and extra low latency is not required.

The approach provided in this paper is applicable to a wide range of tasks related to time series: analysis, efficient storage, approximation, motif discovery, classification, and others. For example, in order to implement a fully functional prototype of the vibration diagnostic system based on models from [16], it is required to have a subsystem able to provide both raw time series functionality and similarity search functionality. The system based on the concept from [17] also requires historical access and advanced time series similarity search capabilities.

It was shown that the chosen methods (symbolic representation + multilevel indexes with approximations at distributed storage, queries are executed in cluster in parallel) greatly outperform standard approaches (standard time series representations, SQL databases) in terms of amount of queries per second and single query execution time.

The ongoing research is targeted at investigation and comparison of concrete algorithms and methods for each stage of the described systems, integration of chosen algorithms, improving them and investigating the means of domain knowledge integration.

The research and implementation of the system based on the concepts from this paper is performed as a part of the project related to time series analysis and visualization in Siemens LLC.

REFERENCES

1. **Lin J., Wei Li et al.** Experiencing SAX: a Novel Symbolic Representation of Time Series, 2007.

2. **Sant'Anna A., Wickstrom N.** Symbolization of time-series: An evaluation of SAX, Persist, and ACA. *Image and Signal Processing, 4th Internat. Congress on*, 2011, Vol. 4, pp. 2223–2228.

3. **Ahmed A.M., Bakar A.A., Hamdan A.R.** Improved SAX time series data representation based

on Relative Frequency and K-Nearest Neighbor Algorithm. *Intelligent Systems Design and Applications, 10th Internat. Conference on*, 2010, pp. 1320–1325.

4. **Pham N.D., Quang Loc Le, Tran Khanh Dang.** Two Novel Adaptive Symbolic Representations for Similarity Search in Time Series Databases. *12th Internat. Asia-Pacific Web Conference APWEB-2010.*

5. **Cui Yu, Bin Cui, Shuguang Wang, Jianwen Su.** Efficient index-based KNN join processing for high-dimensional data. *Inf. Softw. Technol.,* 2007, Vol. 49, No. 4, pp. 332–344.

6. **Kuan J., Lewis P.** Fast k nearest neighbour search for R-tree family. *Information, Communications and Signal Processing, Proceedings of Internat. Conference ICICS-1997*, Vol. 2, pp. 924–928.

7. **Ratnasamy S., Francis P., Handley M., Karp R., Shenker S**. A scalable content-addressable network. *SIGCOMM Comput. Commun.,* 2001, Rev. 31, No. 4, pp. 161–172.

8. **Stoica I., Morris R., Karger D., Kaashoek M.F., Balakrishnan H.** Chord: A scalable peer-to-peer lookup service for internet applications. *SIGCOMM Comput. Commun,* 2001, Rev. 31, No. 4, pp. 149–160.

9. **Mar Yi Yi, Maw Aung Htein.** Tree-based Indexing for DHT-based P2P Systems. *Internat. Journal of Computer Applications*, 2013, Vol. 62, pp. 27–33.

10. **Sorokin A., Selkov G., Goryanin I.** A user-defined data type for the storage of time series data allowing efficient similarity screening. *European Journal of Pharmaceutical Sciences,* 2012, Vol. 46, Iss. 4, pp. 272–274.

11. **Bartunov O., Sigaev T.** Efficient K-nearest neighbour search in PostgreSQL. *Conference talk, PGDay-2010*, Roma.

12. **Dean J., Ghemawat S.** MapReduce: simplified data processing on large clusters. *Commun. ACM,* 2008, Vol. 51, No. 1, pp. 107–113.

13. **Zaharia M., Chowdhury M., Franklin M.J., Shenker S., Stoica I.** Spark: cluster computing with working sets. *In Proceedings of the 2nd USENIX Conference on Hot topics in cloud computing*, USENIX Association, Berkeley, CA, USA, p. 10.

14. GridGain systems GridGain. Available: http://www.gridgain.com

15. **Polyakov V.B.** Organizacija vychislenij v arhitekturah, podderzhivajushhih massovyj parallelism. *Voprosy radioelektroniki.* Moscow: Elektronica Publ., 2008, pp. 33–41. (rus)

16. **Zinkovsky A.V., Potekhin V.V., Ilyin I.Y.** Modelling of vibration activity on a man locomotorium. *Proceedings of the 10th Internat. Congress on Sound and Vibration*, 2003, pp. 3851–3856.

17. **Arsen'ev D.G., Shkodyrev V.P.** Strategija gruppovogo upravlenija robotami na osnove situacionno-celevogo planirovanija. *Izvestija Juzhnogo federalnogo universiteta. Tehnicheskie nauki.* Taganrog, 2010, pp.40–43 (rus)

СПИСОК ЛИТЕРАТУРЫ

1. **Lin Jessica, Wei Li, et al.** Experiencing SAX: a Novel Symbolic Representation of Time Series. 2007.

2. **Sant'Anna A., Wickstrom N.** Symbolization of time-series: An evaluation of SAX, Persist, and ACA // Image and Signal Processing, 4th Internat. Congress on. 2011. Vol. 4. Pp. 2223–2228.

3. **Ahmed A.M., Bakar A.A., Hamdan A.R.** Improved SAX time series data representation based on Relative Frequency and K-Nearest Neighbor Algorithm // Intelligent Systems Design and Applications, 10th Internat. Conf. on. 2010. Pp. 1320–1325.

4. **Pham N.D., Quang Loc Le, Tran Khanh Dang.** Two Novel Adaptive Symbolic Representations for Similarity Search in Time Series Databases // 12th Internat. Asia-Pacific Web Conference. 2010.

5. **Cui Yu, Bin Cui, Shuguang Wang, Jianwen Su.** Efficient index-based KNN join processing for high-dimensional data // Inf. Softw. Technol. 2007. Vol. 49. no. 4. Pp. 332–344.

6. **Kuan J., Lewis P.** Fast k nearest neighbour search for R-tree family // Information, Communications and Signal Processing. Proc. of Internat. Conf. ICICS. 1997. Vol. 2. Pp. 924–928.

7. **Ratnasamy S., Francis P., Handley M., Karp R., Shenker S.** A scalable content-addressable network // SIGCOMM Comput. Commun. 2001. Rev. 31. no. 4. Pp. 161–172.

8. **Stoica I., Morris R., Karger D., Kaashoek M.F., Balakrishnan H.** Chord: A scalable peer-to-peer lookup service for internet applications // SIGCOMM Comput. Commun. 2001. Rev. 31. no. 4. Pp. 149–160.

9. **Mar Yi Yi, Maw Aung Htein.** Tree-based Indexing for DHT-based P2P Systems // Internat. J. of Computer Applications. 2013. Vol. 62. Pp. 27–33.

10. **Sorokin A., Selkov G., Goryanin I.** A user-defined data type for the storage of time series data allowing efficient similarity screening // European J. of Pharmaceutical Sciences. 2012. Vol. 46. Iss. 4. Pp. 272–274.

11. **Bartunov O., Sigaev T.** Efficient K-nearest neighbour search in PostgreSQL // Conf. talk, PGDay-2010, Roma, 2010.

12. **Dean J., Ghemawat S.** MapReduce: simplified

data processing on large clusters // Commun. ACM. 2008. Vol. 51. no. 1. Pp. 107–113.

13. **Zaharia M., Chowdhury M., Franklin M.J., Shenker S., Stoica I.** Spark: cluster computing with working sets // In Proc. of the 2nd USENIX Conf. on Hot Topics in Cloud Computing. USENIX Association, Berkeley, CA, USA. P. 10.

14. GridGain systems GridGain [электронный ресурс] / URL: http://www.gridgain.com

15. **Поляков В.Б.** Организация вычислений в архитектурах, поддерживающих массовый параллелизм // Вопросы радиоэлектроники. М.: Электроника, 2008. С. 33–41.

16. **Zinkovsky A.V., Potekhin V.V., Ilyin I.Y.** Modelling of vibration activity on a man locomotorium // Proc. of the 10th Internat. Congress on Sound and Vibration. 2003. Pp. 3851–3856.

17. **Арсеньев Д.Г., Шкодырев В.П.** Стратегия группового управления роботами на основе ситуационно-целевого планирования // Известия Южного федерального университета. Технические науки. Таганрог, 2010. С. 40–43.

ZOBNIN, Sergey S. *St. Petersburg State Polytechnical University. Siemens LLC.*
195251, Politekhnicheskaya Str. 21, St. Petersburg, Russia.
E-mail: sergey.s.zobnin@gmail.com

ЗОБНИН Сергей Сергеевич – *студент кафедры систем и технологий управления Санкт-Петербургского государственного политехнического университета.*
195251, Россия, Санкт-Петербург, ул. Политехническая, д. 21.
E-mail: sergey.s.zobnin@gmail.com

POLYAKOV, Vadim B. *St. Petersburg State Polytechnical University.*
195251, Politekhnicheskaya Str. 21, St. Petersburg, Russia.
E-mail: vadim7702@yandex.ru

ПОЛЯКОВ Вадим Борисович – *доцент кафедры систем и технологий управления Санкт-Петербургского государственного политехнического университета, доктор технических наук.*
195251, Россия, Санкт-Петербург, ул. Политехническая, д. 21.
E-mail: vadim7702@yandex.ru

UDC 533.9.082.74=111

S.I. Vorobyev, V.V. Potekhin, S.N. Semenov

FILTERING METHOD FOR ANALYSIS OF THE IMAGE RECEIVED BY MICROWAVE PROBING

The article offered a complex of program procedures which can be used to process the point arrays of the space complex amplitudes obtained after reconstructing the images received by microwave probing. The procedures allow creating space «depth maps» of the probed object by filtering and fitting algorithms. The detail level is enough to make a comparative analysis with «depth maps» of video systems based on pairs of stereo video cameras.

INTELLIGENT IMAGE FILTERING SYSTEMS; MICROWAVE TECHNOLOGY; IMAGES OF RADIOWAVE PROBING.

С.И. Воробьев, В.В. Потехин, С.Н. Семенов

МЕТОДИКА ФИЛЬТРАЦИИ АНАЛИЗА ИЗОБРАЖЕНИЙ МИКРОВОЛНОВОГО ЗОНДИРОВАНИЯ

Предложен комплекс программных процедур обработки массивов точек пространства комплексных амплитуд, полученных в ходе процесса восстановления изображений микроволнового зондирования, способных создавать на основе алгоритмов фильтрации и сглаживания пространственные карты глубины зондируемого объекта, с уровнем детализации, достаточным для осуществления сравнительного анализа с картами глубин видеосистем, на основе конструкции стереопары.

ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА ФИЛЬТРАЦИИ ИЗОБРАЖЕНИЙ; МИКРОВОЛНОВАЯ ТЕХНИКА; ИЗОБРАЖЕНИЯ РАДИОВОЛНОВОГО ЗОНДИРОВАНИЯ.

The term «microwave image of the object» is usually understood as a distribution of complex scattered field amplitudes within the object space. This term, however, only slightly matches the classical definition of the image received «in visible light». However, unlike visible light, it fully informs of the characteristics of the electromagnetic field in the inspected area.

Considering all the above-mentioned, we face a question whether it is possible to obtain similar images of the objects in a visible and microwave range, taking into account differences in optical transmission and deflection properties of mediums, where the radiation propagates. This article considers the method of analyzing and reconstructing images of the conducting objects. It is known, that in the microwave frequency range of the electromagnetic radiation the conducting objects have a higher reflection index [3], which depends on the conductive properties of the object.

During the field reconstruction by the method described in the article [1], among all the range of the reconstruction points, the scattering centers will be the only points which have the biggest amplitude. It is worth noting that in the case of a electromagnetic wave and conductor the scattering centers are actually the points on the outer surface of the object. Since scattering the wave of the given frequency range takes place on the skin layer of the conductor, which is usually not more than several microns.

Thus we can say that points of the probed space, where, during the field reconstruction, the amplitudes of the reconstructed signal are maximal, are the points on the surface of the conducting object, i. e. those points, which could be seen by the video system.

Amplitude distribution of the reconstruction points is not a high gradient function. It is likely to be a function with a smoothly falling amplitude value depending on the distance to the scattering center [2]. This can also mean that if there is a noise component in the signal, the maximal amplitude could shift during the reconstruction in vicinity of the scatterer, and that may cause some uncertainty when trying to identify the scatterer's real location.

Filtering Based on Amplitude Distribution. According to the above-mentioned information, the procedure of locating scattering centers should start from identifying areas of probable localization for these centers, e. g. by using amplitude histogram. In order to distinguish which points of the grid belong to the reflecting surface of the object, we can speculate that a field amplitude value when the signal is reflected from the conducting surface is much bigger than at some distance from the surface. For this purpose a histogram of amplitude distribution (more precisely – of common logarithms of the amplitudes) is built for all points of the array as it is shown in the Fig. 1.

Supposing that the distribution of the amplitude logarithm is Gaussian, i. e. normal, the average value of the logarithm $< \lg(\mathrm{Amp}) >$ and standard deviation – are calculated using standard formulas.

$$< \lg(\mathrm{Amp}) >= \sum_{i=1}^{N} \frac{\lg(\mathrm{Amp})_i}{N} \qquad (1)$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (\lg(\mathrm{Amp})_i - < \lg(\mathrm{Amp}) >)^2}{N}} \qquad (2)$$

where N is a number of elements in an array (number of coordinate grid nodes).

Parameter R will be the one showing the contrast of the final microwave image; if this parameter takes high values, borders of the object surface will blur, whereas if the values are low, some surface areas will disappear. The choice of the parameter value depends on the object, its location relatively to emitting antennas and many other conditions. The optimal value for this parameter is determined manually and is based on the imaging results. This filtration is used to turn down further analyzing those points of the coordinate grid, in which the field amplitude is lower than the set value. In order to do this, we use values $< \lg(\mathrm{Amp}) >$, – from the amplitude diagram and R value from the set program parameters.

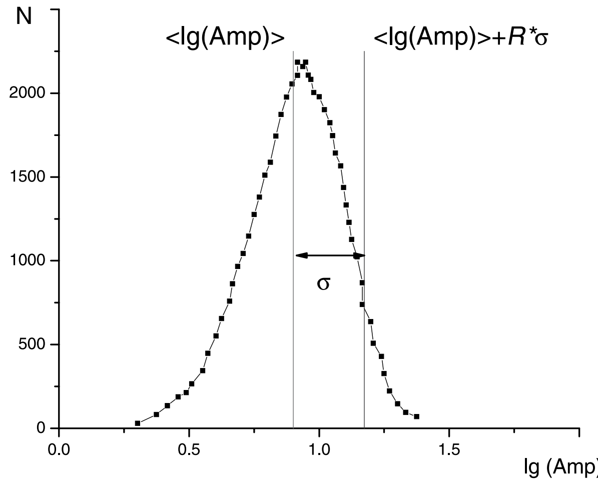


Fig. 1. Histogram of amplitude distribution

The operation algorithm for this filter is the following:

During the cycle all values of the $(\mathrm{Amp})_i$ array are compared with the value $(< \lg(\mathrm{Amp}) > +R\sigma)$.

If $(\mathrm{Amp})_i \geq 10(< \lg(\mathrm{Amp}) > +R\sigma)$, then $(\mathrm{Amp})_i$ saves it value.

If $(\mathrm{Amp})_i < 10(< \lg(\mathrm{Amp}) > +R\sigma)$, then $(\mathrm{Amp})_i = 0$.

Usually R value is 2-2.1 and after this filter is applied, the amplitude array has less than 5 % of the non-zero elements. After the described filtration is applied, those areas around the points are marked, which corresponds to the object surface, but the precise location of the surface borders is yet undetermined. In order to identify precise borders of the object, it is necessary to determine the point which has the maximal amplitude value. The main difficulty is to choose the direction (axis) of the search for this maximal value, because if the direction is wrong, it will result into duality of the scattering center coordinates.

Modelling the scattering processes for the microwave field implies the direction of the search for the maximal values of the reconstructed field intensity. This strongly depends on the shape and size of the effective aperture of the probing emitting array and on the location of the receiving antennas. We used the emitting array consisting of 256 transmitting elements with the square aperture [5]. Each of the elements of the antenna array generated a monochromatic microwave wave with the frequency between 10.5 and 16 GHz. In general, 16 generating frequencies were used. A receiver was placed at the distance of 1 m from the antenna array.

Each side of the square-shaped antenna array is 30 cm. The center of the antenna array (Fig. 2–4) is located in the coordinates (300; 0) along the *X* and *Z* axes accordingly. *Y* axis in this case is not taken into consideration, since Fig. 2–4 show only a two-dimensional (*XZ*) cut of the three-dimensional (*XYZ*) space. For the chosen configuration the scattering center is located on the normal-(beginning in the center of the array)-to-the-antenna array. Fig. 2 shows the reconstructed microwave field for the modeled point scatterer with the coordinates (300; 150).

As it is clear from Fig. 2, the maximal amplitude value and distribution of the secondary maximums of the reconstructed field is done only along the normal line to the receiving antennas array. The direction of the normal line coincides with *Z* axis of the coordinate system.

When the scattering center shifts relatively to the antenna array, the direction of the distribution of the secondary maximums of the scattered field also changes. The direction of the search for the maximal value should change respectively. Hereby the direction of the search will be presented by the line connecting the center of the mass of the antenna array and the scattering center.

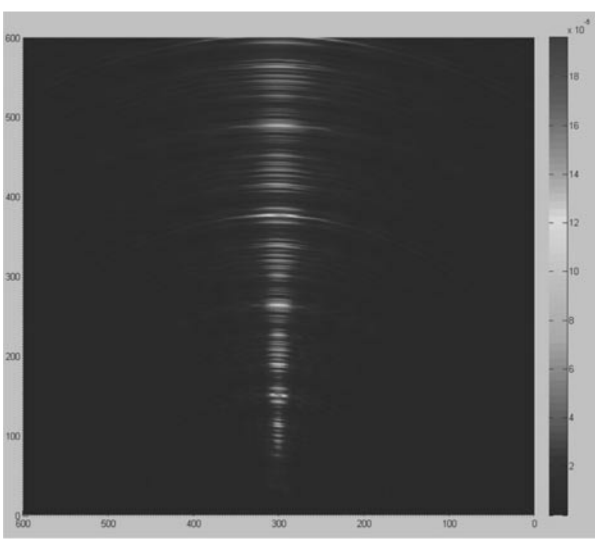


Fig. 2. Distribution of the reconstructed field with the point scatterer located in (300; 150)

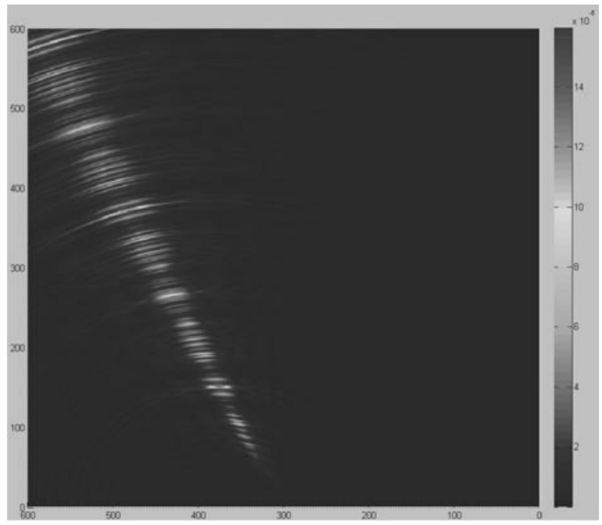


Fig. 3. Distribution of the reconstructed field with the point scatterer located in (380; 150)

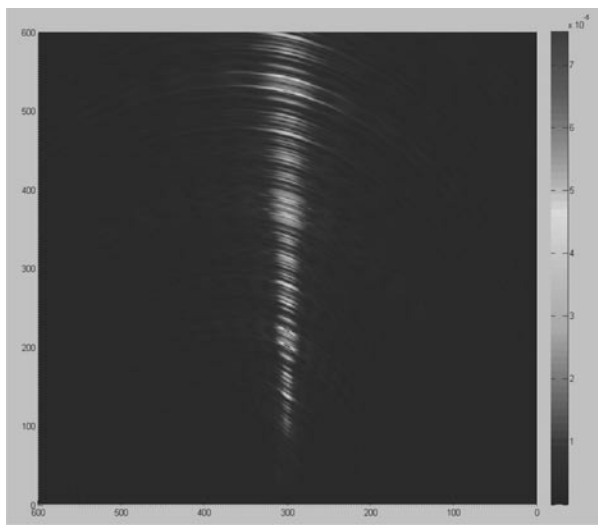


Fig. 4. Distribution of the reconstructed field with two point scatterers in the free space located at the 10 cm distance from each other along *Z* axis

When reconstructing one point object, it is necessary to choose correctly the axis, along which the search for the maximum will be done. This is not an obvious task. Below we consider the example in which the incorrectly chosen direction of the search for the maximum value could cause serious errors. Fig. 4 shows the distribution of the intensities of the reconstructed microwave field for two point scatterers with coordinates (300; 200) and (300; 210) respectively.

If the direction of the search for the maximum value lies along Z axis, we will receive a reconstructed image for only one scatterer instead of two. This may cause serious quality issues (up to the omission of certain elements) on the images of the real objects.

The direction of the search for the maximal values in this case should be determined by the location of the antenna array aperture centers and scattering centers.

When it comes to the practical use of the described method for the reconstruction of the images of the real objects, it is worth noting that the process of the microwave field reconstruction requires a lot of resources and depends directly on the number of the reconstructed points, i. e. on the size of the reconstructed space and required resolution of the object. It was concluded that the direction of the search for the maximal field value should be chosen so that each of the reconstructed points should form a certain direction of the search for the maximal value. This also requires a geometrical fit of a discrete set of the reconstructed points with the location of the receiving-transmitting elements.

After the unambiguous correlation between coordinates of the points with maximal amplitude values is found, filtering of the image does not include only analyzing the field amplitudes, but analyzing a point cloud corresponding to the object surface.

Filtering Based on the Video Image. During further filtering the microwave image is processed together with the video image of the object. The main idea is to eliminate the points of the microwave field, which are not presented on the video image, from the microwave image [4]. The similar data array, which describes the video image, comes from the video image processing module. These data are presented as a 2D-array of the coordinate values of stereo video image points. If some points of the video image with a pair of coordinates x_i and y_i are not available, the corresponding points from the microwave image are also deleted, otherwise the points save their previous values.

Filtering Based on the Method of the Connected Domains. The initial 3D model, which is formed during the image reconstruction, has a certain amount of noises caused by outer electromagnetic noises, imperfect light conditions and shape and texture of the

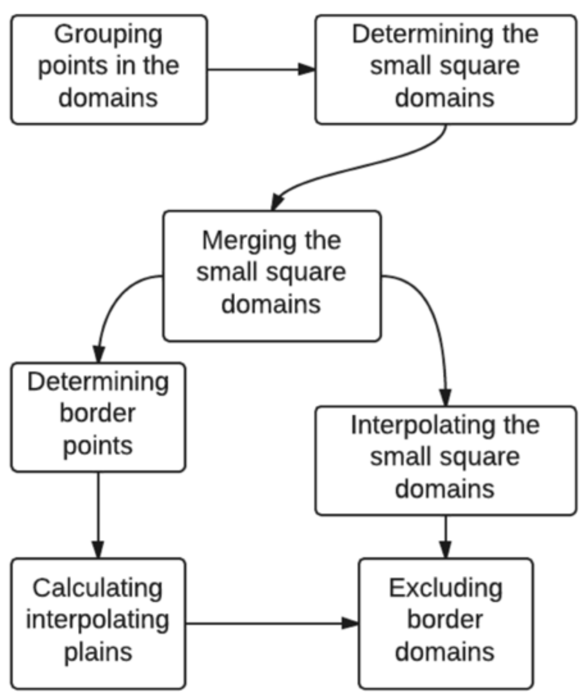


Fig. 5. Diagram of the filtering and interpolation algorithm applied to the reconstructed microwave image

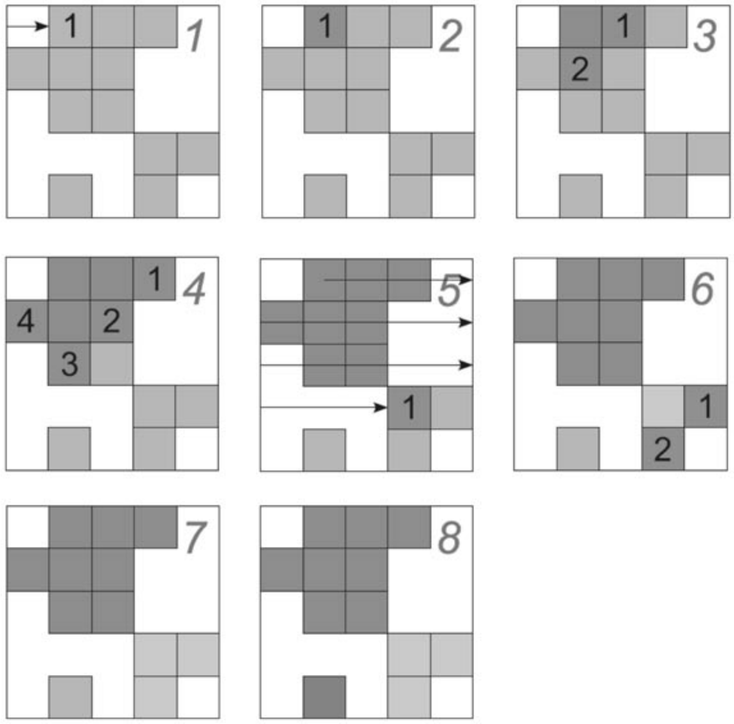


Fig. 6. Method of connected domains

objects. Filtering and interpolation are applied to the 3D model to lessen the influence of the above-said parameters during the following processing steps. The output data are shown as Z matrix with $n_x \times n_y$ dimensions. Values for each *Zij* element are equal to the distance to the point of the object with coordinates (x_i, y_j). During the first step of filtering the connected domains are determined. In other words, if any (*i, j*) point belongs to *K* domain, each neighboring points $(i + \delta i, j + \delta j)$, $\delta i, \delta j \in \{-1, 0, 1\}$. $\left|Z_{i+\delta i, j+\delta j} - Z_{i,j}\right| < \Delta z$, describes conditions of belonging to K domain, where Δz – a threshold value.

For each element, which still belongs to no group, all neighboring elements are determined, which belong to the same group, and the same operation is applied to those elements. This continues while any new neighboring elements, which meet the connecting condition, are available. Then the group is closed and processing goes to the next yet «uncoloured» point. Schematically the process is shown in Fig. 6.

A number of points belonging to each domain is identified. Domains, whose square is less than the certain threshold value, should be interpolated. Two of such domains are merged if they have at least two adjacent bordering points. After merging the domains are interpolated. The interpolation is done with areas of plains located on the edges of the interpolated domain. For each interpolated domain *K*0 all domains *Kj* are chosen for those bordering on it. The points belonging to these domains and bordering on *K*0 are used to calculate the regression coefficient for the plain. The interpolated domain is filled with distance values corresponding to the found plain. Full scheme of processes of «connected domains» filter is shown in Fig. 5.

In this article the author described the filtering method for the microwave images received by probing conducting objects. The method is based on the number of the discrediting methods of the microwave image components which do not belong to the surface of the inspected object. The present method could be considered as a set of main filtering steps, without any description of methods that are used to smooth or increase the image quality. Nevertheless, the method could be used as a basic approach to receive images of this type.

REFERENCES

1. **Zverev V.A.** Fizicheskie osnovyi formirovaniya izobrazheniy volnovyimi polyami. Moscow: IPF RAN, 1998. (rus)

2. **Yaroshlav L.P., Merzljakov N.S.** Tsifrovaya golografiya. St. Petersburg: Nauka Publ.,1982. (rus)

3. **Averianov V., Kuznetsov A.** Method for Remote Inspection of Target in Controlled Area of Space, *Patent Russian Federation*, No. 2294549, 2005. (rus)

4. **Averianov V., Vorobyev S., Gorshkov I.** The Method for Determining the Dielectric Permittivity of the Object, *Patent Russian Federation*, No. 2408005, 2008. (rus)

5. **Vorobyev S.I., Aver'yanov V.P., Osipov M.Yu, Semenov S.N.** The Multi- Building System of Real Time Microwave Imaging. *Proceedings of the 13th Internat. Scientific-Practical Conference Hi-Tech-2012.*

СПИСОК ЛИТЕРАТУРЫ

1. **Зверев В.А.** Физические основы формирования изображений волновыми полями. М.: ИПФ РАН, 1998 .

2. **Ярославский Л.П., Мерзляков Н.С.** Цифровая голография. СПб.: Наука, 1982.

3. **Averianov V., Kuznetsov A.** Method for Remote Inspection of Target in Controlled Area of Space // Patent Russian Federation 2294549, 2005.

4. **Averianov V., Vorobyev S., Gorshkov I.** The Method for Determining the Dielectric Permittivity of the Object // Patent Russian Federation 2408005, 2008.

5. **Воробьев С.И., Аверьянов В.П., Осипов М.Ю., Семенов С.Н.** Многопозиционная система построения микроволнового изображения в режиме реального времени // Сб. трудов XIII Междунар. науч.-практич. конф. Hi-Tech-2012.

VOROBYEV, Stanislav. I. *St. Petersburg State Polytechnical University.*
195251, Polytekhnicheskaya Str. 21, St. Petersburg, Russia.
E-mail: s.vorobjev@gmail.com

ВОРОБЬЕВ Станислав Игоревич – *аспирант кафедры систем и технологий управления Санкт-Петербургского государственного политехнического университета, магистр физики.*
195251, Россия, Санкт-Петербург, ул. Политехническая, д. 21.
E-mail: s.vorobjev@gmail.com

POTEKHIN, Vyacheslav. V. *St. Petersburg State Polytechnical University.*
195251, Polytekhnicheskaya Str. 21, St. Petersburg, Russia.
E-mail: slava.potekhin@gmail.com

ПОТЕХИН Вячеслав Витальевич – *доцент кафедры систем и технологий управления Санкт-Петербургского государственного политехнического университета, кандидат технических наук.*
195251, Россия, Санкт-Петербург, ул. Политехническая, д. 21.
E-mail: slava.potekhin@gmail.com

SEMENOV, Semen. N. *St. Petersburg State Polytechnical University.*
195251, Polytekhnicheskaya Str. 21, St. Petersburg, Russia.
E-mail: ussrcom@mail.ru

СЕМЕНОВ Семен Николаевич – *аспирант кафедры квантовой электроники Института физики, нанотехнологий и телекоммуникаций Санкт-Петербургского государственного политехнического университета.*
195251, Россия, Санкт-Петербург, ул. Политехническая, д. 21.
E-mail: ussrcom@mail.ru

UDC 621.396.62 =111

E.V. Balashov, A.S. Korotkov

FEEDBACK ULTRA-WIDEBAND AMPLIFIER WITH UNBALANCED INPUT AND BALANCED OUTPUT

A new type of UWB LNA presents. The amplifier has got a novel feedback topology and load circuit that allows the designer to obtain a high voltage gain and matching in the wide frequency range from 3.1 GHz to 10.6 GHz without cascading, current reuse technique and an additional matching circuit. The key point of the amplifier is the source degeneration with resistive shunt series feedback and the load circuit based on the combined parallel-series resonance circuit. The voltage gain of the amplifier is 9.7 dB, matching is better than −10 dB, the noise figure lies between 5.4 and 7 dB in the whole frequency range.

CMOS; LNA; ACTIVE BALUN; UWB.

Е.В. Балашов, А.С. Коротков

УСИЛИТЕЛЬ С ОБРАТНОЙ СВЯЗЬЮ, НЕБАЛАНСНЫМ ВХОДОМ И БАЛАНСНЫМ ВЫХОДОМ

Представлен новый тип сверхширокополосного усилителя. Усилитель построен на основе новой топологии обратной связи и цепи нагрузки, что позволяет увеличить коэффициент усиления по напряжению и полосу согласования в широком диапазоне частот от 3,1 до 10,6 ГГц без каскадирования, техники повторного использования тока *current reuse* и без дополнительных согласующих цепей. Ключевым схемотехническим решением, использованным при построении усилителя, является одновременное введение двойной отрицательной обратной связи: индуктивной последовательной по току и резистивной параллельной по напряжению, и цепи нагрузки на основе параллельного и последовательного резонансного контура. Коэффициент усиления по напряжению усилителя 9,7 дБ, согласование лучше чем −10 дБ, коэффициент шума находится между 5,4 и 7 дБ во всем диапазоне частот.

МАЛОШУМЯЩИЙ УСИЛИТЕЛЬ; СВЕРХШИРОКОПОЛОСНАЯ РАДИОСВЯЗЬ.

Ultra-wideband (UWB) is a wireless transmission scheme that occupies bandwidth more than 500 MHz. The use of such bandwidth provides the possibility to reach high data rates for short ranges with less power consumption in comparison with conventional narrowband systems. The UWB signal can occupy bandwidth of 3.1−10.6 GHz. One of the fascinating technologies for implementation of the UWB transceiver is CMOS technology [1]. It provides a way to place RF, base-band and digital parts of the receiver on the same chip. The use of system-on-a-chip (SOC) conception allows the designer to reduce the price of the device. The UWB transceiver is formed as a combination of receiver and transmitter parts [2]. Low noise amplifier is one of the most important blocks of the receiver part as it determines the noise figure and the matching of the receiver to an antenna. The design of the LNA is a difficult task because it should strike a compromise between different characteristics of the amplifier such as power consumption, input matching, gain, noise figure, linearity and stability.

The paper describes a new approach to design the broadband low noise amplifier. This approach makes it possible to design the amplifier without techniques such as cascading and current reuse. The idea of the approach is based on matching of the common source amplifier to source degeneration and resistive shunt series feedback. The broadening of the gain bandwidth with load circuit involves using of a parallel resonance circuit with an additional series resonance circuit. Basic amplifier topologies are presented in the second section. Different feedback topologies, matching techniques and loads are also briefly described in this section. The third section describes a novel broadband inductively degenerated common-source amplifier with shunt-series feedback and two resonance load architecture. The results of simulation are presented in the third section. The paper ends with the conclusion and the discussion of the results.

Wideband Amplifier Architecture

The amplifier should provide impedance matching of the signal source R_s (usually equal to 50 Ohm) to the combination of high-voltage gain A_V. As the CMOS amplifier is usually used in direct conversion or low IF receivers, the output of the amplifier is connected with the next receiver stage (e. g. mixer) that has high-resistance input impedance. So the CMOS amplifiers dispense with the output matching. The matching problem can be solved using different feedback and/or matching circuits. To realize a high-voltage gain the design of the appropriate load circuit is used. The circuit allows reaching high-voltage gain without additional power consumption due to effective conversion of the output current to the output voltage. The cascading leads to high power consumption [3]. The power consumption can be minimized by a current reuse technique [4]. Thus, the cascading and current reuse methods are techniques which can improve the performance of the amplifier. The amplifier presented in this work can be considered as a basic block of any LNAs. It means that the cascading and current reuse methods can be applied to this circuit as well. This is the reason why the methods will not be especially discussed. The UWB LNA can be built using common-gate or common-source circuits which are shown in Fig. 1.

The noise factor of common-gate amplifiers depends on matching conditions and has the theoretical value of 2.2 dB. In practice the minimum noise figure of this LNA type is about 3.3—4.4 dB [3, 5, 6]. Some serious disadvantage of the common-gate LNA is its relatively low transconductance that can't provide the low noise figure and high gain in a whole frequency band. So this type of the amplifier is usually used when the bandwidth does not exceed 2 GHz [5] or as the first stage of multi cascade

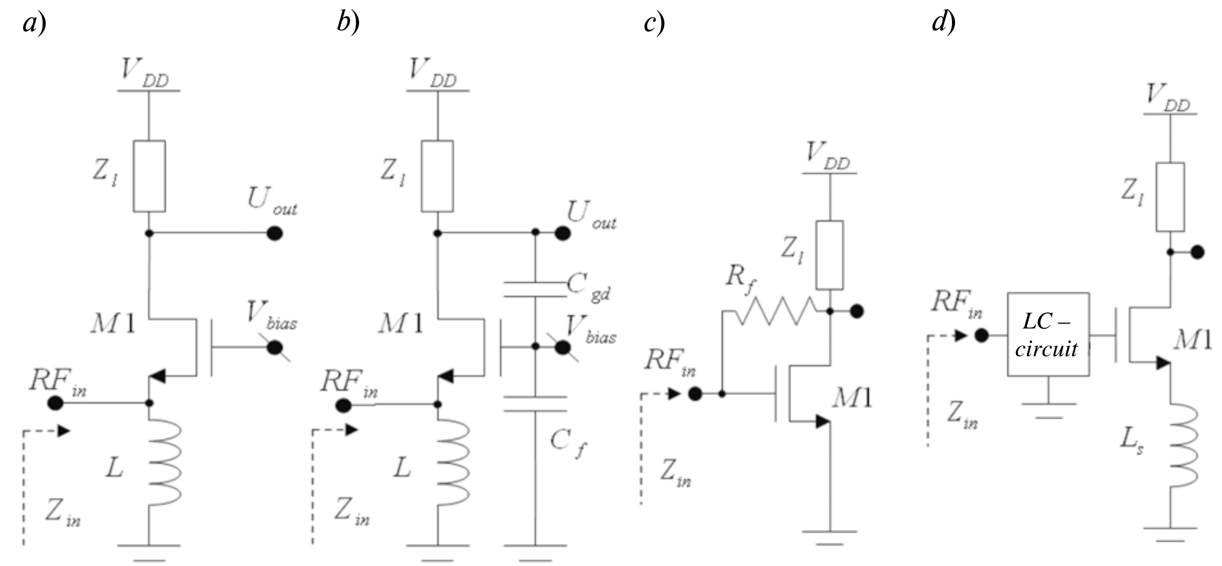


Fig. 1. The UWB LNA topologies:
a — the common-gate amplifier without feedback; *b* — the common-gate amplifier with voltage-voltage feedback;
c — the common-source amplifier with shunt-series feedback; *d* — the common-source inductively degenerated amplifier

amplifiers where the next stages enhance the amplification bandwidth [3, 6].

The additional feedback can be used in common-gate amplifiers to dispose the dependence between the noise factor and match conditions. The common-gate amplifier can provide the real impedance equal to 50 Ohm with the higher value of transconductance by means of the voltage-voltage feedback using capacitive divider (Fig. 1 *b*) [7]. The capacitance divider is formed by capacitor parasitic gate-drain capacitance C_{gd} of the transistor *M*1 and the feedback capacitance C_f. But the frequency range of this circuit does not cover all frequency range since the voltage-voltage feedback is frequency dependent that leads to decrease of the real part of the input impedance with frequency. Another drawback of the circuit is attenuation of the input signal due to the capacitive divider formed by the gate-source parasitic capacitance of the transistor *M*1 and capacitance C_f. Thus the common-gate amplifier with capacitance feedback covers the frequency range only from 3 to 8 GHz.

Common-source stages can be used in LNAs with the so-called noise cancellation structure [8—10] or with an additional feedback circuit [11—13]. The amplifiers with noise cancellation consume much power due to the circuits consist of several stages. The noise figure is quite high due to the noise cancellation in a broad frequency range and element mismatches. The power consumption of these amplifiers is not less than 14 mW while the noise figure is not less than 4.5 dB.

Another way to reach the matching is to add shunt-series feedback (Fig. 1 *c*) using resistor R_f to common-source amplifiers [12]. But the influence of the parasitic capacitance at the output leads to increase of the real part of the input impedance $\mathrm{Re}\{Z_{in}(s)\}$ with frequency. At the same time the input capacitance leads to decrease of the value $\mathrm{Re}\{Z_{in}(s)\}$. So the frequency bandwidth of the amplifier matching is broadened with high value of transconductance despite the frequency dependence of A_V. Unfortunately the bandwidth of matching is usually limited near 5 GHz that is not enough to cover all UWB frequency range. So this type of amplifier can be used at the first stage in multi stage amplifiers due to considerable power consumption [12, 13].

The other way of achieving a broadband input matching is inductively source degenerated cascode LNA. Usually this LNA can be combined with an additional input bandpass LC-filter that neutralizes the reactive part of the input impedance across the wide frequency range [4, 14, 15] (Fig. 1 *d*). The inductively degenerated common source amplifier is obtained by adding the inductor L_s between the transistor *M*1 source and ground that forms a series resonance circuit together with parasitic gate-source capacitance C_{gs} of the transistor *M*1. This series resonance circuit also consists of resistive impedance that is equal to $g_m L_s / C_{gs}$, where g_m is transconductance of the transistor *M*1. The LC-filter at the input consists of a number of reactive elements, which can lead to larger chip area and noise figure degradation in the case of its on-chip implementation, or can lead to additional external components. Also this method suffers from the influence of parasitic capacitances and inductances due to packaging that can not be appreciated beforehand.

The input matching and the flat gain in the wide frequency band can be obtained using the so-called distributed amplifiers as well [16—18]. But it is not possible to use the distributed amplifier in a portable application due to its large power consumption (sometimes it is more than 100 mW). Also its noise figure is quite high. Sometimes it can be more than 8 dB.

Inductively Degenerated Common-Source Amplifier with Shunt-series Feedback and Two Resonance Load Circuit

Designing a low noise amplifier for UWB radio in the frequency range of 3.1—10.6 GHz based on CMOS technology needs choosing a feedback configuration of the LNA that can provide simultaneously good input matching, high gain and low noise. The proposed circuit based on the source degenerated amplifier with shunt series feedback. As it is shown in Fig. 2, the common source amplifier with shunt series feedback provides matching in a low frequency range (Fig. 2 *a*) and the source degenerated amplifier provides matching at the frequency band near the resonant frequency of the series tank by inductance L_g and gate-source

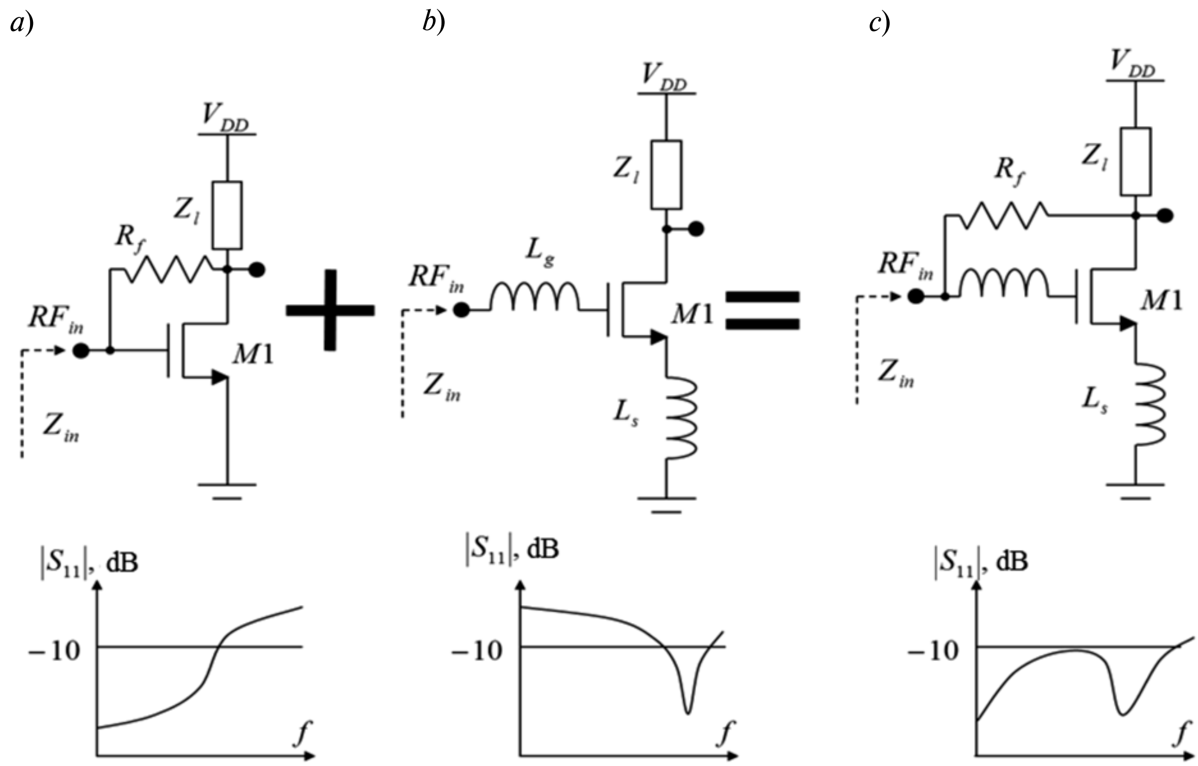


Fig. 2. The amplifier topologies and its *S*11 parameter absolute value vs. frequency:
a — the common-source amplifier with shunt-series feedback; *b* — the common-source inductively degenerated amplifier; *c* — the proposed common-source amplifier with dual feedback

capacitance of transistor *M*1 (Fig. 2 *b*). So the combination of two types of feedback can considerably broaden the frequency band of the amplifier because the matching band broadens as well (Fig. 3 *c*).

So the matching task can be formulated as the following inequality using the reflection coefficient Γ with acceptable level −10 dB

$$|\Gamma| \approx |S_{11}| = \left|\frac{Z_{in}(s,\mathbf{M}) - R_s}{Z_{in}(s,\mathbf{M}) + R_s}\right| < -10\,\text{dB}, \quad (1)$$

where Z_{in} is the input impedance; R_s — source resistance; s — complex frequency; **M** is the vector of the element values.

This inequality should be fulfilled in the desired frequency range and can have a multitude of solutions. But in this case the solution that provides the maximum voltage gain, $A_{V\max}$, should be chosen. So the (1) should be solved together with the next condition:

$$A_{V\max} \approx |S_{21}(\mathbf{M}_{opt})| = \max|S_{21}(s,\mathbf{M})|, \quad (2)$$

where $\mathbf{M}_{opt}$ is the vector of element values which leads to maximum of the voltage gain. The symbolic solution of this set is not possible. Therefore the element values should be obtained using numerical optimization.

Let us consider the full circuit of the amplifier that was designed using the proposed feedback topology. The described procedure provides the possibility to obtain the parameters of the amplifier circuit shown in Fig. 3 *a*. The first stage of the circuit is the amplifier with a dual feedback; the second stage is an active balun with unity voltage gain. The amplifier is designed to work on the capacitance load of the next stage.

It is based on the cascode stage that consists of the transistors *M*1—*M*3. The equivalent width of transistors pair *M*1, *M*2 is two times as large as transistor *M*3 to increase transconductance of transistor pair *M*1, *M*2 and minimize the gate-source capacitance of *M*3. The source degeneration inductance is formed by a parallel connection of the planar inductor *L*1 and *L*2 to decrease the total inductance to cover large frequency band. The

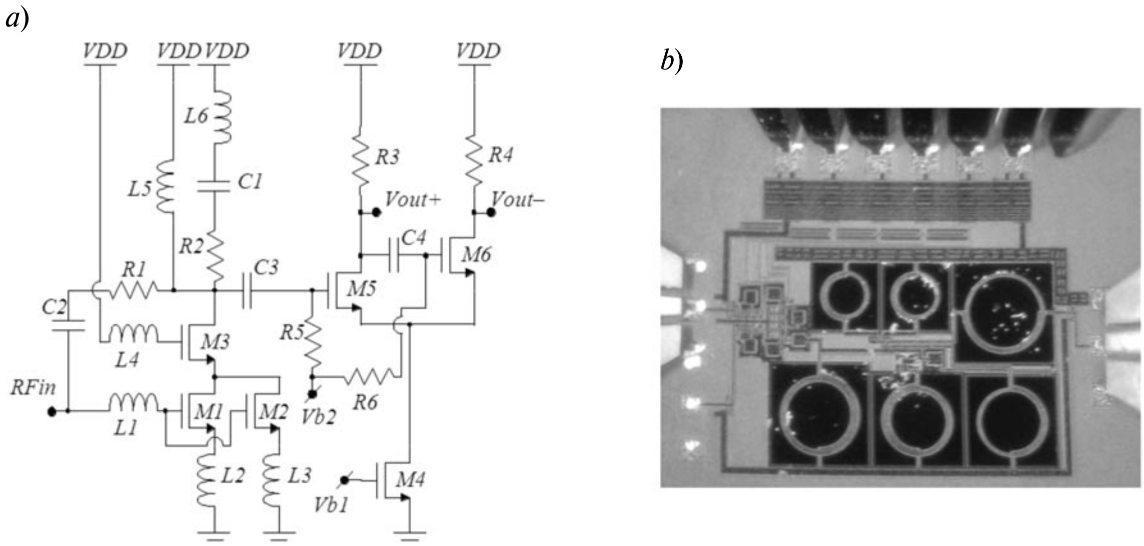


Fig. 3. The proposed feedback ultra-wideband lownoise amplifier:
a — circuit; *b* — chip photography through a microscope

planar inductor *L*1 resonates with the gate-source capacitance of transistors *M*1 and *M*2. The shunt-series resistor *R*1 provides matching in low frequency range. The planar inductor *L*4 boosts the transconductance of the transistor *M*3 in the high frequency range. The circuit load is implemented by means of elements *L*5, *L*6, *C*1, *R*2 to provide flat voltage gain in the whole frequency range. The load is based on inductor *L*5 that resonates with the parasitic capacitance *C* at the output of the first stage of the amplifier. In contrast to the different bandwidth extension techniques, for example described in [19], the inductor *L*5 with the series *RLC*-circuit constructed from elements *L*6, *C*1, *R*2 are used to increase the voltage gain in desired frequency band that is far from zero frequency. This series circuit introduces the second resonance frequency of the load impedance. In addition, *L*5 allows biasing the amplifier without loosing of voltage swing. The transistors *M*4—*M*6 and resistors *R*3—*R*4 form the active balun to convert the single ended signal to differential form. The other passive elements are used for biasing and coupling of ac and dc signals. The load is based on inductor *L*5 that resonates with the parasitic capacitance *C* at the output of the first stage of the amplifier. In contrast to the different bandwidth extension techniques, for example described in [19], the inductor *L*5 with the series *RLC*-circuit constructed from elements *L*6, *C*1, *R*2 are used to increase the voltage gain in desired frequency band that is far from zero frequency. This series circuit introduces the second resonance frequency of the load impedance.

The LNA design procedure should include the influence of the passive element parasitics, because the parasitic elements, especially inductors, cause performance degradation effects of the circuit at high frequency range. So the optimization incorporated in circuit simulation programs should be used. The parameters of optimization are the values and the geometric sizes of passive elements. The optimization task is formulated as follows:

maximization of $|S_{21}|_{max}$ in the frequency range 3.1—10.6 GHz;

fulfillment of $|S_{11}|_{max}<-10$ dB in the frequency range 3.1—10.6 GHz;

fulfillment of the gain flatness $|S_{21}|_{max}-|S_{21}|_{min}<3$ dB in the frequency range 3.1—10.6 GHz, where the $|S_{21}|_{max}$, $|S_{21}|_{min}$ are the highest and lowest absolute values of S_{21} consequently, $|S_{11}|_{max}$ is the highest absolute value of S_{11} in desired frequency range.

The amplifier elements were calculated during the optimization and simulation as equivalent subcircuits. These equivalent subcircuits describe the elements in frequency range till 20 GHz. The amplifier has got the following simulated characteristics: the reflection coefficient is less than −10 dB in the frequency band from 3.1 to 10.6 GHz, the maximum absolute value of the voltage gain is 9.7 dB and the noise figure is ranging from 5.4 to 7.0 dB. The first stage consumes 4.2 mA from the 1.8 voltage source, the second one − 12 mA. The element parameters and characteristics were obtained by optimization and simulation using the Cadence Virtuoso custom design IC platform.

As a result all transistors *M*1−*M*5 were realized as a multi finger structure with an equivalent dimension of the *W*/*L* equal to 105 um/0.18 um. The inductors *L*1, *L*2, *L*3, *L*4, *L*5, *L*6 are spiral planar inductors. The inductors *L*2 and *L*3 have turn number 1.5 and the internal diameter 126 um with the inductance about 0.57 nH. The inductor *L*1 has the inductance about 1.3 nH by turn number 1.5 and internal diameter 227 um. The inductor *L5* and *L*6 have inductances 2.9 and 2.0 nHn consequently by turn number 2.5 and internal diameter 210 um and 200 um consequently. The feedback resistor *R*1 with resistance 300 Ohm and load resistor *R*2 with resistance 200 Ohm are P+ poly resistors. The resistors *R*3 and *R*4 are 50 Ohm P+ poly resistors.

The described amplifier was implemented using UMC CMOS 180 nm as a standalone chip and its chip photography through a microscope is shown on Fig. 3 *b*. For the measurement purpose the ESD protection circuit and two buffers with the gain minus 6 dB were added on the outputs of the amplifier to drive 50 Ohm external load.

The results of the circuit simulation of the chip using the Cadence Virtuoso custom design IC platform, on-wafer measurement using network analyzer Rohde & Schwarz ZVA40, probe station Cascade Microtech EP6RF are shown in Fig. 4.

The parameter |*S*11| is presented in Fig. 4 *a*. As we can see |*S*11| is less then −10 dB in the frequency range 3.0−9.8 GHz that is in a good accordance with the results of simulations. Its minimum is on the frequency of 8 GHz and is equal to −25 dB. The parameter |*S*21| (Fig. 4 *b*) defines the gain of the amplifier. Its maximum value is near 3.4 dB and lies in the frequency 3 GHz. The experimental results show that the matching bandwidth is on the frequency range of 2.4−8.4 GHz by the level −20 dB with the minimum value of |*S*11| −30 dB

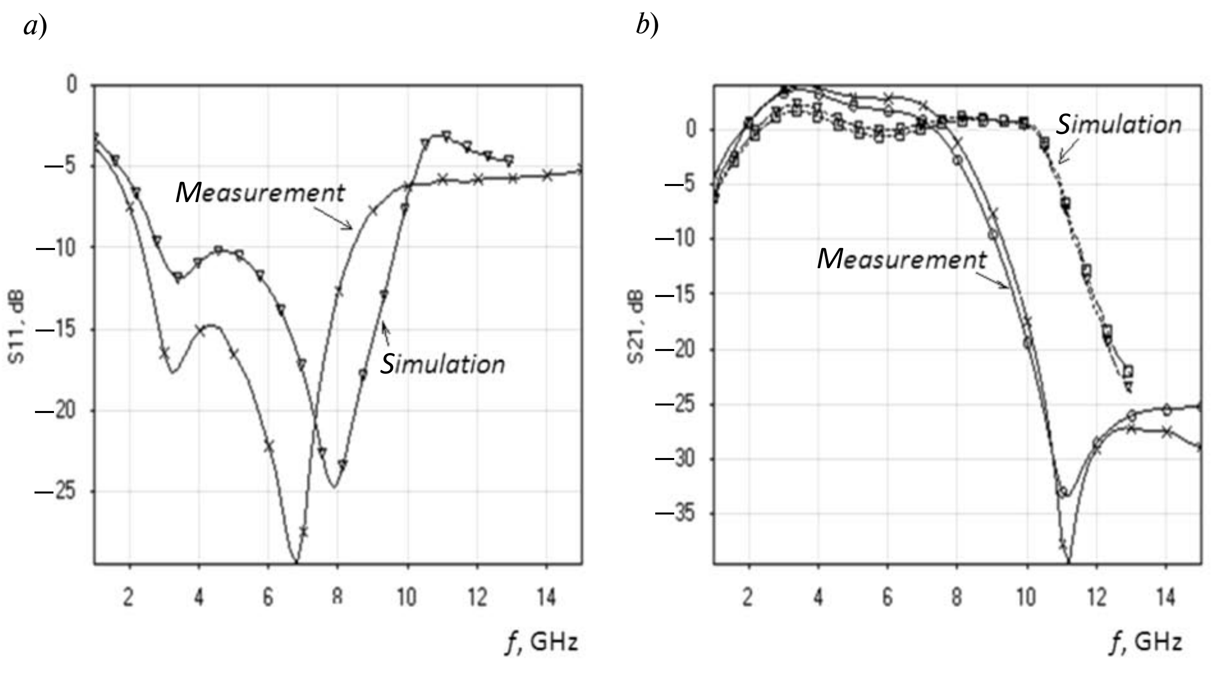


Fig. 4. The measurement and simulation results:
a − the dependence of |S11| vs. frequency; *b* − the dependence of |S21| vs. frequency on the balanced outputs

on the frequency 7.0 GHz. The experimental results show that the maximum of |*S*21| parameter is 4 dB on the frequency 3 GHz and the amplification bandwidth is 2.0–7.5 GHz.

The difference between the measurement and simulation is the result of technological parameter dispersions. Although the parameters |*S*11| and |*S*21| are close to simulations in the frequency band till 8 GHz, it confirms the efficiency of the proposed scheme. It also shows that the gain would be 6 dB higher without matching buffers how it was predicted by simulations. The |*S*11| dependence vs. frequency character with to minimum on the graph (Fig. 4 *b*) that was predicted (expected) also confirms the efficiency of the proposed feedback architecture of the amplifier. As we can see, the |*S*21| dependence vs. frequency measured on the balanced outputs of the amplifier is quite close to each other (Fig. 4 *b*). Thus the use of the active balun is an efficient solution to obtain a balanced signal that is necessary for the frequency conversion at the next stage.

In this paper we come to the conclusion that the proposed LNA differs significantly from the amplifiers published in recent years. The described method is based on the use of the shunt-series feedback to match the LNA in low frequency band and the use of the inductive degeneration to match the LNA in high frequency band. The active balun is used to convert the unbalanced signal to the balanced form. The parameters of the LNA are compared with parameters of low noise amplifiers described earlier in the literature. So the amplifier has got the following characteristics: the reflection coefficient is less than required value –10 dB in the frequency band from 3.1 to 10.6 GHz, the maximum value of the voltage gain is 9.7 dB and the noise figure is ranging from 5.4 to 7.0 dB. The first stage consumes 4.2 mA, while the second stage consumes 12 mA from the 1.8 voltage source. Thus, the results proved by the comparison of simulation and experimental results show that the simulation and measurement are quite close in frequency band till 8 GHz.

REFERENCES / СПИСОК ЛИТЕРАТУРЫ

1. **Abidi A.A.** RF CMOS comes of age. *IEEE Journal of Solid-State Circuits*, 2004, Vol. 39, pp. 549–561.

2. **Lu I.S.-C.**, **Weste N.**, **Parameswaran S.** A digital ultra-wideband multiband transceiver architecture with fast frequency hopping capabilities. *In Proc. IEEE Conference on Ultra-wideband Systems and Technologies Conference*, Piscataway, NJ, USA, 2003, pp. 448 – 452.

3. **Lu Y., Yeo K.S., Cabuk A., Ma J., Do M.A., Lu Zh**. A novel CMOS low-noise amplifier design for 3.1- to 10.6 GHz ultra-wideband wireless receivers. *IEEE Trans. on Circuits and Systems – I.* 2006, No. 8, Vol. 53, pp. 1683–1692.

4. **Lin Y.-J.**, **Hsu Sh. S. H.**, **Jin J.-D.**, **Chan C.Y.** A 3.1-10.6 GHz ultra-wideband CMOS low noise amplifier with current-reused technique. *IEEE Microwave and Wireless Components Letters.* 2007, Vol. 17, No. 3, pp. 232–234.

5. **Razavi B. et al.** A UWB CMOS Transceiver. *IEEE Journal of Solid-State Circuits.* 2005, Vol. 40, No. 12, pp. 2555–2562.

6. **Chen K.-H., Lu J.-H., Chen B.-J., Liu Sh.-I.** An Ultra-wideband 0.4–10-GHz LNA in 0.18-um CMOS. *IEEE Transactions on Circuits and Systems – II.* 2007, Vol. 54, No. 3, pp. 217–221.

7. **Cusmai G., Brandolini M., Rossi P., Svelto F.** A 0.18-um CMOS selective receiver front-end for UWB applications. *IEEE Journal of Solid-State Circuits*, 2006, Vol. 41, No. 8, pp. 1764–1771.

8. **Liao Ch.-F., Liu Sh.-I.** A broadband noise-canceling CMOS LNA for 3.1-10.6 GHz UWB receivers. *IEEE Journal of Solid-State Circuits*, 2007, Vol. 42, No. 2, pp. 329–339.

9. **Wang Ch.-Sh., Wang Ch.-K.** A 90nm CMOS low noise amplifier using noise neutralizing for 3.1-10.6 GHz UWB system. In Proc. *IEEE Conference on Solid-State Circuits*, 2006, pp. 251–254.

10. **Qiang Li, Yue Ping Zhang.** A 1.5-V 2–9.6-GHz inductorless low-noise amplifier in 0.13-um CMOS. *IEEE Journal of Solid-State Circuits*, 2007, Vol. 55, No. 10, pp. 2015–2033.

11. **Ch.-Y. Wu, Fadi Riad Shahroury**. A low-voltage CMOS LNA design utilizing the technique of capacitive feedback matching network. *In Proc. IEEE Conference on Electronics, Circuits and Systems*, 2006, pp. 78–81.

12. **Gharpurey R.** A broadband low-noise front-end amplifier for ultra-wideband in 0.13-mm CMOS. *IEEE Journal of Solid-State Circuits*, 2005, Vol. 40, No. 9, pp. 1983–1986.

13. **Chen Sh.-Ch., Wang R.-L., Kung M.-L., Kuo H.-Ch.** An integrated CMOS low noise amplifier for 3-5 GHz UWB applications. *In Proc. IEEE*

Conference Electron Devices and Solid-State Circuits, 2005, pp. 225–228.

14. **Bevilacqua A., Niknejad A.M.** An ultrawideband CMOS low-noise amplifier for 3.1-10.6 GHz Wireless Receivers. *IEEE Journal of Solid-State Circuits*, 2004, Vol. 39, No. 12, pp. 2259–2268.

15. **Wang Y., Iniewski K.** A low power CMOS low noise amplifier for 3-10 GHz Ultra-wideband wireless receivers. *In Proc. 49th IEEE Internat. Midwest Symp. on Circuits and Systems*, 2006, pp. 353–357.

16. **Ballweber B.M., et al.** A fully integrated 0.5-5 GHz CMOS distributed amplifier. *IEEE Journal of Solid-State Circuits*, 2000, Vol. 35, No. 2, pp. 231–239.

17. **Ahn H.-T., Allstot D.J.** A 0.5-8.5 GHz fully differential CMOS distributed amplifier. *IEEE Journal of Solid-State Circuits*, 2002, Vol. 37, No. 8, pp. 985–983.

18. **Zhang F., Kinget P.R.** Low-power programmable gain CMOS distributed LNA. *IEEE Journal of Solid-State Circuits*, 2006, Vol. 41, No. 6, pp. 1333–1343.

19. **Shekhar S., Walling J.S., Allstot D.J.** Bandwidth extension techniques for CMOS amplifiers. *IEEE Journal of Solid-State Circuits*, 2006, Vol. 41, No. 11, pp. 2424–2439.

20. **Kim Ch.-W., Kang M.- S., Anh Ph.T., Kim H.-T., Lee S.-G.** An ultra-wideband CMOS low noise amplifier for 3-5 GHz UWB system. *IEEE Journal of Solid-State Circuits*, 2005, Vol. 40, No. 2, pp. 544–547.

BALASHOV, Evgeny V. *St. Petersburg State Polytechnical University.*
195251, Polytekhnikheskaya Str. 29, St. Petersburg, Russia.
E-mail: balashov_ev@mail.ru

БАЛАШОВ Евгений Владимирович – *доцент кафедры интегральной электроники Института физики, нанотехнологий и телекоммуникаций Санкт-Петербургского государственного политехнического университета, кандидат технических наук.*
195251, Санкт-Петербург, ул. Политехническая, д. 29.
E-mail: balashov_ev@mail.ru

KOROTKOV, Alexander S. *St. Petersburg State Polytechnical University.*
195251, Polytekhnicheskaya Str. 29, St. Petersburg, Russia.
E-mail: korotkov@rphf.spbstu.ru

КОРОТКОВ Александр Станиславович – *заведующий кафедрой интегральной электроники Института физики, нанотехнологий и телекоммуникаций Санкт-Петербургского государственного политехнического университета, доктор технических наук, профессор.*
195251, Санкт-Петербург, ул. Политехническая, д. 29.
E-mail: korotkov@rphf.spbstu.ru

UDC 621.372.54=111

N.V. Ivanov, A.S. Korotkov

A 180-nm CMOS HIGH-LINEAR COMPLEX G_m-C FILTER FOR RECEIVERS OF SATELLITE NAVIGATION SYSTEMS

Realization of high-linear G_m-C complex filter in 180-nm CMOS technology to be used in satellite navigation system receivers is covered. The filter with 25 MHz bandwidth was designed with THD of about 80 dB.

G_m-C FILTERS; COMPLEX FILTERS; SATELLITE NAVIGATION SYSTEMS RECEIVERS; TRANSCONDUCTANCE AMPLIFIER; 180-nm CMOS.

Н.В. Иванов, А.С. Коротков

КОМПЛЕКСНЫЙ G_m-C ФИЛЬТР С НИЗКИМ УРОВНЕМ НЕЛИНЕЙНЫХ ИСКАЖЕНИЙ НА ОСНОВЕ КМОП-ТЕХНОЛОГИИ С РАЗРЕШЕНИЕМ 180 НМ ДЛЯ ПРИЕМНИКОВ СИСТЕМ СПУТНИКОВОЙ НАВИГАЦИИ

Рассмотрена реализация G_m-C комплексного фильтра с низким уровнем нелинейных искажений для применения в приемниках систем спутниковой навигации. Рассчитан фильтр с полосой пропускания 25 МГц с уровнем третьей гармоники порядка –80 дБ.

G_m-C ФИЛЬТРЫ; КОМПЛЕКСНЫЕ ФИЛЬТРЫ; ПРИЕМНИКИ СИСТЕМ СПУТНИКОВОЙ НАВИГАЦИИ; ТРАНСКОНДУКТИВНЫЙ УСИЛИТЕЛЬ; КМОП 180 НМ

Complex G_m-C filters are widely used for channel selection in the front-end of on-chip receivers. The advantages of G_m-C filters, as compared to conventional active-RC filters and switched capacitor filters, include a higher operating rate and lower power consumption. However, high linearity is a problem for this type of filter structures. It is mostly caused by the characteristics of a transconductance amplifier (TA). If the linearity problem is solved, the advantage of a high operating rate could be of use in satellite navigation system receivers, e. g. GPS and GLONASS systems, where the channel passband in a broadband mode is about 25 MHz.

This paper presents the implementation of a well-known technique which realizes the G_m-C complex filter, described in [1]. It was used for significantly low frequencies – Bluetooth and ZigBee applications. The technique is based on the frequency shift of low-pass filter amplitude response by implementing interconnections. It should be noted that there are just a few known TAs designed for a frequency range of satellite navigation systems. The first part of the paper is devoted to the problem of high-linear TA realization suitable for the proposed application. The second part describes the proposed filter realization. The third part discusses the problem of filter tuning.

Transconductance Amplifier Realization

Linearization techniques for TAs were properly discussed in [2]. Most common are adaptive feedback technique, source degeneration technique and compensation technique. Even without considering the G_m/I ratio, which is a factor of efficiency for TAs, it was shown that the compensation technique, first proposed by authors of [3], is optimal for the high-linear TA design in terms of THD level in equal conditions. Though lower THD could be achieved with a source degeneration technique, it is relevant in exchange for a lower G_m/I ratio, which leads to higher consumption for the same G_m value.

The proposed TA structure is shown in Fig. 1 *a*. It includes saturation region amplifier section, M1-M4, triode region amplifier section M5-M8, feedforward circuit M9-M14

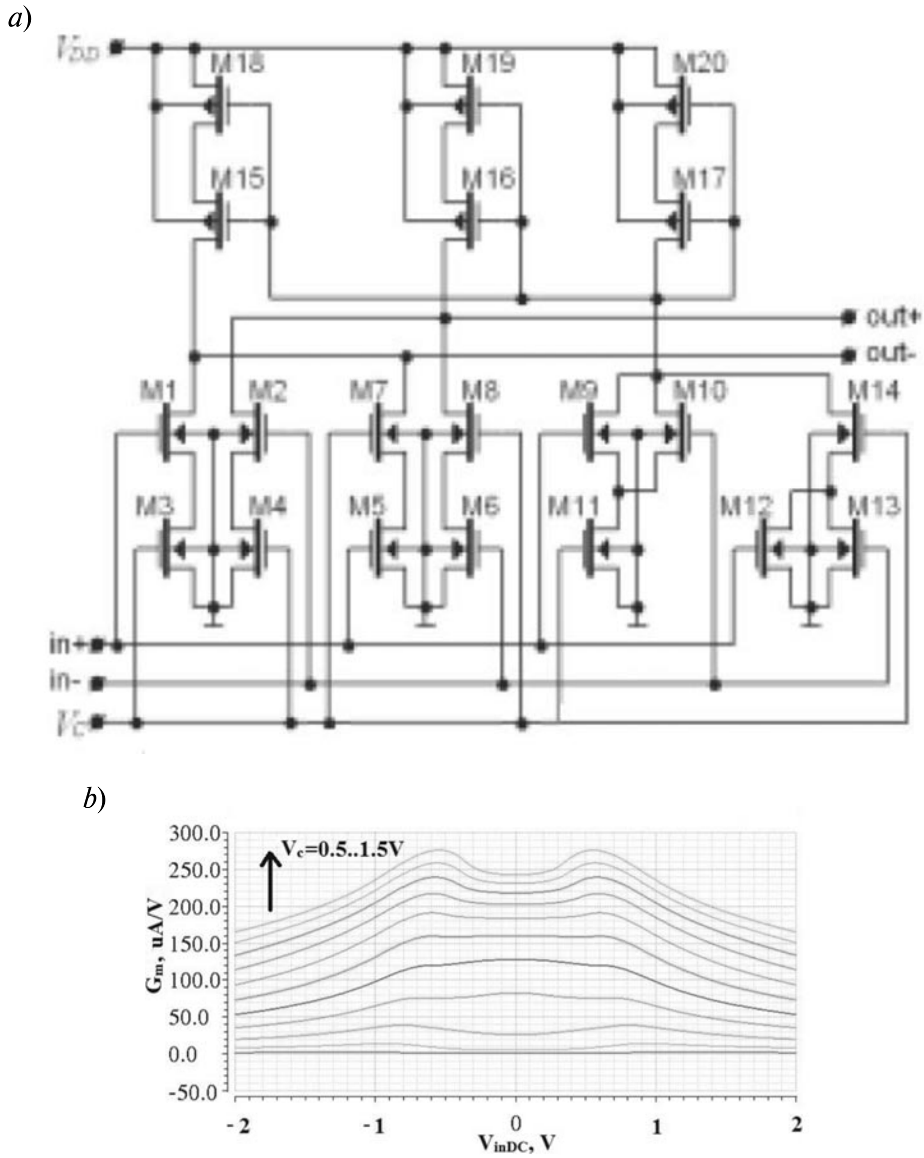


Fig. 1. Transconductance amplifier structure (*a*); DC characteristic of proposed TA (*b*)

Table 1

Comparison of transconductance amplifiers

Transconductance amplifier	[3], 2002	[4], 2010	[5], 2012 (source degen-eration)	This work
Technology, nm	350	180	180	180
Supply voltage, V	2.5	1.8	1.8	1.8
THD, dB	>54 (1 V_{pp} 1 MHz)	51 (1 V_{pp} 50 MHz)	<50 (0.8 V_{pp} 10 MHz)	59.22 (1 V_{pp} 50 MHz)
G_m value, uA/V	68	120–287*	140–340*	159 (130–200)*

* Variation due to V_c change for the fixed linear input amplitude range.

and current mirror as dynamic load M15-M20. Using a parametric analysis in Cadence IC 6.15 for UMC_180 library transistors P_18_MM and N_18_MM, supply voltage V_{dd} of 1.8V and control voltage V_c of 1V, the TA with G_m value of 159 uA/V was designed. The transistor length was set to 240 nm for better operating performance. Resulting transistors sizes are the following: M1-M4, M11: W = 1.95 um, M9-M10: W = 0.975 um, M7-M8, M14: W = 240 nm, M5-M6: W = 2 um, M12-M13: W = 1 um, M15-M20: W = 14.8 um.

In Fig. 1 *b* DC characteristic $G_m = f(V_{inDC})$, where V_{inDC} is the applied voltage difference at the input of the TA, is shown. 1 % G_m variation range is ±672 mV, which is nearly 75 % of the supply voltage limit. Table 1 compares characteristics of the proposed TA versus familiar analogues. The G_m/I ratio, which was mentioned only for [3], is 1.15 for the proposed amplifier. In [3] it was less than 1.

In [2], no information was given for the frequency at which THD was simulated. As it is, for lesser input amplitude of 0.5 V_{pp} THD was certainly less than 59 dB. For better comparison the power consumption and G_m/I ratio should be analysed, but most publications do not provide necessary information.

Complex Filter

The proposed filter structure is shown in Fig. 2 *a*. Number 1 marks the input TA, which realizes voltage-to-current conversion; 2 marks TAs in a resistor mode; 3 marks a block of gyrators, which realizes inductance imitators; 4 marks blocks of grounded capacitors; 5 marks blocks of interconnection gyrators G_{11}-G_{55} for complex reactance transform: each node of a frequency-dependent element, namely capacitor, is connected to its counterpart in other channel. The Frequency shift is defined by the equation:

$$\omega_0 = G_{ii} / C_i,$$

where G_{ii} – value of interconnection TA (in gyrator); C_i – value of the node capacitor.

Calculated values for the capacitors are the following:

$$C_1 = C_5 = 2.50 \text{ pF}, \; C_2 = C_4 = 6.55 \text{ pF},$$
$$C_3 = 8.10 \text{ pF}.$$

Taking into account the effect of parasitic capacitance, after the first simulation the capacitor values were recalculated:

$$C_1 = C_5 = 1.89 \text{ pF}, \; C_2 = C_4 = 4.96 \text{ pF},$$
$$C_3 = 6.13 \text{ pF}.$$

The values of the TAs needed for interconnection gyrators are the following:

$$G_{11} = G_{55} = 173 \text{ uA/V}, \; G_{22} = G_{44} = 453 \text{ uA/V},$$
$$G_{33} = 560 \text{ uA/V}.$$

These TAs were designed by the same method as the main TA. The characteristics do not change linearly with a proportional increase (or decrease) of transistor sizes. Using a parametric analysis, amplifiers with nearly the same linearity and G_m/I ratio were realized.

The complex filter amplitude response $|K(p)|$ is shown in Fig. 2 *b*. The dotted curve refers to a filter response before capacitors recalculation. The solid line refers to a final response of the designed filter. Table 2 compares complex filter characteristics. There are no available works where the bandwidth would be the same, but it is worth noting that even for lower frequencies and narrower bandwidth the linearity of the designed filter is significantly better.

Tuning

It is known that in G_m-C filters the tuning system requires the compensation of the technological variation of parameters. There are different approaches to the problem of building a tuning circuit, e. g. in accordance with [3], it could be a phase-locked loop with the G_m-based VCO and reference oscillator f_{ref}, as shown in Fig. 3 *a*. The idea is always in tuning by changing the control voltage V_c. Taking into account a significant difference in values of interconnection TAs, which could lead to a different change for the same control voltage change, it is imperative to check if the two tuning systems are required.

Fig. 3 *b* shows tuning characteristics of the filter. The control voltage V_c was changed simultaneously for all TAs of the filter structure from 0.9 V (dotted curve) to 1.1 V (dotted-dashed curve). The solid line refers to an initial value of 1 V. The frequency shift is not the

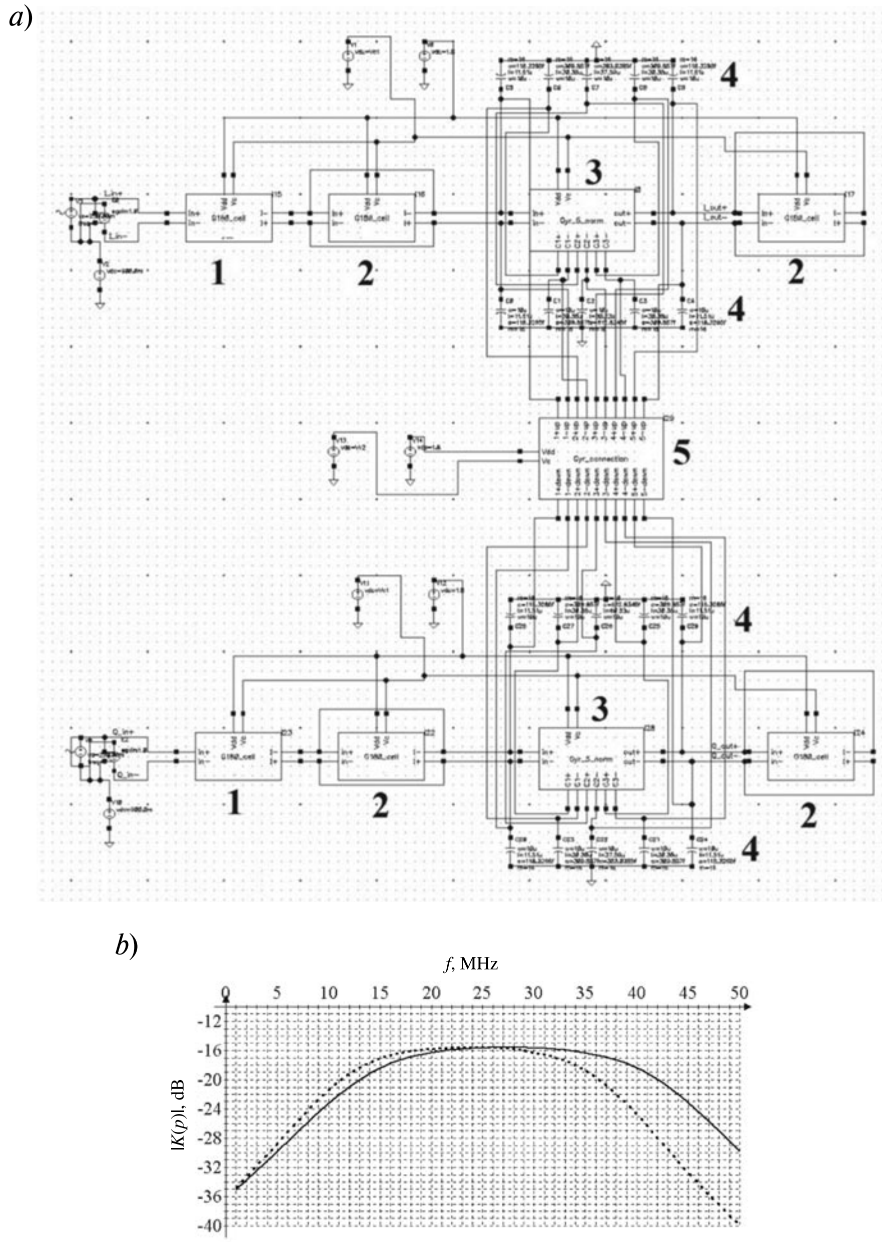


Fig. 2. G_m-C complex filter structure (*a*); frequency response of complex filter (*b*)

Table 2

Comparison of complex filter characteristics

Work	[6], 2006	[7], 2013	This work
Technology	G_m-C, 180 nm	ARC, 65 nm	G_m-C, 180 nm
Prototype	Butterworth, 6th order	6th order	Butterworth, 5th order
Center frequency, MHz	2	4–28	27.5
Bandwidth, MHz	3	4.2–18	25
Image rejection ratio, dB	55	> 31	> 47
Linearity	THD 45 dB (0.2 V_{pp} 2 MHz)	IIP3 25–28 dBm	THD >80 dB (1 V_{pp} 25 MHz) IIP3 33 dBm (100 Ohm)
Power consumption, mW	0.72	8.5–26	≈36

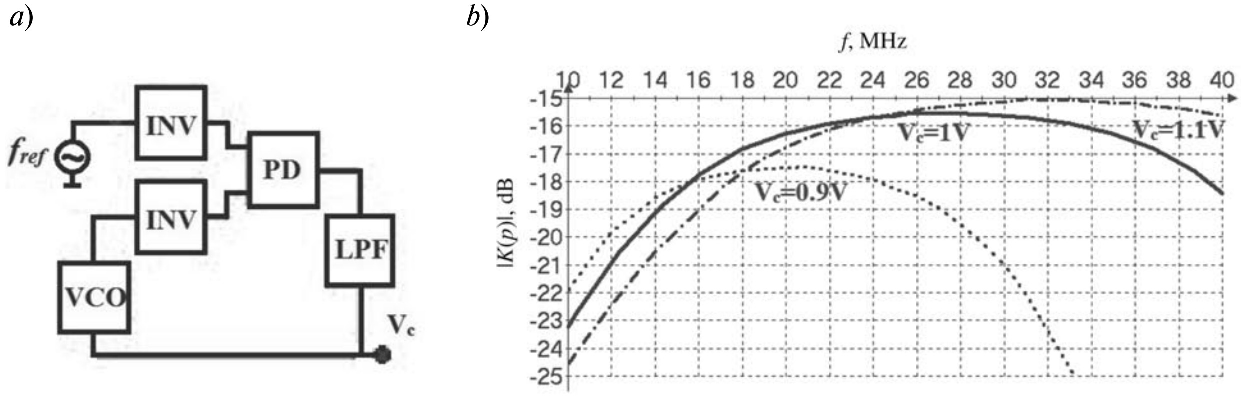


Fig. 3. Block diagram of tuning circuit [3] (*a*); tuning characteristic of proposed filter (*b*)

same for V_c = 1.1 V and V_c = 0.9 V, because of non-linear dependence of G_m versus control voltage, as could be seen from Fig. 1 *b*. A more important thing is that the relative change of the center frequency and bandwidth is not the same (23.63 and 28 %, respectively for 0.9 V and 12.73 and 12 % for 1.1 V). Depending on the target range of the frequency shift and bandwidth change, both the second frequency tuning system and automatic gain control circuit could be required.

A complex 180-nm CMOS G_m-*C* filter with high linearity was designed. The obtained linearity is more than 80 dB THD @1V_{pp} 25 MHz, for 25 MHz bandwidth filter centered at 27.5 MHz. The power consumption is about 36 mW for 5+5 order Butterworth complex filter. The achieved linearity in terms of IIP3 is at least 5 dBm higher. The problem of filter tuning was discussed. It was confirmed that for a relatively high frequency shift more than one tuning system would be required.

REFERENCES / СПИСОК ЛИТЕРАТУРЫ

1. **Phanumas Khumsat, Apisak Worapishet** Design of 1.8-V CMOS Polyphase Filter for Dual-Mode Bluetooth/Zig-Bee Transceiver. *ECTI Transactions on Electrical Eng., Electronics and Communications,* 2006, Vol. 4, No.1, pp. 29—34.

2. **Ha Le-Thai, Huy-Hieu Nguyen, Hoai-Nam Nguyen, Hong-Soon Cho, Jeong-Seon Lee, Sang-Gug Lee** An IF Bandpass Filter Based on a Low Distortion Transconductor. *IEEE Journal Solid-State Circuits*, 2010, Vol. 45, No. 11, pp. 2250—2261.

3. **Korotkov A.S., Morozov D.V., Hauer H., Unbehauen R.** A 2.5-V, 0.35-um CMOS transconductance-capacitor filter with enhanced linearity. *Proc. MWSCAS'02*, Tulsa, 2002, Vol. 3, pp. 141—144.

4. **Vijaya Bhadauria, Krishna Kant, Swapna Banerjee** A Tunable Transconductor With High Linearity. *Proc. IEEE Asia Pacific Conference on Circuits and Systems*, 2010, pp. 5—8.

5. **Shin-Jye Hsu, Che-Yu Lu, Chung-Chih Hung.** 40MHz *G*m-*C* Filter with High Linearity OTA for Wireless Applications. *Proc. Internat. Symp. on VLSI Design, Automation, and Test*, Apr. 2012, pp. 1—4.

6. **Ramin Zanbaghi, Mojtaba Atarodi, Armin Tajalli** An Ultra-Low Power G_m-C Complex Filter for Low-IF Wireless PAN Applications. *Proc. TENCON* 2006, pp. 1—4.

7. **Nan Qi, Zheng Song, Baoyong Chi Albert Wang, Tianling Ren Zhihua Wang.** A Multimode Complex Bandpass Filter With G_m-assisted Power Optimization and I/Q Calibration. *Proc. ISCAS* 2013', pp. 1845—1848.

IVANOV, Nikita V. *St. Petersburg State Polytechnical University.*
195251, Politekhnicheskaya Str. 29, St. Petersburg, Russia
E-mail: ivanovnick@mail.ru

ИВАНОВ Никита Валерьевич — *инженер первой категории кафедры интегральной электроники Санкт-Петербургского государственного политехнического университета, студент шестого курса.*
195251, Россия, Санкт-Петербург, ул. Политехническая, д. 29
E-mail: ivanovnick@mail.ru

KOROTKOV, Alexander S. *St. Petersburg State Polytechnical University.*
195251, Politekhnicheskaya Str. 29, St. Petersburg, Russia.
E-mail: korotkov@rphf.spbstu.ru

КОРОТКОВ Александр Станиславович – *заведующий кафедрой интегральной электроники Института физики, нанотехнологий и телекоммуникаций Санкт-Петербургского государственного политехнического университета, доктор технических наук, профессор.*
195251, Санкт-Петербург, ул. Политехническая, д. 29.
E-mail: korotkov@rphf.spbstu.ru

UDC 004.932.2=111

N.B. Ampilova, I.P. Soloviev, Yu.V. Shupletsov

ON FRACTAL, STATISTICAL AND MORPHOLOGICAL METHODS OF DIGITAL IMAGE ANALYSIS IN MEDICAL RESEARCH

The applicability of some methods of analysis and classification to the images of biomedical preparations is discussed. For images of blood, connective tissue and bone tissue fractal characteristics were calculated. When researching the images of blood obtained by the sensitive crystallization method, statistical, morphological and spectral signs were obtained, which were classified by a software modeled artificial neural network. The results of numerical experiments are given.

IMAGE ANALYSIS; FRACTAL METHODS; NEURAL NETWORKS; BIOMEDICAL PREPARATIONS.

Н.Б. Ампилова, И.П. Соловьев, Ю.В. Шуплецов

О ФРАКТАЛЬНЫХ, СТАТИСТИЧЕСКИХ И МОРФОЛОГИЧЕСКИХ МЕТОДАХ АНАЛИЗА ЦИФРОВЫХ ИЗОБРАЖЕНИЙ В МЕДИЦИНСКИХ ИССЛЕДОВАНИЯХ

Статья посвящена исследованию применимости методов анализа и классификации к изображениям биомедицинских препаратов: жидкой крови, соединительной и костной тканей. Для этих классов изображений вычислены фрактальные характеристики. Также исследованы изображения препаратов крови и экстрактов растений, полученных методом чувствительной кристаллизации. Для этих классов вычислены статистические, морфологические и спектральные признаки, которые затем классифицированы с помощью программно-моделируемой искусственной нейронной сети. Приведены результаты экспериментов.

АНАЛИЗ ИЗОБРАЖЕНИЙ; ФРАКТАЛЬНЫЕ МЕТОДЫ; НЕЙРОННАЯ СЕТЬ; БИОМЕДИЦИНСКИЕ ПРЕПАРАТЫ.

In biology and medicine the possibility of automatic processing large sets of images, determining the type of issue analyzed, revealing tumors and detecting some foreign compositions may have an influence on the research and to assist a physician to form a diagnosis [1].

For complex biological systems we as a rule do not have adequate mathematical models, but one can register the results of their functioning and obtain digital images. Such an approach is quite appropriate when studying the action of ultralow doses of a medicine on a biological system: since it is difficult to describe the results of the effect by using only one main factor (sign), we register the states of a vital functions sign, for example blood. Digital images of the registered states may be analyzed by mathematical methods. Hence, such an image may be considered as the phase portrait of the system studied at a point of time.

As a rule such digital images have rather complex structure and to analyze it the methods of fractal analysis are applied. The computation of fractal dimensions allows us to analyze and classify images of pharmacological preparations, living organism tissue, geological band fractures. In practice the images are unions of different fractals (so called multifractals), being every fractal has its own fractal dimension and all the fractal sets are interpenetrative. Hence the methods of multifractal analysis were worked out that allows obtaining a multifractal spectrum – a set of fractal dimensions.

The results of the research of different classes of biomedical preparations that is based on the using fractal methods is given in [2]. To compute the fractal dimension of an image (the Minkovsky dimension) the modified fractal signature method was used.

The concept of fractal signature and the method were introduced in [3] to analyze and classify textures. The method is based on the Mandelbrot idea about the approximate calculation of the length of coastline by measuring the area of a strip that contains the line and has width 2δ, where δ is a fixed number. Then the length is approximately the area divided by 2δ. It should be noted that the length increases as δ decreases. At the same time Mandelbrot noted that there is an interval for δ in which the value $L(\delta)$ becomes stable.

The authors applied this method to measure the area of a gray level surface constructed by a digital image. The sequence of special blankets is constructed over the surface. For every blanket its volume is calculated, the surface area and the so called «fractal signature» (the ratio of the logarithm of the surface area to the logarithm of the scale) is defined. (It should be noted that the Minkovsky dimension of the area may be easily obtained from the fractal signature.) So we have a sequence of areas and signatures in accordance with the number of blankets. For two images we can compare the obtained signature vectors: the closeness between vectors shows an adjacency of their textures. In [4] this method was applied to the analysis of images of bone tissue for different color scales.

In [5] the authors applied the method to calculate the fractal dimension of a document image and called it the «modified fractal signature method». They used only two consecutive blankets and marked out the areas corresponding background, text and graphics in accordance with the Minkovsky dimension. In such a variant the method demonstrated high reliability and was successfully used in [2] to classify biomedical preparation images. The image was divided into small boxes and the fractal dimension of both the area and boxes was calculated. That allowed us to obtain a «map» of the area and the dependence the surface area from the box size. The graphic of such a dependence was considered as a classification sign.

It is interesting to note that the method gave good results when classifying (defining a focusing degree) SAR (Synthetic Aperture Radar) images [6, 7].

As numerous researches show, the systems with chaotic behavior are very common in nature. Since in such a system one can trace a trajectory only on short time intervals, we should use an appropriate statistical description of the system asymptotic. It is the so called multifractal formalism that helps in deciding the problem. This technique is based on self-similarity property for an image, which gives a possibility to obtain the distribution of a probability measure. Such an approach allows us to associate the fractal properties of the image with the range of nonuniformity of the measure distribution. In the application to digital images to obtain such a distribution means to part the set of points of an image into subsets in which points have close characteristics. For a point one can use density defined as the limit (when the box size tends to zero) of the ratio of the logarithm of the box (containing the point) measure to the logarithm of the box side length. Thereby the image is a union of subsets containing the points with close densities. For each subset its fractal dimension is calculated. The set of these dimensions is called multifractal spectrum, which may be considered as the image characteristic. Numerical results showed that the multifractal spectrum method may be successfully applied to classify both bone tissues (health and affected by osteoporosis) and some classes of histological preparations [8].

It is reasonable that fractal and multifractal methods are applicable to definite classes of images. In medical diagnostic the images of preparations of blood and plant extracts obtained by sensitive crystallization by Pfaiffer method (the addition of a small dose of blood or plant extract to the solution of cuprum chloride) [9] are very important. It is well known that structure peculiarities of obtained blood crystals allow revealing both clinical presentations of various diseases and existence of definite tendencies to their progress. For such images fractal methods are not quite suitable. When analyzing the described images we construct the following feature vectors:

- statistical textural ones obtained by using gray level cooccurence matrix;
- morphological ones (skeleton representation);
- spectral ones (Gabor filters) [10].

Then these vectors were classified by a software modeled artificial neural network. To learn the network the RProp (Resilient Propagation) method was applied [11].

The Methods of Solution

The modified fractal signature method. In what follows we use the terminology of [12]. Let F be a nonempty bounded subset of R^n, Ω – a finite ε – covering of F and $N_\varepsilon(F) = |\Omega|$. Then the following number is called capacity dimension of F:

$$\dim_B F = \lim_{\varepsilon \to 0} \frac{\log_2 N_\varepsilon(F)}{-\log_2 \varepsilon}. \quad (1)$$

This dimension is a kind of so called box-computing (or box-counting) dimensions (the name follows from the definition). In practice the formula (1) is rarely applied because to calculate $N_\varepsilon(F)$ we have to save data about all the elements of the covering, whereas F may occupy only a part of it. Hence it is preferable to calculate the dimension that really is a box-counting one, but may be obtained by a more economic method.

We define δ-parallel body for F (denoted by F_δ) as the set of points which are on the distance no greater than δ from F:

$$F_\delta = \{x \in R^n : |x - y| \le \delta, y \in F\}. \quad (2)$$

Denote n-dimensional volume of F_δ by $\mathrm{Vol}^n(F_\delta)$. If for a constant D (when $\delta \to 0$) the limit of the ratio $\mathrm{Vol}^n(F_\delta) / \delta^{n-D}$ is positive and bounded, then D is called the Minkovsky dimension of the set F and denoted by $\dim_M F$. It is known [12] that the following relation holds (for nonempty bounded sets in R^n): $\dim_B F = \dim_M F$.

Now we consider $F = \{F_{ij},\ i = 0, 1, \ldots, K,\ j = 0, 1, \ldots, L\}$ – the presentation of a gray scale image, where F_{ij} is the intensity of the pixel with the coordinates (i, j). We redefine F in a point with real coordinates (x, y), where $i < x < i + 1, j < y < j + 1$ by the corresponding value F_{ij}. The function F specifies a surface in 3-dimensional space and is called the surface of gray level function. For brevity this surface is also denoted by F. To calculate fractal dimension of this surface the so called «blanket technique» is used. Construct for the surface F the δ-parallel body (blanket) with the help of the defined below top surface of blanket $u_\delta(i, j)$ and bottom surface $b_\delta(i, j)$. Set $u_0(i, j) = b_0(i, j) = F_{ij}$ and define surfaces for $\delta = 1, 2, \ldots$ by the following recurrent relation:

$$u_\delta(i, j) = \max\{u_{\delta-1}(i, j) + 1, \max_{|(m,n)-(i,j)| \le 1} u_{\delta-1}(m, n)\}, \quad (3)$$

$$b_\delta(i, j) = \min\{b_{\delta-1}(i, j) - 1, \min_{|(m,n)-(i,j)| \le 1} b_{\delta-1}(m, n)\}. \quad (4)$$

A point $M(x, y)$ is included in the δ-parallel body, if $b_\delta(i, j) < M(x, y) < u_\delta(i, j)$. It should be noted that in accordance with (3) and (4) the δ-parallel body for a given δ includes $(\delta - 1)$-parallel body. The volume of the δ-parallel body is computed from u_δ and b_δ:

$$\mathrm{Vol}_\delta = \sum (u_\delta(i, j) - b_\delta(i, j)). \quad (5)$$

The area A_δ of the fractal surface (the surface of the gray level function) is computed as

$$A_\delta = \frac{\mathrm{Vol}_\delta - \mathrm{Vol}_{\delta-1}}{2}, \quad (6)$$

and the fractal dimension of the surface (the Minkovsky dimension) is calculated by the formula

$$D \approx 2 - {\log_2 A_\delta}/{\log_2 \delta}. \quad (7)$$

As it was shown in [15], to estimate D one may use only two values of δ, namely δ_1 and δ_2 and the formula (7) is equivalent to the following relation:

$$D \approx 2 - \frac{\log_2 A_\delta - \log_2 A_\delta}{\log_2 \delta_1 - \log_2 \delta_2}. \quad (8)$$

When implementing the method we compute the area of the surface, areas of cells, *a*) *b*)

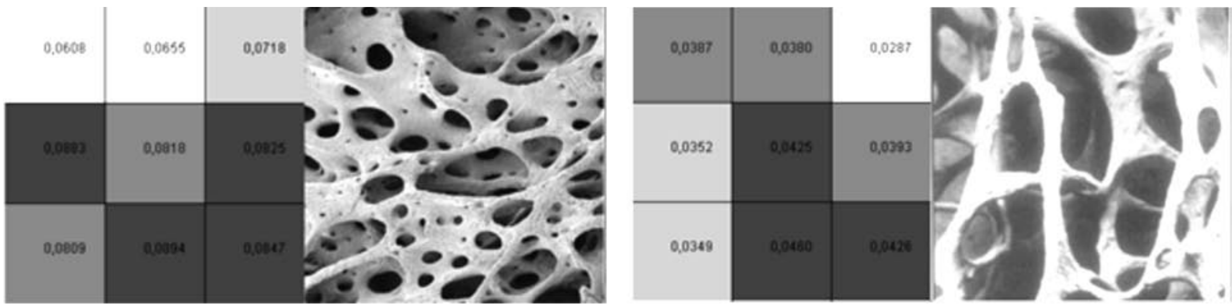


Fig. 1. Health bone tissue – *a*; affected bone tissue (osteoporosis) – *b*
The source image (right) and its map (left). The image size is 270×270, cell size is 90×90

fractal dimension for the surface and fractal dimensions of cells. We also construct a «map» of the image, where for a given cell size the cell areas were shown in a color scale. On Fig. 1 such maps are obtained for the cell size 90×90 (image size 270×270). The dependence both the surface area and the fractal dimension on cell size is shown on Fig. 2. We use such graphics as main features to compare the images of different classes. Generally speaking, it is sufficient to obtain both areas and dimensions only for the whole surface, but similar characteristics for the parts of the image give more obvious form to the results. It is the different fractal dimension values for different parts of an image that allowed the authors of [5] to distinguish text, graphics and background in text documents. In our experiments the method demonstrated good separability of features for all the classes of the preparations classified.

Computing multifractal spectrum. As it was mentioned above, we may consider an image as the support of some distributed measure. Cover this support by nonintersecting boxes (cells) $\{M_i\}$ (with intersections on the box boundaries) with the side length l and suppose that the box measure is l^{α_i}, where α_i are real numbers. Then one may define the point sets E_α (for a given α) such that α_i are close to α. The set $\{\alpha_i\}$ is a set of «densities» of the image points. For each point its density is computed through the box measure and the box side length [13]. Hence we collect the points with similar densities into the sets E_α. For all such sets we compute their fractal dimensions $f(\alpha_i)$. By this means the image is considered as the union of interwoven subsets, being each of them has own fractal dimension. The set of these dimensions forms multifractal spectrum (MFS).

We note that locating of sets E_α means some categorization of the image points, and the defined measure describes this categorization (for digital images measure is naturally defined by

a) *b*)

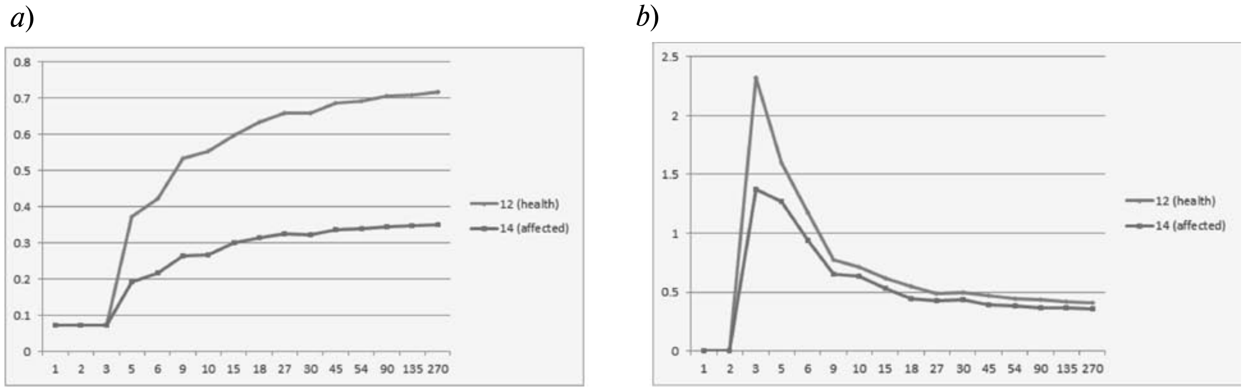


Fig. 2. The dependence the surface area (left) and fractal dimension (right) on cell size for health (number 12) and affected bone tissue (number 14)
The image size is 270×270

a) *b*) *c*)

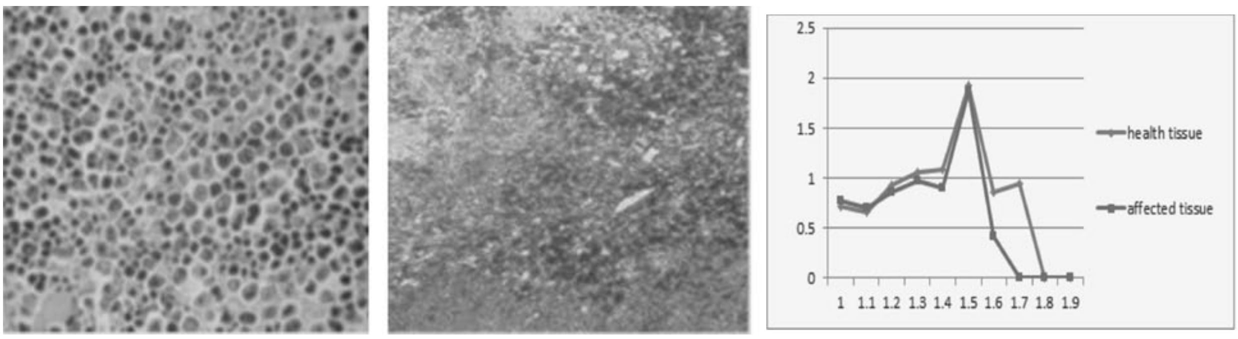


Fig. 3. The health (*a*) and affected (*b*) liver preparations and their MFS spectra (*c*)

pixel intensities [14]). Different categorization methods lead to different multifractal spectra. The method of MFS direct determination was described in [13]. In [8] we applied the method to classify the images of mentioned preparation. It demonstrated reliable results for bone and connective tissues, and was not effective for the images of blood. It is the expected result because these images do not have any fractal structure. The following illustrations show the images of health and affected liver preparations and the plots of their multifractal spectra (Fig. 3). On the *OX* axis the values α are marked, and the fractal dimensions of corresponding sets E_α are shown on *OY*.

Classification of images obtained by the

a) *b*) *c*) *d*) *e*)

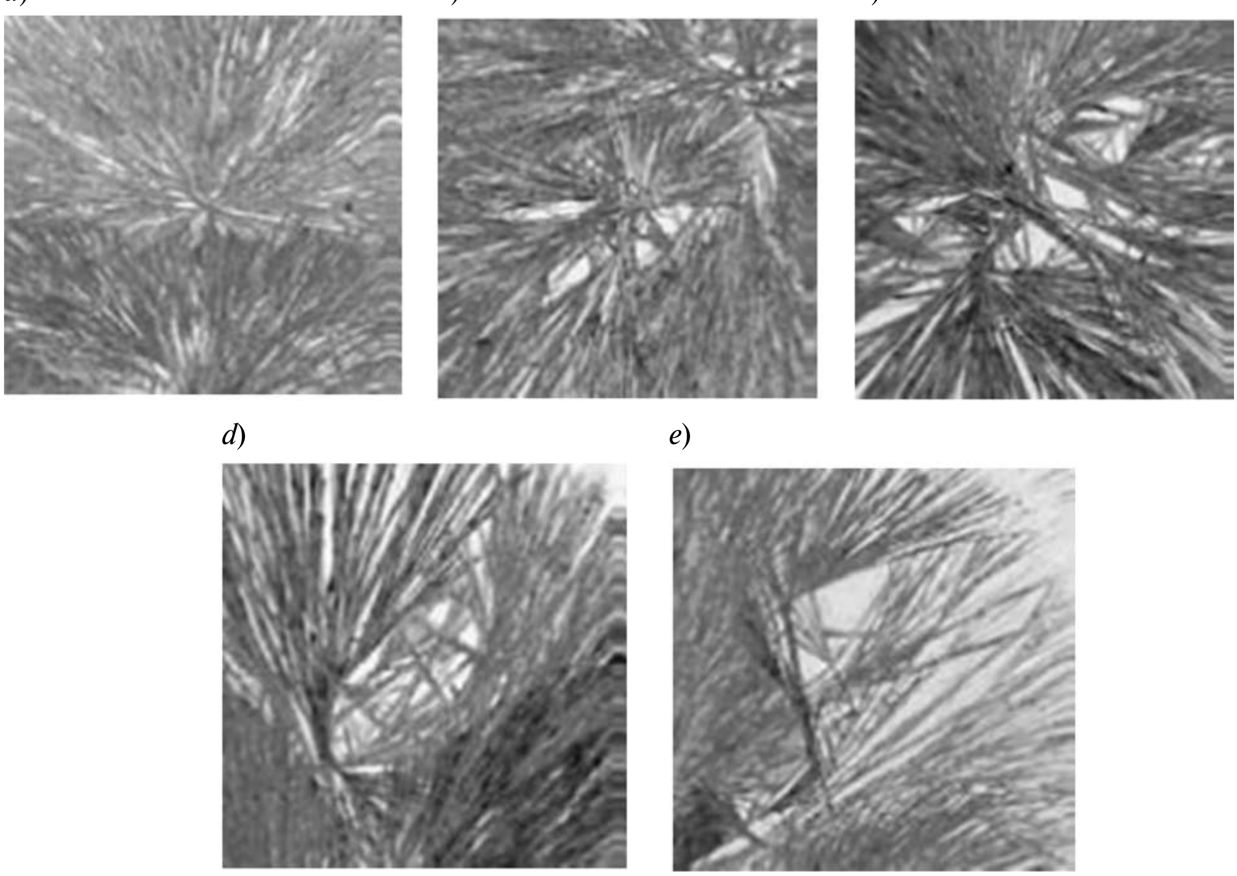


Fig. 4. Patterns of blood crystals obtained by the sensitive crystallization method

sensitive crystallization method. As it was mentioned above, the crystals of blood and plant extracts obtained by the sensitive crystallization method are very important in medical diagnosis. Their crystal structures have various forms which may be considered as classification features for classes of preparation. For blood crystals the star form (typical for acute inflammatory process) and the hole structure (typical for degenerative processes) are the examples of structures. The examples of blood crystals are shown on Fig. 4:

crystallization in the star form, which is typical for acute inflammatory process (Fig. 4 *a*);

crystallization in the star form with hole structure, which is typical for chronic inflammatory process (Fig. 4 *b*);

hole structure of crystals, which is typical for degenerative processes (Fig. 4 *c*);

hollow crystal form – typical for benign tumors (Fig. 4 *d*);

hollow crystal form with transversal structures, typical for malignant tumors (Fig. 4 *e*).

Plant extract crystals have their own features. On Fig. 5 typical plant extract crystals from the classes of images used in numerical experiments are shown.

Statistical textural features. Spatial gray level co-occurrence estimates image properties related to second-order statistics which considers the relationship among pixels or groups of pixels (usually pairs of pixels). To analyze and classify textures of digital images Haralick [15] suggested the use of gray level co-occurrence matrices (GLCMs) which have become one of the most well-known and widely used texture features. This method is based on the joint probability distributions of pairs of pixels. GLCM shows how often each gray level occurs at a pixel located at a fixed geometric position relative to each other pixel, as a function of the gray level. The (1,3) entry in a matrix for right neighbors, for example, would show the frequency or probability of finding gray level 3 immediately to the right of pixel with gray level 1.

GLCM has a parameter – a relative position of pixels, which is defined by the angle and distance. To form vectors of features for an image we use the following statistical texture features obtained on a basis of the normalized GLCM – $P_{i,j}^{\text{Norm}}$.

Contrast $\sum_{i,j=0}^{N-1} P_{i,j}^{\text{Norm}}(i-j)^2$ defines a measure of a contrast between a pixel and its neighbours or between a pixel and the whole image.

Homogeneity $\sum_{i,j=0}^{N-1}(P_{i,j}^{\text{Norm}} / 1+(i-j)^2)$ describes the density of distribution of elements in $P_{i,j}^{\text{Norm}}$ relative to its diagonal.

Correlation $\sum_{i,j=0}^{N-1} P_{i,j}^{\text{Norm}}((i-\mu I)(j-\mu J) / / \sqrt{(\sigma I^2)(\sigma J^2)})$ of the GLCM matrix defines a correlation degree between a pixel and its neighbour or such a degree between a pixel and the whole image. Here μI, μJ, σI^2 and σJ^2

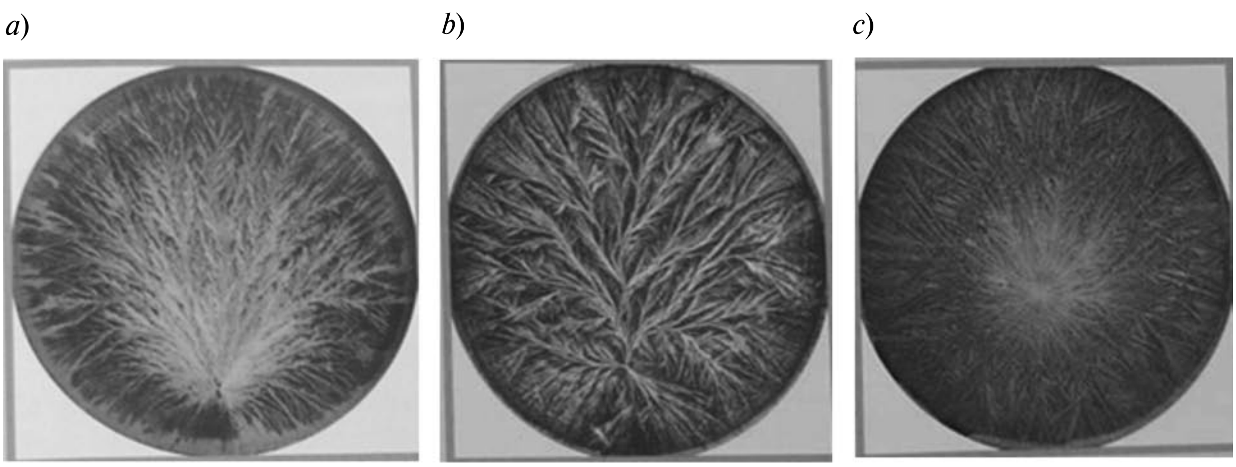


Fig. 5. Patterns of plant extract crystals obtained by the sensitive crystallization method:
a – orange juice; *b* – wheat flour solution; *c* – wheat extract

are mathematical expectations and dispersions of pixel intensities, which are calculated by using a given pixel and its neighbours.

Textural features based on the Gabor filter. We define two-dimensional Gabor's function as the following:

$$g(X,Y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{\left(-\frac{1}{2}\left(\frac{X^2}{\sigma_x^2}+\frac{Y^2}{\sigma_y^2}\right)+2\pi i\omega x\right)}.$$

Then we construct a Gabor's wavelet basing on the following generating functions: $g_{mn}(X,Y) = a^{-m} \cdot g(X^{'},Y^{'})$, $X^{'} = a^{-m}(X\cos\theta + Y\sin\theta)$, $Y^{'} = a^{-m}(X\sin\theta + Y\cos\theta)$, $\theta = n\pi/N$, $a > 1$. Here the integers m and n define the scale and orientation of the wavelet respectively: $m = 0, 1, ..., M$, $n = 0, 1, ..., N$, where M and N are the given numbers of scales and orientations.

Let $I(x,y)$ be a distribution of gray level for a digital image. Define the convolution with the Gabor core g_{mn}:

$$W_{mn}(x,y) = \int I(u,v) g^{*}_{mn}(u-x, v-y)\,du\,dv.$$

In this formula * means complex conjugation, W_{mn} – the result of convolution according to the Gabor core with the scale m and orientaition n. We suppose that the areas with local textures are spatially homogenious, then the defined below values μ_{mn} and σ_{mn} may be considered as the area characteristics for the classification problem:

$$\mu_{mn} = \iint |W_{mn}(x,y)|\,dx\,dy.$$

$$\sigma_{mn} = \sqrt{\iint (|W_{mn}(x,y)| - \mu_{mn})^2\,dx\,dy}.$$

The feature vector to classify images is formed as the following $\bar{f} = [\mu_{00} \quad \sigma_{00} \quad \mu_{01} \quad \sigma_{01} \cdots \mu_{mn} \quad \sigma_{mn}]$.

Morphological features. Mathematical morphology is a method of obtaining structure components of an image that may be useful to represent and describe it. Boundaries, skeletons and convex hulls are examples of such structures. In this work for the purpose of recognition we use skeletons. This approach leads to reducing the task dimension. To mark out the image skeleton the software packages are used [16].

The classifier of signs. In this work three-layered perceptron is used as the classifier. The number of neurons in the input layer of network is defined by the size of a vector of features (depending on the method of the vector construction). The number of neurons in the mean (the second) layer is equal to integer part of one half of the number of neurons in the input layer (an empirical recommendation). In our experiment the number of output neurons (the third layer) is defined by the number of classes of images considered in every group. The algorithm RProp (Resilient Propagation) accelerating the back propagation of error algorithm (Backprop) is applied to setup the network parameters [17, 11].

The classifier testing was performed for three groups of images:

1) images of blood crystals obtained by the sensitive crystallization method;

2) images of plant extracts crystals obtained by the sensitive crystallization method;

3) images of brain tumors.

We considered the following classes of brain tumors – astrocytoma, nevrinoma, oligodendroglioma. As it was shown in [2], for these classes fractal methods give good results. So here we use them to verify the classifier work.

a) *b*) *c*)

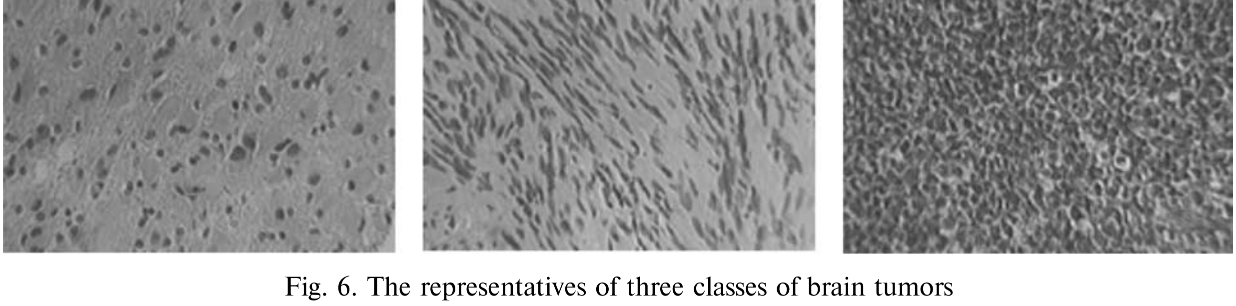
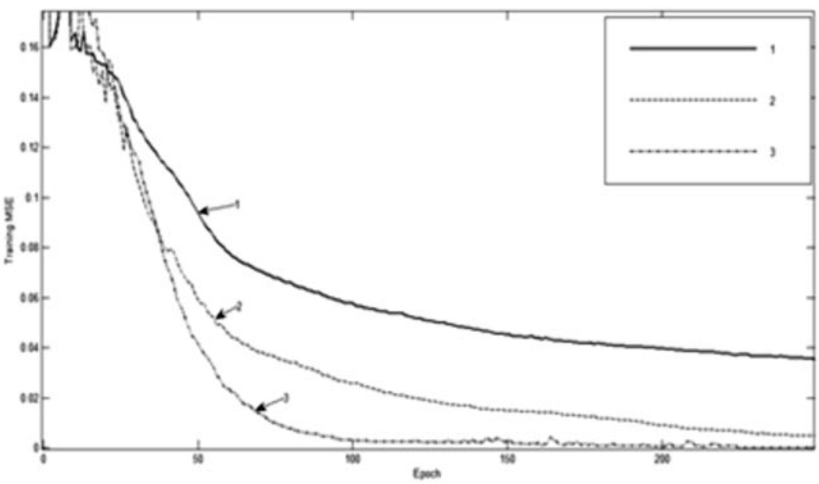


Fig. 6. The representatives of three classes of brain tumors

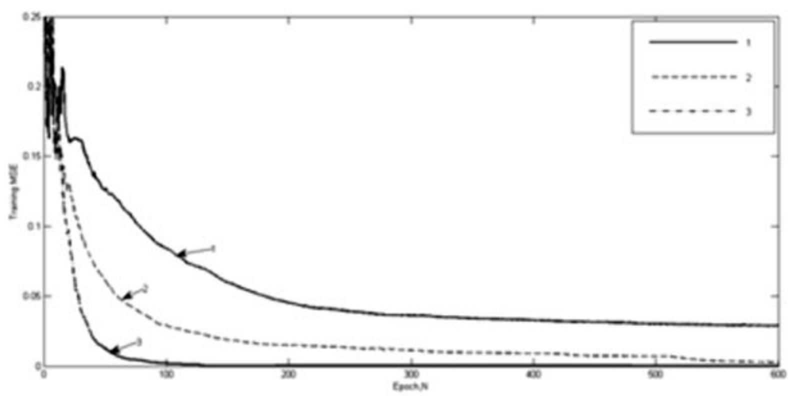
On Fig. 6 typical representatives of the classes mentioned are shown: *a* – astrocytoma; *b* – nevrinoma; *c* – oligodendroglioma.

For every group a vector of features was calculated on a basis of one of the methods mentioned above (statistical, spectral and morphological). Then the vector was input to the neural network to construct the classifier. On Fig. 7 the graphics of the dependence mean-square error on the number of iteration

a)



b)



c)

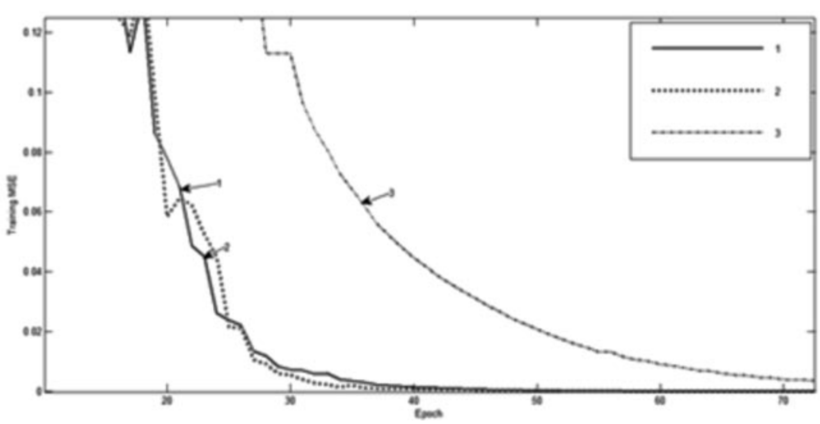


Fig. 7. The graphics of dependence mean-square error on the number of iteration:
a – for statistical texture features; b – for texture features based on Gabor's filter; *c* – for morphological features

in the cycle of weighting coefficients adjusting are shown. The digits 1—3 denote the group of images described above.

The results of experiments show that the model gives good results for images of different classes. The best results (minimal time) for the classification of blood crystal were obtained for morphological signs. For brain tumors statistical signs and Gabor's filters were more preferable. This result is in the conformity with [2], because both the fractal method and statistical characteristics are based on using pixel intensities.

Analysis and interpretation of characteristics obtained on a basis of mathematical methods are key factors for successful work of specialists in biology and medicine. The investigations made by the authors demonstrated the applicability textural, fractal, spectral and morphological methods to a classification of biomedical preparation images. These results may become a reliable basis to design and implement the tools assisting experts in their practical activity.

The authors express their gratitude to D. Koshechkin, medical chief executive of «Organic product» limited company, for many stimulating discussions and D. Vorobiev for participation in experiments.

The work was partially supported by RFBR grant No. 13-01-00782.

REFERENCES

1. **Ampilova N., Gurevich E., Soloviev I.** Texture recognition as a quantitative analysis of underdoses in pharmacology and medicine. *Proc. 5 Internat. Conference CEMA10*, 2010, Athens, Greece, pp. 78—81.

2. **Ampilova N., Gurevich E., Soloviev I.** Application of Modified Fractal Signature & Regny Spectrum Methods to the Analysis of Biomedical Preparations Images. *Proc. 6 Internat. Conference CEMA11*, 2011, Sofia, Bulgaria, pp. 96—100.

3. **Peleg Sh., Naor J., Hartley R., Avnir D.** Multiple Resolution Texture Analysis and Classification. *IEEE Trans. Pattern Analysis and Machine Inteligence*, 1984, Vol. PAMI-6, No. 4.

4. **Ampilova N., Soloviev I., Shupletzov Yu.** On some aspects of the fractal signature method. *Proc. 8 Internat. Conference CEMA13*, 2013, Sofia, Bulgaria. pp. 80—84.

5. **Tang Y.Y., Ma Hong, Xi Dihua, Mao Xiaogang, Suen C.Y.** Modified Fractal Signature (MFS): A New Approach to Document Analysis for Automatic Knowledge Acquisition. *IEEE Trans. Knowledge and Data Eng.* 1997, Vol. 9, No. 5, pp. 742—762.

6. **Pandis C., Malamou A., Stefaneas P., Frangos P.** Applying the modified fractal signature method to image classification: some preliminary results for ISAR radar images. *Proc. 7 Internat. Conference CEMA12*, 2012, Athens, Greece, pp. 50—52.

7. **Malamou A., Pandis C., Karakasiliotis A., Stefaneas P., Kodokostas D., Frangos P.** Application of an autofocusing algorithm for SAR image focusing improvement and application of the modified signature method for SAR image classification for the case of real radar data. *Proc. 8 Internat. Conference CEMA13*, 2013, Sofia, Bulgaria, pp. 108—111.

8. **Ampilova N., Soloviev I., Shupletzov Yu.** Multifractal spectrum as a classification sign for biomedical preparations images. *Proc. 7 Internat. Conference CEMA12*, 2012, Athens, Greece, pp. 98—101.

9. **Waldburger B.** Die Empfindliche Kristallisation. *Forschungsinstitut der Naturwissenschafliche Sektion am Goetheanum*, Dorhach, 2011.

10. **Vorobiev V., Gribunin V.** Teoria i practica wavelet preobrazovanii. St. Petersburg, 1999, 284 p. (rus).

11. **Riedmiller M., Braun H.** A direct adaptive method for faster backpropagation learning: The RPROP algorithm. *In Proc. of the IEEE Internat. Conference on Neural Networks,* San Francisco, 1993.

12. **Falconer K.** Fractal Geometry. Mathematical Foundations and Applications. John Wiley & Sons, 1990, 289 p.

13. **Xu Y., Ji H., Fermьller C.** Viewpoint Invariant Texture Description Using Fractal Analysis. *Internat. Journal of Computer Vision,* 2009, No. 83, pp. 85—100.

14. **Gonzalez R., Woods R.** Obrabotka cifrovih izobrazhenii, Moscow: Technosfera, 2006, 1085 p. (rus)

15. **Haralick R.** Statistical and Structural Approaches to Texture. *Proc. of the IEEE,* 1979, Vol. 67, pp. 786—804.

16. Available http://www.mathworks.com/help/toolbox/images

17. **Haykin S.** Neural Networks: A Comprehensive Foundation, 2 ed., Prentice Hall, 1999, 823 p.

СПИСОК ЛИТЕРАТУРЫ

1. **Ampilova N., Gurevich E., Soloviev I.** Texture recognition as a quantitative analysis of underdoses in pharmacology and medicine // Proc. 5 Internat. Conf. CEMA10, Athens, Greece. 2010. Pp. 78–81.

2. **Ampilova N., Gurevich E., Soloviev I.** Application of Modified Fractal Signature & Regny Spectrum Methods to the Analysis of Biomedical Preparations Images // Proc. 6 Internat. Conf. CEMA11, Sofia, Bulgaria. 2011. Pp. 96–100.

3. **Peleg Sh., Naor J., Hartley R., Avnir D.** Multiple Resolution Texture Analysis and Classification // IEEE Trans. Pattern Analysis and Machine Inteligence. 1984. Vol. PAMI-6. No. 4.

4. **Ampilova N., Soloviev I., Shupletzov Y.** On some aspects of the fractal signature method // Proc. 8 Internat. Conf. CEMA13, Sofia, Bulgaria. 2013. Pp. 80–84.

5. **Tang Y.Y., Ma Hong, Xi Dihua, Mao Xiaogang, Suen C.Y.** Modified Fractal Signature (MFS): A New Approach to Document Analysis for Automatic Knowledge Acquisition // IEEE Trans. Knowledge and Data Eng. 1997. Vol. 9. No. 5. Pp. 742–762.

6. **Pandis C., Malamou A., Stefaneas P., Frangos P.** Applying the modified fractal signature method to image classification: some preliminary results for ISAR radar images // Proc. 7 Internat. Conf. CEMA12, Athens, Greece. 2012. Pp. 50–52.

7. **Malamou A., Pandis C., Karakasiliotis A., Stefaneas P., Kodokostas D., Frangos P.** Application of an autofocusing algorithm for SAR image focusing improvement and application of the modified signature method for SAR image classification for the case of real radar data // Proc. 8 Internat. Conf. CEMA13, Sofia, Bulgaria. 2013. Pp. 108–111.

8. **Ampilova N., Soloviev I., Shupletzov Y.** Multifractal spectrum as a classification sign for biomedical preparations images // Proc. 7 Internat. Conf. CEMA12, Athens, Greece. 2012. Pp. 98–101.

9. **Waldburger B.** Die Empfindliche Kristallisation. Forschungsinstitut der Naturwissenschafliche Sektion am Goetheanum, Dorhach, 2011.

10. **Воробьёв В., Грибунин В.** Теория и практика вейвлет-преобразования. СПб., 1999. 284 с.

11. **Riedmiller M., Braun H.** A direct adaptive method for faster backpropagation learning: The RPROP algorithm // In Proc. of the IEEE Internat. Conf. on Neural Networks. San Francisco, 1993.

12. **Falconer K.** Fractal Geometry. Mathematical Foundations and Applications. John Wiley & Sons, 1990. 289 p.

13. **Xu Y., Ji H., Fermьller C.** Viewpoint Invariant Texture Description Using Fractal Analysis // Internat. J. of Computer Vision. 2009. No. 83. Pp. 85–100.

14. **Гонзалес Р., Вудс Р.** Обработка цифровых изображений. М.: Техносфера, 2006. 1085 с.

15. **Haralick R.** Statistical and Structural Approaches to Texture // Proc. of the IEEE. 1979. Vol. 67. Pp. 786–804.

16. [Электронный ресурс] / URL: http://www.mathworks.com/help/toolbox/images

17. **Haykin S.** Neural Networks: A Comprehensive Foundation. 2 ed. Prentice Hall, 1999. 823 p.

AMPILOVA, Natalya B. *St. Petersburg State University.*
199034, Universitetskaia emb. 7/9, St. Petersburg, Russia.
E-mail: n.ampilova@spbu.ru

АМПИЛОВА Наталья Борисовна – *доцент кафедры информатики Санкт-Петербургского государственного университета, кандидат физико-математических наук.*
199034, Россия, Санкт-Петербург, Университетская наб., д. 7/9.
E-mail: n.ampilova@spbu.ru

SOLOVIEV, Igor P. *St. Petersburg State University.*
199034, Universitetskaia emb. 7/9, St. Petersburg, Russia.
E-mail: i.soloviev@spbu.ru

СОЛОВЬЕВ Игорь Павлович – *доцент кафедры информатики Санкт-Петербургского государственного университета, кандидат физико-математических наук.*
199034, Россия, Санкт-Петербург, Университетская наб., д. 7/9.
E-mail: i.soloviev@spbu.ru

SHUPLETSOV, Yuri V. *St. Petersburg State University*.
199034, Universitetskaia emb. 7/9, St. Petersburg, Russia.
E-mail: yury.shupletsov@gmail.com

ШУПЛЕЦОВ Юрий Владимирович – *аспирант кафедры информатики Санкт-Петербургского государственного университета*
199034, Россия, Санкт-Петербург, Университетская наб., д. 7/9.
E-mail: yury.shupletsov@gmail.com

UDC 519.876.5=111

V.V. Okolnishnikov, S.V. Rudometov

A SYSTEM FOR COMPUTER SIMULATION OF TECHNOLOGICAL PROCESSES

A new simulation system for computer simulation of technological processes is described. The paper contains a brief description of the simulation system and its characteristics, in comparison to analogous commonly used systems. Also there is a description of existing and perspective applications for this simulation system that proves its usability and reliability.

SIMULATION; TECHNOLOGICAL SYSTEMS; AUTOMATED PROCESS CONTROL SYSTEMS.

В.В. Окольнишников, С.В. Рудометов

СИСТЕМА ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ ТЕХНОЛОГИЧЕСКИХ ПРОЦЕССОВ

Описана новая система имитационного моделирования технологических процессов. Приведено описание системы, ее характеристики, сравнение с существующими аналогичными системами моделирования. Также описаны существующие приложения, подтверждающие простоту использования и надежность этой системы моделирования.

ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ; ТЕХНОЛОГИЧЕСКИЕ СИСТЕМЫ; АВТОМАТИЗИРОВАННЫЕ СИСТЕМЫ УПРАВЛЕНИЯ ТЕХНОЛОГИЧЕСКИМИ ПРОЦЕССАМИ.

Design Technological Institute of Digital Techniques, Siberian Branch of the Russian Academy of Sciences (DTIDT) produces a number of hardware and software for different industries, like coal and oil mining. These products are Automated Process Control Systems (APCS) including both Programmable Logic Controllers (PLC) and Process Control Software (PCS). PLC are mostly used for in-place control of Technological Objects (TO) like conveyors in coal mining, whereas PCS is used to control the whole enterprise, including, but not limited by, an energy consumption control, personnel control, etc.

The nature of TO under control implies they are extremely dangerous in running. It means, that APCS must be very reliable, and do not allow potentially dangerous situations to appear. On the other hand, these systems must be «trained» to react adequately in the case of any emergency. To achieve this, computer-based simulation was introduced. The primary goal of such simulation was to check that the PLC and PCS cannot be a source of dangerous situations, and these systems are reliable and safe. The other goal is to show, how the control programs will work in non-standard situations. And finally, the simulation can be used to train personnel that will use new APCS.

To achieve these goals the simulation system MTSS (Manufacturing and Transportation Simulation System) was created. There are also other applications of this simulation system which are not limited by applications inside the DTIDT. These applications are the simulation for an automated warehouse [1] and the simulation model for a complex computing network.

Requirements for MTSS

Non-formal requirements. MTSS supposes that creation of simulation model will be done by subject matter experts (SME), but not by specialist in simulation. Simulation models in MTSS are created and analyzed visually, without application of the simulation theory. Later these models can be used by their authors in their work, without any interference of the specialist in simulation.

Non-formal requirements are:

• to create the simulation models by subject matter experts who are not specialists in simulation;

• to create the simulation models of complex technological systems combining simulation models of technological objects in these systems;

• to hide any aspects of simulation from a final user (only the possibility to tune parameters must be presented; these parameters must be clearly understandable for the final user (SME));

• to create the simulation models fast, allowing users to focus on the problem solving but not on development of simulation itself (and then on a simulation model);

• to control simulation run visually, with ability to pause simulation at any time and to examine or change simulation parameters;

• to examine statistical parameters at any time during simulation, in a form that is clearly understandable for the final user.

Formal requirements for MTSS. These requirements are given below in a form that can be applied to technological systems:

• visual interactive interface for simulation creation and execution;

• usage of graphical tools for model creation;

• support for fast model creation;

• simulation model is created from existing, ready-to-use simulation models;

• simulation model creation by SME;

• simulation model and simulation system must be able to be connected to any external systems;

• simulation model works in two-dimension (2D) and possibly in three-dimension (3D) views;

• simulation results are presented as complete analysis, without necessity of any additional analysis;

• any statistical data can be exported and analyzed externally.

The system MTSS was realized in accordance with all of these requirements.

Architecture of MTSS

Elemental models and disposition. Technological systems (TS) consist of technological objects (TO) which can be defined by their type. TS have input products; these products are processed, and then leave TS. Each TO in TS can be controlled by a supervising (control) program which is usually a part of modern TS.

Elemental model (EM) is a simulation model of a TO type. Each item of EM in the model presents an item of TO in a real system.

Each TO in TS interacts with some other TO in this system. It means that in simulation there must be possibility for EM items to interact. This interaction will simulate the correspondent interaction in real TS. Such connection can be presented by «port ontology» [2].

Port ontology in its original form is a way to organize communications between EM items

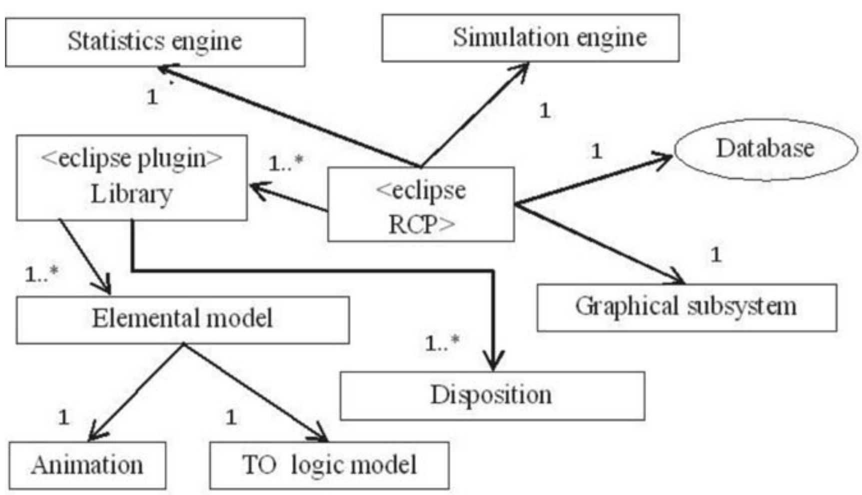


Fig. 1. Architecture of MTSS

encapsulating different approaches to build the simulation. Later the items of the simulation model that are created with port ontology can be connected to each other. MTSS uses this feature for possible EM items to be connected visually. Such connection allows different EM items to communicate (to interact) to each other in the simulation model.

Important part of EM is its visual image. It is used to identify EM item in simulation model, both when model is developed and run.

Control programs in original TS must also have its presentation in the simulation model. MTSS allows to create such models as a disposition level. It means, the MTSS has two-tier architecture: the lower tier is EM items and the upper tier is disposition.

Structure. Fig. 1 is a graphical presentation of architecture of MTSS. The figure uses in UML notation. Next major components of MTSS are presented on this figure.

Simulation engine is used for encapsulation of the simulation aspects. Different ways can be used to create this engine – classical or distributed simulation, etc. All aspects of the computer-based simulation must be encapsulated in this component.

Graphical subsystem is designed for visual presentation of the simulation model, during its creation process and running. MTSS implies that the simulation model will be created and run in 2D. The 3D can be primarily used (or not used) for the goals of presentation.

Statistics engine is used for gathering, storing and processing of any statistics generated during the simulation.

Database is a way to connect to external systems, by data transfer between MTSS and external systems.

Library is a way to unite a number of EMs and dispositions. Later MTSS can be exported as a software product that contains one or many similar libraries. This allows to produce any problem-oriented software products.

Animation and TO logic model are the major parts of any EM. Animation is used to show EM states during simulation run, and also in visual model creation. TO logic model must be implemented to simulate behaviour of original TO.

MTSS was created by using Eclipse platform and Java programming language.

Perfomance of MTSS

One of the main possibilities of MTSS is to

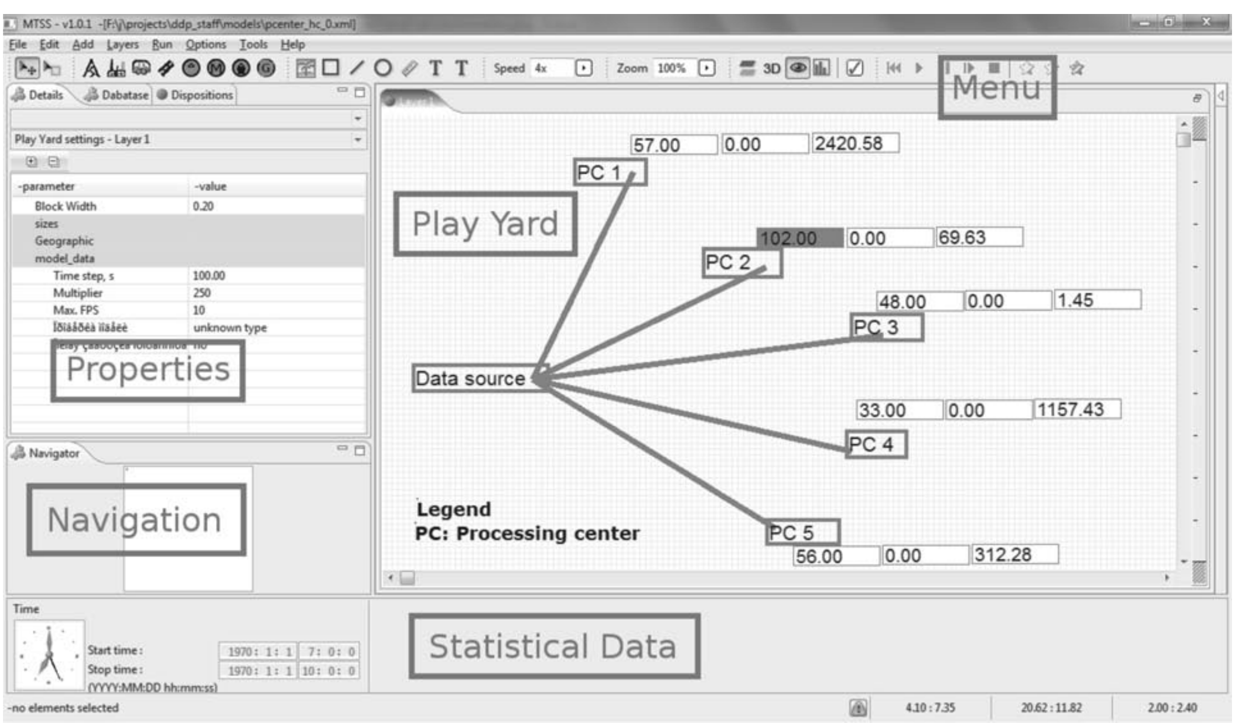


Fig. 2. User interface of MTSS

create the simulation models with SME fast. MTSS defines a structure of EMs that developers of EMs must follow. MTSS can be connected to external systems, to allow simulation to operate with data from real systems or create data for such systems.

The approach of creation of simulation models in MTSS allows users to create detailed and complex models. This fact, and possibility to connect simulation models to external systems, possibly, will allow to use MTSS even as a subsystem for APCS.

Simulation engine in MTSS is built as a modification of Mera system [3]. According to architecture of a simulation system, Simulation Engine is a module, and that is why one of the possibilities of MTSS is a usage of different implementations of simulation engines.

User interface of MTSS. Fig. 2 shows the user interface of MTSS. Main window of MTSS consists of following areas:

Play Yard is an area for model creation and running. Users of MTSS should place EM items here and connect them visually to other EM items which are already on Play Yard.

Properties are used for changing of the model parameters. Also, if EM item is selected in Play Yard, the properties for this item can be changed.

Navigation is an area for fast navigation in a model. This small view contains all its elements in it (while Play Yard can contain just a part of EM items in a model).

Statistical data area is used for output of common statistical parameters that depend on libraries used.

Menu is an application menu and a toolbox. This area allows selecting EM to add it to the model.

Creation of simulation model. Users build simulation models from ready-to-use EMs connecting the EM items visually. This allows the user to speed up models creation, to make them obvious for the users who are not familiar with aspects of computer-based simulation. For this purpose the port ontology is used.

In MTSS, ports have graphical representation and physical coordinates. Each EM can have more than one port. This allows the users to connect visually different ports to different EM items in the model, and to check visually correctness of the model. Ports in MTSS are connected to each other by dragging one port to another. Ports are counter connected if their coordinates overlay and EM implementation will allow to do it (by its internal logic). After connecting the ports, the information can be used during simulation run, in order to organize communication between different EM items.

Play Yard has its fixed physical size that can be seen in scaled Play Yard view. Navigation area can be used to move between parts of Play Yard. This allows user to build simulation models of TS, for example, buildings or big areas. MTSS also supports setting of geographical coordinates for Play Yard.

Model can be either stopped before end time of simulation, paused or stepped. The speed of model execution can be changed at any time of model running. Such handling gives the user, who carries out the simulation, a full control over the simulation process, the ability to examine the situations occur in the simulation model in details.

Animation in MTSS is asynchronous (independent on the simulation speed). Animation can be completely turned off to speedup model execution. It means that graphical engine is independent on simulation speed. If animation requires many resources, it can be delayed without need of whole simulation to be turned off. Animation can be turned off at any time during the simulation run.

The speed of the simulation can be varied during simulation run.

Statistics gathering and presentation occur during simulation run, and do not require additional calculations. During the simulation run the user will receive important information (speed, energy consumption, product flow) and some calculated values (e. g. energy consumption during the time period, system performance etc.). Using this data the user can decide to stop simulation before the actual simulation end time. Commonly-used tools are also presented in the system (charts, data export tool, and etc.).

Control of simulation run is mainly visual using the navigation tools and tools for simulation time management.

SME will not operate with any aspects of computer-based simulation theory (except «model time»). MTSS does not provide any possibilities for the final user that leads to aspects like queries, model decomposition etc.

In MTSS, SME will build simulation model fast, from ready-to-use EMs. Simulation experiments can also be carried fast, primarily visually and simulation results will be also achieved fast. There are no other specialists who can be involved in such investigation, there is no need to search for any specific information. Everything what SME needs, MTSS will provide. It is the major distinction between MTSS and other known simulation systems.

Creation of libraries with EM. MTSS allows programmers (specialists in simulation) to unite EMs and disposition programs in libraries. Programmers are free to use a common scheme for creation of simulation model in MTSS. The main distinction is that simulation model will be created not for the whole TS, but just for one TO type.

The programmer must keep in mind that the exact simulation model, the exact TS under simulation, and the exact set of simulation experiments are unknown when library of EMs is created. That is why the programmer must imply algorithms in models of logic for Ems more detailed, predict more use-cases for EM than that was originally specified. This can make creation of EMs complex. It means, that detailed investigation of TO and TS simulated must be done. But such specification turns the original task to wider one, and allows the final user to use EMs in more tasks than it was originally requested.

There are no special graphical tools in MTSS that may «help» to imply the logic of TO models as such implementation is complex. It means that advantages of graphical presentation of algorithm will be lost. According to different estimates up to 80 % of code will be written «by hands», even if special graphical tools exist in a simulation system.

Commands and states. MTSS allows (but does not insist on) the following approach to decompose the logic of EM. This approach will allow the programmer to simplify and formalize the model of the logic in EM. This logical part must be presented as a set of *states*; the transition between them must be done by *commands.*

State is an abstraction that allows the programmer to denote the current activity of EM. It is considered in MTSS that each EM item in each model time is in one state.

Command of EM is a rule to change the state of EM. The command includes the condition to start, initial state, end state, and algorithm for translation between states of EM. The command can be sent by EM itself, by the user, by another EM item, or a disposition. Each command will be added to the queue of commands existing in any EM item.

MTSS has an interface with Databases (Relational Database Management Systems – RDBMS). This allows creating simple and uniform way to connect to the external system. Java allows to connect to any RDBMS using its unified interface.

The usage of such interface can be not enough. That is why the programmers can use their own styles for accessing data from external systems. The only limitation is that this connection must be created in the disposition.

Comparison with Existing Simulation Systems

Computer-based simulation was invented as a way to solve problems that was unable to solve analytically (mostly mathematically). The original problem must be decomposed to the form that allows it to be solved by some existing approach (queuing theory, Petri nets, etc.).

Systems for computer simulation can be divided into classes:

simulation languages and libraries;

systems for visual composition of simulation models.

This classification is not exact: simulation languages can have a visual graphical environment, or systems for visual composition will translate simulation models to some existing simulation language.

But most important is that all of these systems can be used just by specialists in simulation. Only these specialists were able to made a decomposition of the original problem, create simulation experiments, validate and verify models, and most important thing is to made a backward decomposition of the simulation results to the final user (SME).

Over the past ten years a number of simulation systems proved that SME can be used [4, 5]. But these systems based on existing are built, previous versions of the same products can be appropriate only for specialists in simulation. As a rule, the interface of such system is untouched, only some subject-oriented components are added. And these systems are still based on decomposition of original TS, like in a classic simulation. But today computers are not limited in memory and processor resources, they are very good with user interfaces, and they allow to create simulation experiments with very detailed simulation models, with detailed and informative graphical representation.

Another disadvantage of such systems is that they try to be universal. This leads to overloaded interfaces and knowledge area of simulation system itself.

SME does not need anything of this. They just need a tool with elements they are familiar with, that will allow to build simulation models immediately, without any preparation like decomposition or data preparation. They even do not need to know if this is a simulation model. Such systems (and our experience shows it) SME are easier to understand and they are more preferable. MTSS was created to satisfy these requirements exactly.

Applications with MTSS

There is a number of existing applications built using MTSS [6–8].

One of the latest and perspective developments is the simulation of a network of calculation centres processing data that comes from different sources. This data can arrive from meteorological or seismic stations distributed over a large territory or even in space (satellites). Each node in this network has a complex architecture including data storage and queries, a control part and a computational part. The goal of simulation is:

- to predict loading of each node in such network: network topology, node performance and other parameters (some part of them might not be specified yet but will be added later) can be varied;
- to propose load balancing during planned upgrades of a set of nodes;
- to propose load balancing in a case of a number of nodes malfunction;
- to propose network behaviour when input data flow changes its behaviour.

The input from each data source is supposed to be Gamma-distributed. Original network of calculation centres is decomposed to a number of EMs:

1. EM for calculation centre. This complex EM contains parts simulating data storage, data queries, and data processing.
2. EM for wire between two calculation centres or data sources; such wire can be «broken» and will also simulate the delay when transmitting huge amount of data from the source to the destination.
3. EM for data source. This EM can be adjusted for various types of data, which will (depending on the data type) have a length, speed of transmission and frequency of generation. Frequency can be specified also by selecting the distribution and time interval.

This library of EMs is tested in order to be adequate in a certain ideal case that has analytical solution. Test results will be upper-approach analytical results for average and maximum value of node loading, and covariant matrix.

MTSS is the simulation environment that makes it possible to create very detailed and reliable simulation models by subject matter experts who are not familiar with computer-based simulation.

MTSS started as a simulation environment for technological systems. But its applications are not limited only by TS itself – using the same approach, the specialist in simulation can build libraries of EMs for wider range of systems. The example is the library for computer simulation of complex computational network.

This work is supported by the Russian Foundation for Basic Research (Project 13-07-98023 p_siberia_a).

REFERENCES / СПИСОК ЛИТЕРАТУРЫ

1. WirthLogistic GMBH [Online]. Available: www.wirthsim.com (Accessed 19.10. 2011).

2. **Liang V.C., Paredis C.J.J.** Foundations of multiparadigm modelling and simulation: a port ontology for automated model composition. *Proc. of the 2003 Winter Simulation Conference*, pp. 613–622.

3. **Rudometov S.V., Okolnishnikov V.V.** Development of Distributed Simulation System. *Proc. of the 7th Internat. Conference Parallel Computing Technologies.* Nizhni Novgorod, Russia, 2003, pp. 524–527.

4. **Harrell C., Ghosh B.K., Bowden R.** Simulation using ProModel. McGraw Hill, 2004, 733 p.

5. WITNESS simulation [Online]. Available: http://www.lanner.com (Accessed 12.02.2011).

6. **Okolnishnikov V.V., Rudometov S.V., Zhuravlev S.S.** Monitoring System Development Using Simulation. *Proc. of 2010 IEEE Region 8 Internat. Conference on Computational Technologies in Electrical and Electronics Engineering.* Irkutsk, Russia, 2010, pp. 736–739.

7. **Okolnishnikov V.V., Rudometov S.V., Zhuravlev S.S.** Simulation environment for industrial and transportation systems. *Proc. of Internat. Conference on Modelling and Simulation.* Prague, Czech Republic, 2010, pp. 337–340.

8. **Rudometov S.V.** Workflow for Rapid Simulation of Complex Distribution Centers. *Proc. of Internat. Conference on Modelling and Simulation.* Prague, Czech Republic, 2010, pp. 374–377.

OKOLNISHNIKOV, Victor V. *Design Technological Institute of Digital Techniques, Siberian Branch of the Russian Academy of Sciences.*
630090, Rzhanova Str. 6, St. Novosibirsk, Russia.
E-mail: okoln@mail.ru

ОКОЛЬНИШНИКОВ Виктор Васильевич – *ведущий научный сотрудник Конструкторско-технологического института вычислительной техники СО РАН, доктор технических наук.*
630090, Россия, г. Новосибирск, ул. Академика Ржанова, д. 6.
E-mail: okoln@mail.ru

RUDOMETOV, Sergey V. *Design Technological Institute of Digital Techniques, Siberian Branch of the Russian Academy of Sciences.*
630090, Rzhanova Str. 6, St. Novosibirsk, Russia.
E-mail: rsw@academ.org

РУДОМЕТОВ Сергей Валерьевич – *научный сотрудник Конструкторско-технологического института вычислительной техники СО РАН, кандидат технических наук.*
630090, Россия, г. Новосибирск, ул. Академика Ржанова, д. 6.
E-mail: rsw@academ.org

UDC 164=111

N.O. Garanina

ELEUSIS: PERFECT RECALL FOR INDUCTIVE REASONING

This paper formalizes a variant of inductive game Eleusis. A model of the game is an interpreted system with perfect recall agents for the players and the dealer. The pecularity of Eleusis multi-agent system is that the agents have to guess the behavior of the system, rather than some static information about the system. We express some Eleusis rules and properties of the system by the formulas of propositional knowledge logic and branching time *Act*-CTL-K_n.

MULTIAGENT SYSTEMS; EPISTEMIC LOGIC; ELEUSIS; PERFECT RECALL; INTERPRETED SYSTEMS.

Н.О. Гаранина

ЭЛЕВСИН: АБСОЛЮТНАЯ ПАМЯТЬ ДЛЯ ИНДУКТИВНОГО ВЫВОДА

Формализован вариант индуктивной карточной игры Элевсин. Моделью игры является интерпретированная система с агентами с абсолютной памятью, реализующими игроков и раздающего. Особенностью мультиагентной системы Элевсин является то, что агенты должны вычислить поведение самой системы, а не какую-либо статическую информацию. Некоторые правила игры Элевсин и свойства построенной мультиагентной системы выражены формулами логики знаний и действий *Act*-CTL-K_n.

МУЛЬТИАГЕНТНЫЕ СИСТЕМЫ; ЛОГИКА ЗНАНИЙ; ЭЛЕВСИН; АБСОЛЮТНАЯ ПАМЯТЬ; ИНТЕРПРЕТИРОВАННЫЕ СИСТЕМЫ.

There are a lot of puzzles that give clear comprehension of various notions from a theory of epistemic logics. For example, the Coordinated Attack Problem [3] shows unreachability of common knowledge in the environment with time gaps; in Muddy Children Puzzle [3] knowledge is based on the quantity of observations, while in Dining Cryptographers Problem [9] knowledge is based on the quality of observations; Mars Robot Puzzle [14] illustrates knowledge taken by double-checking observations; False Coin Puzzle [11] demonstrates knowledge acquisition. In the last two puzzles and in the puzzle from this paper perfect recall for agents is crucial[1].

To the best of our knowledge, epistemic logics was not used for a definition of inductive knowledge of agents before. In this paper we formalize inductive game Eleusis. This game was introduced by Robert Abbot in 1956. We give a description of the first variant of the game from [4]:

Eleusis (pronounced ee-loo-sis) is a game for three or more players. It makes use of the standard deck of playing cards. Players take turns at being the dealer, who has no part in the actual play except to serve as a kind of umpire. He deals to the other players until one card remains. This is placed face up in the center of the table as the first card of the «starter pile». To make sure that players receive equal hands, the dealer must remove a certain number of cards before dealing. For three players (including the dealer, who of course does not get a hand) he removes one card; for four players, no cards; five players, three cards, and so on. The removed cards are set aside without being shown.

After the cards are dealt and the «starter card» is in place, the dealer makes up a secret rule that determines what cards can be played on the starter pile. It is this rule that corresponds to a law of science; the players may think of the dealer as Nature, or, if they prefer, as God. The dealer writes his rule on a piece of paper,

[1] For the Dining Cryptographers this notion is not so important because substantial information (even or odd number of differences) could be coded by a boolean variable.

which he folds and puts aside. This is for later checking to make sure that the dealer does not upset Nature's uniformity by changing his rule. For each player the object of the game is to get rid of as many cards as possible. This can be done rapidly by any player who correctly guesses the secret rule.

An example of a very simple rule is: «If the top card of the starter pile is red, play a black card. If the top card is black, play a red card».

There are many studies of this game, but they refer to such aspects as induction algorithms for players discovering the rule [2], machine learning technique [10] or the computational complexity of various decision problems that arise in Eleusis [7]. Multi-agent systems and epistemic logics were not in the focus of the game researchers before.

We study Eleusis in trace-based synchronous perfect recall multi-agent systems. In such systems agents have memory: their knowledge depends on the states passed and on the previous actions. We can describe this kind of agents because semantics of knowledge is defined on traces, i. e. finite sequences of states and actions, and every agent can distinguish traces with different sequences of information available for it. Each element of the trace represents a state of the system at some moment of time. A model of this game differs from other multi-agent systems as in this system the agent has complete information about the system and tries to guess modus operandi of another agent or of the environment.

For formalization of Eleusis world we use a Propositional Logic of Knowledge and Branching Time *Act*-CTL-K_n [12] and a notion of an interpreted system [6, 3] enriched by perfect recall [8]. Properties of these systems could be checked by some model checker of multi-agent systems with perfect recall, for example, MCMAS [9].

Background Logics and Models

First, we would like to give definitions to combined Propositional Logic of Knowledge and Branching Time *Act*-CTL-K_n from [12] briefly. This logic is a fusion of Propositional Logic of Knowledge (PLK) [3] and Computational Tree Logic (CTL) [1] extended by action symbols. Usually, semantics of *Act*-CTL-K_n is defined in terms of a satisfiability relation $\models$ in environments that are a special kind of labeled transition systems. In this paper we prefer to use interpreted systems which are a specific form of environments.

Let us reformulate the definition of interpreted systems from [6]. We denote a set of agents by $A = \{1, ..., n\}$ ($n \in N$), for each agent $i \in A$ a set of local states by L_i and possible actions by $Acts_i$, and a set of local states and actions for the environment by L_e and $Acts_e$ respectively. Let $Act = Acts_1 \times ... \times Acts_n \times Acts_e$ be the set of joint actions. We assume that a set of protocols $P_i : L_i \to 2^{Acts_i}$, for $i \in A$, represents the behavior of every agent, and a protocol $P_e : L_e \to 2^{Acts_e}$ for the environment.

A formal definition of interpreted systems is as follows:

Definition 1. (*of interpreted systems*). Given a set of agents $A = \{1, ..., n\}$ *an interpreted system* is a tuple $M = (G; t; P; \sim_1, ..., \sim_n; \iota; V)$, where

(1) G is the set of the global states for the system $G = L_1 \times ... \times L_n \times L_e$; $\iota \in G$ is the initial state;

(2) a transition function $t : G \times Act \to G$ models computations in the system;

(3) a joint protocol $P : G \to 2^{Act}$ defines the behavior of the system such as $P = \langle P_1, ..., P_n, P_e \rangle$, where P_i and P_e are protocols of agents and environment;

(4) $\sim_i \subseteq G \times G$ ($i \in A$) is an epistemic indistinguishability relation for each agent $i \in A$ defined by $s \sim_i s'$ iff $l_i(s') = l_i(s)$, where the function $l_i : G \to L_i$ returns the local state of agent i from a global state s;

(5) $V : G \to 2^{Prp}$ is a valuation function for a set of propositional variables Prp such as *true* $\in V(s)$ for all $s \in G$. V assigns to each state a set of propositional variables that are assumed to be true at that state.

For every $a \in Act$ an *a-run* is a maximal sequence of states $gs = s_1 ... s_j s_{j+1} ...$ such as $t(s_j, a) = s_{j+1}$ for all $j > 0$.

Every indistinguishability relation expresses the fact that an agent has incomplete information about system states.

Definition 2. (*of Act-CTL-K_n syntax*). *Syntax* of *Act*-CTL-K_n consists of formulas that are constructed from Boolean constants {*true*, *false*}, propositional variables, connectives $\neg$,

$\wedge$, $\vee$, and the following modalities. Let $i \in A$, $a \in Act$, φ and ψ be formulas. Then formulas with the modalities are:

- knowledge modalities: $K_i\varphi$ and $S_i\varphi$ (they are read as `an agent *i* knows' and `an agent *i* supposes');
- action modalities: $AX^a\varphi$, $EX^a\varphi$, $AG^a\varphi$, $EG^a\varphi$, $AF^a\varphi$, $EF^a\varphi$, $A\varphi U^a\psi$, and $E\varphi U^a\psi$ (*A* is read as 'for all futures', *E* – 'for some futures', *X* – 'next time', *G* – 'always', *F* – 'sometime', *U* – 'until', and a sup-index *a* – 'in a-run(s)').

Syntax of *Act*-CTL-K$_n$ combines modalities of PLK [3] and CTL [1] with action symbols. Semantics of *Act*-CTL-K$_n$ follows semantics of these logics.

Definition 3. (*of Act-CTL-K$_n$ semantics*). A satisfiability relation $|=$ between interpreted systems, states, and formulas is defined inductively with respect to a structure of formulas. For Boolean constants, propositional variables, and connectives a satisfiability relation is standard. For the action modalities semantics is almost the same as for the standard CTL-modalities, but is formulated for *a*-runs. For the knowledge modalities we define the semantics as follows. Let $s \in G$, $i \in A$, φ be a formula, then

- $s \models_M (K_i\varphi)$ iff for every s' : $s \sim_i s'$ implies $s' \models_M \varphi$;
- $s \models_M (S_i\varphi)$ iff for some s' : $s \sim_i s'$ and $s' \models_M \varphi$.

Semantics of a formula φ in an interpreted system *M* is the set of all worlds of *E* that satisfies this formula φ: $M(\varphi) = \{s \mid s \models_M \varphi\}$.

Further we consider just *Act*-CTL-K$_n$ normal formulas in which negation is used in literals only. Every *Act*-CTL-K$_n$ formula is equivalent to some normal formula due to «De Morgan» laws.

We use trace-based perfect recall synchronous (PRS) interpreted systems generated from background interpreted systems. In PRS environments (1) states are sequences of states of initial interpreted systems with a history of generating actions; (2) there are transitions from one sequence to another with an action *a* made by extending the sequence with a state reachable by *a* from the last state of the sequence; (3) perfect recall protocols take into account local traces of agents when modelling their actions; (4) an agent does not distinguish such sequences if the background system performs the same sequence of actions, and the agent cannot distinguish the sequences state by state; (5) propositional variables are evaluated at the last state of sequences with respect to their evaluations in the background interpreted systems.

Definition 4. (*of a PRS-system*). Let *M* be an interpreted system $M = (G; t; P; \sim_1, ..., \sim_n; \iota; V)$. A trace-based *Perfect Recall Synchronous interpreted system* generated by *M* is another interpreted system $prs(M) = (G_{prs}; t_{prs}; P_{prs}; \approx_1, ..., \approx_n; \iota_{prs}; V_{prs})$[2]:

(1) G_{prs} is the set of all pairs (gs, as), where non-empty $gs \in G^*$, $as \in Act^*$, $|gs| = |as| + 1$, and $t(gs_j, as_j) = gs_{j+1}$ for every $j \in [1..|as|]$; $ls_i(gs) \in L_i^*$ is a sequence of local states of agent $i \in A$ with $(ls_i)_j = l_i(gs_j)$; $\iota_{prs} \in G_{prs}$ is the initial state;

Let us assume that $(gs, as), (gs', as') \in G_{prs}$:

(2) a transition function $t_{prs} : G^{prs} \times Act \to G_{prs}$ is defined as follows: for every $a \in Act$: $t_{prs}((gs, as), a) = (gs', as')$ iff $as' = as\hat{\ }a$, $gs' = gs\hat{\ }s'$, and $t(g_{|gs|}, a) = s'$;

(3) a joint protocol $P_{prs}(M) : G_{prs} \to 2^{Act}$ such as $P_{prs} = \langle (P_1)_{prs}, ..., (P_n)_{prs}, (P_e)_{prs} \rangle$, where every $(P_i)_{prs}$: $L_i^* \to 2^{Acts}i$ does not depend on P_i[3] and $(P_e)_{prs} : L_e^* \to 2^{Acts}e$ such as $(P_e)_{prs}(ls_e \hat{\ } s_e) = P_e(s_e)$;

(4) for every $i \in A$: $(gs, as) \approx_i (gs', as')$ iff $as = as'$ and $gs_j \sim_i gs'_j$ for every $j \in [1..|gs|]$;

(5) for every $p \in Prp$: $(gs, as) \in V_{prs}(p)$ iff $gs_{|gs|} \in V(p)$.

In PRS-systems agents have some kind of memory because an awareness expressed by an indistinguishability relation depends on the history of the system evolution. Actions of perfect recall agents are knowledge-based because protocols of agents take into account the previous states and actions.

Eleusis Model

In order to define the background Eleusis interpreted system *E-M* we have to determine the following features: (1) agents and an environment; (2) their local states and

[2] In the definition, for every set *S* let *S** be the set of all finite sequences over *S* and the operation ^ stand for the concatenation of finite words.

[3] We model behavior according to information history, but not to the single local state.

generated global states; (3) local agent actions and generated global actions; (4) a transition function; (5) protocols of agents and the environment; (6) indistinguishability relations for agents; (7) propositional variables and their evaluation.

(1) Agents and an environment. Let players of the game and the dealer be agents. The environment may not be specified in our case. Without loss of generality, we consider two players – Alice and Bob, and the dealer Clare.

(2) Local and global states. It is not reasonable to fix special local states for agents because they have complete information about the system. We define the global state using the following local variables:

Value = {2, 3, 4, 5, 6, 7, 8, 9, 10, *Kn*, *Q*, *K*, *A*} is a set of card values;

Suit = {♠, ♣, ♦, ♥} is a set of suits;

Cards = *Value* × *Suit* is a set of cards;

Obs ∈ *Cards* is the last observable card in the dealer's pile;

Prt ∈ *Cards* ∪ {∅} is a pretender card to satisfy the rule put by some player;

$Deck_A$, $Deck_B \subset$ *Cards* are the decks of Alice and Bob which they have to get rid of;

Turn ∈ {*A*, *B*} determines whose move is next;

Moves ∈ [0..*n*] is a number of dealer replies. Let *n* be big enough to discover the rule by players.

Then every global state is $(Obs, Prt, Deck_A, Deck_B, Turn, Moves) \in G$, where $G = Cards \times (Cards \cup \{\varnothing\}) \times 2^{Cards} \times 2^{Cards} \times \{A, B\} \times [0..n]$.

We set $\iota = (Obs, \varnothing, Deck_A, Deck_B, Turn, 0)$, where $Obs \in Cards$, $Deck_A, Deck_B \subset Cards \setminus \{some\ card\}$[4] as initial state, and *Turn* ∈ {*A*, *B*}. Further we use the following notation: for every card *C* ∈ *Cards* let *C.Val* be a value of *C* and *C.Suit* be a suit of *C*.

(3) Local and global actions. Let us define the following actions of agents:

turn is an action of agent-players. The result of this action is a new pretender card;

accept is an action of the agent-dealer used for validation that the pretender card satisfies the rule. The result is a new observable card and a decreased deck of a player that gives the pretender;

reject is an action of the agent-dealer used for disproof that the pretender card satisfies the rule. The result is passing the move.

Let us note that in this system agents cannot act parallelly because their actions depend on each other. Hence global actions are rather a PDL-program [5] of actions *move* = *turn*;(*accept* ∪ *reject*) (to make an experiment then to give a result) than a product *turn*×{*accept*, *reject*}. It is easy to make a product by the PDL-program move introducing lazy action *skip* which does nothing for every agent.

(4) A transition function. The transition function for actions is given in terms of pre- and post-conditions of local state variables: $(pre_1, ..., pre_5) \to (post_1, ..., post_5)$, where for every $j \in [1..5]$ (1) pre_j is a precondition for the corresponding variable; the precondition of omitted variable means that the corresponding variable has any value; (2) $post_j$ is a postcondition for the corresponding variable; the postcondition of omitted variable means that the corresponding variable has the same value as in the precondition. Let *card* ∈ *Cards* and *other_cards* ⊂ *Cards*. We describe turn and reject for Alice only, because for Bob these actions are similar.

1. *turn*:

$(Prt = \varnothing,\ Deck_A = \{card\} \cup other\_cards,\ Turn = A) \to (Prt = card,\ Deck_A = other\_cards)$;

2. *accept*:

$(Prt = card) \to (Obs = card,\ Prt = \varnothing,\ Moves = Moves + 1)$;

3. *reject*:

$(Prt = card,\ Deck_A = other\_cards;\ Turn = A) \to$

$(Prt = \varnothing,\ Deck_A = \{card\} \cup other\_cards,\ Turn = B,\ Moves = Moves + 1)$.

(5) Protocols of agents. In our model protocol for players is easy: they perform any action in the state. They try an arbitrary card they have. We choose this simple behavior of players because the definition of actions corresponds to reasoning too cumbersome for this short paper. This is a topic for our future work. Nevertheless, it is possible to formulate and analyse some interesting properties of the Eleusis system.

[4] In order to provide our agents with an equal quantity of cards, the dealer has to remove one card from the initial deck.

The protocol for dealer actions depends very much on the chosen rule. In many cases it must be a perfect recall protocol because the dealer relies on the previously accepted cards when accepting or rejecting cards. Some of these protocols are given in next section.

(6) Indistinguishability relations. As game agents have complete information about the system in our variant of Eleusis, we assume that every indistinguishability relation is equality.

(7) Propositional variables and their evaluation. In our case a set of propositional variables strictly depends on formulas to be verified. Hence we define them later using values of local variables. Thus a valuation function V is defined naturally.

The corresponding perfect recall interpreted system for Eleusis $E\text{-}prs(M)$ can be made from $E\text{-}M$ by Definition 4. Perfect recall is really needed for agent-players because in order to discover the rule they have not only to consider accepted cards, but to remember rejected cards. Several perfect recall protocols for the dealer depending on the rules are given in the section below.

Eleusis Rules

The most interesting thing in Eleusis and its formalization is the accepting rule. We consider the rules that use card descriptions only[5], i. e. their values and suits (colors and faces are subsets of suits and values). There are variants of Eleusis where rules could be defined as a function from sequences to boolean values [7]. Time modalities next $\boldsymbol{X}$, globally $\boldsymbol{G}$ and until $\boldsymbol{U}$ are sufficient for a wide class of Eleusis rules[6]. We formulate several rules from [4] using these time modalities. The corresponding dealer protocols are described too. Let us define the following sets of cards: $Even = \{card \mid card.Val \in \{2, 4, 6, 8, 10, Q\}\}$, $Odd = \{card \mid card.Val \in \{3, 5, 7, 9, Kn, K, A\}\}$, $Black = \{card \mid card.Suit \in \{\spadesuit, \clubsuit\}\}$, and $Red = \{card \mid card.Suit \in \{\diamondsuit, \heartsuit\}\}$. Let $v \in Value$ and $s \in Suit$.

In dealer actions the above PDL-program $move = turn; (accept \cup reject)$ follows the Dealer protocol corresponding to the rule of the game. The dealer protocol is a function from sequences of the system states to a set of dealer actions enabled at these sequences. For sequences we use the following regular notation. Let p be a propositional condition for system variables that uses set inclusion, subsets, equalities, inequalities of system variables, and boolean connectives. For simplicity we consider p to be a «propositional» state from a set of system states that satisfies with this propositional condition p:

- a propositional state p is a sequence of states;
- $p;q$ is a sequence of states which is concatenation of sequences of states from p and q;
- p^* is a sequence of states which is a finite repeat of states from p.

Let $state$ be a propositional state from a complete set of system states G. Let us assume that by the definition of propositional states in sequence $state^*$ states can differ. It is easy to see that the set of states in which the dealer rejects a pretender card is the complement to the set of states in which the dealer accepts the pretender card. We shall describe both cases.

1) Alternatively even and odd cards:

$\boldsymbol{AG}^{move}((Obs \in Even \to \boldsymbol{AX}^{move}\ Obs \in Odd) \wedge (Obs \in Odd \to \boldsymbol{AX}^{move} Obs \in Even))$.

The dealer protocol:

$state^*; ((obs \in Even \wedge prt \in Odd) \vee (obs \in Odd \wedge prt \in Even)) \to accept$;

$state^*; ((obs \in Even \wedge prt \in Even) \vee (obs \in Odd \wedge prt \in Odd)) \to reject$.

2) The card played must have either the same suit or the same value as the card on top of the pile:

$\boldsymbol{AG}^{move}((Obs.Val = v \wedge Obs.Suit = s) \to \boldsymbol{AX}^{move}(Obs.Val = v \vee Obs.Suit = s))$.

The dealer protocol:

$state^*; (Obs.Val = prt.Val \vee Obs.Suit = prt.Suit) \to accept$;

$state^*; (Obs.Val \neq prt.Val \wedge Obs.Suit \neq prt.Suit) \to reject$.

3) If the top two cards are of the same color, play a card from ace to 7. If they are of different colors, play a card from 7 to king:

[5] It is possible to devise rules taking into account some external features like names of players or the color of their hair. These rules correspond to the experiments in which the effect of a experimenting person and his equipment is considerable.

[6] We do not need F in future because Eleusis rule determines all cards in sequences step by step.

$\boldsymbol{AG}^{move}(((Obs \in Black) \to \boldsymbol{AX}^{move}((Obs \in Black) \wedge \boldsymbol{AX}^{move}(Obs.Val \in [A..7])) \vee$
$(Obs \in Black) \to \boldsymbol{AX}^{move}((Obs \in Red) \wedge \boldsymbol{AX}^{move}(Obs.Val \in [8..K]))) \vee ((Obs \in Red) \to \boldsymbol{AX}^{move}((Obs \in Red) \wedge \boldsymbol{AX}^{move}(Obs.Val \in [A..7])) \vee (Obs \in Red) \to \boldsymbol{AX}^{move}((Obs \in Black) \wedge \boldsymbol{AX}^{move}(Obs.Val \in [8..K]))))$

The dealer protocol:

$state^*$; $obs \in Black$; $(obs \in Black \wedge prt.Val \in [A..7]) \to accept$;

$state^*$; $obs \in Red$; $(obs \in Red \wedge prt.Val \in [A..7]) \to accept$;

$state^*$; $obs \in Black$; $(obs \in Red \wedge prt.Val \in [8..K]) \to accept$;

$state^*$; $obs \in Red$; $(obs \in Black \wedge prt.Val \in [8..K]) \to accept$;

$state^*$; $obs \in Black$; $(obs \in Black \wedge prt.Val \in [8..K]) \to reject$;

$state^*$; $obs \in Red$; $(obs \in Red \wedge prt.Val \in [8..K]) \to reject$;

$state^*$; $obs \in Black$; $(obs \in Red \wedge prt.Val \in [A..7]) \to reject$;

$state^*$; $obs \in Red$; $(obs \in Black \wedge prt.Val \in [A..7]) \to reject$.

4) If the second card from the top is red, play a card with a value equal to or higher than this card. If the second card is black, play a card of equal or lower value:

$\boldsymbol{AG}^{move}(((Obs \in Red \wedge Obs.Val = v) \to \boldsymbol{AX}^{move}(Obs \in Cards) \wedge \boldsymbol{AX}^{move}(Obs.Val \geq v)) \vee ((Obs \in Black \wedge Obs.Val = v) \to \boldsymbol{AX}^{move}(Obs \in Cards) \wedge \boldsymbol{AX}^{move}(Obs.Val \leq v)))$

The dealer protocol:

$state^*$; $(obs \in Red \wedge Obs.Val = v)$; $(prt.Val \geq v) \to accept$;

$state^*$; $(obs \in Black \wedge Obs.Val = v)$; $(prt.Val \leq v) \to accept$;

$state^*$; $(obs \in Red \wedge Obs.Val = v)$; $(prt.Val \leq v) \to reject$;

$state^*$; $(obs \in Black \wedge Obs.Val = v)$; $(prt.Val \geq v) \to reject$.

All above rules from [4] are applicable over a fixed number of cards. The next rule uses indefinite, but finite number of cards.

5) Change the color of cards after an ace:

$\boldsymbol{AG}^{move}((Obs \in Black)\ \boldsymbol{AU}^{move}(Obs.Val = A \wedge \boldsymbol{AX}^{move}\ Obs \in Red) \vee (Obs \in Red)\ \boldsymbol{AU}^{move}\ (Obs.Val = A \wedge \boldsymbol{AX}^{move}\ Obs \in Black))$.

The dealer protocol:

$(obs \in Black)^*$; $(obs \in Black \wedge Obs.Val = A)$; $(prt \in Red) \to accept$;

$(obs \in Red)^*$; $(obs \in Red \wedge Obs.Val = A)$; $(prt \in Black) \to accept$;

$(obs \in Black)^*$; $(obs \in Black \wedge Obs.Val = A)$; $(prt \in Black) \to reject$;

$(obs \in Red)^*$; $(obs \in Red \wedge Obs.Val = A)$; $(prt \in Red) \to reject$.

After expressing the accepting rule *Rule* we can formulate the features of the Eleusis model. For example:

1) $\boldsymbol{AF}^{move}\ (K_A Rule \vee K_B Rule)$ – there is a moment in the future when some player knows the rule;

2) $\boldsymbol{AF}^{move}(K_A Rule \wedge \neg K_B Rule)$ – there is a moment in the future when Alice knows the rule earlier than Bob;

3) $(Turn = A) \to \boldsymbol{AF}^{move}\ (K_A Rule \wedge \neg K_B Rule)$ – if Alice moves first then she is a winner;

4) $\neg(K_A Rule \vee K_B Rule)\ \boldsymbol{AU}^{move}\ (K_A Rule \vee K_B Rule) \wedge (Moves < m)$ – the rule can be discovered in less than *m* steps.

Let us define that time complexity of the model checking these formulas depends on the complexity of the rule. The dealer can take into account a finite part of the sequence which determines accepting cards: in this case the model checking the complexity corresponds to the linear complexity of checking *Act*-CTL-K_n formulas in forgetful semantics [13]. But when the dealer makes the decision on pretender cards considering all sequence, then it corresponds to true perfect recall semantics and the model checking properties of the system has non-elementary complexity [12].

In this paper we consider a simple variant of the inductive game Eleusis. A model of the game is an interpreted system with perfect recall agents for the players and the dealer. This is an example of a multi-agent system in which agents have to guess the behavior of the system rather than the information about it. We express some Eleusis rules by formulas of branching time logic Act-CTL and some properties of the system by formulas of *Act*-CTL-K_n with knowledge modalities. Model checking of these properties depends on the complexity of the Eleusis rule. It can be linear or non-elementary according to the size of Eleusis model.

We do not give protocols of players guessing the rule because it is too complex for a short paper. In the future we plan to develop a simple protocol and try to do model checking of the corresponding Eleusis system with some perfect recall model checker.

The research has been supported by Russian Foundation for Basic Research (grant 13-01-00643) and by Siberian Branch of Russian Academy of Science (Integration Grant n.15/10 «Mathematical and Methodological Aspects of Intellectual Information Systems»).

REFERENCES / СПИСОК ЛИТЕРАТУРЫ

1. **Clarke E.M., Grumberg O., Peled D.** Model Checking. MIT Press, 1999.

2. **Dietterich T.** Applying General Induction Methods to the Card Game Eleusis. *Proc. of the 1st Annual National Conference on Artificial Intelligence.* Stanford University. AAAI Press/MIT Press, 1980, pp. 218–220.

3. **Fagin R., Halpern J.Y., Moses Y., Vardi M.Y.** Reasoning about Knowledge. MIT Press, 1995.

4. **Gardner M.** The Second Scientific American Book of Mathematical Puzzles and Diversions. University of Chicago Press Edition, 1987, 251 p.

5. **Harel D.** First-Order Dynamic Logic. *Lecture Notes in Computer Science*, 1979, Vol. 68.

6. **Kacprzak M., Lomuscio A., Penczek W.** Verification of Multi-agent Systems via Unbounded Model Checking. *Proc. of the 3 Internat. Joint Conference on Autonomous Agents and Multi-agent Systems*, 2004, Vol. 2, pp. 638–645.

7. **Kurzen L.** Eleusis: Complexity and Interaction in Inductive Inference. *Proc. of the 2 ILCLI Internat. Workshop on Logic and Philosophy of Know-ledge, Communication and Action*, Donostia – San Sebastian, Spain. 2010. The University of the Basque Country Press, 2010.

8. **Lomuscio A. R., van der Meyden R., Ryan M.** Knowledge in Multi-agent Systems: Initial Configurations and Broadcast. *ACM Transactions on Computational Logic*, 2000, Vol. 1, No. 2, pp. 247–284.

9. **van der Meyden R., Su K.** Symbolic Model Checking the Knowledge of the Dining Cryptographers. *Proc. of the 17th IEEE Workshop on Computer Security Foundations*, 2004, pp. 280–291.

10. **De Raedt L., Van Laer W.** Inductive Constraint Logic. *Proc. of 6th International Conference Algorithmic Learning Theory,* Fukuoka, Japan, 1995, Springer, Lecture Notes in Computer Science, 1995, Vol. 997, pp. 80–94.

11. **Shilov N.V., Yi K.** How to find a coin: propositional program logics made easy. *In Current Trends in Theoretical Computer Science, World Scientific*, Vol. 2, 2004, pp. 181–213.

12. **Shilov N.V., Garanina N.O., and Choe K.-M.** Update and Abstraction in Model Checking of Knowledge and Branching Time. *Fundameta Informaticae*, 2006, Vol. 72, No. 1–3, pp. 347–361.

13. **Shilov N.V., Garanina N.O.** Combining Knowledge And Fixpoints. Preprint, Novosibirsk, 2002, No. 98, 50 p.

14. **Shilov N., Garanina N., Bodin E.** Multi-agent approach to a Dijkstra problem. *In Proc. of Workshop on Concurrency, Specification and Programming,* Helenenau, 2010. Humboldt-Universitat zu Berlin. Informatik-Bericht, No. 237, Vol. 1, pp. 73–84.

GARANINA, Natalia O. *Laboratory of Theoretical Programming, A.P. Ershov Institute of Informatics Systems, Siberian Branch of the Russian Academy of Sciences.*
630090, Acad. Lavrentjev Ave. 6, Novosibirsk, Russia.
E-mail: garanina@iis.nsk.su

ГАРАНИНА Наталья Олеговна – *научный сотрудник лаборатории теоретического программирования Института систем информатики имени А.П. Ершова Сибирского отделения РАН.*
630090, Россия, г. Новосибирск, пр. Акад. Лаврентьева, д. 6.
E-mail: garanina@iis.nsk.su

UDC 621.311.24=111

V.V. Potekhin, A.L. Loginov, D.V. Akulov, I.G. Poletaev, M.A. Konishchev

INTELLECTUAL CONTROL ALGORITHM FOR REDUCING THE AERODYNAMIC INTERACTION AT AUTONOMOUS WIND FARM

The article deals with the problem of wind farm intellectual control. It describes an intellectual control algorithm developed by the authors, and considers the aerodynamic interaction of wind turbines which can decrease total energy output. An autonomous wind farm has been modelled in Matlab/Simulink. Wind farm work with the application of an intellectual control algorithm is compared with the conventional one.
INTELLECTUAL CONTROL ALGORITHM; WIND TURBINE; AUTONOMOUS WIND FARM.

В.В. Потехин, А.Л. Логинов, Д.В. Акулов, И.Г. Полетаев, М.А. Конищев

ИНТЕЛЛЕКТУАЛЬНЫЙ АЛГОРИТМ УПРАВЛЕНИЯ СНИЖЕНИЕМ АЭРОДИНАМИЧЕСКОГО ВЛИЯНИЯ В АВТОНОМНОЙ ВЕТРОЭЛЕКТРОСТАНЦИИ

Описана проблема интеллектуального управления ветропарками. Разработан интеллектуальный алгоритм управления, учитывающий аэродинамические взаимодействия ветровых турбин между собой, которые снижают выработку электроэнергии. Автономная ВЭС смоделирована в среде Matlab/Simulink. Проведено сравнение двух режимов работы автономной ВЭС: с применением интеллектуального алгоритма управления и без применения.
ИНТЕЛЛЕКТУАЛЬНЫЙ АЛГОРИТМ УПРАВЛЕНИЯ; ВЕТРОУСТАНОВКА; АВТОНОМНАЯ ВЕТРОЭЛЕКТРОСТАНЦИЯ.

The ways of transforming wind energy into mechanical energy have been known since the ancient times. However, only at the beginning of the XX century V.P. Vetchinkin developed the theory of an ideal wind wheel based on the theory of an ideal propeller and introduced the term «power efficiency coefficient» (*Cp*). Later the German physicist Betz and the Russian scientist N.E. Zhukovsky proved that an ideal wind turbine cannot produce power greater than 16/27 (*Cp* = 0.593) of wind power [1].

From the practical point of view, Professor G.H. Sabinin suggested the most complete description of the «theory of ideal wind turbine». According to him, the ideal wind turbine's power efficiency coefficient is equal to 0.687.

In 2001 the complex analytical model (GGS) based on Navier–Stokes equations was developed. This model states that *Cp* = 61 %. Besides estimation, it was proved on the supercomputer based on the finite element method that the value of *Cp* is inside the interval [0.593; 0.61].

Power, taken from the wind, depends on the power efficiency coefficient and is calculated by the formula [2]:

$$C_p = \frac{P_{WF}}{P_U}, \tag{1}$$

where P_{WF} – wind turbine power; P_U – airflow power.

Thus, power efficiency coefficient C_p is a variable, which depends on two control wind turbine parameters – its blade pitch angle β and rotor's angular velocity ω. Therefore, we may achieve optimal values of airflow power taken for any wind speed: by regulating both control parameters: ω и angle β[3]:

$$C_p = f(\omega, \beta). \tag{2}$$

For low-power wind turbines (up to 50 kW) regulation of the angle is energetically unfavorable, since it requires additional actuators. For them, the immediate challenge is to ensure the optimal rotor speed constant at a fixed angle blade [4].

Current Problems

Wind turbines interrelation is another important problem [4, 5]. Wind farm efficiency

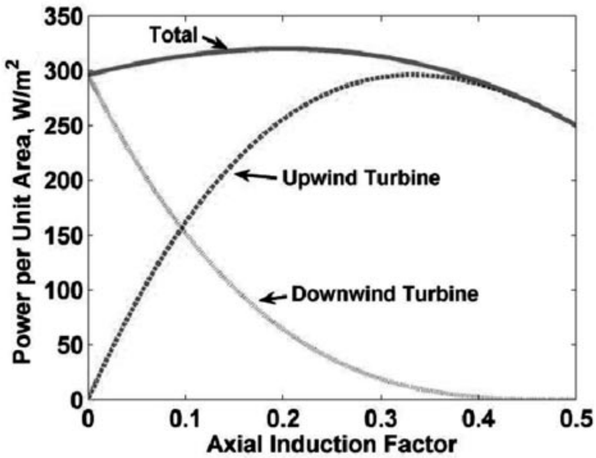


Fig. 1. Wind-produced power versus axial induction factor for a two-turbine array in which one turbine is downwind of the other

is calculated as a ratio of real power produced by the farm to the expected sum of all powers produced by each wind turbine:

$$\eta = \frac{W_{WF}}{W_{WT_i}}, \quad (3)$$

where W_{WF} — total power produced by wind farm; W_{WTi} — power produced by the turbine with index *i*.

Thus, power generated by the successive couple of turbines depends directly on an axial induction factor of the first turbine e(Z). As presented in Fig. 1, maximum power of the wind turbine matches the optimal value of an axial induction factor e1 = 0.33, and the power generated by the second wind turbine falls down. And the maximum produced power may be achieved when e1 = 0.2 [6].

This fact proves that wind farm optimization should involve the optimization of each turbine in connection with others. Wind turbines should be considered as a distributed system with strong interactions and each turbine should be optimized as part of the complex system. So the wind farm power regulation requires a special intellectual control algorithm.

Intellectual Control Algorithm

In our work we have designed the system which takes into account the interaction between turbines. However, our approach is not focused on creating the concrete separate algorithms to optimize the work of a turbine in accordance with aerodynamic interaction. In our work we include this important aspect into a global high-level control algorithm for small and medium sized f-grid wind farms, which have the major potential for Russian decentralized regions.

Two major goals are taken into account in this algorithm:

maximizing wind farm's power generation;

providing frequency (50 ± 1Hz) and voltage (380 ± 5 %) stability in off grid power systems [7, 8].

As part of the first objective, we developed a model which includes aerodynamic interactions in the process of calculating the generated power. In our work we suggest applying aerodynamic interaction of wind turbines by evaluating losses. First wind speed behind each turbine is determined as

$$V = u\left(1 - (1 - \sqrt{1-m})\left(\frac{D}{(D+2kX)^2}\right)\right), \quad (4)$$

where u — wind initial speed before first turbine; V — airflow speed behind the turbine on distance X; D — wind turbine wheel's diameter; m — wind turbine moment of wheel; $k \approx 0.05 - 0.075$ — whirlwind collapse constant.

The parameter m depends on mechanical parameters $m = f(T_c, \beta)$, where T_c — mechanical moment of resistance; β — blade installation angle.

So the effectiveness of a wind turbine as part of the wind farm determent is the following function with m on control parameters:

$$\eta_{WF} = f(\vec{m}, \vec{\varphi}, \vec{\gamma}) \rightarrow \eta_{WF} = f(\vec{T}_c, \vec{\beta}, \vec{\gamma}), \quad (5)$$

where γ — air flow angle relatively rotor's axis.

Power generation effectiveness will be included into a formula of generated power:

$$\begin{aligned} W_{WF} &= \int P_{WF}(u, T_c, \beta, \vec{n}_{WT}, \vec{m}, \vec{\varphi}_t, \vec{\gamma}, \vec{P}_t, \vec{P}_W, \\ q, L, I, U) &= \int [\sum_{i=1}^{n} (N_{WTi}(U_i)\eta_{reg}(T_c, \beta) \times \\ &\times \eta_{or}(\gamma_{ti})\eta_{gen}(P_e)\eta_{gear}(q, P_t)\eta_{el}(P_b))] \times \\ &\times \eta_{tr}(L, I, U)\eta_{diss}(P_w, \vec{\varphi}_t)\eta_{WF}(\gamma_t, \vec{m}). \end{aligned} \quad (6)$$

The formula contains other coefficients which determine other losses such as η_{reg} — loss-

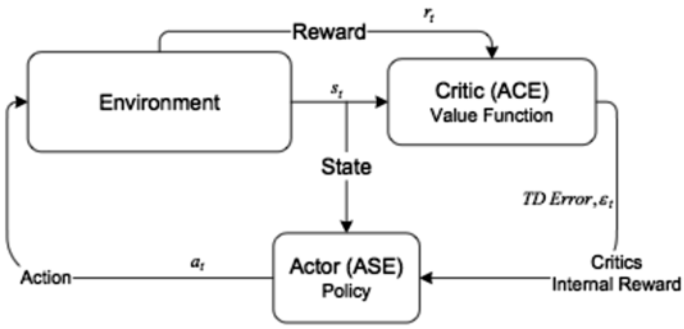


Fig. 2. Adaptive Heuristic Critic (AHC) model

es of rotor's frequency regulation, η_{or} – orientation regulation losses, etc. The final criteria will be as follows:

$$\begin{cases} W_{WF}(\vec{T}_c, \vec{\gamma}_t) \to \max \\ f = (50 \pm 1)Hz \\ U = (380 \pm 18)V \end{cases}. \quad (7)$$

These criteria will be integrated into an algorithm of the agent. This algorithm is based on an adaptive critic design. The structure will be based on two components:

the adaptive critic;

the executor.

The critic's aim is to evaluate the system state and come up with the inner reward, based on this evaluation. In general, it is some sort of the evaluation of function's approximation according to current strategy, which is translated to the executor. Both the executor and the critic may learn at the same time. The executor tries to choose *t*-optimal strategy (strategy which is optimal while time is *t*) based on the value, which is obtained from the critic. At the same time the critic tries to set the evaluation function according to the current strategy, formulated by the executor. The model of an adaptive heuristic critic is shown in Fig. 2.

The multiagent system suggested has a hierarchical structure, which is divided into multiagent groups. Each multiagent group has an agent-parent, which controls the behavior of the whole group by gathering the information about the group using local and group situational vectors and by modifying its behavior using local and group plan vectors [9].

A multiagent system for autonomous wind farm control has a hierarchical structure. The wind farm control panel is situated in the root of the hierarchy and carries out observation and control over the whole system. On the next levels of the hierarchy there are nodes which implement control over wind turbines, then wind turbines. The model of energy consumers is situated on the top level of the hierarchy. The hierarchy of the multiagent system can be

Energy production in two work modes

Wind speed 5 m/s			Wind speed 10 m/s		
	Without control, kWh	Intellectual algorithm, kWh		Without control, kWh	Intellectual algorithm, kWh
Turbine 1	0.238	0.472	Turbine 1	2.311	4.813
Turbine 2	0.052	0.321	Turbine 2	2.235	3.375
Turbine 3	0.052	0.285	Turbine 3	2.128	2.953
Turbine 4	0.016	0.235	Turbine 4	2.056	2.714
Turbine 5	0.013	0.239	Turbine 5	1.999	2.650
$\sum P$	0.371	1.552	$\sum P$	10.729	16.505

viewed as a graph-tree with the only distinction between two direct descendants of one node in the hierarchy, so the additional channels of energy transmission can be introduced. These additional channels can be used for balancing the amount of electrical energy and as an alternative channel for electrical energy delivery in case of main channels failure. [10]

The multiagent system can be viewed as a directed weighted graph. In this graph nodes are intelligent agents and edges of the graph are channels for electrical energy delivery.

Wind speed and direction known at least at one point influence the input data. Output data includes optimal wheel frequency, interpreted by setting electromechanical moment on a turbine generator's shaft regulated by PWM (pulse-width modulation). For blade-rotating turbines frequency is regulated by blade pitch rotation. Data for learning/control grouped into sectors

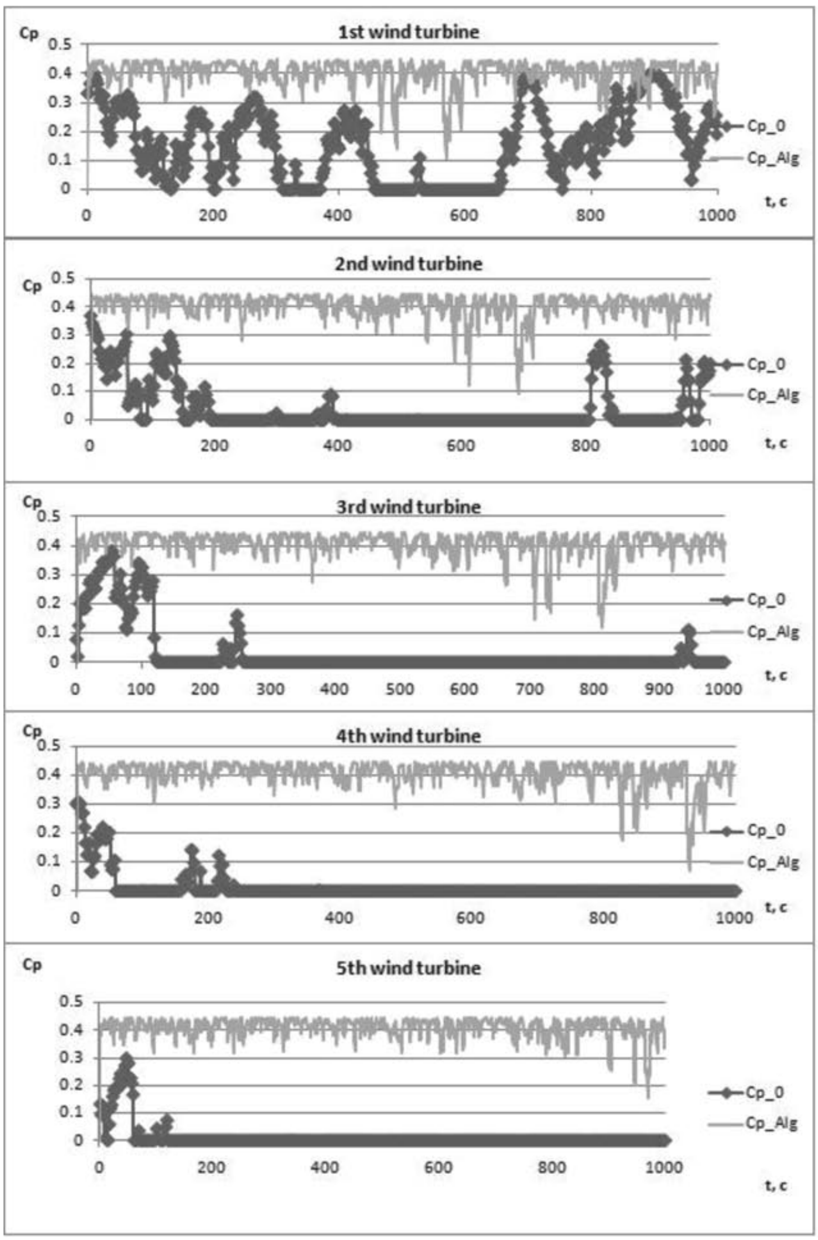


Fig. 3. Comparison of power efficiency coefficient *Cp* using intellectual control algorithm and without control algorithm. Wind speed 5 m/s

(12 units) depends on wind speed. Each control parameter matches its sector. At the same time, during the learning process the delay between the data received from measurement stations and wind farm reaction is considered.

The mathematical model of 100 kW-power wind farm and wind farm control algorithm were implemented in the simulation environment Matlab/Simulink. The wind farm modelled consists of 5 identical wind turbines, 20 kW each. The experiments were conducted to make the comparative analysis.

Wind turbines placements and wind direction were selected in the way the turbines interrelations took place.

Wind farm worked in two modes: with the application of the intellectual control algorithm and without control. Working without control implies turbines rotated at a constant velocity.

Then wind speed was set at two different levels: 5 m/s and 10 m/s.

The goal was to compare the total output power and the power generation of each wind turbine in two control modes.

The intellectual control algorithm allows to increase the output wind farm power in comparison with wind farm work without control. At wind speed of 5 m/s the algorithm efficiency is 318 %, at wind speed of 10 m/s the algorithm efficiency is 53.8%. The results are presented in Table and Fig. 3.

The wind farm power increase was reached by raising power efficiency coefficient and minimizing wind turbines interrelations. The intellectual control algorithm provides stabilized wind farm work even with varying wind speed, wherein every turbine will rotate with optimal rotor's angular velocity. Experiments results show high efficiency of the intellectual control algorithm and the expedience of using it at real wind farms.

REFERENCES

1. **Bogdanov K. Yu.** Science synthesis – weapon of learning in XXI century. *Physics*, 2006, No. 9.

2. **Pandey K., Tiwari A.** Maximum power point tracking of wind energy convertion system with permanent magnet synchronous generator. *Internat. Journal of Engineering Research & Technology,* 2012, Vol. 1, Iss. 5.

3. **Azouz M., Shaltout A., Elshafei M.A.L.** Fuzzy Logic Control of Wind Energy Systems, *Proceedings of the 14th Internat. Middle East Power Systems Conference, Cairo Univ., Egypt,* 2010.

4. **Johnson K., Naveen T.** Wind farm control: addressing the aerodynamic interaction among wind turbines, *American Control Conference*, 2009 .

5. **van Dam F., Gebraad P., van Wingerden J.-W.** A Maximum Power Point Tracking Approach for Wind Farm Control, *Proc. of the 'The Science of Making Torque from Wind' Conference, Oldenburg, Germany,* 2012.

6. **Marden J., Ruben S., Pao L.** A Model-Free Approach to Wind Farm Control Using Game Theoretic Methods, *IEEE Transactions On Control Systems Technology*, Vol. 21, No. 4.

7. **Li T., Feng A.J., Zhao L.** Neural Network Compensation Control for Output Power Optimization of Wind Energy Conversion System Based on Data-Driven Control. *Journal of Control Science and Engineering*, 2012.

8. **Bayat M., Sedighizadeh M., Rezazadeh A.** Wind Energy Conversion Systems Control Using Inverse Neural Model Algorithm. *Internat. Journal of Engineering and Applied Sciences*, 2010.

9. **Shkodyrev V.P.** Innovatsionnyie tehnologii v zadachah upravleniya bolshimi raspredelennyimi sistemami [Innovative technologies in the problems of managing large distribution systems]. *Nauchno-tekhnicheskie vedomosti SPbGPU. Nauka i obrazovanie.* Ustoichivoe razvitie i energetika. St. Peterburg: SPbGPU Publ., 2012, No. 3, Pp. 73–76. (rus)

10. **Arseniev D.G., Shkodyrev V.P., Potekhin V.V., Kovalevsky V.E.** Multiagent approach to creating an energy consumption and distribution system. *Proc. of the VII Internat. Conference International Cooperation in Engineering Education,* St. Petersburg, Russia, 2012.

СПИСОК ЛИТЕРАТУРЫ

1. **Bogdanov K. Yu.** Science synthesis – weapon of learning in XXI century // Physics. 2006. No. 9.

2. **Pandey K., Tiwari A.** Maximum power point tracking of wind energy convertion system with permanent magnet synchronous generator // Internat. Journal of Engineering Research & Technology (IJERT). 2012. Vol. 1. Iss. 5.

3. **Azouz M., Shaltout A., Elshafei M.A.L.** Fuzzy Logic Control of Wind Energy Systems // Proc. of the 14th Internat. Middle East Power Systems Conf. (MEPCON'10), Cairo Univ., Egypt, 2010.

4. **Johnson K., Naveen T.** Wind farm control:

addressing the aerodynamic interaction among wind turbines // American Control Conference. 2009 .

5. **van Dam F., Gebraad P., van Wingerden J.-W.** A Maximum Power Point Tracking Approach for Wind Farm Control // Proc. of the Science of Making Torque from Wind Conf., Oldenburg, Germany. 2012.

6. **Marden J., Ruben S., Pao L.** A Model-Free Approach to Wind Farm Control Using Game Theoretic Methods // IEEE Transactions On Control Systems Technology. Vol. 21, No. 4.

7. **Li T., Feng A.J., Zhao L.** Neural Network Compensation Control for Output Power Optimization of Wind Energy Conversion System Based on Data-Driven Control // Journal of Control Science and Engineering. 2012.

8. **Bayat M., Sedighizadeh M., Rezazadeh A.** Wind Energy Conversion Systems Control Using Inverse Neural Model Algorithm // Internat. Journal of Engineering and Applied Sciences. 2010.

9. **Шкодырев В.П.** Инновационные технологии в задачах управления большими распределенными системами // Научно-технические ведомости СПбГПУ. Наука и образование. Устойчивое развитие и энергетика. СПб.: Изд-во СПбГПУ, 2012. № 3-1(154). С. 73–76.

10. **Arseniev D.G., Shkodyrev V.P., Potekhin V.V., Kovalevsky V.E.** Multiagent approach to creating an energy consumption and distribution system // Proc. of the VII Internat. Conf. International Cooperation in Engineering Education. St. Petersburg, 2012.

POTEKHIN, Vyacheslav. V. *St. Petersburg State Polytechnical University.*
195251, Polytekhnicheskaya Str. 21, St. Petersburg, Russia.
E-mail: slava.potekhin@mail.ru

ПОТЕХИН Вячеслав Витальевич – *доцент кафедры систем и технологий управления Санкт-Петербургского государственного политехнического университета, кандидат технических наук.*
195251, Россия, Санкт-Петербург, ул. Политехническая, д. 21.
E-mail: slava.potekhin@gmail.com

LOGINOV, Anatoly L. *St. Petersburg State Polytechnical University.*
195251, Polytekhnicheskaya Str. 21, St. Petersburg, Russia.
E-mail: a_loginov@inbox.ru

ЛОГИНОВ Анатолий Леонидович – *доцент кафедры систем и технологий управления Санкт-Петербургского государственного политехнического университета, кандидат технических наук.*
195251, Россия, Санкт-Петербург, ул. Политехническая, д. 21.
E-mail: a_loginov@inbox.ru

AKULOV, Dmitrii. V. *St. Petersburg State Polytechnical University.*
195251, Polytekhnicheskaya Str. 21, St. Petersburg, Russia.
E-mail: dmitryab87@ya.ru

АКУЛОВ Дмитрий Вадимович – *инженер кафедры систем и технологий управления Санкт-Петербургского государственного политехнического университета.*
195251, Россия, Санкт-Петербург, ул. Политехническая, д. 21.
E-mail: dmitryab87@ya.ru

POLETAEV, Ivan. G. *St. Petersburg State Polytechnical University.*
195251, Polytekhnicheskaya Str. 21, St. Petersburg, Russia.
E-mail: polivager@mail.ru

ПОЛЕТАЕВ Иван Георгиевич – *ассистент кафедры систем и технологий управления Санкт-Петербургского государственного политехнического университета.*
195251, Россия, Санкт-Петербург, ул. Политехническая, д. 21.
E-mail: polivager@mail.ru

KONISHCHEV, Mikhail A. *St. Petersburg State Polytechnical University.*
195251, Polytekhnicheskaya Str. 29, St. Petersburg, Russia.
E-mail: mak@cef.spbstu.ru

КОНИЩЕВ Михаил Анатольевич – *Директор учебно-научного центра технологии преобразования энергии возобновляемых источников Санкт-Петербургского государственного политехнического университета.*
195251, Россия, Санкт-Петербург, ул. Политехническая, д. 29.
E-mail: mak@cef.spbstu.ru

UDC 004:005=111

D.A. Tyzhnenko, A.V. Vasilyeva, P.M. Valov, V.V. Potekhin

INTELLIGENT INFORMATIONAL-MEASURING SYSTEM FOR MONITORING AND OPTIMIZATION OF POWER CONSUMPTION

An intelligent informational-measuring system for monitoring and optimization of power consumption was developed for the use by housing and communal services.

AUTOMATED INFORMATIONAL-MEASURING SYSTEM FOR ACCOUNTING OF POWER CONSUMPTION; AUTOMATED SYSTEM FOR ACCOUNTING AND CONTROL OF POWER RESOURCES; INTELLIGENT INFORMATIONAL-MEASURING SYSTEM; INTELLIGENT EQUIPMENT FOR ACCOUNTING OF POWER RESOURCES; ANALYSIS AND OPTIMIZATION OF POWER CONSUMPTION.

Д.А. Тыжненко, А.В. Васильева, П.М. Валов, В.В. Потехин

ИНТЕЛЛЕКТУАЛЬНАЯ ИНФОРМАЦИОННАЯ СИСТЕМА МОНИТОРИНГА И ОПТИМИЗАЦИИ ПОТРЕБЛЕНИЯ ЭНЕРГОРЕСУРСОВ ДЛЯ ЖИЛИЩНО-КОММУНАЛЬНОГО ХОЗЯЙСТВА

Разработана интеллектуальная информационная система мониторинга и оптимизации потребления энергоресурсов для жилищно-коммунального хозяйства.

АВТОМАТИЗИРОВАННАЯ ИНФОРМАЦИОННО-ИЗМЕРИТЕЛЬНАЯ СИСТЕМА УЧЕТА ЭНЕРГОПОТРЕБЛЕНИЯ; АВТОМАТИЗИРОВАННАЯ СИСТЕМА КОНТРОЛЯ И УЧЕТА ЭНЕРГОРЕСУРСОВ; ИНТЕЛЛЕКТУАЛЬНЫЕ ИНФОРМАЦИОННО-ИЗМЕРИТЕЛЬНЫЕ СИСТЕМЫ; ИНТЕЛЛЕКТУАЛЬНЫЕ УСТРОЙСТВА УЧЕТА ЭНЕРГОРЕСУРСОВ; АНАЛИЗ И ОПТИМИЗАЦИЯ ЭНЕРГОПОТРЕБЛЕНИЯ.

The project task was to develop an informational-measuring system for monitoring and accounting of power consumption and a decision support system for power resource optimization and increase of power efficiency in economy of municipal formations and communal services. The work was achieved through an international cooperation with a scientific research organization from the Republic of South Africa (RSA). The final product is an intelligent informational-measuring system which is used to obtain complete information about the parameters of power consumption. After that the collected information is analyzed for decision support of power use optimization at the farm level (in the RSA), on the one hand, and at the level of small-scale enterprises and management companies (in the Russian Federation), on the other hand.

One of the priorities of the energy strategy is the social orientation in the fuel-power complex development, or, in other words, the increase in level of the population life. At the same time, new market conditions do not emphasize large-scale raise of power source production rate, but more efficient ways of power utilization, which is saving. The support of power infrastructure is carried out by the private management company or the homeowner's association (the HOA). Private households, such as country houses, cottages, town houses, flats, as well as small and middle-size business enterprises, may act as individual consumers [1—3].

Thus, there are two levels of influence upon infrastructure:

level of management company;

separate households owned by private individual.

Let us consider an organization as a system. The system power consumption is formed of the following major items: power consumption by the infrastructure of the system, power consumption for public purposes, power utilization by individual consumers and losses that appear during transfer of electric power:

$$E_{\text{syst}} = E_{\text{infr}} + E_{\text{pp}} + E_{\text{ic}} + E_{\text{losses}}, \qquad (1)$$

where E_{syst} – total power consumption by the system, kilowatt-hour (kwh); E_{infr} – power consumption for infrastructure support, kwh; E_{pp} – power consumption for public purposes, kwh; E_{ic} – power consumption by individual consumers, kwh; E_{losses} – electric power losses, kwh.

Optimization problem of power consumption can be put in two ways: either minimization of power consumption or cost of power minimization.

All systems of parameter acquisition and control are built according to the same principles. First of all, it is the acquisition of equipment parameters, which are usually presented as electrical signals. After that data are recorded. Finally, the identification of inadmissible parameters is performed followed by a corresponding indication or elaboration of control action. The acquisition system of flow data consists of hardware connected to the technological setting and software, which realizes procedures for recording and storing the acquired data, including error handling and control of characteristics values. In case of multiple settings hardware items may be duplicated, therefore every hardware item receives a unique identification number or address [4–8].

The multiagent approach was proposed as a method of power distribution control.

Multiagent systems can be successfully applied to solving various control problems. These systems can be viewed as a set of intelligent agents that can interact with each another. The effective work organization is gained by development and implementation of algorithms, which allows a group of agents to fulfill tasks, which cannot be fulfilled by agents individually.

The situation at a point of time t_i is formed by states of all system variables which they have at a point t_i and which they had in the past (at points of time $t_j, j \in \{0, ..., i-1\}$). The situation may also be defined as a logical term that is made of a starting situation (usually defined as S_0) and all situation, which result from the starting situation through application of some actions to it. The situation constitutes the fact that a combination of events is true or false at the same time in a certain place. For example, the situation S can be defined as $S = e_i \wedge \overline{e_j}$. The situation instances group into the situation classes, where $S_{CL} = \{S_{\perp}, S_0, ..., S_k\}$ is a set of situation classes. The situation class is a classification of situations into semantically equivalent groups. For example, the situation class S can be defined as $S = e_i \wedge (e_j \vee \overline{e_j})$, if e_j does not correspond to any of the actions that are possible in this situation class.

A group of agents can «feel» similar events by monitoring the state of the system and forming probabilistic distributions of origins for certain events and situations. An identical set of plans P, containing a special plan P_i for every situation in the class S_i is assigned to each group member. The plan $P_{\perp}$, that corresponds to the class $S_{\perp}$, forms a situation, when all agents in the group do not react to events. Plans are organized into a hierarchy. At the top level of the hierarchy are the most common situations, whereas specific situations are at the bottom levels of the tree structure.

Function $U : P \times S_{CL} \rightarrow R$ defines usefulness of a plan execution command for a situation class. As a situation class is unique for all system positions in the working environment (WE) and does not depend on other locations, we can consider only one position of the system in the WE. Therefore, in order to simplify mathematical notation, we delete, where it is possible, the index, that is responsible for the location of the system in the WE. Thus, the function U must satisfy the following constraints:

$$s \in S_i \text{ and } s \in S_j \text{ and } S_i \subseteq S_j \rightarrow U(P_i, S_i) \geq U(P_j, S_i); \qquad (2)$$

$$s \notin S_i \text{ and } s \in S_j \text{ and } S_i \subseteq S_j \rightarrow U(P_i, S_j) < U(P_j, S_j); \qquad (3)$$

$$U(P_{\perp}, S_{\perp}) = 0. \qquad (4)$$

Let R denote a group of N autonomous agents R_j $(j = \overline{1, N})$, , which cooperate in a specified environment E. State of each agent $R_j(t) \in R(j = \overline{1, N})$ at the time instant t is defined by the following vector-function:

$$R_j(t) = [r_{1,j}(t), r_{2,j}(t), \ldots, r_{h,j}(t)]^{\mathrm{T}}, \quad (5)$$

where $r_{i,j}(t)$ $(i = \overline{1,h})$ are state variables of agent j.

The state of an agent group is obtained by the following vector-function:

$$R(t) = (R_1(t), R_2(t), \ldots, R_N(t)). \quad (6)$$

State of the WE around agent j at the time instant t is obtained by the following equation:

$$E_j(t) = [e_{1,j}(t), e_{2,j}(t), \ldots, e_{w,j}(t)]^{\mathrm{T}}, \quad (7)$$

where $e_{i,j}(t)$ $(i = \overline{1,w})$ are parameters of the environment that surrounds agent j.

Hence, state of the WE, where intelligent agent operate at the time instant t, provided that the environment is stationary, is defined by the following expression:

$$(t) = (E_1(t), E_2(t), \ldots, E_N(t)). \quad (8)$$

If agents are absent, then $E_i(t) = E_i =$ $=$ const.

The WE and agents, cooperating within the WE, together comprise a system called the «agents-environment group». State of the «agents-environment group» system at a time instant t can be viewed as a system state, defined by a couple $S_c = (R, E)$.

States of the «agents-environment group» system can be described as points in a $(N(h+w))$ – dimensional state space $\{S_c\}$. The initial system state can be considered as the situation $S_c^0 = (R^0, E^0)$, that is defined by the following vector-functions:

$$R^0 = R(t_0), E^0 = E(t_0), \quad (9)$$

which corresponds to the time instant t_0 of the work start of the agent group. Thereafter the final state of the system is denoted as $S_c^f = (R^f, E^f)$ and is described by the following expressions:

$$R^f = R(t_f), E^f = E(t_f), \quad (10)$$

which corresponds to the time instant t_f of the end of work of the agent group.

The «agents-environment group» state at a present instant t' is called the current system state. This state $S_c^{t'} = (R^{t'}, E^{t'})$ is defined by the following equations:

$$R^{t'} = R(t'), E^{t'} = E(t'), \quad (11)$$

which corresponds to the current time instant t'.

A group of agents R is acting somehow translating the system from the initial state into the ending state. It is assumed, that every agent $R_j \in R$ $(j = \overline{1,N})$ is capable of performing some actions that can be described by the following vector-functions:

$$A_j(t) = [a_{1,j}(t), a_{2,j}(t), \ldots, a_{m,j}(t)]^{\mathrm{T}}. \quad (12)$$

A set of actions that are available for the agent $R_j \in R$ can be described as points in a m-dimensional action subspace $\{A_j\}$.

A set of actions that can be executed by a group of agents is the union of all actions that can be executed by each agent individually

$$\{A_c\} = \{A_1\} \cup \{A_2\} \cup \ldots \cup \{A_n\}. \quad (13)$$

A set of actions $A_j(t)$ that is necessary to achieve the common goal for a group of agents is the control action of the agent group R. Actions that are executed by a group of agents at the time instant t can be described by the following expression:

$$A_c(t) = (A_1(t), A_2(t), \ldots, A_N(t)). \quad (14)$$

Changes in the state of the system can be described by a system of differential equations of the following form:

$$S_c(t) = f_c(S_c(t), A_c(t), g(t), t). \quad (15)$$

In this case, states and actions of the agent group can be constrained as follows (in general case):

$$S_c(t) \in \{S_c^p(t)\} \subset \{S_c\}, \quad (16)$$

where $\{S_c^p(t)\}$ is a set of admissible states of the system «agents-environment group» at the time t, and

$$A_c(t) \in \{A_c^p(t)\} \subset \{A_c\}, \quad (17)$$

where $\{A_c^p(t)\}$ is a set of admissible actions of the system «agents-environment group» at the time t.

Taking into consideration the above definitions, the control task of the agents group comes to finding in a time interval $[t_0, t_f]$ such actions $\overline{A_j}(t)$ for every agent $R_j \in R$ that satisfies the system of relations (15), initial conditions (9), final conditions (10), constraints (16), (17) and an extreme is provided for the functional:

$$Y_c = \Phi(S_c^f, t_f) + \int_{t_0}^{t_f} F(S_c(t), A_c(t), g(t), t)dt, \quad (18)$$

$$Y_c = \Phi(R_1^f, R_2^f, \ldots, R_N^f, E^f, g(t_f), t_f) + \Psi, \quad (19)$$

$$\Psi = \int_{t_0}^{t_f} F(R_1(t), R_2(t), \ldots, R_N(t), E(t), A_1(t), A_2(t), \ldots, A_N(t), g(t), t)dt, \quad (20)$$

which defines the goal of agents group work and quality of control process.

Actions $\overline{A_j}(t)$ $(j = \overline{1, N})$ are optimal for the group of agents R to achieve the common goal.

It is obvious, that existence of the condition for the control action, that translates the system from one state to the other, is not enough to execute a control problem of a group of agents, which works in a dynamic nondeterministic environment. It is also necessary for the control action to be found within such a short interval that guarantees the system state $S_c(t) = (R(t), E(t))$ does not considerably change it. Therefore, in addition to the system agents state and the action constraints there must be specified time constraints, which determine the amount of time available for the search of action $\overline{A_j}(t)$ $(j = \overline{1, N})$. Put it another way, a condition $t_p \le \tau_p$ must be satisfied, where t_p is the necessary time for the search of the control action, and τ_p is the maximum time that is allotted for the search of the control action. Time τ_p depends on many factors. Firstly, it depends on the speed of the processes that take place in the environment. Secondly, the maximum time τ_p depends on the state changes of the agents group. Finally, changes in the environment influence the maximum possible time for searching for the control action. An agent group is called controlled if there is such a solution to the problem of a group control which described above. It can be found in time $t_p \le \tau_p$ [9]

The developed multiagent system for the control of power consumption and distribution, just as a physical system of the power distribution, holds a hierarchical structure. At the root of the structure there is a management company, which performs monitoring of the whole system and controls its work. Nodes of the next levels of the structure are responsible for power distribution from the management company to the private consumers of the power. In the following individual consumers of the electric power are located from the root level of the hierarchy. The alternative power can be supplied at any level of the system structure, starting from the power distributing management com-

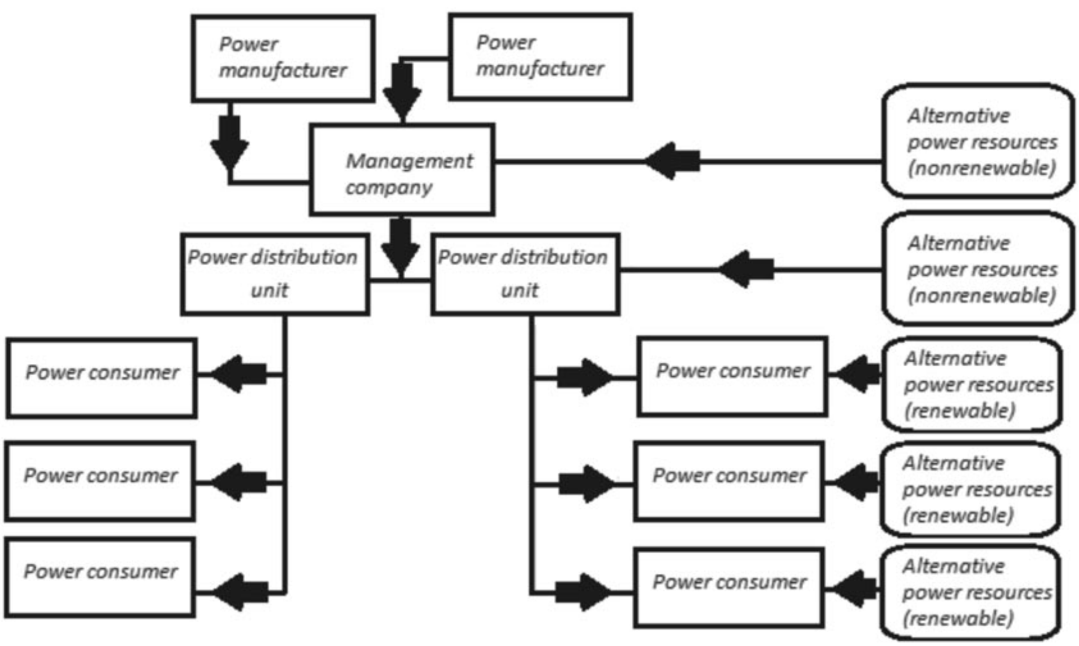


Fig. 1. Hierarchical structure of the system

pany and finishing at the individual consumers level. Trunk electricity is provided only through the management company, which is distributed between all other system members (Fig. 1).

The hierarchy of the multiagent system is almost tree-like with the only difference of additional channels of power transfer between direct descendants of one agent. The extra channels are used either to balance the volume of the power or as reserved transfer channels in case of failure of the main power channels.

The WE of the system is deterministic. The actions of the agents in the environment have a fixed effect and there is no need for an agent to check the steps of its action or to check the fullness of the action every time, when a certain action has been performed.

Moreover, the WE of the system is dynamic, so it can be changed not only by actions of the agents, but also by external forces. Dynamism of the system is shown through appearance of random disturbances in the work of the system, such as the outage of power channels and agents.

Furthermore, the WE of the system is continuous. Thus, the system has an infinite number of states, and it complicates the functioning of the agents. This problem is solved by the approximation of an infinite number of states into a finite set of states that can be perceived by agents in the system.

Finally, the WE is a real-time environment, therefore time characteristics of the system are the key features that are considered while estimating both agents and the whole systems efficiency. During the work of the system time constrains are imposed on the agents response time and on the query time [10].

In the given multiagent system the following agent roles can be singled out:

power manufacturer;

access point to the electricity trunk;

alternative nonrenewable power manufacturer;

alternative renewable power manufacturer;

power distributor;

local area network (LAN) switch;

metropolitan area network (MAN) switch;

power consumer.

The structure of the intelligent agent depends a lot on the roles that are assigned to it during the system development. Every intelligent agent possesses only hardware for the control. In this project control hardware is based on the industrial programmable logical controllers (PLC), manufactured by Siemens. The PLC models in use are Simatic S7-200/300/400.

All agents have the identical basic structure and functionality (Fig. 2). On the one hand, every agent is capable of decision making and can forecast its future volumes of power consumption. On the other hand, the agents can execute different actions depending on their place in the system hierarchy.

System can be divided into two main parts: central control body (CCB) and agents. Every agent cooperated with the local PLC and with the CCB. Agents basically read data from PLC and send it to the main system. Moreover, they have a local database for data storage. The

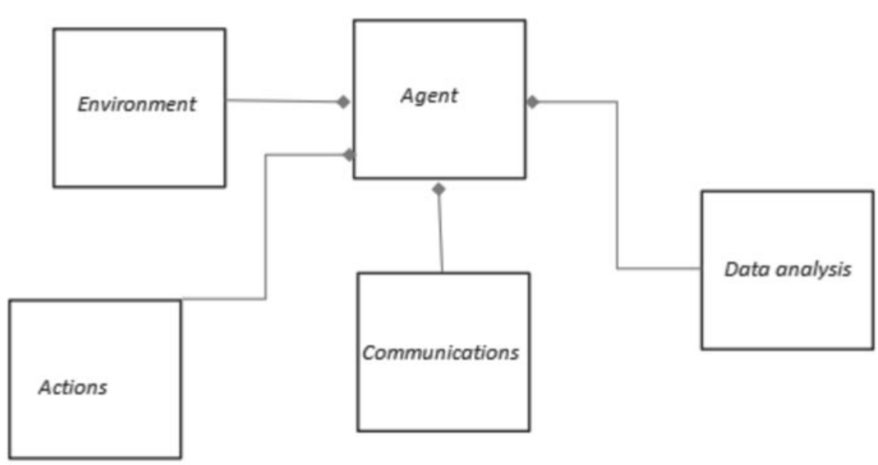


Fig. 2. Basic agent structure

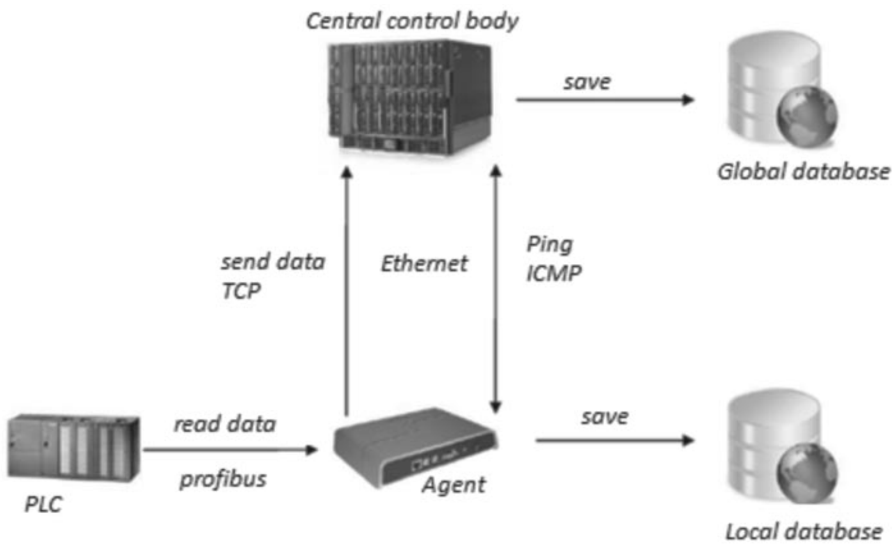


Fig. 3. General structure of an agent

general structure of the agent is depicted in Fig. 3.

The central control body receives data from the agents and stores them in a global database. In addition to this, CCB can check for the link to the agents and can produce reports about possible errors.

The agents as well as the central control body were developed in C# programming language, hence the target platform is the Windows operating system (OS). Nevertheless, it is also possible to use a Linux system provided the specific software called Mono is installed. It allows to run extern C# programs under Linux OS. Communication programming was achieved through Zero Message Queue (0MQ) platform for TCP utilization and with the use of LIBNODAVE library for communication with PLC. The user interface was developed in CSS-framework TwitterBootstrap.

The decision support system structure is presented in Fig. 4. The user of the electric power (operator) interacts with the system via a mobile application or a web-interface and, after the operator the information has been processed by the analyst, he or she receives feedback with

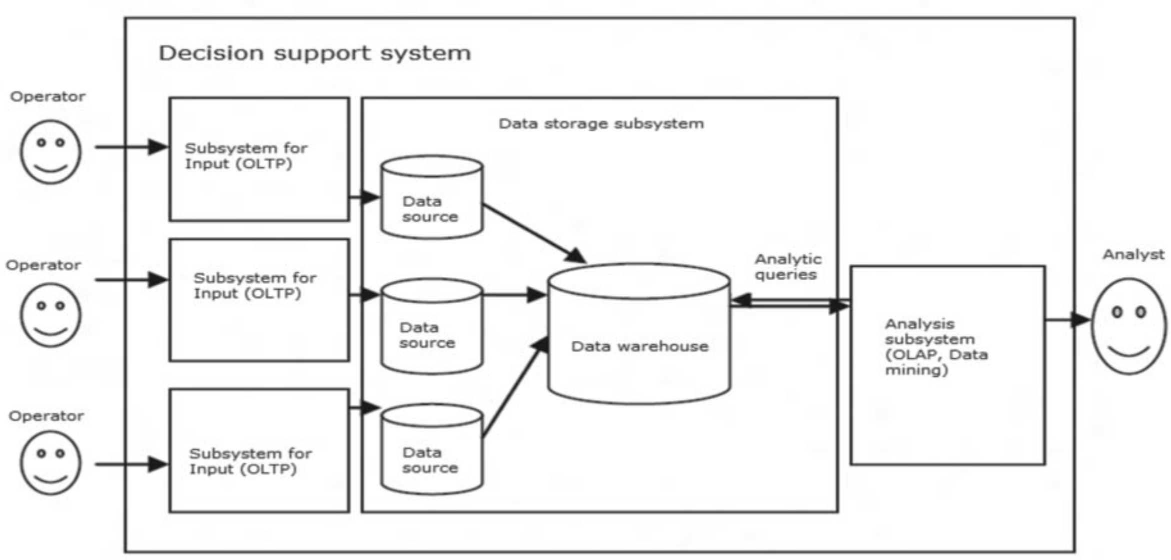


Fig. 4. Decision support system

possible solutions that are directed to minimize the overall payment for electricity.

There are a few interfaces available in the system for the end user:

add/edit device record;

add power consumption data;

report settings (visualization settings);

power consumption reports;

forecast of power consumption/saving.

Development and introduction of the intelligent informational-measuring system for monitoring and optimization of power consumption, which can control distribution, consumption and accounting of electric power, is a promising task.

Within the project we conducted experiments, completed the test bed and patented it as a know-how.

REFERENCES

1. **Wenzel T.P.** Energy Data Source Book for the U.S. Residential Sector, Sept. 1997.
2. **Denisenko V.V.** Kompyuternoye upravleniye tekhnologicheskim protsessom, eksperimentom, oborudovaniyem. Moscow: Goryachaya liniya — Telekom Publ., 2009, 608 p. (rus)
3. **Mylnikov L.A.** Mikroprotsessornyye sistemy sbora i obrabotki dannykh dlya avtomatizatsii kontrolya bystroizmenyayushchikhsya parametrov tekhnologicheskikh protsessov. Perm: PGTU Publ., 2009. (rus)
4. **Barsegyan A.A., Kupriyanov M.S., Stepanenko V.V., Kholod I.I.** Tekhnologii analiza dannykh: Data Mining, Visual Mining, Text Mining, OLAP. St. Petersburg: BKhV-Petersburg Publ., 2008. (rus)
5. **Codd E.F., Codd S.B., Salley C.T.** Providing OLAP (on-line analytical processing) to User-Analysts: An IT mandate. Technical Report, 1993.
6. **Larson B.** Razrabotka biznes-analitiki v Microsoft SQL Server 2005, per. s angl. Moscow: Piter Publ., 2008, 688 p. (rus)
7. **Berger A., Gorbach I.** Microsoft SQL Server 2005 Analysis Services: OLAP i mnogomernyy analiz dannykh. St. Petersburg: BKhV-Petersburg Publ., 2007, 908 p. (rus)
8. Process Automation Handbook. London: Springer, 2007, pp. 351—363.
9. **Loginov A.L., Potekhin V.V., Akulov D.V., Khritonenkov A.K., Poletayev I.G.** Intellektualnaya sistema upravleniya avtonomnoy vetroelektrostantsiyey. *Nauchno-tekhnicheskiye vedomosti SPbGPU. Informatika. Telekommunikatsii. Upravleniye.* St. Petersburg: SPbGPU Publ., 2013, No. 5(181), pp. 50—57. (rus)
10. **Shkodyrev V.P.** Innovatsionnyye tekhnologii v zadachakh upravleniya bolshimi raspredelennymi sistemami. *Nauchno-tekhnicheskiye vedomosti SPbGPU. Nauka i obrazovaniye. Ustoychivoe razvitie i energetika.* St. Petersburg: SPbGPU Publ., 2012, No. 3-1(154), pp. 73—77. (rus)

СПИСОК ЛИТЕРАТУРЫ

1. **Wenzel T.P.** Energy Data Source Book for the U.S. Residential Sector. Sept. 1997.
2. **Denisenko V.V.** Компьютерное управление технологическим процессом, экспериментом, оборудованием. М.: Горячая линия—Телеком, 2009. 608 с.
3. **Мыльников Л.А.** Микропроцессорные системы сбора и обработки данных для автоматизации контроля быстроизменяющихся параметров технологических процессов. Пермь: Изд-во ПГТУ, 2009.
4. **Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И.** Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAPM. СПб.: БХВ-Петербург, 2008.
5. **Codd E.F., Codd S.B., Salley C.T.** Providing OLAP (on-line analytical processing) to User-Analysts: An IT mandate. Technical Report. 1993.
6. **Ларсон Б.** Разработка бизнес-аналитики в Microsoft SQL Server 2005. Пер. с англ. М.: Питер, 2008. 688 с.
7. **Бергер А.** Microsoft SQL Server 2005 Analysis Services: OLAP и многомерный анализ данных. СПб.: БХВ-Петербург, 2007. 908 с.
8. Process Automation Handbook. London: Springer, 2007. С. 351—363.
9. **Логинов А.Л., Потехин В.В., Акулов Д.В., Хритоненков А.К., Полетаев И.Г.** Интеллектуальная система управления автономной ветроэлектростанцией // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. СПб.: Изд-во СПбГПУ, 2013. № 5(181). С. 50—57.
10. **Шкодырев В.П.** Инновационные технологии в задачах управления большими распределенными системами, Научно-технические ведомости СПбГПУ. Наука и образование. Устойчивое развитие и энергетика. СПб.: Изд-во СПбГПУ, 2012. № 3-1(154). С. 73—77.

TYZHNENKO, Denis A. *St. Petersburg State Polytechnical University.*
195251, Polytekhnicheskaya Str. 21, St. Petersburg, Russia.
E-mail: tyzhnenkod@mail.ru

ТЫЖНЕНКО Денис Александрович – *студент кафедры систем и технологий управления Санкт-Петербургского государственного политехнического университета.*
195251, Россия, Санкт-Петербург, ул. Политехническая, д. 21.
E-mail: tyzhnenkod@mail.ru

VASILYEVA, Anastasiia V. *St. Petersburg State Polytechnical University.*
195251, Polytekhnicheskaya Str. 21, St. Petersburg, Russia.
E-mail: vasilyeva.anastasia@bk.ru

ВАСИЛЬЕВА Анастасия Владимировна – *студентка кафедры систем и технологий управления Санкт-Петербургского государственного политехнического университета.*
195251, Россия, Санкт-Петербург, ул. Политехническая, д. 21.
E-mail: vasilyeva.anastasia@bk.ru

VALOV, Pavel M. *University of Waterloo.*
200 University Ave. West Waterloo, ON, Canada N2L 3G1.
E-mail: valov.pm@gmail.com

ВАЛОВ Павел Михайлович – *аспирант кафедры David R. Cheriton School of Computer Science University of Waterloo, магистр.*
N2L 3G1Canada, ON, Waterloo, 200 University Ave. West.
E-mail: valov.pm@gmail.com

POTEKHIN, Vyacheslav V. *St. Petersburg State Polytechnical University.*
195251, Polytekhnicheskaya Str. 21, St. Petersburg, Russia.
E-mail: slava.potekhin@gmail.com

ПОТЕХИН Вячеслав Витальевич – *доцент кафедры систем и технологий управления Санкт-Петербургского государственного политехнического университета, кандидат технических наук.*
195251, Россия, Санкт-Петербург, ул. Политехническая, д. 21.
E-mail: slava.potekhin@gmail.com

UDC 681.3.016=111

S.V. Mescheryakov, D.A. Shchemelinin

ANALYTICAL OVERVIEW OF ZABBIX INTERNATIONAL CONFERENCE 2013

Zabbix International Conference is a growing annual meeting of professionals from various countries and IT companies where Zabbix enterprise-class automated control system is used. Zabbix Conference doesn't have its own proceedings published; only abstracts and presentation media are available on Zabbix web site [1]. This article is an analytical overview of six top presentations that are the most interesting, innovative and valuable for computer science and business.

AUTOMATED CONTROL; DATABASE; BIG DATA; CLOUD COMPUTING; DISTRIBUTED ENVIRONMENT.

С.В. Мещеряков, Д.А. Щемелинин

АНАЛИТИЧЕСКИЙ ОБЗОР МЕЖДУНАРОДНОЙ КОНФЕРЕНЦИИ ЗАББИКС-2013

Международная конференция по системе Заббикс является быстро развивающимся ежегодным собранием профессионалов из различных стран и IT компаний, где используется автоматизированная система управления масштаба предприятия Заббикс. На Конференции не издаются тезисы докладов, на интернет-сайте Заббикс доступны только краткие аннотации и медиапрезентации [1]. Данная статья содержит аналитический обзор шести наиболее интересных докладов, имеющих научную новизну и практическую ценность.

АВТОМАТИЗИРОВАННОЕ УПРАВЛЕНИЕ; БАЗА ДАННЫХ; БОЛЬШИЕ ДАННЫЕ; ОБЛАЧНЫЕ ВЫЧИСЛЕНИЯ; РАСПРЕДЕЛЕННОЕ ОБОРУДОВАНИЕ.

Zabbix automated control system is the enterprise-class solution based on both hardware and software, which can be used for real-time monitoring, alerting, troubleshooting, computer-aided control, capacity analysis and other business purposes in a large distributed production environment [2]. Zabbix can be effectively used for monitoring performance and automated control of multiple hosts in a cloud computing infrastructure. Object-relational databases from different vendors, for example MS SQL, Oracle, and others, can be used to store big data and analyze historical trends.

Zabbix control system is an open source tool and it is free to install at any enterprise. Nevertheless, every implementation of the Zabbix system meets certain problems, such as monitoring data delay, out of memory outage, network storm, database performance bottlenecks, events visibility, etc. So, any feedback getting from all Zabbix experienced users and the cooperative discussion of production issues are always helpful for further improvement of Zabbix interface and internal architecture.

In 2013, the Zabbix International Conference took place on September 6-7 in Riga, Latvia. About 150 registered participants arrived from more than 20 countries, including USA, Brazil, Japan, UK, Germany, France, Spain, Italy, Austria, Netherlands, Poland, Russia, Baltic and Nordic countries, and some other Europeans. The most active countries are UK (13 attendees), Russia (12), Poland (8), Netherlands (6), Lithuania, Italy and Germany (5 per each). The Conference combined 23 presentations in different subjects – database architecture, performance, integration, capacity, and monitoring experience. The sections below are focused on repeatable, recurring and chronic problems of Zabbix enterprise implementation – active control of external objects, database performance, system scalability, integrated interface and dashboards.

Zabbix: Where We Are. The current status of the Zabbix production system, the progress since last year and future plans are announced by Alexey Vladishev, the founder and CEO of Zabbix SIA (Riga, Latvia).

Major features are introduced in the new product release Zabbix 2.2 and they are declared as follows:

1. Given modern tendency towards virtualization and a growing number of virtual machines (VM) populated in enterprise data centers, now VMs monitoring is supported in Zabbix system, particularly VMWare, vCenter, vSphere, including auto-discovery of guest VMs, built-in checks and Zabbix pre-defined templates.

2. Zabbix system performance is improved by a factor of 2 to 5, depending on the number of monitoring hosts, metrics and database size. This advantage can be reached by means of extending Zabbix cache and less number of update transactions on DB server.

Fig. 1. Zabbix Enterprise Appliance ZS-5200

3. Reaction to Zabbix events became faster and the triggers are able to act on the events even when a certain host is disabled and the items are in the unsupported status for some reasons.

4. Zabbix 2.2 enterprise control system has been supported for five years with fixing product bugs, critical, security, and other issues.

The new Zabbix enterprise appliance ZS-5200, made in Japan (Fig. 1), allows for monitoring up to 20K items from external hosts, network devices, and other hardware. That is

Fig. 2. Zabbix Partnership Map

an alternative of using existing Zabbix server with web interface and automatic configuration instead of manual setting up.

The Zabbix enterprise-class automated control system is still an open source product, but not open core, and it is widespread all over the world and it has partnerships with 68 business companies (Fig. 2).

This presentation is available on Zabbix and YouTube web sites [1, 3].

Perobbix + Zabbix DB Monitoring. A new approach for database monitoring is presented by Julio Cesar Hegedus, Senior Linux Consultant, Yenlo BV (Amsterdam, the Netherlands).

Perobbix stands for Perl + Oracle + Zabbix. The idea is to monitor DB servers using Perl scripts running SQL queries against relational databases, such as Oracle, MySQL, etc. Perobbix solution is implemented in Yenlo enterprise infrastructure consisting of 800 hosts, 130 DBs, 80K items.

Perobbix architecture looks like Fig. 3. The approach has the following advantages:

1. There is no Zabbix agent required to install on a host under Zabbix monitoring.

2. Not only a DB host itself but the databases, that is DB integrity, DB sizes, performance of DB transactions, etc., can be monitored.

3. Different methods, drivers, and engines can be used for DB connection. A read-only access is enough that is more safety.

4. There are no limitations for the number of queries in a batch to be executed against DB.

5. Performance of the Perobbix system does not depend on the Zabbix data delay if it happened.

A full presentation is available on Zabbix and YouTube web sites [1, 4].

When It Comes to Scalability. The new distributed architecture of the Zabbix monitoring system is proposed by Leo Yulenets, Operations Tools Lead, RingCentral (USA) [5]. NoSQL solution based on MongoDB open-source document-oriented data warehouse [6] for storing historical data is described.

RingCentral private cloud infrastructure is the biggest one ever monitored by the Zabbix system, consisting of more than 4K hosts total and about 0.5M metric values per minute. Moreover, RingCentral multi-service for over 300K US customers is a fast growing IP-telecommunication industry, having up to 40 % annual increment. Sooner or later, the performance degradation and, as a result, the Zabbix data delay became a showstopper for further scalability in a rapidly growing enterprise environment.

Reducing the number of monitoring items and/or extending polling time intervals is of a big manual effort, which is not a good idea. Adding more servers, proxies, high-performance storages, or other hardware is a

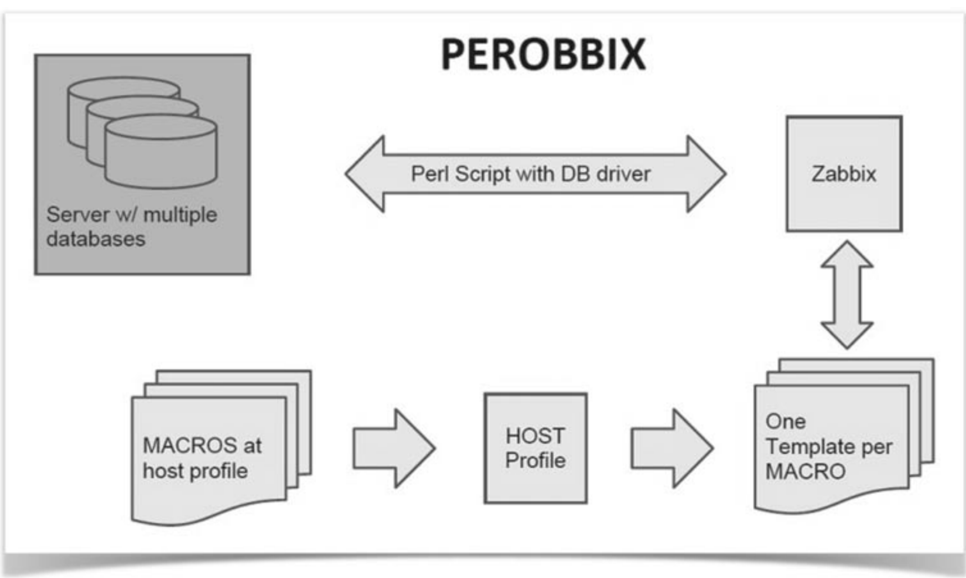


Fig. 3. Perobbix Architecture

too expensive solution and it does not help a lot because MySQL DB is the main bottleneck due to multiple read-write transactions executed simaltaneously.

As a temporary workaround, Zabbix is split into two monitoring systems, each working with a separate database. Special reports and dashboards are created to observe events, alarms, and other control data from both locations on a single monitoring console. As the next step, historical data should be consolidated for more efficient troubleshooting and analysis. But there can be the risk of data delay depending on partitioning between the real-time data amount and history size. Monitoring data delays and reporting gaps for at least 6 min may finally lead to a missed customer service outage that can be of really high cost for the company business.

The alternative Zabbix architecture, named Octopus, is proposed to reach horizontal scalability with a rather low cost (Fig. 4). The Octopus architectural solution is developed and is currently being implemented at RingCentral. Several Zabbix monitoring systems are consolidated on different levels to meet the enterprise monitoring requirements, including data delivery with no delays and gaps, calculation of items and triggers within a short period of time, keeping long period of history data, providing events and data visibility on all levels.

One day statistics is kept as real time data in MySQL relational database, while all the

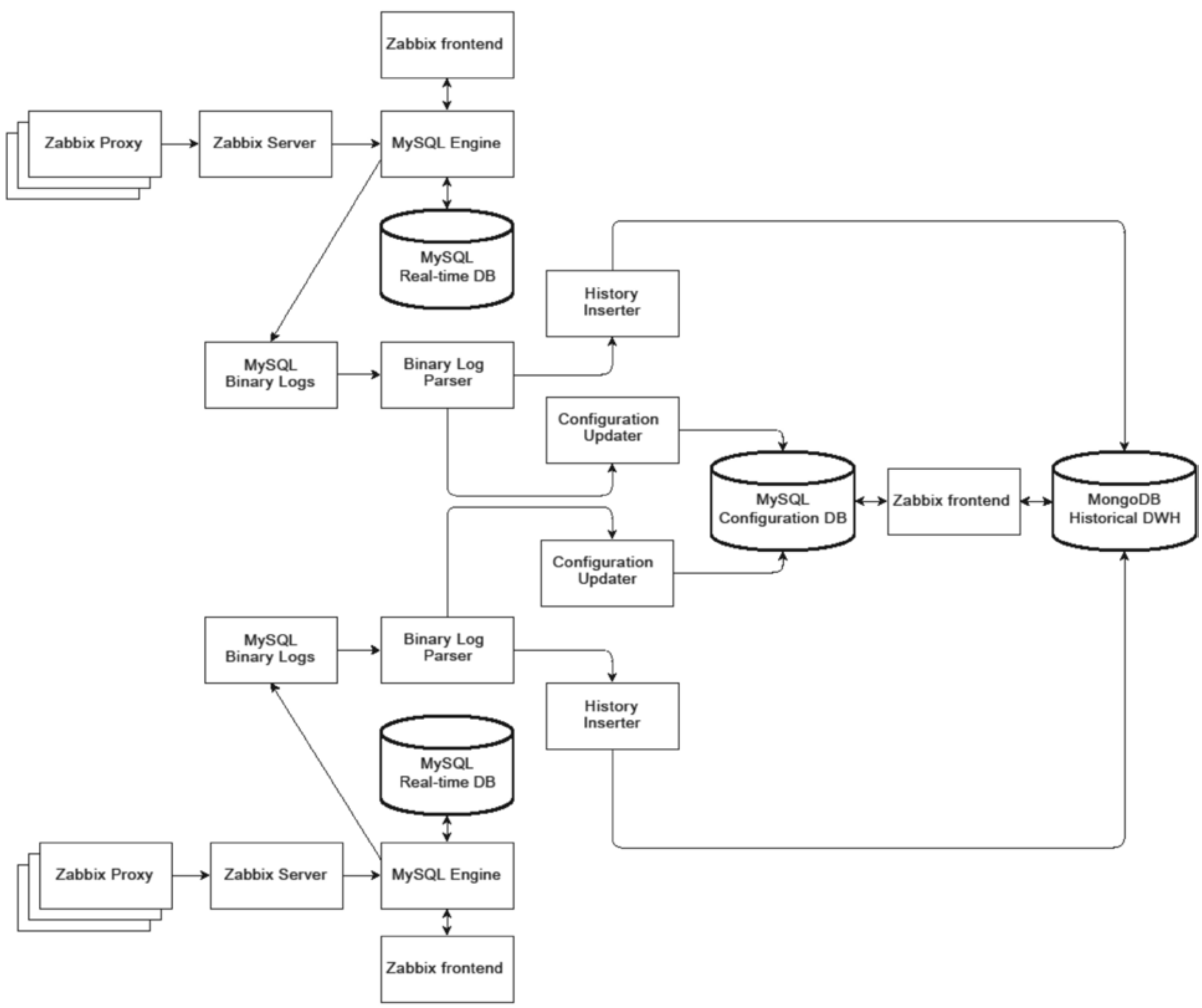


Fig. 4. Octopus Distributed Architecture

other history and the trends are transferred to MongoDB noSQL data warehouse. Real time data does not require separate network storage and can be stored on a local drive, or even in memory. Zabbix system events and alerts are consolidated on a single dashboard using web API.

The following benefits of the new approach and Octopus distributed architecture are expected:

1. Scalability improvement, allowing as many Zabbix monitoring systems as needed.

2. There are no monitoring data delays and no reporting gaps, even in peak time with high workload.

3. The reads from and the writes to history DB can be separated for better performance.

4. Extending historical data retention from 3 to 12 months, or even longer period of time.

The main disadvantage of the Octopus distributed architecture is that programming skills and resources are required to synchronize different Zabbix instances, create consolidated custom dashboards with real time data and the reports with historical data.

Presentation slides and video are available on Zabbix and YouTube official web sites [1, 7].

Which Database is Better for Zabbix? PostgreSQL vs. MySQL. The results of measuring performance of the new product release Zabbix 2.2 along with the latest versions of MySQL and PostgreSQL are presented by Yoshiharu Mori, Consultant, SRA OSS (Japan).

The two categories of the performance stress tests, simple and partitioning ones, are carried out in the cloud computing environment, having 600 hosts, 26K monitoring items, and 10K Zabbix triggers. The number of values that Zabbix server was able to process per second is used as the main metric to estimate the performance of both MySQL and PostgreSQL databases.

Some results are shown in Fig. 5. All the others are provided in the presentation that is available on Zabbix web site [1] and YouTube video [8].

Generally, PostgreSQL is more stable under high input/output load and the Zabbix performance is better than MySQL though the difference is not significant. One more conclusion is that DB tuning (buffering, partitioning, transaction logs) is required in each case of certain implementation.

Integrated Dashboard Design. A new design of the integrated Zabbix dashboard is presented by Lukasz Lipski, IT Operations Specialist, Nordea Bank Polska S.A. (Poland).

An integrated dashboard provides a real business value for the company management. This presentation demonstrates some ways to display Zabbix reports along with external data, such as JIRA, SharePoint, BMC Remedy, and the third party software, with minimal development effort. Real world scripts, written in Perl, Ruby, Twitter Bootstrap, for getting information from external data sources are provided.

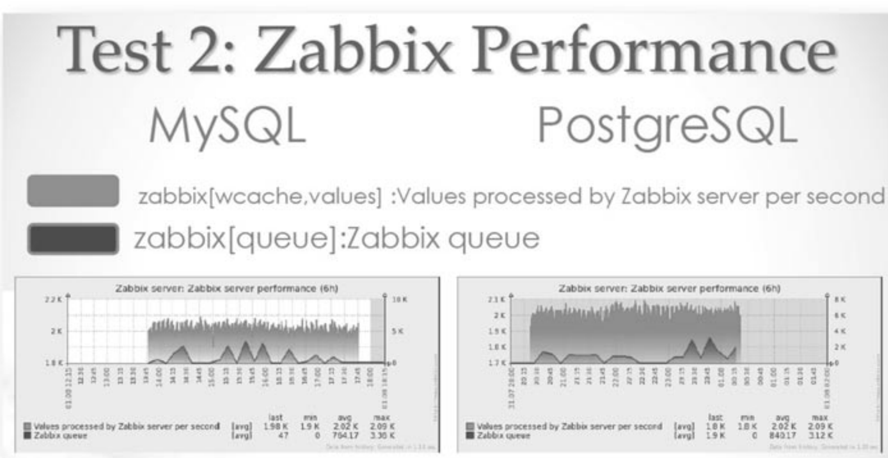


Fig. 5. Zabbix Performance Test

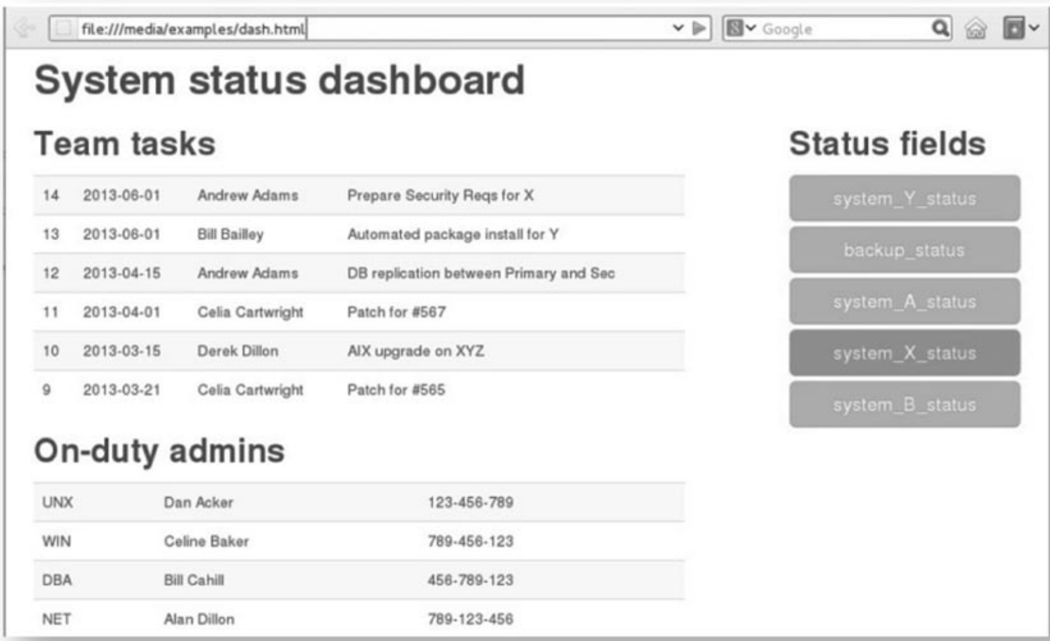


Fig. 6. Example of Integrated Dashboard

One example of the integrated dashboard is shown in Fig. 6. Specific dashboards are implemented in Nordea Bank Polska e-banking system, having 300 hosts and 40K items under Zabbix monitoring:

1. Central monitoring system that shows Zabbix events and alerts from different monitoring systems.

2. Task list, which consolidates Zabbix data with JIRA tasks.

3. Contact list, which combines Zabbix reports with the detailed information from SharePoint.

More examples of integrated dashboards are available in the presentation slides on the Zabbix web site [1] and YouTube video [9].

Complete Log Infrastructure with Zabbix Alerting. This is a conceptual presentation of using Zabbix for log capturing and alerting by Pieter Baele, Linux System Architect, ICTRA NMBS-Holding (the ICT provider of the National Railway Company of Belgium).

The big cloud computing infrastructure needs flexible and powerful tools for log analysis. The Zabbix system is a perfect platform for monitoring resources and alerting, but it is not

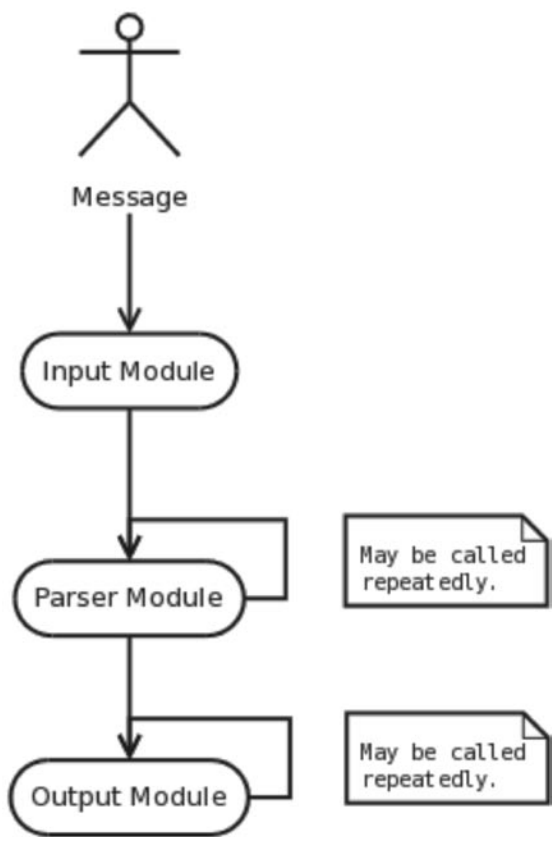


Fig. 7. Log Parsing

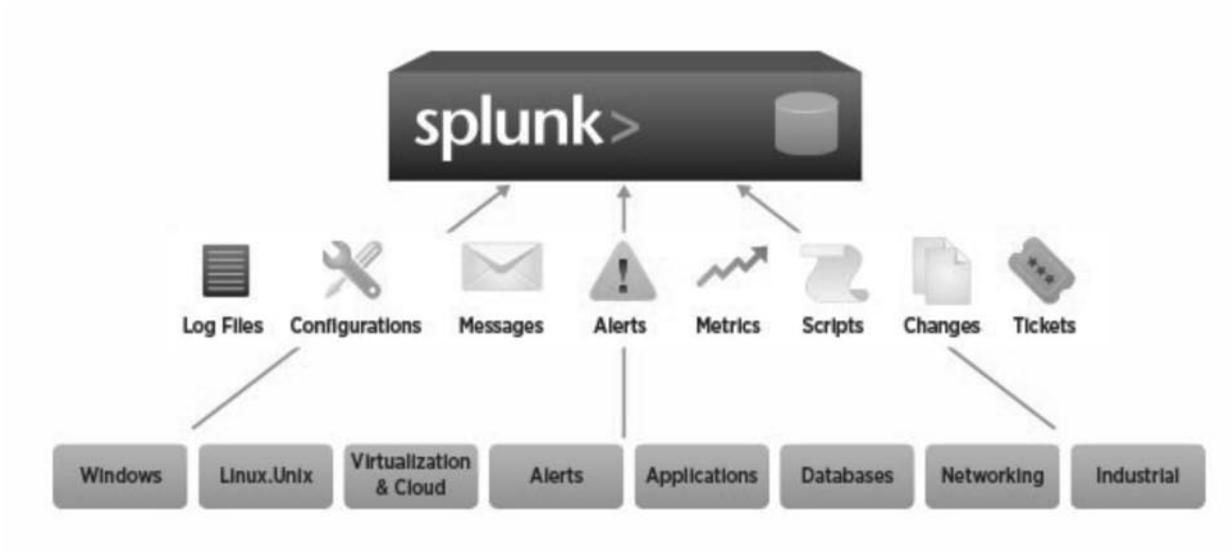


Fig. 8. Splunk Architecture

aimed at storing, transformation, analyzing a huge number of log files. Specific tools are required to capture the logs in a big distributed production environment.

The following concepts of the log infrastructure are proposed depending on real enterprise purposes:

1. Due to popular demand, file content and log file parsing using regular expressions has been implemented in Zabbix 2.2. Returning a part of a string can significantly reduce the workload on the Zabbix database.

2. Rsyslog is an open source logging tool having application level reliability, central log repository, filtering, high-precision timestamps, and configurable outputs.

3. Logstash is easy to deploy and it is used to collect logs, parse and store for later analysis (Fig. 7). Various log formats and filters are supported.

4. ElasticSearch is a distributed search, indexing and analytics technology. It is horizontally scalable and supports archiving on storage. Kibana interface is used as the web frontend.

5. Graylog is a logging tool using Lightweight Directory Access Protocol (LDAP) to distributed information services over the Internet. LDAP integration is good for applications and syslogs.

6. Octopus is an open source log management having LDAP feature and a lot of templates included. It can be integrated with Zabbix alert senders and it is nice for enterprise usage.

7. ELSA stands for Enterprise Log Search and Archive. It uses LDAP and MySQL though the query language is specific. Email alerts to Zabbix server are possible.

8. Splunk is non-free scalable product (Fig. 8). It is easy to install and can be integrated with Zabbix using Zabbix sender.

9. Fluentd seems to have better performance in the line of other tools, but it is not tested enough. The largest user is currently collecting 5 TB of daily logs from 5000 servers with a rate of 50,000 messages per second.

The above approaches, except the 9th, are tested and have experience of monitoring hundreds of Linux servers. Practical examples of custom scripts for both the OS and web applications are shown on presentation slides [1, 10].

The benefit of annual Zabbix International Conference is to share knowledge and experience of using the Zabbix enterprise-class system between IT professionals in order to improve and automate monitoring the multi-host production environment.

SQL, Perl, or other type of scripts is a universal technique that can be used for any external check of the network environment where Zabbix agent is hard to install.

Database is always a bottleneck in the cloud computing infrastructure. DB tuning is required for better performance with no reporting data delay in each case of certain implementation.

The distributed Zabbix architecture is valuable when an enterprise system grew up to its scalability threshold in terms of hosts and monitoring metrics. In the case of the Zabbix database split, the integrated dashboard should be created to consolidate monitoring data from separate sources into one report.

The accuracy of conclusions is also approved by similar investigation [11].

REFERENCES / СПИСОК ЛИТЕРАТУРЫ

1. Zabbix International Conference 2013. Agenda and presentation abstracts. Riga, 2013. Available: http://www.zabbix.com/conf2013_agenda.php

2. Zabbix official web site. Available: http://www.zabbix.com/product.php

3. **Vladishev A.** Zabbix: Where We Are. *Zabbix Internat. Conference*, 2013. Available: http://youtu.be/c8K83MJS-jg

4. **Hegedus J.C.** Perrobix + Zabbix DB Monitoring. *Zabbix Internat. Conference*, 2013. Available: http://youtu.be/pJCV_ui0orQ

5. RingCentral official web site. Available: http://www.ringcentral.com/

6. MongoDB official web site. Available: http://www.mongodb.org/

7. **Yulenets L.** When It Comes to Scalability. *Zabbix Internat. Conference*, 2013. Available: http://youtu.be/1Eq-9q16AOU

8. **Mori Y.** Which Database is Better for Zabbix? PostgreSQL vs. MySQL. *Zabbix Internat. Conference*, 2013. Available: http://youtu.be/LHitd_GTC-w

9. **Lipski L.** Integrated Dashboard Design. *Zabbix Internat. Conference*, 2013. Available: http://youtu.be/_gy4qzyZf_o

10. **Baele P.** A Complete Log Infrastructure with Zabbix Alerting. *Zabbix Internat. Conference*, 2013. Available: http://youtu.be/f5IfBJdj_Xk

11. **Mescheryakov S.** Zabbix Conference Overview. RingCentral Inc., USA, 2013. Available: http://wiki.ringcentral.com/display/OPS/Zabbix+Conference+Overview

---

MESCHERYAKOV, Sergey V. *St. Petersburg State Polytechnical University.*
195251, Politekhnicheskaya Str. 29, St. Petersburg, Russia.
E-mail: serg-phd@mail.ru

МЕЩЕРЯКОВ Сергей Владимирович – *профессор кафедры автоматов Санкт-Петербургского государственного политехнического университета, доктор технических наук, доцент.*
195251, Россия, Санкт-Петербург, ул. Политехническая, д. 29.
E-mail: serg-phd@mail.ru

SHCHEMELININ, Dmitry A. *RingCentral Inc.*
1400 Fashion Island Blvd., San Mateo, CA, USA 94404.
E-mail: dshchmel@gmail.com

ЩЕМЕЛИНИН Дмитрий Александрович – *руководитель департамента развития и эксплуатации облачных IT платформ компании RingCentral, кандидат технических наук.*
1400 Fashion Island Blvd., San Mateo, CA, USA 94404.
E-mail: dshchmel@gmail.com

UDC 681.3.016=111

S.V. Mescheryakov, D.A. Shchemelinin

INTERNATIONAL CONFERENCE FOR THE PERFORMANCE EVALUATION AND CAPACITY ANALYSIS BY CMG

The CMG International Conference for the resource management and performance evaluation of enterprise computing systems is held in the USA annually since 1975. CMG Conference meets together technical experts to share ideas and experiences for the performance and capacity analysis in various areas of industry. The conference proceedings are published by CMG every year, though not for all presentations, and are available on CMG web site [1] for registered members only. CMG papers older than 5 years can be downloaded for free. This article is a review of CMG organization as a whole and CMG 2013 Conference in particular.

PERFORMANCE; CAPACITY; BIG DATA; CLOUD COMPUTING; DISTRIBUTED ENVIRONMENT.

С.В. Мещеряков, Д.А. Щемелинин

АНАЛИЗ МЕЖДУНАРОДНОЙ КОНФЕРЕНЦИИ CMG ПО ОЦЕНКЕ ПРОИЗВОДИТЕЛЬНОСТИ И НАГРУЗКИ

Международная конференция CMG по управлению ресурсами и оценке производительности вычислительных систем масштаба предприятия проводится в США ежегодно, начиная с 1975 года. В Конференции принимают участие технические эксперты, чтобы поделиться идеями и опытом анализа производительности и нагрузки в различных отраслях промышленности. По результатам конференции CMG публикуются тезисы докладов, однако они доступны на интернет-сайте CMG [1] не по всем презентациям и только для зарегистрированных членов CMG. Статьи старше 5 лет находятся в свободном доступе. Данная статья содержит обзор организации CMG в целом и Конференции 2013 года в частности.

ПРОИЗВОДИТЕЛЬНОСТЬ; НАГРУЗКА; БОЛЬШИЕ ДАННЫЕ; ОБЛАЧНЫЕ ВЫЧИСЛЕНИЯ; РАСПРЕДЕЛЕННОЕ ОБОРУДОВАНИЕ.

The Computer Measurement Group (CMG, Inc.) [2] is a not-for-profit, worldwide organization focused on the insurance and efficiency of IT services delivered to the enterprise through performance improvement, capacity analysis and forecasting. Over the past decade, CMG is known as a leading organization for information exchange among enterprise computing professionals.

CMG has 4 independent levels of publications – the CMG Journal [3], MeasureIT [4], the CMG Bulletin [5], and the CMG Conference Proceedings [1]. The CMG Journal is published at least 3 times per year, and some papers from the Journal might be presented at the CMG Conference. MeasureIT is a free electronic monthly newsletter, written by and for computer professionals, including the best papers from the most recent CMG Conference. The CMG Bulletin is also a periodical publication but does not include articles, only CMG news and CMG items.

Each year the CMG International Conference is organized in a different place of the USA. It attracts up to 3000 attendees from various countries and business companies. CMG has many representative groups all over the world – in Europe, Asia, Australia, North and Latin America – where smaller regional meetings take place several times per year.

Everybody can submit a paper and/or presentation slides to the CMG by using Editor's Assistant (EDAS) as a web-based conference and journal management system [6]. The CMG Conference has its own annual life cycle with time deadlines, including abstract and paper submission in late spring, reviewing and editing in summer, final acceptance in early autumn,

Fig. 1. Netflix Starz Page

and presentation to the CMG Conference at the end of a year. Each paper without authors' names is blind reviewed by 6 anonymous referees simultaneously. A paper is accepted for the CMG presentation if it is approved by all referees.

In 2013, the 39th International Conference for the Performance and Capacity by CMG was held on November 4-8 in La Jolla, CA, USA [7]. The Conference program is available in EDAS system [8] and consists of the following parallel sections:

• Application Performance Management (APM)
• Capacity Planning (CP)
• IT Service Management (ITSM)
• Performance Engineering and Testing (PET)

In addition to regular sessions, there are technical forums, workshops across popular platforms and exhibitions from CMG vendors. Some keynote presentations are described below.

Cloud Native Capacity, Performance and Cost Optimization Tools and Techniques

This workshop is provided by Adrian Cockcroft, Director, Cloud Architecture at Netflix Inc. (USA) [9, 10].

Netflix is the world's leading IT service for provisioning stream movies and TV shows over the Internet (Fig. 1).

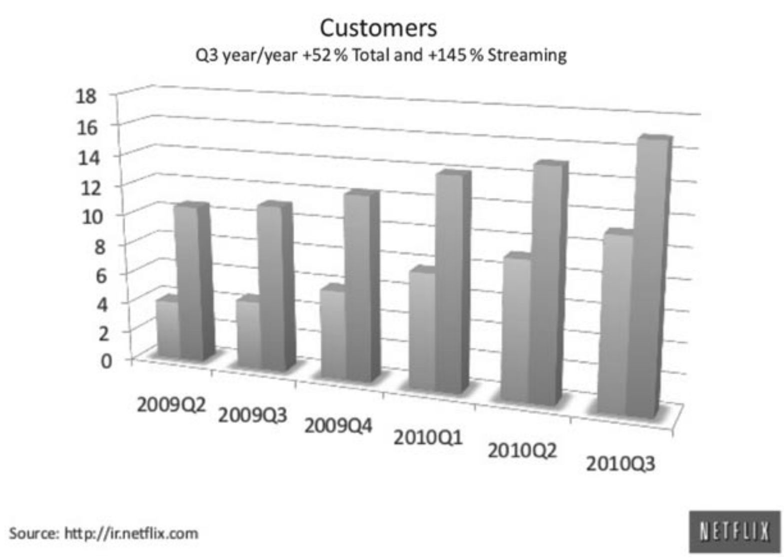


Fig. 2. Netflix Capacity Growth

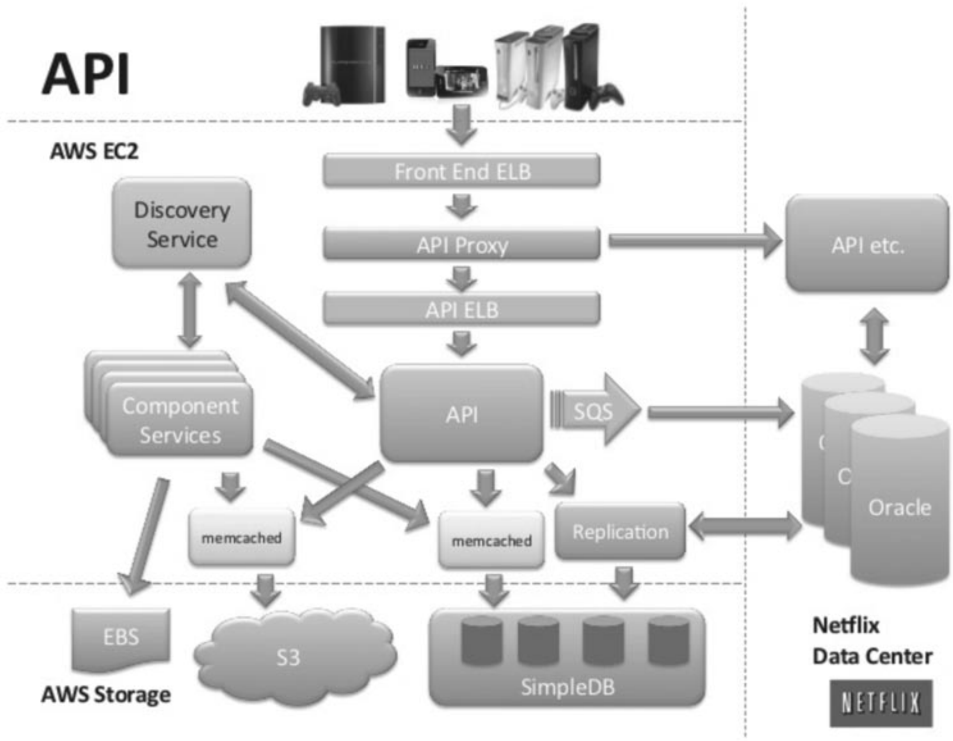


Fig. 3. Netflix Systems Architecture

There are more than 16 million subscribers in the United States and Canada, but the capacity growth rate is accelerating and unpredictable (Fig. 2).

Netflix cloud architecture is shown in Fig. 3. Netflix uses Hadoop technology across the biggest cloud environment including about 10K hosts, dozen thousands of videos, terabytes of daily transfer rate. Oracle distributed database stores information about subscribers and customers metadata. Zone aware routing is used for extended load balancing (ELB). API proxies help in the vertical scalability for geographically concentrated clients. Keynote, AppDynamics, Epic and Nimsoft NMS are introduced for automated real time service monitoring and cloud alerting in the data center [11].

Due to the big data and high logging rate, the internal solutions are developed for the log analysis. Most of the Netflix internal API tools are available as open source downloads [12] and have also been integrated with other open source software of the third party companies.

Main problems of the current cloud infrastructure and ways to resolve them are defined as follows:

1. Central SQL database is risky to fail. The solution is to decentralize the deployment of database instances and use the distributed NoSQL storage based on Cassandra.

2. To maintain high service availability in a fast growing business, auto scaling user groups with on-demand savings is introduced.

3. Cost optimization of hardware resources is required for handling unexpected peak demands. In cloud based systems, benchmarking is extremely effective because large configurations can be created and tested quickly with a relatively low cost.

4. Different monitoring tools work with different data sources. Integration of custom dashboards in a single monitoring portal is needed.

5. To go global to international markets, launch new data centers in Europe, Brazil, and other countries (Fig. 4).

The Future of Code Production, and of Capacity Planning

Typical modern capacity problems and future prediction for human coding and software maintenance are presented by Ron

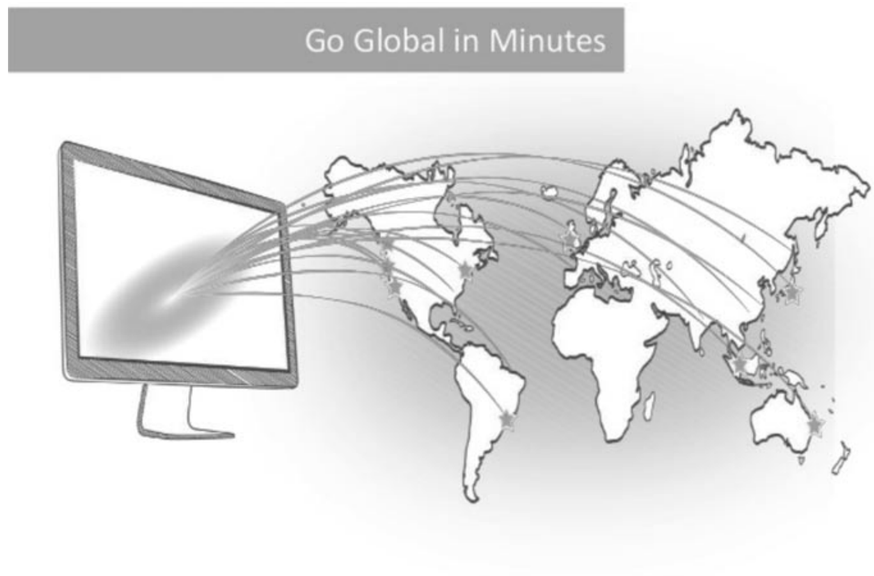


Fig. 4. Netflix Geographical Zones

Kaminski, Kimberley–Clark, Mullen Award Winner – 2003.

Within decades the vast clouds of hardware and the huge network bandwidth have been observed in many firms. The system computing resources (CPU, memory utilization, network traffic, etc.) are accelerating faster than human coding experience and applications quality. As a result, the decisions on the software development and maintenance are made by using the logic «cost per hour» rather than coding skills and quality.

Most humans make coding mistakes. Training humans to code well is a difficult and slow process, so attempting to improve all the skills of all coders on the planet is impossible. We can instead try to recognize the common patterns of applications failure and automatically detect and correct the issues at the early phase. Hereby capacity planning will become much easier.

Typical modern capacity problems are defined as follows:

1. Applications now run on shared machines and use common resources due to virtualization. In many cases of bad implementation, an application causes early failure due to the chronic overuse of hardware resources. When an application goes slow down or fails with an «out of memory» error, the immediate IT request is for more CPU or RAM to be installed (even though 92 % of the time caused by some process pathology, code issue or overloaded IO path). The only temporary help is to add the hardware, ignore the real problems, though the applications do not speed up after the changes.

2. Business applications move from «in house» to the cloud, causing additional concern about security, networking, user locations and distances from big pipes and data centers.

3. More and more functions get outsourced

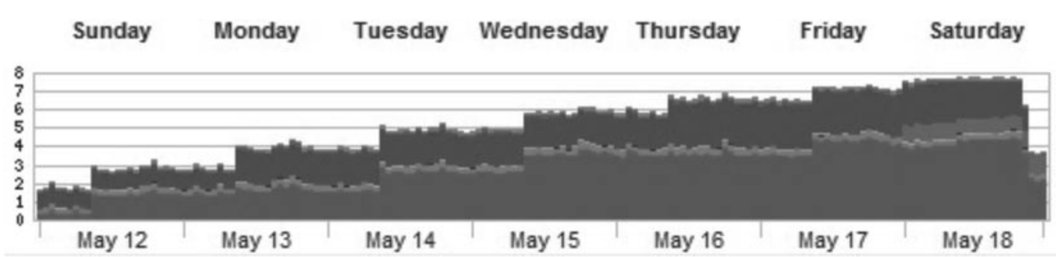


Fig. 5. Textbook Example of Increasing Usage of Computing Resources

to cheaper part-time labor, or become purchased web services of the third party vendors.

Fig. 5 shows a textbook example, which is repeated on thousands of cloud hosted servers running web-based applications. They use an increasing number of computing resources over a week until troubleshooting is escalated or service is restarted as a part of the pre-planned maintenance.

What has to be changed in capacity planning:

1. Along with virtualization, the actual measurement from a non-virtual aware OS is simply impossible. So the measures of 93 % of systems are a lot less accurate than they used to be.

2. Automated monitoring and analytical tools for cloud infrastructure from external vendors are really too expensive and thus helpless.

3. More precise measurements of what has happened on a machine are needed. A random sample is statistically valid where collectors use less than 1—3 % of system resources and accurate in the high 98 % range.

4. The data from all sources may not be perfect. However, we don't need the perfect data to track, for example, 3600 machines with an average utilization of 4—6 %.

5. Need to fully automate detecting critical issues, including code looping and high CPU utilization, which the applications are prone to.

6. Data collectors for all modern devices, not only limited to tablets, smartphones, other popular handhelds and OSes, are required.

7. Need to demand more sophisticated capacity reports that offer deeper insights into what is really running, to give us better and faster ways to spot problems.

Summary of Other Presentations

The goal of the CMG annual Conference [1] is to glean information and experience on the latest technologies from experts, academicians, consultants and vendors on measuring the performance and capacity of computing systems. The purpose of the CMG 2013 Conference [7] is to evaluate the impact of virtualization, cloud computing and big data. The main subjects of the CMG 2013 Conference are as follows:

1. Applications performance management, including measurement and tuning. Everything in this area starts with measurement, tuning and optimization. More issues are usually assessing business service support to ensure their maintenance. Papers for this area include discussions of what data to gather, how and where to keep it and how effectively to analyze and report on it. Some solutions are applied across all environments [13—15], others are specific to particular operating systems or storage subsystems [16, 17].

2. Capacity planning, including modeling and statistics. This subject includes issues of managing the available capacity, determining future demands, estimating cases when the current capacity will no longer be sufficient and cost to increase it. Mathematical approaches as well as forecasting from trends in business and resource utilization are introduced in [18—20].

REFERENCES / СПИСОК ЛИТЕРАТУРЫ

1. The CMG International Conference official web site. Available: http://www.cmg.org/conference/

2. The Computer Measurement Group organization official web site. Available: http://www.cmg.org/

3. The CMG Journal official web site. Available: http://www.cmg.org/national/journal.html

4. Official web site of the Measure IT journal by CMG. Available: http://www.cmg.org/measureit/

5. Official web site of the CMG Bulletin. Available: http://www.cmg.org/national/publications.html#Bulletin

6. EDAS Conference and Journal Management System. Available: https://edas.info/

7. The 39th International Conference for the Performance and Capacity by CMG, Inc. La Jolla, USA, 2013. Available: http://www.cmg.org/conference/cmg2013/

8. Program for Performance and Capacity 2013 by CMG. La Jolla, USA, 2013. Available: http://edas.info/p14745

9. **Cockcroft A.** Cloud Native Capacity, Performance and Cost Optimization Tools and Techniques. *Proc. of the 39th International Conference for the Performance and Capacity by CMG.* La Jolla, USA, 2013. Available: http://www.slideshare.net/adrianco/cmg-workshop

10. **Cockcroft A.** Netflix in the Cloud. Available: http://www.slideshare.net/adrianco/netflix-on-cloud-combined-slides-for-dev-and-ops

11. Epic NMS Cloud Monitoring and Alerting. Available: http://epicnms.com

12. Netflix Open Source Tools. Available: http://github.com/netflix

13. **Levine C.** The Challenges of Measuring Database Performance in the Cloud. *Proc. of the 39th Internat. Conference for the Performance and Capacity by CMG.* La Jolla, USA, 2013. Available: http://www.cmg.org/conference/

14. **Podelko A.** Agile Aspects of Performance Testing. *Proc. of the 39th Internat. Conference for the Performance and Capacity by CMG.* La Jolla, USA, 2013. Available: http://www.cmg.org/conference/

15. **Johnson P.** Java Performance Analysis/Tuning. *Proc. of the 39th Internat. Conference for the Performance and Capacity by CMG.* La Jolla, USA, 2013. Available: http://www.cmg.org/conference/

16. **Gelb I.** System z Performance, Capacity & TCO Q&A. *Proc. of the 39th Internat. Conference for the Performance and Capacity by CMG.* La Jolla, USA, 2013. Available: http://www.cmg.org/conference/

17. **Schwartz J.** Windows System Performance Measurement and Analysis. *Proc. of the 39th Internat. Conference for the Performance and Capacity by CMG.* La Jolla, USA, 2013. Available: http://www.cmg.org/conference/

18. **Salsburg M.** Modelling & Forecasting. *Proc. of the 39th Internat. Conference for the Performance and Capacity by CMG.* La Jolla, USA, 2013. Available: http://www.cmg.org/conference/

19. **Mescheryakov S.** Capacity Management of Java-based Business Applications Running on Virtualized Environment. *Proc. of the 39th Internat. Conference for the Performance and Capacity by CMG.* La Jolla, USA, 2013. Available: http://www.cmg.org/conference/

20. **Ruj A., Murty J.** Building a Predictive Capacity Model from Historical Usage Data. *Proc. of the 39th Internat. Conference for the Performance and Capacity by CMG.* La Jolla, USA, 2013. Available: http://www.cmg.org/conference/

MESCHERYAKOV, Sergey V. *St. Petersburg State Polytechnical University.*
195251, Politekhnicheskaya Str. 29, St. Petersburg, Russia.
E-mail: serg-phd@mail.ru

МЕЩЕРЯКОВ Сергей Владимирович – *профессор кафедры автоматов Санкт-Петербургского государственного политехнического университета, доктор технических наук, доцент.*
195251, Россия, Санкт-Петербург, ул. Политехническая, д. 29.
E-mail: serg-phd@mail.ru

SHCHEMELININ, Dmitry A. *RingCentral Inc.*
1400 Fashion Island Blvd., San Mateo, CA, USA 94404.
E-mail: dshchmel@gmail.com

ЩЕМЕЛИНИН Дмитрий Александрович – *руководитель департамента развития и эксплуатации облачных IT платформ компании RingCentral, кандидат технических наук.*
1400 Fashion Island Blvd., San Mateo, CA, USA 94404.
E-mail: dshchmel@gmail.com

UDC 681.3.016=111

D.E. Bortyakov, S.V. Mescheryakov, D.A. Shchemelinin

INTEGRATED MANAGEMENT OF BIG DATA TRAFFIC SYSTEMS IN DISTRIBUTED PRODUCTION ENVIRONMENTS

An effective management of a cloud-based distributed production infrastructure with multiple service equipment and big data traffic is impossible without a centralized automated control system. The purpose of integrated control includes on-line monitoring of the current state of the entire distributed infrastructure according to key performance indicators, automatic alerting in emergency cases or outages and, when possible, auto-restoring of production services. All these question described in this paper are based on the work of a particular company.

AUTOMATED CONTROL; DATABASE; BIG DATA; INTEGRATION; CLOUD COMPUTING; DISTRIBUTED ENVIRONMENT.

Д.Е. Бортяков, С.В. Мещеряков, Д.А. Щемелинин

ИНТЕГРИРОВАННОЕ УПРАВЛЕНИЕ БОЛЬШИМИ ДАННЫМИ ТРАНСПОРТНЫХ СИСТЕМ В РАЗЛИЧНЫХ ПРОИЗВОДСТВЕННЫХ УСЛОВИЯХ

Эффективное управление обширной производственной инфраструктурой с многочисленным сервисным оборудованием и большими потоками данных невозможно без централизованной системы автоматического контроля. В задачи интегрированного управления входит непрерывный мониторинг текущего состояния всей распределенной инфраструктуры по ключевым производственным критериям, автоматическое оповещение в случае критических ситуаций или отказов и, по возможности, автоматическое восстановление работоспособности сервисов. Все эти вопросы рассмотрены в данной статье на примере конкретного предприятия.

АВТОМАТИЗИРОВАННОЕ УПРАВЛЕНИЕ; БАЗА ДАННЫХ; БОЛЬШИЕ ДАННЫЕ; ИНТЕГРАЦИЯ; ОБЛАЧНЫЕ ВЫЧИСЛЕНИЯ; РАСПРЕДЕЛЕННОЕ ОБОРУДОВАНИЕ.

Overview of RingCentral Cloud Infrastructure

RingCentral (RC) is the international IT company with a central office in Silicon Valley (San Mateo, CA, USA) provisioning voice over Internet protocol (VoIP), mobile platforms, short message service (SMS), email, fax, conference and other IP communication services for more than 300,000 business customers in the USA, Canada and Europe [1].

RC production system is a cloud based multi-component infrastructure with big data traffic across all the distributed environments. Fig. 1 shows RC a scalable architecture located in 4 data centers on the West Coast and East Coast of the USA and in Western Europe. RC environments both production and stress test are rapidly growing along with customer demand amounting to 40 % annual rate. So now it consists of more than 3500 hosts including hardware (HW) but mainly they are the virtual machines (VM) for more efficient IT maintenance and saving computing resources.

All production servers are grouped into about 60 pools and RC components (Table 1), each provisioning a particular custom service and functionality or connecting with an external public switched telephone network (PSTN) and other third party providers. The example of workflow for the registration of RingCentral Mobile (RCM) devices in RC system is shown in Fig. 2.

Registration is the initial mandatory action to authenticate in RC system and grant access to the entire set of RC services for mobile users. RCM registration consists of the following steps:

1. RCM sends the authentication request via IP-network to RC system. HTTP request is

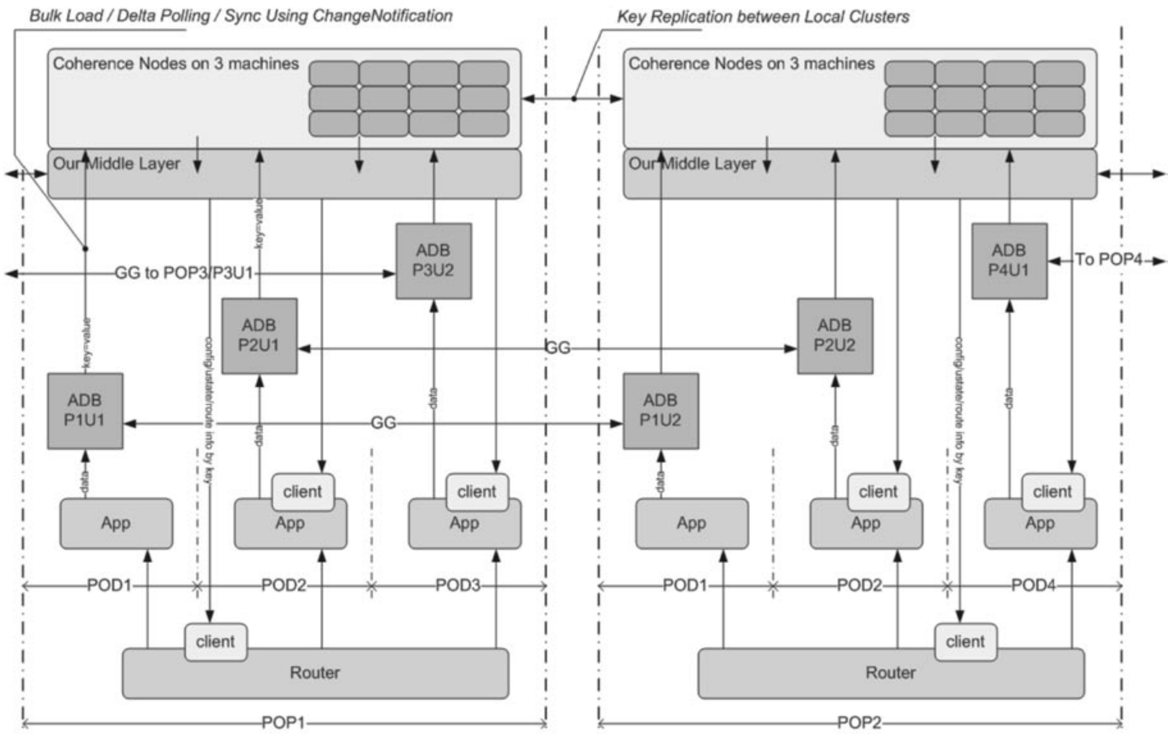


Fig. 1. RingCentral Scalable Architecture

Table 1

Definitions of Some RC Partitioning and Service Components

Abbreviation	Full Name	Definition
ADB	Account Database	Separate database to limit the clients count per POD for better performance and scalability
CDB	Common Database	DB storing common client's data for all PODs in POP
GUD	Global User Directory	Storing all clients global data and routing requests from Common Layer to a particular POD
ISR	Inbound Telco SIP Proxy Server	Routing inbound telecommunication requests to a particular POD based on GUD query results
JWR	JEDI Proxy Router	Routing inbound HTTP requests to JWS of a particular POD based on GUD query results
JWS	JEDI Server	Windows based server running RC Java applications provisioning RC web services such as signup, RCM registration to the clients
POD	Part of Data	The part of environment providing all RC services to clients of a particular ADB for better performance and scalability
POP	Point of Presence	RC environment, either active or standby, located physically in one data center
PWS	Platform Web Server	Windows based server running RC Java applications provisioning all RC services to mobile users
TAS	Telephony Access Server	Windows based server running RC applications provisioning telephony calls, inbound faxes and other VoIP services to RC clients

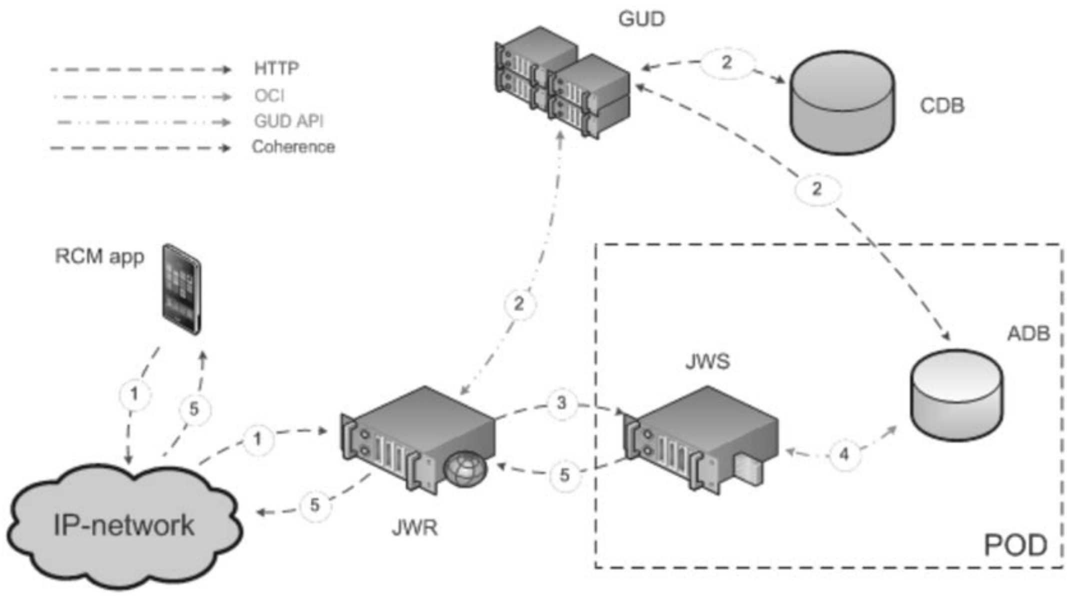


Fig. 2. Workflow for RCM Registration Service

transmitted and processed by JWR router.

2. JWR sends a query against GUD to determine the destination POD number of that particular RCM client. GUD global data is always up-to-date due to Golden Gate (GG) replication between CDB and all ADB databases via Oracle Coherence Interface (OCI).

3. HTTP request for authentication is routed via JWR through JWS server to a particular POD based on GUD query results.

4. JWS sends a query against local ADB in POD to grant or deny access to RC system for that particular request.

5. If the account exists in ADB and it is successfully registered, JWS server compiles corresponding XML configuration file and sends it back to RCM end device.

Zabbix Integrated Monitoring System

Zabbix enterprise-class monitoring system [2] is introduced as an open source integrated solution to manage such a huge multi-host distributed cloud infrastructure. Zabbix is effectively used for real-time monitoring, alerting, troubleshooting, an automated control, capacity analysis and other purposes.

A simple monitoring architecture consists of Zabbix Web Server, SQL DB server and a series of proxies to get counters from external RC hosts with a regular polling time interval and store performance data into DB (Fig. 3). There are different ways to send the results of measurement to Zabbix:

1. Zabbix agent installed on a host is automatically configured for predefined system metrics like CPU utilization, memory usage, disk free space, network packets loss, server ping, service availability, etc.

2. Zabbix trapper that is helpful for implementing any custom script on a host is sending the results to Zabbix via Simple Network Management Protocol (SNMP).

3. An external check which doesn't require Zabbix agent installation and allows executing any custom SQL queries or other code script against a host is returning the results back to Zabbix. Some examples of Perl scripts are given [3].

All the events, the latest metric values and historical trends are integrated in Zabbix DB. Any database of object-relational type (MySQL, PostgreSQL, Oracle, etc.) can be used as Zabbix DB.

Zabbix configuration includes:

1. The list of external hosts being under monitoring.

2. The descriptions of metrics for measuring availability and the performance of hosts, system services and business applications running on servers.

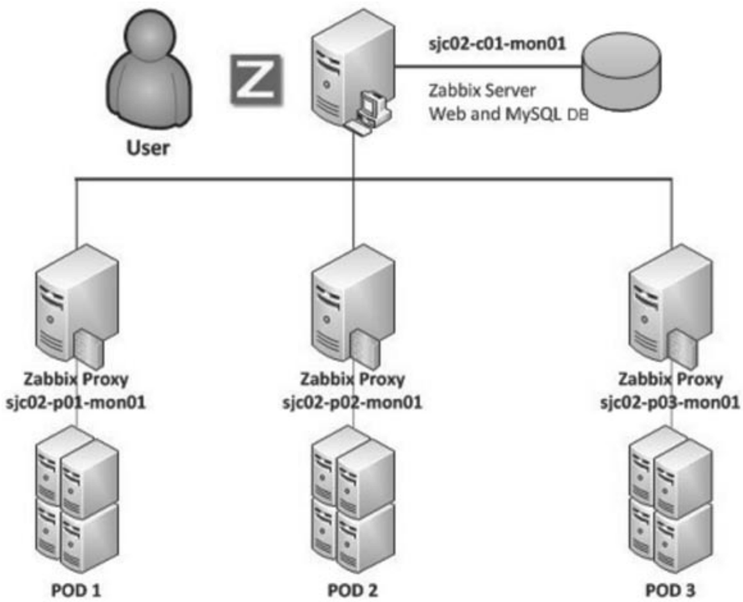


Fig. 3. Zabbix Monitoring System Architecture

Table 2

Example of Zabbix Items for RC Host

Name	Last check	Last value	Change	Interval	History
ICMP Availability (3 Items)					
ICMP ping	1 Dec 2013 12:22:30	1	-	90	180
ICMP packet loss	1 Dec 2013 12:22:34	0 %	-	90	180
ICMP ping sec	1 Dec 2013 12:22:32	0.28 msec	-0.04 msec	90	180
JAVA CMS Memory (5 Items)					
Java CMS Memory heap used	1 Dec 2013 12:22:37	233.5 Mbytes	+5.97 Mbytes	1800	365
Java CMS Perm Gen pool used	1 Dec 2013 12:22:47	145.01 Mbytes	+368 bytes	1800	365
Java CMS Thread count	1 Dec 2013 12:22:49	161	-1	1800	365
Java CMS Memory heap free	1 Dec 2013 12:22:55	293.42 Mbytes	+3.56 Mbytes	1800	365
Java CMS Perm Gen memory free	1 Dec 2013 12:22:58	261.71 MB	-	1800	365
JMX_App_ServiceWeb (6 Items)					
JMX_JEDI (7 Items)					
JMX: JEDI: JDBC pool connection wait time	1 Dec 2013 12:22:56	0	-	60	365
JMX: JEDI: JDBC statement execution time	1 Dec 2013 12:22:56	0 seconds	-	60	365
JMX: JEDI: JDBC statement prepare time	1 Dec 2013 12:22:56	0	-	60	365
JMX: JEDI: JDBC pool leased connections	1 Dec 2013 12:22:56	0	-	60	365
JMX: JEDI: HTTP requests processing time	1 Dec 2013 12:22:56	35	-429	60	365
JMX: JEDI: HTTP requests per sec	1 Dec 2013 12:22:56	0.04	-0.16	60	365
JMX: JEDI: Web processing active thread count	1 Dec 2013 12:22:56	0	-	60	365
- **other** - (69 Items)					

Table 3

Example of Zabbix Triggers for RC Host

Severity	Name	Expression
Information	POD_SWS_v3:JBoss needs to be restarted	{sjc01-p01-sws01:java.cms.memory.free.**max**(300)}<{$CMS_PERMGEN_MEM_FREE_CRITICAL}
Critical	Template_Java_CMS_Memory:Free CMS memory is too low	{sjc01-p01-sws01:java.cms.memory.free.**max**(300)}<{$CMS_MEM_FREE_CRITICAL}
Warning	Template_Java_CMS_Memory:Free CMS memory is too low	{sjc01-p01-sws01:java.cms.memory.free.**max**(600)}<{$CMS_MEM_FREE_WARNING}

3. The triggers that fire on the events when the predefined thresholds are exceeded. Triggers have different severities such as information, warning and critical.

4. The graphs to analyze performance data and historical trends.

The sample lists of Zabbix configuration items and triggers defined for RC hosts are shown in Tables 2 and 3 respectively.

In addition to standard system metrics, which are predefined in Zabbix agent and if they are installed on a host, custom items and business oriented metrics are implemented. To monitor user's activity and server's resources of Java Enterprise Development Implementation (JEDI) special metrics using Java Management Extensions (JMX) shown in Table 2 are introduced [4—6]. JMX technique provides remote access to internal objects, classes, services and other resources of a Java application that allows measuring actual workload in a server pool. The example of daily statistics in Fig. 4 shows that 20 database connections are permanently established with a host, but they are not actively used in threads (4 as maximum). It means that the capacity of a server pool for this particular RC component can be reduced.

Automatic Alerting on Critical Events

Zabbix triggers fire alarms automatically when a certain metric item exceeds the specified threshold value. Fig. 5 shows the example of Zabbix graph where the degradation of Java memory is observed in many instances, each time triggering an alert in Zabbix system when Java free memory is below critical threshold 5 MB.

Java memory leak is a well-known and it is a

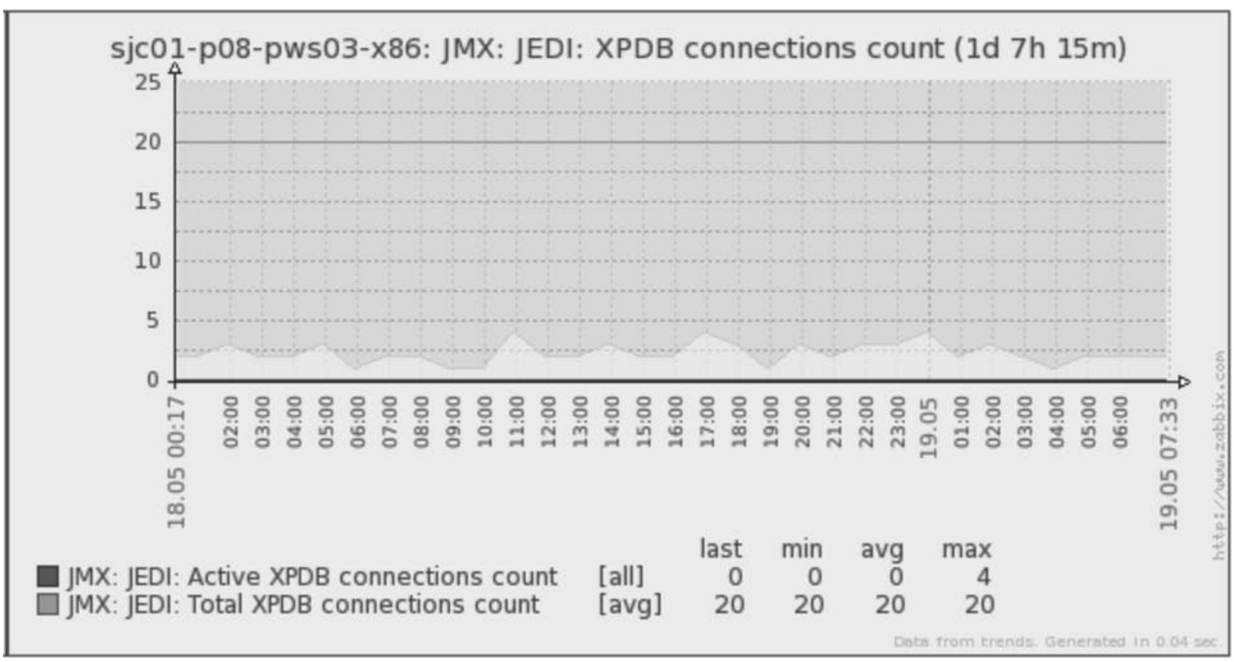


Fig. 4. Example of JMX Metrics for DB Connections

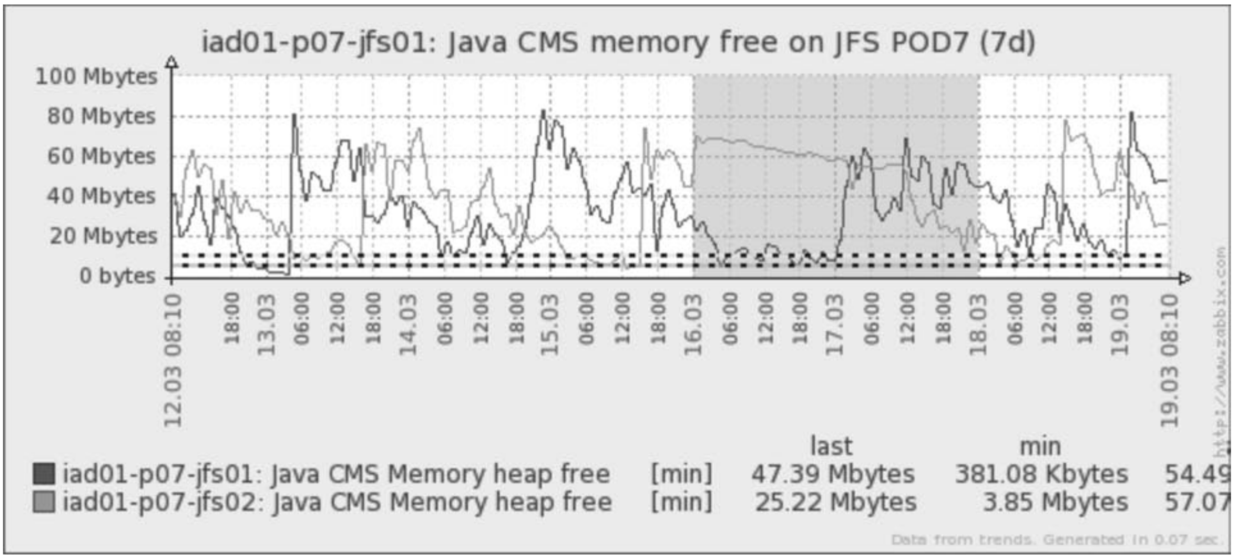


Fig. 5. Sample Graph with Multiple Triggering on Critical Java Free Memory

chronic problem all over the world. The objects, classes and their instances are dynamically created in Java applications that cause virtual memory degradation which depends on user's activity. Traditionally Garbage Collector (GC) is used to partially free up Java memory and reduces the impact of memory leak. The example in Fig. 5 figures out a capacity redundancy in a server pool because critical alarms take place too often while GC is not enough to restore Java memory.

All the alarms, detected in Zabbix monitoring system, are sent via email to an appropriate operation team and are also reflected on the specially designed integrated dashboard.

Integrated Operations Dashboard

An integrated dashboard can be created with a minimal development effort and is really valuable for business services management. The example of RC integrated dashboard or operations monitoring console is shown in Fig. 6. It displays all the events and alerts from different locations having 3500 hosts and over 200K items under Zabbix monitoring with 2 min refresh rate. Additional details for every alert such as JIRA link, time and a person to acknowledge, workaround to fix, etc., can be provided for more efficient troubleshooting.

There are a lot of examples of real world scripts for getting information from external data sources and creating integrated dashboards [7].

Automatic Restoring Anomalies

In addition to monitoring and automatic alerting, Zabbix monitoring system can be configured for automatic actions to resolve the anomalies detected on a host. A good example of such solution is the auto-remediation procedure implemented on JEDI hosts to restart JBoss service when Java virtual free memory is

	Last change	Host	Name	Acknowledged	Acknowledged by
Warning	06 Dec 2013 09:49:43	total.pod6	Free activated numbers in 866 NPA is less than 7%	12 m 15 s	Jayn Layson
Warning	06 Dec 2013 09:49:38	total.pod1	Free activated numbers in 866 NPA is less than 7%	12 m 20 s	Jayn Layson
Critical	05 Dec 2013 09:04:43	sjc01-p09-tas01	TAS has been blacklisted by MORE THAN one ISR during last 10 minutes	13 h 31 m 2 s	Kathlyn Mercado
Critical	05 Dec 2013 08:51:46	sjc01-p09-tas02	TAS has been blacklisted by MORE THAN one ISR during last 10 minutes	13 h 43 m 59 s	Kathlyn Mercado
Critical	05 Dec 2013 08:40:01	sjc01-p09-TAS	At least 2 TAS hosts blacklisted by sjc01-c01-isr05-06 during 10 mins	13 h 55 m 44 s	Kathlyn Mercado

Fig. 6. Example of RC Integrated Operations Dashboard [1]

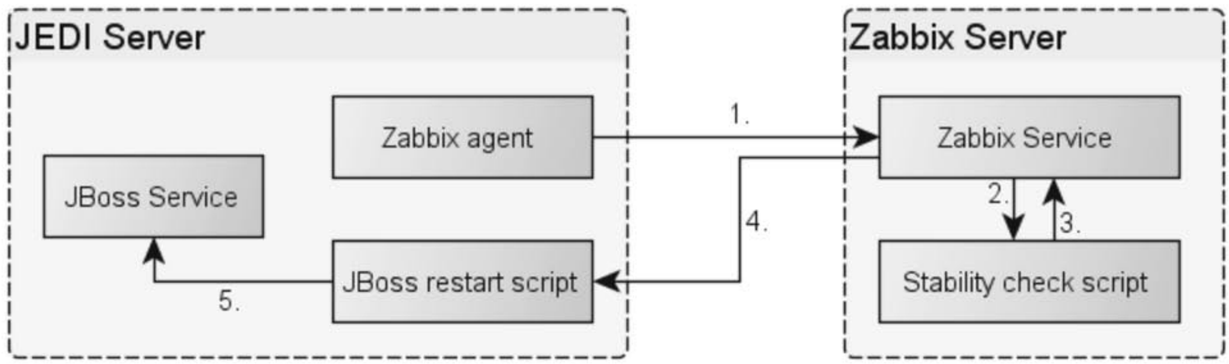


Fig. 7. Auto-remediation Procedure for JEDI Hosts

below critical threshold in spite of GC which are periodically issued on a host. JMX metrics and triggers like those listed in Tables 2 and 3 and described above are used to measure the allocated Java objects and remaining virtual memory resources more accurately than standard system items predefined in Zabbix. Prior to restart Java service, an additional check of the availability of the other VMs in a pool is needed to prevent business service outage. So the logic of auto-remediation procedure is rather complicated and consists of the following steps (Fig. 7):

1. Zabbix agent sends monitoring data from JEDI host to Zabbix server determining that Java free memory is too low and activating the trigger «JBoss which needs to be restarted».
2. Zabbix server executes the stability check script which is specially designed and located locally on Zabbix server to verify the availability of the other JEDI servers in the same pool as a problematic host.
3. If the capacity in a server pool is enough to provide a stable business service, the confirmation is returned back to Zabbix server stating that JEDI host can be safely restarted.
4. Zabbix server initiates the specially designed script which is located on JEDI host to restart JBoss service locally.
5. Finally, after JBoss is restarted, the availability and Java memory resources of JEDI host are monitored by Zabbix as usual.

Scalability and Capacity Analysis

Zabbix enterprise-class system allows monitoring up to 3000 hosts with 200,000 items total from HW and VM, network devices, databases and other cloud environment. High availability of Zabbix system strongly depends on Zabbix DB size, HW performance, the amount of monitoring hosts and items and their polling frequency. Data delay in monitoring, if it happens, may lead to missed service outage that is not acceptable on production environment.

In case of capacity growth and limitations caused by user's workload, the scalability of Zabbix monitoring system can be reached using one way or the combination of the following ways:

1. Reducing the number of monitoring items and/or extending the polling time interval is not a good idea. This might be applicable as a short term workaround if other solutions suggested below can't be implemented.
2. Installing more Zabbix servers, proxies, high performance storages or other HW devices shown in Fig. 3 can help to reduce the impact of monitoring data delay and prevent potential service outages. But this is too expensive solution and doesn't help a lot because the database productivity is the main bottleneck due to multiple read-write transactions executed against the same DB table in parallel.
3. Zabbix splits into 2 or more independent monitoring systems each working with a separate database seems like a permanent workaround. On the other hand, special reports and integrated dashboards should be created half manually to observe consolidated alarms, historical trends and the other data from all

locations on a single monitoring console. The risk of data delay still exists and depends on partitioning between the amount of real-time data and the history.

4. Alternative Zabbix architecture, named Octopus, which is based on MongoDB noSQL open-source document-oriented data warehouse [8], is proposed to reach scalability with a relatively lower cost [9]. The main advantage of the Octopus architecture is that a real-time data is stored in MySQL relational database on a local drive or even in memory due to small size, while all the other history with almost unlimited data retention and the forecasting trends are transferred to MongoDB noSQL data warehouse on network storage.

The main disadvantage of the Octopus distributed architecture is that human resources and web API programming skills are required to consolidate real-time and historical data in a single dashboard and Zabbix reports.

Octopus noSQL approach is currently being under construction at RingCentral Company [1] and has been presented at the annual Zabbix Conference [10].

Recap of Integrated Management Solutions

In multi-host virtualized cloud based distributed production infrastructure like RingCentral [1], the performance data and servers availability can be effectively monitored and managed by Zabbix enterprise-class integrated system.

On JEDI hosts, in particular, independently of HW or VM, JMX metrics [4—6] are introduced for measuring more accurately the actual user's workload and virtual memory resources, more effective monitoring Java applications and automatic alerting in case of the critical degradation.

It is highly recommended to create the integrated operations dashboard to display consolidated events and triggers on a centralized monitoring console, since it provides a lot of benefits for managing a distributed production system with big data traffic with minimal development efforts using SQL, Perl, Ruby or other code.

Besides monitoring and alerting, many critical anomalies concerning system resources in general and Java virtual memory in particular can be detected and automatically fixed if we configure Zabbix and custom scripts for that purpose.

In a cloud computing infrastructure monitoring the database is often a bottleneck which depends on the amount of hosts, metrics, polling time interval, the size of real-time and historical data. So for better DB performance and no reporting data delay, several approaches to distributed Zabbix monitoring architecture are proposed to reach acceptable scalability and capacity.

REFERENCES

1. RingCentral official web site. Available: http://www.ringcentral.com/

2. Zabbix official web site. Available: http://www.zabbix.com/product.php

3. **Hegedus J.C.** Perrobix + Zabbix DB Monitoring. *Zabbix International Conference*, 2013. Available: http://youtu.be/pJCV_ui0orQ

4. Java Management Extensions (JMX) — Best Practices. Oracle Technology Network. Available: http://www.oracle.com/technetwork/java/javase/tech/best-practices-jsp-136021.html

5. JMX Monitoring. *Zabbix 2.0 official documentation*. Available: http://www.zabbix.com/documentation/2.0/manual/config/items/itemtypes/jmx_monitoring?s[]=jmx

6. **Mescheryakov S., Shchemelinin D.** Capacity Management of Java-based Business Applications Running on Virtualized Environment. *Proc. of the 39th Internat. Conference for the Performance and Capacity by CMG*. La Jolla, USA, 2013. Available: http://www.cmg.org/conference/

7. **Lipski L.** Integrated Dashboard Design. *Zabbix Internat. Conference*, 2013. Available: http://youtu.be/_gy4qzyZf_o

8. MongoDB official web site. Available: http://www.mongodb.org/

9. **Yulenets L.** When It Comes to Scalability. *Zabbix Internat. Conference, 2013.* Available: http://youtu.be/1Eq-9q16AOU

10. Agenda and presentation abstracts. *Zabbix Internat. Conference 2013.* Riga, 2013. Available: http://www.zabbix.com/conf2013_agenda.php

СПИСОК ЛИТЕРАТУРЫ

1. Официальный Интернет-сайт RingCentral [электронный ресурс]/URL: http://www.ringcentral.com/
2. Официальный интернет-сайт Заббикс [электронный ресурс]/URL: http://www.zabbix.com/product.php
3. **Hegedus, J.C.** Перробикс + Заббикс мониторинг БД [электронный ресурс]//Доклад Междунар. конф. Заббикс. Рига, 2013. URL: http://youtu.be/pJCV_ui0orQ
4. Java Management Extensions (JMX) – Практические рекомендации [электронный ресурс]//Oracle Technology Network. URL: http://www.oracle.com/technetwork/java/javase/tech/best-practices-jsp-136021.html
5. Мониторинг с использованием JMX [электронный ресурс]//Официальная документация Zabbix 2.0. URL: http://www.zabbix.com/documentation/2.0/manual/config/items/itemtypes/jmx_monitoring?s[]=jmx
6. **Мещеряков С., Щемелинин Д.** Управление потенциалом бизнеса Java приложений на виртуальном оборудовании [электронный ресурс] // Тез. докл. 39 Междунар. конф. CMG по оценке производительности и потенциала вычислительных систем. Ла Хойя, США, 2013. URL: http://www.cmg.org/conference/
7. **Lipski L.** Проектирование интегрированных отчетов [электронный ресурс] // Доклад Междунар. конф. Заббикс. Рига, 2013. URL: http://youtu.be/_gy4qzyZf_o
8. Официальный интернет-сайт MongoDB [электронный ресурс] / URL: http://www.mongodb.org/
9. **Юленец, Л.** Когда приходит время масштабирования [электронный ресурс] // Доклад Междунар. конф. Заббикс. Рига, 2013. URL: http://youtu.be/1Eq-9q16AOU
10. Перечень докладов и аннотации [электронный ресурс] // Междунар. конф. Заббикс-2013. Рига, 2013. URL: http://www.zabbix.com/conf2013_agenda.php

---

BORTYAKOV, Danil E. *St. Petersburg State Polytechnical University.*
195251, Polytekhnicheskaya Str. 29, St. Petersburg, Russia.
E-mail: bortyakov@mail.ru

БОРТЯКОВ Данил Евгеньевич – *доцент кафедры транспортных и технологических систем Санкт-Петербургского государственного политехнического университета, кандидат технических наук.*
195251, Россия, Санкт-Петербург, ул. Политехническая, д. 29.
E-mail: bortyakov@mail.ru

MESCHERYAKOV, Sergey V. *St. Petersburg State Polytechnical University.*
195251, Polytekhnicheskaya Str. 29, St. Petersburg, Russia.
E-mail: serg-phd@mail.ru

МЕЩЕРЯКОВ Сергей Владимирович – *профессор кафедры автоматов Санкт-Петербургского государственного политехнического университета, доктор технических наук.*
195251, Россия, Санкт-Петербург, ул. Политехническая, д. 29.
E-mail: serg-phd@mail.ru

SHCHEMELININ, Dmitry A. *RingCentral Inc.*
1400 Fashion Island Blvd., San Mateo, CA, USA 94404.
E-mail: dshchmel@gmail.com

ЩЕМЕЛИНИН Дмитрий Александрович – *руководитель департамента развития и эксплуатации облачных IT платформ компании RingCentral, кандидат технических наук.*
1400 Fashion Island Blvd., San Mateo, CA, USA 94404.
E-mail: dshchmel@gmail.com

NEW2AN 2014: The 14th International Conference on Next Generation Wired/Wireless Advanced Networks and Systems

http://www.new2an.org/

August 27 – 29, 2014, St. Petersburg, Russia

Highlights – **The 14th International Conference on Next Generation Wired/Wireless Advanced Networks and Systems (NEW2AN 2014) co-located with the 7th International Conference on Internet of Things and Smart Spaces (ruSMART 2014).**

The proceedings will be published in Springer's LNCS and indexed by relevant databases.

Acceptance rate is 35 % over the last 5 years.

Authors of selected best papers will be invited to submit extended versions of their papers to be published in a special issue of a journal.

Important dates

- **Paper submission: May 05, 2014**
- **Notification of acceptance: June 10, 2014**
- **Camera-ready version: June 21, 2014**

I. About the Conference

Future mobile networks will evolve out of existing wired and wireless networks of different nature by means of convergence, fusion and mutual penetration; next-generation multimedia services in turn pose new requirements and challenges on treatment of traffic within communication systems – a certain degree of multilayer co-operation has to be achieved. This forms a framework on harmonization of results obtained separately in different areas of network research like network performance evaluation, architecture and protocol design, and low layer network design. The Conference goal is in the identification, investigation and integration of new algorithms, approaches, architectures, methods and mechanisms to enable proper and efficient operation of a next-generation IP-based wireless network. Therefore, wireless networks and systems and their interaction with wired networks and systems shall be widely examined and addressed throughout the Conference.

The history of NEW2AN dates back to 1993, today it is an established conference featured by proceedings published by Springer LNCS and distinguished keynote speakers.

The NEW2AN 2014 will be held in St. Petersburg, Russia – the meeting point of East and West. Through its stunning architecture and wonderful art galleries and museums, among a number of other attractions, the wealth of history of East and West can be seen and felt in this beautiful city.

II. Topics of Interest and Paper Submission

Topics of interest include but are not limited to:

1. Fundamentals and PHY
2. MAC design and Cross-Layering
3. Fixed, Mobile and Wireless Networks
4. Services and Applications
5. Nano Communications

The authors are encouraged to submit full papers describing original, previously unpublished, complete research, not currently under review by another conference or journal, addressing state-of-the-art research and development in all areas of computer networking and data communications.

All papers will be reviewed and accepted papers will appear in the conference proceedings, published by Springer-Verlag in the LNCS series.

Papers must be submitted electronically in the Conference site. The maximum size of papers should be 5000 words including tables and figures. Please adhere to the formatting standard for a 12-page manuscript of Springer-Verlag LNCS. The cover page must contain an abstract of about 150 words, 3-5 keywords, name and affiliation of author(s) as well as the corresponding author's e-mail and postal address. Selected best papers will appear in a journal special issue.
Submission link: http://edas.info/newPaper.php?c=16597

For further and up-to-date information please visit official web page of the Conference: http://www.new2an.org

Scientific journal

ST. PETERSBURG STATE POLYTECHNICAL UNIVERSITY JOURNAL.
TELECOMMUNICATIONS AND CONTROL SYSTEMS
No. 1 (188) 2014

The founder of the journal is the Federal State Budgetary Educational Institution
of Higher Professional Education – St. Petersburg State Polytechnical University.

The journal is registered with the Federal Service for Supervision in the Sphere of Telecom,
Information Technologies and Mass Communications (ROSKOMNADZOR).
Certificate ПИ № ФС77-51457 issued October 19, 2012.

Phone: 8(812)552-62-16

E-mail: infocom@spbstu.ru

http://ntv.spbstu.ru/telecom/

Digital typesetting by *A.N. Smirnov*

The Polytechnical University Publishing House is headed by *A.V. Ivanov*

Approved for print February 28, 2014. Paper size: 60×84 1/8 Print run: 1000
Offset printing; Convent. Print. Sheets – 13.95; Account. & Publ. lists – 13.95

St. Petersburg State Polytechnical University.
Polytechnical University Publishing House,
member of the Publishing-Polygraphic Association of Universities of Russia.
University and publishing house address:
29 Politekhnicheskaya Str, St. Petersburg, 195251, Russia.

НАУЧНОЕ ИЗДАНИЕ

«НАУЧНО-ТЕХНИЧЕСКИЕ ВЕДОМОСТИ САНКТ-ПЕТЕРБУРГСКОГО ГОСУДАРСТВЕННОГО ПОЛИТЕХНИЧЕСКОГО УНИВЕРСИТЕТА. ИНФОРМАТИКА. ТЕЛЕКОММУНИКАЦИИ. УПРАВЛЕНИЕ»

«ST. PETERSBURG STATE POLYTECHNICAL UNIVERSITY JOURNAL. COMPUTER SCIENCE. TELECOMMUNICATIONS AND CONTROL SYSTEMS»

№ 1 (188) 2014

Учредитель – Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Санкт-Петербургский государственный политехнический университет»

Журнал зарегистрирован Федеральной службой по надзору в сфере информационных технологий и массовых коммуникаций (Роскомнадзор). Свидетельство о регистрации ПИ № ФС77-51457 от 19.10.2012 г.

Телефон редакции (812)552-62-16, 297-18-21

E-mail: infocom@spbstu.ru

Компьютерная верстка *А.Н. Смирнов*

Директор Издательства Политехнического университета *А.В. Иванов*

Лицензия ЛР № 020593 от 07.08.97

Подписано в печать 28.02.2014. Формат 60×84 1/8. Бум. тип. № 1.
Печать офсетная. Усл. печ. л. 13,95. Уч.-изд. л. 13,95. Тираж 1000. Заказ

Санкт-Петербургский государственный политехнический университет
Издательство Политехнического университета
член Издательско-полиграфической ассоциации университетов России
Адрес университета и издательства: 195251, Санкт-Петербург, ул. Политехническая, д. 29.

«St. Petersburg State Polytechnical University Journal. Computer Science. Telecommunication and Control Systems»

INSTRUCTIONS FOR AUTHORS

Aim and Scope

The «St. Petersburg State Polytechnical University Journal. Computer Science. Telecommunication and Control Systems» publishes papers on COMPUTER SCIENCE, INFORMATION AND COMMUNICATION TECHNOLOGIES, SIGNAL PROCESSING, ELECTRONICS, AUTOMATION CONTROL SYSTEMS, ROBOTICS.

Manuscript submission

Please refer to the Instructions for Authors at E-Submission system for details and additional information on how to prepare your manuscript to meet the journal's requirements and how to use the online submission system.

Please note that if the paper has two or more authors, only the corresponding author is required to be logged in the system. Information about other authors may be provided while submitting a manuscript by using the online submission system.

Publication language

Manuscripts are accepted in Russian or English.

After submission

All submitted papers are reviewed and revised by using the online submission system. Once a paper is accepted for publication, the authors receive a notification and information about the scheduled journal issue and publication date.

After paper acceptance the authors will be asked to sign a publishing licensing agreement.

Authors of rejected papers will receive the reviewers' or editor's notes explaining the reason of paper rejection. Authors are advised not to resubmit earlier rejected papers. Such papers will be rejected by the editor without sending them to reviewers.

Submitted manuscripts that do not meet the formatting or the subject area requirements may not be published.

In case a large number of papers are submitted, admission may stop before the stated issue deadline.

Journal Indexing

The journal is abstracted/indexed in the VINITI RAS Abstract Journal (Referativnyi Zhurnal), VINITI RAS Scientific and Technical Literature Collection, Ulrich's Periodicals Directory, and Russian Science Citation Index.

Contacts

Web: https://ntv.spbstu.ru
Phone: +7(812) 552-62-16
E-mail: infocom@spbstu.ru