Studies in Systems, Decision and Control 171

Vyacheslav Kharchenko
Yuriy Kondratenko
Janusz Kacprzyk *Editors*

Green IT Engineering: Social, Business and Industrial Applications

Springer

Studies in Systems, Decision and Control

Volume 171

Series editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland
e-mail: kacprzyk@ibspan.waw.pl

The series "Studies in Systems, Decision and Control" (SSDC) covers both new developments and advances, as well as the state of the art, in the various areas of broadly perceived systems, decision making and control–quickly, up to date and with a high quality. The intent is to cover the theory, applications, and perspectives on the state of the art and future developments relevant to systems, decision making, control, complex processes and related areas, as embedded in the fields of engineering, computer science, physics, economics, social and life sciences, as well as the paradigms and methodologies behind them. The series contains monographs, textbooks, lecture notes and edited volumes in systems, decision making and control spanning the areas of Cyber-Physical Systems, Autonomous Systems, Sensor Networks, Control Systems, Energy Systems, Automotive Systems, Biological Systems, Vehicular Networking and Connected Vehicles, Aerospace Systems, Automation, Manufacturing, Smart Grids, Nonlinear Systems, Power Systems, Robotics, Social Systems, Economic Systems and other. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution and exposure which enable both a wide and rapid dissemination of research output.

More information about this series at http://www.springer.com/series/13304

Vyacheslav Kharchenko · Yuriy Kondratenko
Janusz Kacprzyk
Editors

Green IT Engineering: Social, Business and Industrial Applications

Springer

Editors
Vyacheslav Kharchenko
Department of Computer Systems, Networks and Cybersecurity
National Aerospace University "KhAI"
Kharkiv, Ukraine

and

Center for Safety Infrastructure-Oriented Research and Analysis
Public Company RPC Radiy
Kropyvnytskyi, Ukraine

Yuriy Kondratenko
Department of Intelligent Information Systems
Petro Mohyla Black Sea National University
Mykolaiv, Ukraine

Janusz Kacprzyk
Systems Research Institute
Polish Academy of Sciences
Warsaw, Poland

ISSN 2198-4182 ISSN 2198-4190 (electronic)
Studies in Systems, Decision and Control
ISBN 978-3-030-00252-7 ISBN 978-3-030-00253-4 (eBook)
https://doi.org/10.1007/978-3-030-00253-4

Library of Congress Control Number: 2018953316

© Springer Nature Switzerland AG 2019
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Green IT engineering (green information technologies engineering) is a special kind of computer technology based on energy saving and modern efficient information technologies. It could be considered as dual-purpose technology for (a) improving energy efficiency of computer systems and (b) increasing productivity, safety, and environmental performance of different industrial technological processes as well as information processing in social, business, and other applications.

The purpose of this book is to present a systematic account of research directions, results, and challenges, and a description and analysis of applications of green IT engineering and green IT approaches in different fields of human activity. Special attention is paid to the implementation of the green specialized software, programmable and hardware components, communications, cloud and IoT-based systems, and IT infrastructures to the engineering, business, social, and economy/management domains.

The monograph is a "de-facto" Volume 3 of the three volume sets in which the first book has been "Green IT Engineering: Concepts, Models, Complex Systems Architectures" and the second book has been "Green IT Engineering: Components, Networks and Systems Implementation", both are published by Springer in 2017. The present third volume, with the same first part of the title, i.e., "Green IT Engineering", aims at motivating researchers and practitioners from different IT domains to follow and advocate the ideas of green values for social, business, and industrial applications.

The first versions of all contributions were discussed at the workshops and seminars (United Kingdom, Italy, Portugal, Sweden, Ukraine) held within the international TEMPUS-project GreenCo and the ERASMUS + project ALIOT, as well as at the international conferences DESSERT, IDAACS, ICTERI during 2013–2018.

The book has a chapter-oriented structure according to the "vertical" structure of the green IT, from design and optimization of green IT systems to their applications in social, educational, business, and engineering fields.

The chapters are prepared according to a general paradigm and a unified scheme of contents, and describe step-by-step implementation processes of green IT in economy, management, and engineering, to just name a few.

In terms of structure, the 25 chapters of the book, presented by authors from different countries from all over the world: Estonia, Italy, Germany, Greece, Malaysia, Norway, Poland, Russia, Slovakia, Ukraine, the United Kingdom, and the USA, are grouped into four parts: (1) development and optimization of Green IT systems, (2) modeling and experiments with green IT systems, (3) industry and transport Green IT systems application, and (4) social, educational, and business aspects of Green IT systems application.

The chapters have been meant to provide an easy to follow introduction to the topics that are addressed, including the most relevant references, so that anyone interested in them can start his/her introduction to the topic through these references. At the same time, all of them correspond to different aspects of a work in progress being carried out in various research groups throughout the world and, therefore, provide information on the state of the art of some of these topics.

The first part, "Development and Optimization of Green IT Systems", includes six contributions: Chapter "Including Software Aspects in Green IT: How to Create Awareness for Green Software Issues", by E. Kern, A. Guldner, and S. Naumann, introduces the idea of a green software and its engineering. Complementary to the definitions and models in the addressed field, a more practical insight is given by presenting illustrative examples of energy measurements of software products. While these aspects show that the research effort is increasingly dealing with software issues of Green IT, these mainly scientific ideas have hardly reached practical relevance, at least so far. Following the life cycle perspective of software products, the authors present examples of concepts on how to increase the awareness of green software, proposing (a) to implement continuous energy efficiency measurements during the development phase and (b) to create an eco-label for software products to inform about their environmental friendliness. The aim of the chapter is to point out solutions that may lead to a more environmentally friendly way of developing software, a well-informed procurement behavior regarding software products, and a more sustainable user behavior concerning information and communication technology (ICT).

E. D. Stetsuyk, D. A. Maevsky, E. J. Maevskaya, and R. O. Shaporin, in their Chapter "Information Technology for Evaluating the Computer Energy Consumption at the Stage of Software Development", describe an approach based on the absolute value estimation of the computer devices energy consumption. It can help to choose an optimal energy saving solution at the software development stage. There are three stages for the energy consumption estimation in the proposed information technology. The first stage is the scanning of assembler code, determination of assembler instruction, and their volume in bits. At the second stage, the method for the estimated power consumption based on RAM types is applied. The third stage boils down to the estimation of the full energy consumption. For the implementation of information technology, the program tool is developed. This tool helps to create a green software for checking the energy consumption of computer devices at all stages of development.

In their Chapter "Green IT Engineering: Green Wireless Cooperative Networks", S. Fu and J. Wu consider the methods for securely performing the calculations required for the fundamental modular arithmetic operations, namely, the multiplication and exponentiation using mobile, embedded, remote, or distant computational resources which offer the possibility of development of the green information processing systems. These methods are targeted at the distributed paradigms of cloud computing resources and the Internet of Things (IoT) applications. They provide security by avoiding the disclosure to the cloud resource of either the data or the user secret key. Simultaneously, environmental effects of processing are minimized by the simplifications of the operations and by transferring demanding calculations to energy-efficient data centers. The developed modular multiplication algorithm provides a faster execution time on low complexity hardware in comparison with the existing algorithms and is oriented toward the variable value of the modulus, especially with the software implementation on microcontrollers and smart cards whose architectures include a small number of bits.

A. Drozd, S. Antoshchuk, J. Drozd, K. Zashcholkin, M. Drozd, N. Kuznietsov, M. Al-Dhabi, and V. Nikul, in Chapter "Checkable FPGA Design: Energy Consumption, Throughput and Trustworthiness", consider the green FPGA design processes based on the energy efficiency and safety for the critical applications. The array structures that are traditionally used in digital components of safety-related systems reduce a checkability of circuits, creating a problem of hidden faults which can be accumulated in a normal mode and reduce the fault tolerance of the circuit and the safety of system in the emergency mode. Soft and cardinal ways of array structure reduction are proposed based on the development of the truncated arithmetical operations implemented in reduced array structures and on the parallelization of calculations in serial codes with the use of bitwise pipelines. The comparative analysis in complexity, throughput, and energy consumption of the iterative array and bitwise pipeline multiplier is executed experimentally with use of Altera Quartus II.

In Chapter "A Prospective Lightweight Block Cipher for Green IT Engineering", A. Andrushkevych, Y. Gorbenko, O. Kuznetsov, R. Oliynykov, and M. Rodinko consider general requirements for modern block ciphers, especially for the implementation at lightweight cryptographic transformations for critical distributed environment applications with a green IT conformance. The overview of the well-known block ciphers and lightweight primitives PRESENT and CLEFIA, defined at ISO/IEC 29192-2, as well as a specification of the lightweight block cipher Cypress are presented. The Cypress performance in software is approximately three times higher than the AES one based on Windows, Linux, and Android platforms.

In Chapter "Lightweight Stream Ciphers for Green IT Engineering", O. Kuznetsov, O. Potii, A. Perepelitsyn, D. Ivanenko, and N. Poluyanenko introduce the symmetric cryptographic transformations, in particular, the stream ciphers. The development of an efficient synchronous stream cipher is reduced to the construction of a pseudo-random sequence generator with defined cryptographic properties. Limited physical parameters, low power consumption, low computing power, and other characteristic attributes of the "green" IT engineering forces the use of new approaches for designing cryptographic protection tools. The authors propose new methods and hardware/software tools for a lightweight stream encryption that meet the current requirements of the "green" IT engineering.

The second part, "Modelling and Experiments with Green IT Systems", includes eight contributions. In Chapter "Semi-Markov Availability Model for Infrastructure as a Service Cloud Considering Energy Performance", O. Ivanchenko, V. Kharchenko, B. Moroz, L. Kabak, I. Blindyuk, and K. Smoktii present a stochastic approach based on the semi-Markov process modeling in order to determine the availability level and energy performance for the Infrastructure as a Service (IaaS) cloud. Today, the cloud service providers offer diverse IT services using a large number of physical machines (PMs). The authors propose to use information about technical and information state of a control system in order to improve the monitoring process for the IaaS cloud. Special attention is paid to (a) the development of a monolithic semi-Markov availability model for the IaaS cloud and (b) the evaluation of power consumption for non-failed, that is available, PMs.

G. Kuchuk, A. Kovalenko, I. E. Komari, A. Svyrydov, and V. Kharchenko, in Chapter "Improving Big Data Centers Energy Efficiency: Traffic Based Model and Method", discuss the green aspects of data centers (DCs) which provide the appropriate hosting services and data storage. Various approaches to transmission control of integrated data streams, which are the result of traffic aggregation, being transmitted in a DC related to Big Data processing, are analyzed. The authors develop several models and a method which help reduce the DC's resources usage for data processing by 3%. The application of the proposed models allows the use of resources of the DC in a more efficient way providing the reduction of the transmission time and, correspondingly, an increased energy efficiency of the system.

Chapter "A Markov Model of IoT System Availability Considering DDoS Attacks, Patching and Energy Modes", by M. Kolisnyk and V. Kharchenko, is focused on an important task in the conditions of the rapid spread of the IoT, that is the assessment of its availability under the impact of various attacks. The authors (a) analyze the energy modes of network devices operating system in the smart business center (SBC) and specify the organization of an energy-efficient IoT-SBC subsystem; (b) develop and study Markov models of the SBC availability taking into account the energy regimes, the impact of cyberattacks, the reliability of SBC components, and patches on the firewalls vulnerabilities; and (c) analyze possible methods for reducing the power consumption of the SBC devices estimating the availability factor of SBC in the conditions of successful attacks.

In Chapter "Assessing the Impact of EEE Standard on Energy Consumed by Commercial Grade Network Switches", J. El Khoury, E. Rondeau, J.-P. Georges, and A.-L. Kor note that the reduction of the power consumption of network equipment has been both driven by a need to reduce the ecological footprint of the cloud as well as the immense power costs of data centers. As data centers, core networks, and—consequently—the cloud constantly increase in size, their power consumption should be mitigated. The Ethernet, the most widely used access network, still remains the biggest communication technology used in core networks and cloud infrastructures. As statistics shows, the average utilization rate of the Ethernet links is 5% on desktops and 30% in data centers, so that the power saving potential of EEE could be immense. Experimental results show that the base power consumption of switches does not significantly increase with the size of the switches. Doubling the size of the switch between 24 and 48 ports increases power consumption by 35.39%. EEE has a greater effect on bigger switches, with a power (or energy) gain on the EEE-enabled 48-port switch compared to 2 × EEE-enabled 24-port switch.

C. V. Vasile, C. Pattinson, and A.-L. Kor, in Chapter "Mobile Phones and Energy Consumption", present an analysis of different relevant concepts and parameters that may have impact on energy consumption of Windows Phone applications. This operating system was chosen because there is limited research even though there are related studies for the Android and iOS operating systems. The results for group of experiments are discussed taking into account the duration of the experiment, the battery consumed in the experiment, the expected battery lifetime, and the energy consumption, while the charts display the energy distribution based on the main threads: UI thread, application thread, and network thread.

Chapter "Energy-Efficient Multi-fragment Markov Model Guided Online Model-Based Testing for MPSoC", by J. Vain, L. Tsiopoulos, V. Kharchenko, A. Kaur, M. Jenihhin, J. Raik, and S. Nõmm, describes an efficient online testing method of autonomous mission critical systems (AMCS) in service modes. The multi-fragment Markov models (MFMM) are used for specifying the system reliability and quality-related behavior on high level of abstraction, and the more concrete state and timing constraints are specified explicitly using Uppaal probabilistic timed automata (UPTA). To interrelate these models, the authors demonstrate how the MFMM is mapped to UPTA. The second contribution is the test case selection mechanism for the online identification of AMCS service modes by a model-based conformance testing. The efficiency of active mode sampling is achieved by serializing the test cases for sampling session using hypotheses provided by MFMM.

In Chapter "Decrease of Energy Consumption of Transport Telecommunication Networks Using Stage-by-Stage Controlling Procedure", G. Linets and S. Melnikov propose a new approach to the development of an advanced automatic monitoring system and adaptive control of a transport telecommunication network which make it possible to reduce the energy consumption of switching centers during the analysis and identification of faults and failures in the equipment operation. The energy reduction is attained due to the use of a procedure of finding out abnormal situations, error optimization of the identification of system states, the use of a multi-agent setting, and the use of a permission for agents dependent on a hierarchy level of the transport telecommunication network.

I. Turkin and O. Vdovitchenko, in their Chapter "Model and Methods of Human-Centered Personal Computers Adaptive Power Control", consider two main approaches for solving the problem of reducing the computers' power consumption at the hardware and at the operating system levels, by (a) a dynamic change of the processor voltage and frequency; (b) an advanced configuration's interface and power management technology. The chapter is devoted to the development of the user's activity model and method of its identification in the system of PC's power consumption.

The third part "Industry and Transport Green IT Systems Application" includes six contributions. Chapter "Green Assurance Case: Applications for Internet of Things" by V. Sklyar and V. Kharchenko discusses the "green assurance case," in particular, the assurance case methodology for its implementation for the green IT domain related to the Internet of Things. Fundamentals of the concept of an assurance case are explained as the initial part of the article. Safety and security requirements are considered as a part of a typical assurance case. The green assurance case for the IoT includes the following parts: sustainability assurance part, safety and security assurance part, Green IT principles, power consumption parameters part, as well requirements to green features of of the four IoT layers. An example of the proposed green assurance case methodology for a medical IoT system is represented by an application.

O. Kozlov, G. Kondratenko, Z. Gomolka, and Y. Kondratenko, in their Chapter "Synthesis and Optimization of Green Fuzzy Controllers for the Reactors of the Specialized Pyrolysis Plants", present the developed generalized method for the synthesis and optimization of a green fuzzy controllers (FC) for an automatic control systems (ACS) of the reactor's temperature of a specialized pyrolysis plants (SPP). The proposed method gives the opportunity to synthesize and optimize the Mamdani-type green FCs that provide (a) high values of accuracy and quality indicators of temperature control, (b) low energy consumption in the process of functioning, as well as (c) relatively simple software and hardware implementation. The design of the Mamdani-type green FC for the temperature ACS of the pyrolysis reactor of the experimental SPP has been carried out in order to study and validate the effectiveness of the developed method. The developed green FC has a relatively simple hardware and software implementation and makes it possible to attain high values of quality indicators of temperature modes control at a sufficiently low energy consumption that confirms the high efficiency of the proposed method.

In Chapter "Models and Algorithms for Prediction of Electrical Energy Consumption Based on Canonical Expansions of Random Sequences", I. Atamanyuk, V. Kondratenko, Y. Kondratenko, V. Shebanin, and M. Solesvik discuss the development peculiarities and the use of models and algorithms as elements of the green technology to predict the electric energy consumption based on the mathematical apparatus of canonical expansions of random sequences. The developed calculation method does not impose any limitations on the qualities of random sequences of the change of electric energy consumption (exemplified by the requirements of linearity, Markovian behavior, monotonicity, stationarity, etc.). A comparative analysis of the results of a numerical experiment with the use of the

Kalman filter and the linear prediction method confirms the high efficiency of the developed models and algorithms (the relative error of prediction of electric energy consumption is 2–3%).

Chapter "On Quantitative Assessment of Reliability and Energy Consumption Indicators in Railway Systems", by D. Basile, F. Di Giandomenico, and S. Gnesi, addresses cross-validation on a case study from the railway domain by modeling and evaluating it using different formalisms and tools. Stochastic activity networks models and stochastic hybrid automata models of railroad switch heaters, developed for the purpose of evaluating the energy consumption and reliability indicators, are evaluated with the Mobius and Uppaal SMC. The authors compare the obtained results to improve their trustworthiness and to provide insights into the design and analysis of energy saving cyber-physical systems.

A. Orekhov, A. Stadnik, and V. Kharchenko, in Chapter "Cooperative Ecology Human–Machine Interfaces for Safe Intelligent Transport Systems: Cloud-Based Software Case", analyze the available solutions regarding the cooperative intelligent transport systems (ITS) and their human–machine interfaces (HMI). An approach and a design technique of HMIs for ITS based on designing the cooperative HMIs for intelligent transport systems are suggested. Special attention is paid to the architecture of cloud-based HMI for intelligent transport systems that consist of three main parts (the client end, server end, and the core project) and includes data models for the communication protocol and the common utility functions.

Chapter "Optimisation of Bi-directional Pulse Turbine for Waste Heat Utilization Plant Based on Green IT Paradigm", by Y. Kondratenko, S. Serbin, V. Korobko, and O. Korobko, is devoted to the environment preservation problem which is closely connected with issues of rational use of energy and material resources. By analyzing the state of the art of modern energy saving technologies, it is possible to conclude that there is a lack of effective and efficient solutions for utilizing waste low-temperature and cryogenic heat resources. The thermoacoustic heat machines (TAHM) can effectively use such energy sources. TAHM are characterized by high reliability, relative low cost, and environmental safety. Suchan approach will enable the creation of thermoacoustic turbine generators of a high power. The chapter presents the results of experimental research of the "bi-directional turbine—electric generator—TAHM" complex. A special computer-based control system was developed for the registration and analysis of experimental data. The system includes (a) MCU-based pulse-phase control subsystem for thermoacoustic engine heater, (b) data acquisition subsystem, and (c) SCADA-based PC software for data storage and analysis. The control system, designed using the green IT approach paradigm, is described in detail. The results of physical experiments make it possible to determine the boundary conditions for the computational fluid dynamics (CFD) simulation of working processes in the bi-directional turbine. The structural optimization of a turbine is considered based on the CFD modeling results.

The fourth part "Social, Educational and Business Aspects of Green IT Systems Application", including five contributions. V. Hahanov, O. Mishchenko, T. Soklakova, V. Abdullayev, S. Chumachenko, and E. Litvinova, in Chapter "Cyber-Social Computing", discusses structures of cyber-social computing which are considered as components of cloud-driven technologies for an exact monitoring and moral governance of the society. The main trends in the development of the cyber-physical structure presented in Gartner's Hype Cycle 2017 are described with the aim to apply them in science, education, transport, industry, and state structures. An expanded description of technologies is focused on the development of the smart digital world, green cities, and 5G telecommunications. The recommendations are given for leveraging the top 10 technologies of 2017 in business, scientific, and educational processes of higher education. Special attention is paid to the technologies of emotional–logical computing, metrics for measuring social relations along horizontal and vertical connections, rules and simulation of human behavior, and cyber-physical model of a green statehood for the metric management of resources and citizens.

In Chapter "Architecture for Collaborative Digital Simulation for the Polar Regions", M. Solesvik and Y. Kondratenko focus on offshore oil and gas operations which continuously move to the High North and to the Arctic which implies that problems of possible ship accidents and catastrophes are extremely dangerous for the nature. The authors analyze the recent literature on the new digital simulation technology, propose new ways for the utilization of this novel technology in the maritime industry, present fresh insights related to this architecture, software and hardware related to joint digital simulation platforms for planning safe maritime operations in the Arctic, and effective search and rescue operations. The utilization of joint digital simulation platforms will allow to minimize pollution effects during the navigation and accidents.

Chapter "Computer Support for Decision-Making on Defining the Strategy of Green IT Development at the State Level", by I. Shostak, I. Matyushenko, Y. Romanenkov, M. Danova, and Y. Kuznetsova, presents a complex estimation method of an innovative potential of the national economy in order to determine the prospects of the development of the green IT technology. The methodological basis of the method is cluster analysis. Data analysis is used for predicting the development of the green IT technologies on the dynamic time series. As a predictive model, the authors use the Brown model and a special structural parametric synthesis method that implements the identification of a time series by a preliminary analysis and processing of initial data, and identification of the trend and formation via interval estimates. The results of forecasting may be used, along with other initial data, for a foresight-type research on the definition of the green IT prospective directions of development. Special attention is paid to the generalized structure of interactive decision support system for the development of the green IT at the state level.

K. Czachorowski, M. Solesvik, and Y. Kondratenko, in their Chapter "The Application of Blockchain Technology in the Maritime Industry", present fresh insights related to the application of the novel blockchain technology for reducing pollutions, analyze recent literature on blockchain technology, and propose ways of the utilization of blockchain technology in the maritime industry. The technology has a broad range of applicability making it possible to connect the supply chain more efficiently, providing the exchange and visibility of time-stamped proofed data, and decreasing the industry operational costs with intermediaries and increasing security. It also allows a full visibility for all parties involved with a proof of work, facilitating the class societies inspections, port state control, and audits compliance. The results of the study also show that cases of blockchain application in other fields increase the industry willingness to its application in the maritime industry. Having blockchain implementations, specialized third parties would increase the implementation possibility and the industry willingness due to reduced costs and friction.

In Chapter "Software Engineering Sustainability Education in Compliance with Industrial Standards and Green IT Concept", I. Turkin and Y. Vykhodets present a scientific approach to building the course "Software Engineering Sustainability" to introduce the sustainable concept into educational programs for software engineers. The authors consider the peculiarities of the course design with description of the course content and visible positioning green IT.

To briefly summarize, the chapters selected for this book provide an overview of some crucial problems in the area of green IT engineering and of approaches and techniques that relevant research groups within this area are employing to try to solve the problems considered.

We would like to express our deep appreciation to all authors for their contributions as well as to reviewers for their timely and interesting comments and suggestions. We certainly look forward to working with all contributors again in nearby future.

We also wish to thank Dr. Tom Ditzinger, Dr. Leontina di Cecco, and Mr. Holger Schaepe from Springer for their dedication and help to implement and finish this publication project on time maintaining the highest publication standards.

Kharkiv, Ukraine — Prof. Dr. Sc. Vyacheslav Kharchenko
Mykolaiv, Ukraine — Prof. Dr. Sc. Yuriy Kondratenko
Warsaw, Poland — Prof. Dr. Sc. Janusz Kacprzyk
April 2018

Contents

Part I
Development and Optimization of Green IT Systems

Including Software Aspects in Green IT: How to Create Awareness for Green Software Issues

Eva Kern, Achim Guldner and Stefan Naumann

Abstract While counteracting the increasing demand for natural resources and especially energy of ICT, first successes have become apparent by activities comprised by the term “Green IT”. Nowadays, many of the current activities lay emphasis on the hardware side of Green IT. However, software issues play a significant role in defining system and hardware requirements as well as the amount of energy consumed by ICT devices and the underlying infrastructure. Thus, the following chapter introduces the idea of green software and its engineering. Complementary to definitions and models in the addressed field, a more practical insight is given by illustrating exemplary energy measurements of software products. While these aspects show that the research is increasingly dealing with software issues of Green IT, these mainly scientific ideas have hardly reached practical relevance, so far. Hence, following the life cycle perspective of software products, we present two exemplary concepts on how to increase awareness of green software: addressing especially software engineers, we propose to implement continuous energy efficiency measurements during the development phase. With regards to software users, we propose to create an eco-label for software products to inform about their environmental issues and thus create more transparency in this context. To do so, we present and evaluate criteria and indicators, upon which a label could be based. The chapter concludes with a summary and proposes future activities in the addressed field. Overall, the aim of the chapter is to point out solutions that

E. Kern (✉)
Leuphana University of Lüneburg, Universitätsallee 1, 21335 Lüneburg, Germany
e-mail: mail@nachhaltige-medien.de

E. Kern
University of Applied Sciences Trier, Environmental Campus Birkenfeld, Birkenfeld, Germany

A. Guldner · S. Naumann
University of Applied Sciences Trier, Environmental Campus Birkenfeld, Institute for Software Systems, Birkenfeld, Germany
e-mail: a.guldner@umwelt-campus.de

S. Naumann
e-mail: s.naumann@umwelt-campus.de

© Springer Nature Switzerland AG 2019
V. Kharchenko et al. (eds.), *Green IT Engineering: Social, Business and Industrial Applications*, Studies in Systems, Decision and Control 171,
https://doi.org/10.1007/978-3-030-00253-4_1

might lead to a more environmentally friendly way of developing software, a well-informed procurement behavior regarding software products, and a more sustainable user behavior concerning ICT.

Keywords Green software · Energy measurements · Energy efficiency Eco-label · Green software development · Environment

1 Introduction

Regarding a sustainable development, as firstly published in the Brundland Report [1] and nowadays put straight into 17 United Nations' Sustainable Development Goals (SDGs),[1] information and communication technologies (ICT) play a special part: On the one hand, ICT consumes a lot of resources during its production and usage [2] and will consume even more in the future [3]. On the other hand, these technologies can support environmental processes and even make processes more sustainable in many different branches [4]. The latter is referred to by the term “Green by IT”.

In order to counteract the rising resource consumption by ICT, activities in the field of Green IT found its way into research, industry, and private households. While they are mainly focused on hardware aspects, we emphasize the software side, since “software characteristics determine which hardware capacities are made available and how much electric energy is used by end-user devices, networks, and data centers” [5]. Thus, “referring to software energy consumption or efficiency, we are actually referring to the energy consumed by the hardware when induced by the software to perform some task(s)” [6]. Referring to a sustainable development, Maevsky et al. [7] show the economic relevance of considering the software side of Green IT by presenting a corresponding calculation model. However, we focus on the environmental issues of sustainable software within the following chapter. Additionally, they claim that the “amount of electrical energy consumed by the computer is largely dependent on the program compilation quality and Software optimization degree.” [7]

Within the Green IT community, as well as the groups specialized in Green Software, research activities are increasing during the last years. This is, for example, shown by raising numbers of publications, conferences, information events, workshops, and individually reported in sustainability and environmental reports of the big players in the worldwide IT market. However, especially referring to the software side, the awareness on environmental aspects could be higher in industry and in daily (private) routines. Different studies show that even if the interviewed person knows Green IT strategies, the transfer towards practical

[1]https://sustainabledevelopment.un.org.

implementation and into daily routines is missing—on the professional as well as on the private side [8–10].

Thus, after introducing the research field "green software engineering" (Sect. 2), we will present approaches on how to support especially software developers and users in considering environmental issues of software (Sect. 3). The approaches are linked to the life cycle of software products, based on the "from cradle to cradle" principle. Generally, the concepts can contribute to reducing the overall carbon footprint of ICT and to the United Nations' SDG 12 "Ensure sustainable consumption and production patterns". The chapter concludes with a brief summary and an outlook on potential future research activities (Sect. 4).

2 State of the Art: Green Software and Its Field of Research

The following section will introduce the topic of green software in general and present definitions for central terms, since "[defining] and developing adequate support requires a commonly accepted definition of what sustainability means in and for software engineering" [11]. However, we do not aim at presenting a comprehensive literature overview with the following section. This can be found e.g. in [12, 13].

2.1 Definitions for Green and Sustainable Software: Overview, Commonalities, and Distinctions

Similarly to Green IT and Green by IT (see Introduction), in combining sustainability issues to the software side, "the differentiation can be made by distinguishing software (engineering) for sustainability, which is related to the absolute definition, and sustainable software (or sustainability in software engineering), which is related to the relative definition" [14]. While the so-called "absolute definition" refers to approaches that reduce the environmental impact of a process by using software, Taina [15] talks about "Software Engineering for the Planet (SEP)" referring to the "sustainable by software" dimension. According to their understanding, "Green IT belongs to SEP [and] includes other areas as well". These are [15]: (1) Software support for green education, (2) Green metrics and decision support, (3) Lower IT energy consumption, and (4) Support for better climate and environment models.

Concepts on "sustainable (in) software" deal with possibilities to improve the software product itself. Even if referring to sustainable software in the sense of the Brundlandt Report, many approaches rank environmental aspects first (e.g. [12, 16, 17]). In addition to environmental, Lago et al. [18] describe five dimensions that could be interpreted as five perspectives on sustainable software:

- Social sustainability: Which effects do software systems have on the society (e.g. communication, interaction, government…)?
- Environmental sustainability: How does software affect the environment during, inter alia, development and maintenance?
- Technical sustainability: How can software be created so that it can easily adapt to future changes?
- Economic sustainability: How can software systems be created so that the stakeholders' long-term investments are as safe as possible from economic risks?
- Individual sustainability: How can software be created and maintained in a way that enables developers to be satisfied with their job over a long period of time?

Dividing the concept of sustainability into different dimensions—let it be three, four or five—or applying criteria to it makes the sustainability of a product feasible. Hence, several researchers developed corresponding criteria (e.g. [15, 16, 19, 20]). While Taina [15] distinguishes between feasibility, (carbon footprint, energy, travel, etc.), efficiency (CPU intensity, idleness, etc.), and sustainability (reduction, beauty, etc.), Calero et al. [20] divide sustainability into the three sub characteristics energy consumption, resource optimization, and perdurability. In our project "Sustainable Software Design"[2] we focus on the first of the proposed categories and will publish a set of criteria for sustainable software products. It is hierarchically structured in the main criteria "resource efficiency", "potential hardware operating life", and "user autonomy". The criteria are operationalized by appropriated indicators. In this context, Betz et al. [21] differentiate between "resource oriented indicators", taking environmental aspects of sustainability into account, and "well-being oriented indicators", referring to "human needs aspects of sustainability".

Bringing the different approaches together, the sustainability of software products can be interpreted as a non-functional requirement [21, 22] and thus "the sustainability assessment would be considered as another quality aspect to be taken into account by developers, in accordance with the priorities and requirements imposed for the product being developed" [22]. Seeing it as a quality criterion for software, sustainability is an improvement of software products. Taking this approach, literature talks about "green/sustainable software engineering/development". The aim of this process is the "development of the software product for long living systems that can meet the needs of the present to the future generations with the integration of the three pillars sustainability concept i.e. (environment, economic, social) as to fulfill the requirements in a timely basis." [23]. Thus, "sustainability of a software product [can be defined] as the capacity of developing a software product in a sustainable manner" [20].

Summarizing, the commonalities in the understanding of the term "(green and) sustainable software" are [24]: "(1) environmental and resource protection as well as (2) supporting sustainable development including the three pillars of

[2] http://green-software-engineering.de/en/projects/ufoplan-ssd-2015.html.

sustainability, but setting priorities." None of the definitions stands in total contrast to the other ones but each of them represents a specific perspective that must be kept in mind while talking about green, or rather, sustainable software, since it is strongly dependent on the currently defined view [11]. For the purpose of this chapter, we refer to our definition of green and sustainable software: "[Green and sustainable] software is software, whose impacts on economy, society, human beings, and environment that result from development, deployment, and usage of the software are minimal and/or which have a positive effect on sustainable development." [16]

We lay emphasis on environmental issues of sustainability and use the terms "green software" and "sustainable software" in an analogous manner. A more comprehensive engagement with the terminology can be found in e.g. [17, 25].

2.2 *Green Software Models: Quality Models, Development Models, and Reference Model*

Based on a mutual understanding of green and sustainable software, different models comprise even more aspects that belong to the addressed research field. Our GREENSOFT Model (Fig. 1), presented in [16], takes a broad look at green software and its engineering.

The reference model comprises four parts: the life cycle of software products, sustainability criteria and metrics, procedure models, and recommendations and

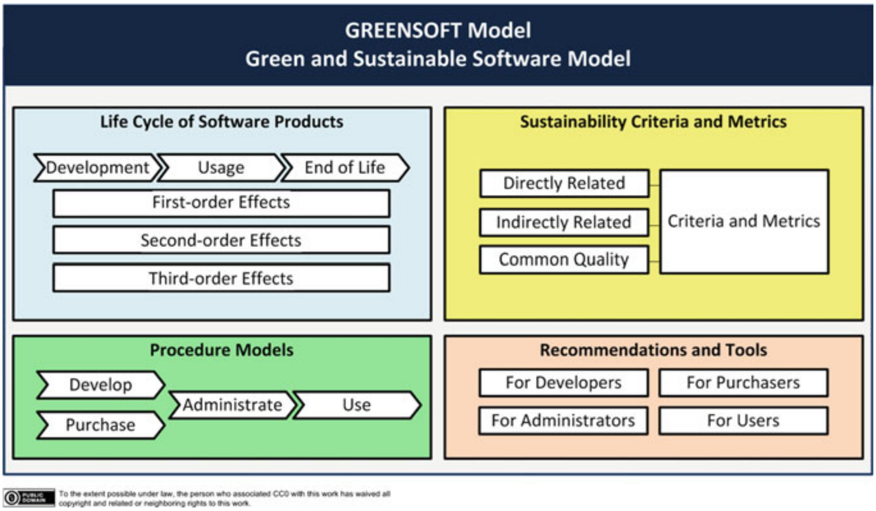


Fig. 1 The GREENSOFT model, a reference model for green and sustainable software engineering [27]

tools. Thus, it structures concepts, strategies and processes in the field of Green (by) IT, focusing on software aspects. Following the approach of a reference model, e.g. as presented in [26], the idea is to present an overview of the field and to structure several aspects in the context of green and sustainable software engineering. Hence, other models can be assigned to the GREENSOFT Model.

Existing life cycle approaches in the context of green software (being the first part of the model), from which criteria and metrics can be derived, will be presented in the next section. Generally, this approach considers ecological, social, and economic aspects of products over their whole life cycle.

The second model part “Sustainability Criteria and Metrics” covers general criteria and metrics regarding quality of software and allows a classification of the sustainability of the products. Based on the approach by Berkhout et al. [28], we distinguish between first-order (“direct environmental effects of the production and use of ICTs”), second-order (“indirect environmental impacts related to the effect of ICTs on the structure of the economy, production processes, products and distribution systems”), and third-order effects (“indirect effects on the environment, mainly through the stimulation of more consumption and higher economic growth by ICTs (‘rebound effect’), and through impacts on life styles and value systems”). Calero et al. [29] use the modelling approach to go deeper into characterization. They take the ISO Model 25010 as starting point and identify characteristics to be included in the model in order to explicitly integrate sustainability aspects of software. Possible characteristics of green and sustainable software products are also summarized within our quality model, presented in [19]. Both models [19, 29] can be placed in the second part of the GREENSOFT Model. Even more possibilities on how to include green software characteristics in existing standard software quality models, are proposed by Gordieiev et al. [30]. They analyze well-known software quality models with regards to green and reliability issues and the evaluation of these characteristics over time. In conclusion, they “assume that the next general [Software Quality Model] will include [Green Software] characteristics in an explicit form” [30].

Lami et al. [31] distinguish between the “software lifecycle” and “software processes” and identify sustainability factors belonging to different processes. They present a model to enable an “evaluation of the degree of process sustainability of an organization” and, in a next step, improve the sustainability of soft-ware processes. Thus, referring to Fig. 1, the model presented by Lami et al. belongs to “Procedure Models”. Here, procedure models refer to the development and purchasing processes of software as well as user support while administrating and using the product. In this context, verification regarding efforts and costs is another important aspect of developing software. Thus, we recommend logging the corresponding efforts and costs of the process, especially to be able to contrast sustainable software engineering with “standard procedures”. One example how to do so is the “Process Model for Green and Sustainable Software Engineering”, that includes a “Sustainability Review and Preview” as presented in [32].

Finally, the GREENSOFT Model mentions “Recommendations and Tools”. Thereby, it points out the importance to integrate different roles into the activities aiming at creating a more sustainable ICT and aims at putting the model into practice: developers, purchasers, administrators, and users need context-specific knowledge, in form of recommendations, tools, and methods to support a sustain-able way of developing, purchasing, administrating, and using software. Exemplarily for the whole software life cycle, we present tools to create awareness and, in the long run, deeper knowledge, of environmental issues of software and thus a way of developing and using software in a green and sustainable manner.

3 Creating Awareness of Green Software: Life Cycle Perspective

The proposed life cycle for software products orients towards a cradle-to-grave approach [33]. The development phase considers impacts on a sustainable development directly or indirectly caused by activities belonging to the software development. Examples for those impacts are the energy consumption caused by ICT devices that are used to program the software and by commuting of the software developers that results in CO_2 emissions. Additionally, also social aspects find their way into this consideration, e.g. working conditions [34]. Following the development phase, the distribution of the product plays a role regarding a sustainable software life cycle. Here, resource consumption caused by the chosen data medium, or rather the download size, are examples for first-order effects. The us-age phase results especially in environmental effects: software induced energy and resource consumptions, hardware requirements and portability with regards to durability are some of the relevant aspects. However, considering issues like transparency and accessibility also relates this phase to social sustainability. Looking at the end of a software's life, deactivation (in regard to e.g. backup size, long term storage of data) and disposal (e.g. packaging, data medium) have an influence on the sustainability of the considered product.

The objective of the life cycle of software products included in the GREENSOFT Model "is to enable stakeholders to assess impacts on [sustainable development]" [16]. Thus, the question "Who are the stakeholders in context of a life cycle of (sustainable) software products?" arises. While Penzenstadler et al. [35] identify stakeholders, who are important for successfully implementing sustainability issues, Herzog et al. [36] list actors developing and actors supporting innovations in the context of Green IT. Concerning our research, we especially take into account stakeholders of the development and the usage phase. This is due to the fact that if it is possible to create or increase their awareness of green software, this can yield the biggest impact, since they are directly connected to software products—either in an active (developing) or a passive (using) role. In both cases, tools to support a more sustainable way of developing and especially using software products are required. Thus, we present two possible tools within the following section.

Here, we lay emphasis on the development and usage phase of the software life cycle, since an "early awareness of green software could save a significant amount of costs compared to refactoring the software at a later stage." [37]. Maevsky et al. [7] and Chemeris et al. [38] also stress the importance of the software development phase and advances in terms of power consumption when adopted early in the design process. Chemeris et al. [38] show the positive effect of program optimization in case of program loops on the resulting power consumption.

However, in most cases, the programmers do not even know about their influence or rather lack of knowledge on energy and sustainability issues regarding software [8, 9, 39]. Sustainability of software is, according to Groher et al. [39], mainly related to "maintainability and long-term usability". Overall, "[technical] and organizational aspects are clearly in focus, followed by economic aspects. Environmental considerations are missing." [39] If practitioners have knowledge in these contexts, they do just minimally take them into account during software development [8], and just connect them to specific kinds of software products (e.g. mobile development [8]). Generally, the transfer of knowledge to practice is missing [9, 40]. Even if the different studies do not agree on the existing amount of awareness on energy, or rather sustainability issues in software engineering, they agree that the potential is still not exhausted, so far. In total, the "lack of attention to software energy consumption is an issue of priorities" [8].

The same applies for users of software products: 51.1% of the respondents of a user study, we conducted in 2016, can imagine taking "certification of environmental issues of the software product" as one possible buying criterion, if there are products with the same functionality, some being "green" and others being "not green" [41]. Overall, the sample of our survey comprised 712 completed questionnaires, primarily answered by German users (Table 1). The survey participants set the priority on the functionality of a software product.

Similarly to software engineers, users might have some ideas on green computing [10, 42], energy [6], or rather sustainability issues of software [43], but awareness, information, and education are missing [6, 8, 10, 42, 43]. Ferreira [6] provides the hypothesis "Given the clear awareness, respondents are taking energy

Table 1 Key data of the survey on the awareness of and the interest in environmental issues of software, conducted in 2016

Method	Online survey
Structure of the survey	Demographic characteristics of the participants Evaluation of criteria for green software products Aspects influencing the acceptance of an eco-label for software products
Participants	German speaking Internet user
Sample	n = 854
Completed questionnaires	712 (revised data set)
Survey period	16/08 to 05/10/2016

consumption into consideration when buying or developing software applications". In this context, Selyamani et al. [42] point out that "[awareness] is a need to practice an idea of the new things. Without awareness, the practice does not do appropriately and adequately and the other way around."

3.1 Development: Continuous Energy Efficiency Measurements

One way to improve awareness of sustainable software, especially with developers and users, is to provide a means for comparable measurements. While other references talk about "software development life cycles" [44, 45], we focus on tools that can be directly integrated in existing development processes and thus support developers in producing sustainable software products. According to Anwar et al. [13], current research activities in the context of green software engineering are missing supporting tools, e.g. "energy aware testing tools". They especially mention tools that evaluate the energy consumption during development as one example of research gaps that need to be closed. Such kinds of tools are required to produce green software products. Currently, "software practitioners prefer to use static analysis and energy profiling in order to point out energy problems during development" [40].

Mahmoud et al. [44] also follow the idea of identifying how software causes energy and resource consumption but take a broader view and thus present a theoretical approach. Shenoy et al. [45] lay the focus on more general guidelines and recommendations for developing software in a sustainable manner. In comparison to both, we outline a more practical approach and its application.

Taking one step back towards defining sustainability goals for the product to be developed, Mahaux et al. [46], Becker et al. [47], and others lay emphasis on requirements engineering and the design of sustainable software.

In order to get an impression of the environmental impact of the software product and to take sustainability issues into account while developing software, we propose to continuously integrate energy efficiency measurements into the development process. The tool presented below can both be integrated in the development phase or used for stand-alone measurements. In that way, it is possible to provide "actionable timely information, to make useful trade-offs between energy efficiency and other quality attributes" as claimed by Anwar et al. [13]. Over time, providing information about the energy consumption of software can help developers to get a feeling for the energy values of the products they are working on [40]. Thus, this information might positively influence the awareness of software engineers regarding environmental issues of software. Additionally, our approach focuses "on rating energy efficiency during the development process on an on-going basis" and is "based on well-known software testing approaches and [continuous integration], to take energy efficiency into account during the daily work of a software developer" [48]. Hence, in order to meet the requirements of practitioners [39], our approach for energy efficiency measurements can be adapted in existing working structures very well. Furthermore, the approach can also be used to compare the energy consumption of two or more software products that perform the same task (e.g. standard software like word processors or databases).

The approach is based upon a measurement setup that follows ISO/IEC 14756, as introduced in [49]. It allows the recording of the energy consumption of a system under test (SUT), which runs the software that is to be measured. A power meter measures the energy consumption of the SUT while it runs the software. Then, a workload generator provides and controls the tasks performed by the software on the SUT in a usage scenario. During the scenario, the SUT can additionally monitor its hardware usage statistics (CPU, RAM, HDD, Network traffic, etc.). If the time of all three systems that collect data (SUT, power meter and workload generator) is synced, the data that is collected this way can easily be aggregated and evaluated by means of timestamps attached to the data. Figure 2 depicts the measurement setup.

Using this setup, we can easily measure the energy consumption and hardware utilization of software and compare two software products that perform the same task in a usage scenario. These scenarios need to be tailored to the software product to be examined (e.g. for interactive software like browsers different scenarios are needed than for non-interactive software like databases) and the questions to be answered (e.g. the comparison of two or more software products vs. the changes of one software product over time vs. measuring individual functionalities, etc.). Furthermore, if the SUT that is measured is a continuous integration server and the usage scenario consists of the unit tests of a software under development,

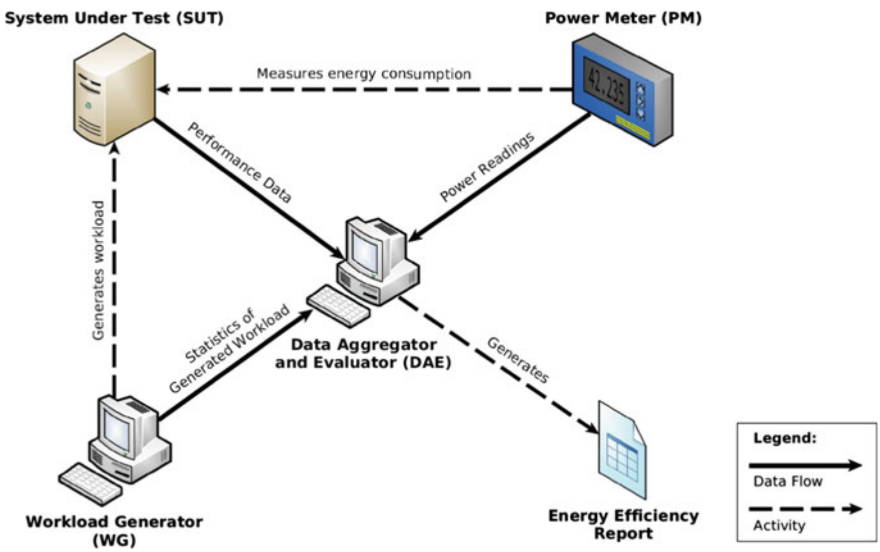


Fig. 2 Setup for measuring power usage and hardware utilization of software (based on [27])

the developers can be provided with a continuous history of the power and hardware consumption during the development phase. An approach for this idea is presented in depth in [50, 51].

To provide an idea of the measurement method and results, in the following we describe an exemplary measurement that we conducted to compare two software products, in this case word processors. The main goal of the measurement was to create a measurement that allowed us to compare the energy consumption and hardware usage of the two products. Hence, we created a usage scenario for this purpose. This scenario is supposed to emulate a user of the software product as realistically as possible. Thus, we first analyzed the tasks that are usually carried out with the products (e.g. typing and formatting text, inserting graphics, creating and updating references and a table of contents, etc.). Then, we checked these tasks against functionalities which may, according to expert opinion, induce high energy demand or high resource utilization (e.g. saving and loading, search and replace, spell-checking, etc.). Additionally, we needed to ensure that all selected actions could be executed with both software products and that the result (in this case the exported.pdf document) is identical for both products. We then measured the energy consumption and hardware utilization of the SUT while performing the scenario (duration approx. 10 min), as described above. Thus, the same work was done with each software product and we could compare the average energy consumption and hardware utilization for this scenario. To ensure that we really measured the energy consumption of the software products, we additionally switched off all possible background processes (like virus scanners, updates, indexing processes, etc.). Furthermore, we repeated the measurements 30 times and took the average to mitigate the impact of outliers. As an example, Fig. 3 shows the resulting graph of the energy measurement.

From the measured average power consumption, we calculated the consumed electrical energy (Wh). Here, it is obvious that word processor 2 consumed more

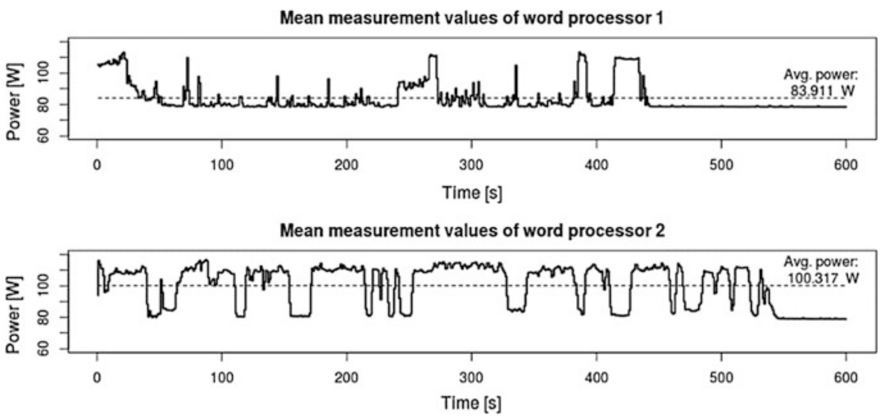


Fig. 3 Comparison of word processors: per second average of power consumption

energy, which was also confirmed by the fact that it took longer to generate the wanted.pdf and that the hardware utilization (especially CPU) during the scenario was higher as well. In this case, word processor 2 used on average 16.72 Wh and word processor 1 only 14.43 Wh.

In general, the measurement method is one possible tool in the field of green software engineering and, thus, belongs to the fourth part of our GREENSOFT Model (Fig. 1). Furthermore, the software energy measurements could create an informative basis for many stakeholders, like programmers, users, purchasers, etc. Just to give an example how the measurements can be interesting for programmers: Maevsky et al. present an experiment to show that the "power consumption increase [s] in a mobile device after its being infected by malicious software." [52] Our idea is to use the measurement results as part of a set of criteria, an eco-label for software products could rely on. Thus, this opportunity for use seems to be especially interesting for purchasers, but also for users. The following section will present the labelling approach in a more detailed way.

3.2 Distribution and Usage: Labelling Green Software Products

Eco-labels are supposed to provide information on the environmental friendliness of products to purchasers and users of these products. Different studies show that these approaches can be considered to be successful [53, 54]. Referring to the GREENSOFT Model, depicted in Fig. 1, eco-labels are attributed to "Sustainability Criteria and Metrics" and "Recommendations and Tools". Depending on the point of view, even the "Life Cycle of Software Products" might be relevant in the context of eco-labelling software products.

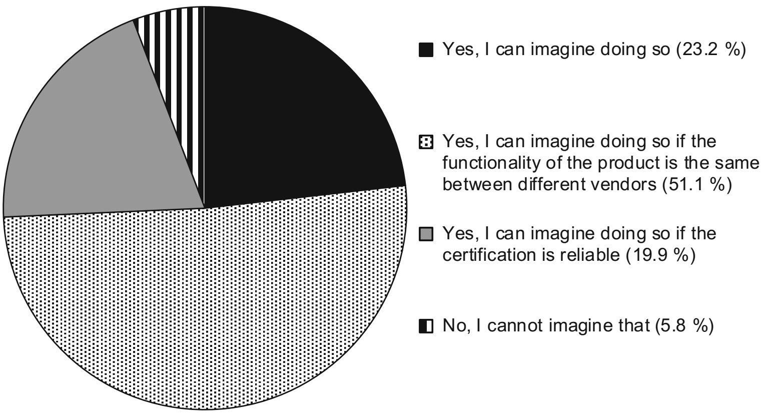


Fig. 4 Results of the question "can you imagine taking 'grading of the environmental impact of the product' as a criterion while searching for a software product?"

However, as far as we know, there is no standardized label or any similar form of characterization for green software products, so far. According to our survey we conducted in 2016 (Table 1), the interest in such an eco-label is given (see Fig. 4, further details can be found in [41]). The potential of such a label is also present: in the USA, energy savings comparable to the energy consumption of ten million households have been achieved by purchasing and using of products certified by the ENERGY STAR in 2002 [53].

In order to develop a label for green and sustainable software products, specific criteria, upon which a label could be based, are required [24]. Figure 5 depicts the procedure to create such a set of criteria.

Since there are Green IT labels referring to hardware products (e.g. ENERGY STAR, Blue Angel, TCO Certified), we analyzed the corresponding awarding guidelines, in order to find possibilities to transfer them to software products (Fig. 5, Step 1: "Collection of hardware criteria" and Step 2: "Deduction of software criteria"). Additionally, we reviewed existing literature on criteria for sustainable software products and added further ideas for criteria. We focused on environmental aspects and the usage phase of a software product. Thus, the frame for the eco-label in mind can be summed up as follows:

1. The label should be awarded to environmentally friendly products, i.e. the underlying criteria should evaluate environmental impacts of the product but not include social or economic aspects.
2. It should be awarded to the product itself, i.e. without the development and distribution process, the surroundings, and the developing organization.
3. It should be awarded because of "green in software" aspects. This means the product itself should be as environmentally friendly as possible. Products that support environmentally friendly processes, or rather make them more environmentally friendly ("green by software"), but are not green themselves should not be included.

Starting with these three principles, we adjusted the criteria collection (Fig. 5, Step 3: "Selection of criteria"). Additionally, as proposed by Mazijn et al. [55], we evaluated the resulting criteria collection by their relevance (Is the aspect relevant for software products?), measurability (Is it possible to measure the aspect through currently existing methods?), feasibility (Is it possible to provide information on the fulfillment of the aspect for software products?), and discrimination (Is the aspect different to other aspects?).

The selection process resulted in the following list of possible criteria for green and sustainable software products:

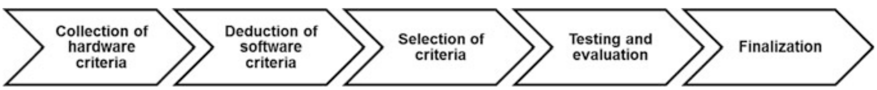


Fig. 5 Procedure approach for developing sustainability criteria for software

- Backward compatibility
- Energy efficiency
- Hardware efficiency
- Hardware sufficiency
- Independence of outside resources
- Maintenance functions
- Platform independence
- Portability
- Quality of product information
- Transparency
- Uninstallability

The next step (Fig. 5, Step 4) was to evaluate these criteria ideas. This was done by (1) evaluating the social interest in the proposed criteria and (2) proof of applicability. The social evaluation was done by our user survey, addressing users of software products [41]. As a result, especially the efficiency aspects (77.5% say "energy efficiency should be labelled", 62.5% say "hardware efficiency should be labelled") and platform independence (56.9%) raise the user's interests. The result goes well together with the statement by Kharchenko et al. [17]: "Energy saving and energy efficiency are the main characteristics of green computing."

In order to prove the relevance and verifiability of the criteria ideas, we evaluated the software products by measurements as described in Sect. 3.1. Besides black-box measurements (e.g. in case of resource and energy consumption), we used existing product information, research, and visual inspection for the evaluation of the criteria. The analysis of the tests includes a calculation of indicators of the proposed criteria, an evaluation of the selected information, a comparison within the product group, and a statistical presentation of the results.

After the evaluation of the criteria collection, we structured, described, and operationalized the criteria. The result of the criteria development process is a set of criteria that is a basis for awarding an eco-label for software products.[3]

4 Conclusion and Outlook

Our contribution describes several approaches and procedure models, how green software can be classified and developed, and how different stakeholders like developers, purchasers, and end users can be supported. In doing so, we showed that the increasing demand of energy by information and communication technology (ICT) is caused by hardware, but software and algorithms induce this consumption. Consequently, and looking at the complete life cycle of a software product, all phases from development, usage, and also disposal have to be considered. Besides several methodological and practical procedures, a deeper

[3]http://green-software-engineering.de/criteria-catalog.

awareness for the role of software regarding energy and resource consumption is necessary, as well. In our contribution, we presented two exemplary concepts on how to increase this awareness for green software. Addressing especially software engineers, we propose to implement continuous energy efficiency measurements during the development phase. With regards to software users, we propose to create an eco-label for software products to inform about environmental issues of software products and thus create more transparency in this context. To be more concrete, we presented and evaluated criteria and indicators, upon which a label could be based.

Future work will test further software products and also modules regarding their "greenness". The overall goal is that these developed criteria are applied on a regular basis in software development processes like other non-functional requirements. This will reduce the energy and resource footprint of ICT and helps additionally to make software more efficient in several cases.

References

1. United Nations General Assembly.: Report of the World Commission on Environment and Development: Our Common Future. UN Document No. A/42/427 English, New York (1987)
2. van Heddeghem, W., Lambert, S., Lannoo, B., et al.: Trends in worldwide ICT electricity consumption from 2007 to 2012. Comput. Commun. **50**, 64–76 (2014)
3. Andrae, A.S.G., Edler, T.: On global electricity usage of communication technology: trends to 2030. Challenges **6**(1), 117–157 (2015)
4. Vickery, G., Mickoleit, A.: Greener and smarter: information technology can improve the environment in many ways. In: Noam, E.M., Pupillo, L.M., Kranz, J.J. (eds.) Broadband Networks, Smart Grids and Climate Change. Springer, pp. 33–37 (2013)
5. Hilty, L., Lohmann, W., Behrendt, S. et al.: Green Software: Final Report of the Project: Establishing and Exploiting Potentials for Environmental Protection in Information and Communication Technology (Green IT). Report Commissioned by the Federal Environment Agency, Berlin, Förderkennzeichen 3710 95 302/3(23) (2015)
6. Ferreira, M.: Green software awareness survey: results. In: Presented at Report Workshop Green Software Architecture, Tuesday 7 June 2011, Amsterdam, Netherlands, Amsterdam (2011)
7. Maevsky, D.A., Maevskaya, E.J., Stetsuyk, E.D.: Evaluating the RAM energy consumption at the stage of software development. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Concepts, Models, Complex Systems Architectures, pp. 101–121. Springer (2016)
8. Pang, C., Hindle, A., Adams, B., et al.: What do programmers know about software energy consumption? IEEE Softw. **33**(3), 83–89 (2016)
9. Chitchyan, R., Becker, C., Betz, S. et al.: Sustainability design in requirements engineering: state of practice. In: Proceedings of the 38th International Conference on Software Engineering Companion, pp. 533–542 (2016)
10. Dookhitram, K., Narsoo, J., Sunhaloo, M.S. et al.: Green computing: an awareness survey among university of technology, mauritius students. In: Conference Proceeding of International Conference on Higher Education and Economic Development, Mauritius. Available from http://tec.intnet.mu/pdf%20downloads/confpaper/confpaper091224.pdf (2012)
11. Penzenstadler, B.: What does sustainability mean in and for software engineering? In: Hilty, L.M., Aebischer, B., Andersson, G. et al. (eds.) ICT4S ICT for Sustainability: Proceedings of the First International Conference on Information and Communication Technologies for

Sustainability, ETH Zurich, February 14–16, 2013. ETH Zurich, University of Zurich and Empa, Swiss Federal Laboratories for Materials Science and Technology, Zürich (2013)
12. Penzenstadler, B., Raturi, A., Richardson, D. et al.: Systematic Mapping Study on Software Engineering for Sustainability (SE4S)—Protocol and Results, Irvine (2014)
13. Anwar, H., Pfahl, D.: Towards greener software engineering using software analytics: a systematic mapping. In: 2017 43rd Euromicro Conference on Software Engineering and Advanced Applications (SEAA), pp. 157–166 (2017)
14. Penzenstadler, B.: Towards a definition of sustainability in and for software engineering. In: Proceedings of the 28th Annual ACM Symposium on Applied Computing, pp. 1183–1185 (2013)
15. Taina, J.: Good, bad, and beautiful software—in search of green software quality factors. In: Lopez-Lopez, J.-C., Sissa, G., Natvig, L. (eds.) Green ICT: Trends and Challenges, XII, pp. 22–27 (2011)
16. Naumann, S., Dick, M., Kern, E., et al.: The GREENSOFT model: a reference model for green and sustainable software and its engineering. SUSCOM **1**(4), 294–304 (2011). https://doi.org/10.1016/j.suscom.2011.06.004
17. Kharchenko, V., Illiashenko, O.: Concepts of green IT engineering: taxonomy, principles and implementation. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Concepts, Models, Complex Systems Architectures. Springer, pp. 3–19 (2016)
18. Lago, P., Koçak, S.A., Crnkovic, I., et al.: Framing sustainability as a property of software quality. Commun. ACM **58**(10), 70–78 (2015)
19. Kern, E., Dick, M., Naumann, S. et al.: Green software and green software engineering—definitions, measurements, and quality aspects. In: Hilty, L.M., Aebischer, B., Andersson, G. et al. (eds.) ICT4S ICT for Sustainability: Proceedings of the First International Conference on Information and Communication Technologies for Sustainability, ETH Zurich, February 14–16, 2013. ETH Zurich, University of Zurich and Empa, Swiss Federal Laboratories for Materials Science and Technology, Zürich, pp. 87–94 (2013)
20. Calero, C., Bertoa, M.F., Moraga, M.Á.: Sustainability and quality: icing on the cake. In: RE4SuSy@RE (2013)
21. Betz, S., Caporale, T.: Sustainable software system engineering. In: 2014 IEEE Fourth International Conference on Big Data and Cloud Computing (BdCloud), pp. 612–619 (2014)
22. Calero, C., Bertoa, M.F., Angeles Moraga, M.: A systematic literature review for software sustainability measures. In: 2nd International Workshop on Green and Sustainable Software (GREENS), pp. 46–53 (2013)
23. Ahmad, R., Baharom, F., Hussain, A.: A systematic literature review on sustainability studies in software engineering. In: Proceedings of KMICe (2014)
24. Kern, E., Dick, M., Naumann, S. et al.: Labelling sustainable software products and websites: ideas, approaches, and challenges. In: Johannsen, V.K., Jensen, S., Wohlgemuth, V. et al. (eds.) Proceedings of EnviroInfo and ICT for Sustainability 2015: 29th International Conference on Informatics for Environmental Protection (EnviroInfo 2015) and the 3rd International Conference on ICT for Sustainability (ICT4S 2015). Copenhagen, September 7–9, 2015, pp. 82–91. Atlantis Press, Amsterdam (2015)
25. Calero, C., Piattini, M.: Introduction to green in software engineering. In: Calero, C., Piattini, M. (eds.) Green in Software Engineering, pp. 3–27. Springer (2015)
26. Schütte, R.: Grundsätze ordnungsmäßiger Referenzmodellierung: Konstruktion konfigurations-und anpassungsorientierter Modelle, vol. 233. Springer (2013)
27. Research Project "GREENSOFT".: Website: Research Project "Green Software Engineering"—Downloads. http://www.green-software-engineering.de/en/downloads.html (2014)
28. Berkhout, F., Hertin, J.: Impacts of Information and Communication Technologies on Environmental Sustainability: Speculations and Evidence: Report to the OECD. http://www.oecd.org/dataoecd/4/6/1897156.pdf. Accessed 02 Mar 2011 (2001)
29. Calero, C., Bertoa, M.F.: 25010+S: a software quality model with sustainable characteristics. Sustainability as an element of software quality. In: Proceeding of the GIBSE (2013)

30. Gordieiev, O., Kharchenko, V., Fusani, M.: Software quality standards and models evolution: greenness and reliability issues. In: International Conference on Information and Communication Technologies in Education, Research, and Industrial Applications, pp. 38–55 (2015)
31. Lami, G., Fabbrini, F., Fusani, M.: Software sustainability from a process-centric perspective. In: Winkler, D., O'Connor, R.V., Messnarz, R. (eds.) EuroSPI 2012, CCIS 301, pp. 97–108. Springer (2012)
32. Kern, E., Naumann, S., Dick, M.: Processes for green and sustainable software engineering. In: Calero, C., Piattini, M. (eds.) Green in Software Engineering. Springer (2015)
33. Tischner, U., Dietz, B., Maßelter, S., et al.: How to Do EcoDesign? A Guide for Environmentally and Economically Sound Design. Verlag form, Frankfurt am Main (2000)
34. Dick, M., Naumann, S.: Enhancing software engineering processes towards sustainable software product design. In: Greve, K., Cremers, A.B. (eds.) EnviroInfo 2010: Integration of Environmental Information in Europe: Proceedings of the 24th International Conference on Informatics for Environmental Protection, October 6–8, 2010, Cologne/Bonn, Germany. Shaker, Aachen, pp. 706–715 (2010)
35. Penzenstadler, B., Femmer, H., Richardson, D.: Who is the advocate? stakeholders for sustainability. In: Green and Sustainable Software (GREENS): 2nd International Workshop on Green and Sustainable Software, pp. 70–77 (2013)
36. Herzog, C., Lefévre, L., Pierson, J.-M.: Actors for innovation in green IT. In: Hilty, L.M., Aebischer, B. (eds.) ICT Innovations for Sustainability: Advances in Intelligent Systems and Computing, pp. 49–67. Springer, Switzerland (2015)
37. Jagroep, E., Broekman, J., van der Werf, J.M.E. et al.: Awakening awareness on energy consumption in software engineering. In: Proceedings of the 39th International Conference on Software Engineering: Software Engineering in Society Track, pp. 76–85 (2017)
38. Chemeris, A., Lazorenko, D., Sushko, S.: Influence of software optimization on energy consumption of embedded systems. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Components, Networks and Systems Implementation. Springer, pp. 111–134 (2017)
39. Groher, I., Weinreich, R.: An interview study on sustainability concerns in software development projects. In: 2017 43rd Euromicro Conference on Software Engineering and Advanced Applications (SEAA), pp. 350–358 (2017)
40. Manotas, I., Bird, C., Zhang, R., et al.: An empirical study of practitioners' perspectives on green software engineering. In: Proceedings of the 38th International Conference on Software Engineering Companion, pp. 237–248 (2016)
41. Kern, E.: Green Computing, Green Software, and Its Characteristics: Awareness, Rating, Challenges. In: Otjacques, B., Hitzelberger, P., Naumann, S., et al. (eds.) From Science to Society. Springer, pp. 263–273 (2018)
42. Selyamani, S., Ahmad, N.: Green computing: the overview of awareness, practices and responsibility among students in higher education institutes. J. Inf. Syst. Res. Innov. (2015)
43. Torre, D., Procaccianti, G., Fucci, D., et al.: On the presence of green and sustainable software engineering in higher education curricula. In: Proceedings of the 1st International Workshop on Software Engineering Curricula for Millennials, pp. 54–60 (2017)
44. Mahmoud, S.S., Ahmad, I.: A green model for sustainable software engineering 2013. Int. J. Soft. Eng. Appl. **7**(4), 55–74 (2013)
45. Shenoy, S.S., Eeratta, R.: Green software development model: an approach towards sustainable software development. In: India Conference (INDICON), 2011 Annual IEEE, pp. 1–6 (2011)
46. Mahaux, M., Canon, C.: Integrating the complexity of sustainability in requirements engineering engineering. In: Svensson, R.B., Berry, D., Daneva, M., et al. (eds.) 18th International Working Conference on Requirements Engineering: Foundation for Software Quality: Proceedings of the Workshops RE4SuSy, REEW, CreaRE, RePriCo, IWSPM and the Conference Related Empirical Study, Empirical Fair and Doctoral Symposium, pp. 28–32 (2012)

47. Becker, C., Betz, S., Chitchyan, R., et al.: Requirements: the key to sustainability. IEEE Softw. **33**(1), 56–65 (2016)
48. Kern, E., Dick, M., Naumann, S., et al.: Impacts of software and its engineering on the carbon footprint of ICT. Environ. Impact Assess. Rev. **52**, 53–61 (2014)
49. Dirlewanger, W.: Measurement and Rating of Computer Systems Performance and of Software Efficiency: An Introduction to the ISO/IEC 14756 Method and a Guide to Its Application. Kassel University Press, Kassel (2006)
50. Drangmeister, J., Kern, E., Dick, M., et al.: Greening Software with Continuous Energy Efficiency Measurement. In: Workshop Umweltinformatik zwischen Nachhaltigkeit und Wandel, Koblenz 2013, pp. 940–951 (2013)
51. Kern, E., Dick, M., Drangmeister, J., et al.: Integrating aspects of carbon footprints and continuous energy efficiency measurements into green and sustainable software engineering. In: Page, B., Fleischer, A., Göbel, J., et al. (eds.) EnviroInfo 2013—Environmental Informatics and Renewable Energies: 27th International Conference on Informatics for Environmental Protection. Proceedings of the 27th EnviroInfo 2013 Conference, Hamburg, Germany, September 2–4, 2013. Shaker Verlag, Aachen, pp. 300–308 (2013)
52. Maevsky, D.A., Maevskaya, E.J., Stetsuyk, E.D., et al.: Malicious software effect on the mobile devices power consumption. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J., (eds.) Green IT Engineering: Components, Networks and Systems Implementation. Springer, pp. 155–172 (2017)
53. Governing Concil of the United Nations Environment Programme.: Background Paper for the Ministerial Level Consultations: Promoting Sustainable Consumption and Production Patterns, Nairobi (2003)
54. Lago, P., Jansen, T.: Creating environmental awareness in service oriented software engineering. In: International Conference on Service-Oriented Computing, pp. 181–186 (2010)
55. Mazijn, B., Doom, R., Peeters, H., et al.: Ecological, Social and Economic Aspects of Integrated Product Policy: Integrated Product Assessment and the Development of the Label 'Sustainable Development' for Products. CP/20. SPSD II—Part I—Sustainable Production and Consumption Patterns (2004)

Information Technology for Evaluating the Computer Energy Consumption at the Stage of Software Development

E. D. Stetsuyk, D. A. Maevsky, E. J. Maevskaya and R. O. Shaporin

Abstract This chapter describes an information technology and practical implementation of the absolute value estimation of the computer devices energy consumption. The evaluation based on the program source code and can help to choose the most optimal solution from the viewpoint of energy saving at the Software development stage. When estimating, the energy that is consumed by the RAM and CPU is determined. These two components are a necessary equipment for the all computer devices, especially for the various devices of the Internet of Things. For estimation of the computer devices energy consumption is using source assembler code of the program and the types of CPU and RAM. There are three stages for energy consumption estimation in proposed information technology. First stage is the scanning of assembler code and determination types of each assembler instruction and their volume in bits. On the second stage, the method for computer's estimating power consumption based on RAM types is applying. The third stage is estimating of full energy consumption. For the implementation of information technology, the program tool is developed. This tool helps to create a green software checking the computer devices energy consumption on all stages of development.

Keywords Green software · Energy consumption · Malicious software
Malicious detection · Smartphone power consumption · Spy program detection

E. D. Stetsuyk · D. A. Maevsky (✉) · E. J. Maevskaya · R. O. Shaporin
Odessa National Polytechnic University, 1 Boulevard of Shevchenko, Odessa, Ukraine
e-mail: Dmitry.A.Maevsky@opu.ua

E. D. Stetsuyk
e-mail: HelMaevskaya@gmail.com

E. J. Maevskaya
e-mail: E.J.Maevskaya@opu.ua

R. O. Shaporin
e-mail: shaporin@opu.ua

© Springer Nature Switzerland AG 2019

V. Kharchenko et al. (eds.), *Green IT Engineering: Social, Business and Industrial Applications*, Studies in Systems, Decision and Control 171,
https://doi.org/10.1007/978-3-030-00253-4_2

1 Introduction

The first volume of the monograph "Concepts, Models, Complex Systems Architectures Green IT Engineering" was published a little over a year ago, in January 2017 [1]. The sixth section of this monograph, which is named as "Evaluating the RAM Energy Consumption at the Stage of Software Development" [2], is devoted to the presentation of the theoretical and methodological foundations for solving a very interesting and important problem—the task of choosing the optimal program code in terms of power consumption. Of course, the program code itself does not consume any power from the electrical network. Power is consumed by a computer, tablet, smartphone, or any other device on which this code is running. The program code only controls the operation of such components as the computing device like CPU, RAM, hard disk, input/output devices, and so on. Exactly these devices consume power from the electrical network during of their work. Their optimal use leads to a reduction of energy consumption for all computing devices. Therefore, we can say, that the program code, mediated through the hardware part, can affect the power consumption of the entire device. Therefore, we will receive more efficient use of electrical energy, while using a more efficient computational algorithm.

So, what is the "efficiency of the algorithm" in terms of energy consumption? The concept of "computational efficiency" and methods for evaluating the computational efficiency of algorithms are well developed [3, 4]. But, effective algorithm in terms of volume of calculations will not necessarily be effective in terms of power consumption [5]. That's why, minimizing, for example, the number of floating point operations, will not automatically mean minimizing the power consumption. The authors proposed and substantiated the approach for determining the efficiency of algorithms from the point of view of the power consumed by the computing device [2].

The approach can be summarized as follows:

1. The effectiveness of the algorithm should be determined at the stage of software development. This is because the choice of this or that algorithm is usually done at this stage. Changing the algorithm of the operation of any part of the program can be difficult or impossible, when the program is already created.
2. The program itself does not yet exist at the development stage, therefore its launch and direct measurements of the power consumption are impossible. There are only source codes for individual fragments of the future software product and its individual algorithms at this stage. Therefore, energy efficiency assessment should be based on the source code of the algorithm using specially developed software tools.

3. It should take the absolute (in watts) value of power consumption as a criterion for choosing the most optimal algorithm for energy consumption. This criterion is the most informative and most universal.
4. It is necessary to develop mathematical models and calculation methods for calculating the absolute value of the power consumption.

2 State of the Art

One of the main contradictions of our era is the contradiction between two global trends in the development of mankind. The first trend is a clear understanding of the fact that we have come to the point where the increase in the generation of electrical energy will lead to irreversible negative consequences for the whole planet [6]:

- Sea levels are expected to rise between 18 and 59 cm by the end of the century and continued melting at the poles could add between 10 to 20 cm.
- Hurricanes and other storms are likely to become stronger.
- Floods and droughts will become more common. Rainfall in Ethiopia, where droughts are already common, could decline by 10% over the next 50 years.
- Less fresh water will be available.
- Some diseases will spread, such as mosquito-borne malaria (and the 2016 resurgence of the Zika virus). Ecosystems will change: Some species will move farther north or become more successful; others won't be able to move and could become extinct.

The second trend is a significant increase in the number of electronic devices that consume electricity. The number of smartphones and tablets is increasing from year to year. Their capabilities are increasing, which, in turn, requires increasing the consumed electricity. The increase in electricity consumption is also expected due to the widespread introduction into our life of devices from Internet of things. This is an objective factor, and no one in the world has yet set the task of reducing the number of such devices. On the contrary, only an increase in their number is predicted. So, according to the Internet resource Statista [7], the total sold mobile devices and devices of the Internet of things will reach about 40 billion by 2020 that is more than five devices per one inhabitant of the Earth. Forecasted data for some types of devices are given in Table 1. The data in this table are derived from the following sources: Desktop PCs and Laptop—[7] (calculated from shipment forecast), Tablets—[7], Smartphones—[8], IoT devices—[9]. This table represents, that

Table 1 Amount forecast of IT-devices to 2020

No.	Device type	Amount (Billion units)
1	Desktop PC	0.2
2	Laptop	0.32
3	Tablets	1.46
5	Smartphone	6.1
6	IoT devices	50

the world will mainly use mobile devices—smartphones and Tablets PC, as well as devices of the Internet of things by 2020.

The problem of saving electricity has become topical because of the significant growth in the number of devices consuming it. But to save consumed electricity, you need to learn how to evaluate it. And here we are again faced with problems.

2.1 The Problems of Estimating the Consumed Electricity

To assess the energy efficiency of computer devices and the Internet of things that are increasingly used worldwide from year to year, we must be able to assess the level of their consumption of electrical energy. Only based on this level we can talk about the efficiency and expediency of using one or another technology to reduce energy consumption. The fact is that the development and implementation of any new technology by itself requires the use of certain resources, including energy. Therefore, any new energy-saving technology will only be appropriate if the effect of its use exceeds the cost of its implementation.

Estimation of electricity consumption level by itself is a complicated and ambiguous task. Especially ambiguity becomes noticeable when we try to estimate the energy consumption of computer devices and the Internet of things. Let's understand more about why this ambiguity arises.

The power of all consumers around the world is eventually due to the network of variable sinusoidal voltage. Such voltage is generated by generators of all power plants. The voltage that our smartphone consumes is initially produced by rectifying and stabilizing the variable of sinusoidal voltage. When we talk about tension—everything is simple—it is variable and sinusoidal. Unfortunately, everything goes harder with the current. Its format is far from always sinusoidal. There are many reasons for it. Pulsed stabilizers and voltage converters are often nonlinear. Due to this nonlinearity, the shape of the current is distorted and deviates from the sinusoidal one. Devices of computer technology by their nature operate in a pulse signal mode when the current is not sinusoidal. The transition mode is a normal operating mode for computer devices. In this case, it is inappropriate to talk about sinusoidal current. Therefore, we have a situation where the form of voltage and current at the input of power sources is not sinusoidal and differ from each other.

As we know from theoretical electrical engineering, in this case only the so-called instantaneous power is uniquely determined. This power is defined as the product of the voltage value at some point in time at the current value at the same moment. But this instantaneous power changes over time and cannot be an indicator of electricity consumption.

Of course, when we talk about the power consumed by this or that device, we mean the so-called active power that can perform a useful job. Most measuring devices show exactly the active power, while doing so with given accuracy only for sinusoidal currents and voltages. But in addition to the active power, there is reactive power, distortion power and total power in electric circuits. This is a big problem in the definition and interpretation of all these capacities.

All devices of the computer technology and the Internet of things can be divided into two groups.

Devices in the first group must always be connected to the electrical network and cannot function without such connection. These devices include desktops and their monitors. Let's call these devices non-autonomous devices.

The devices in the second group do not require a permanent connection to the network. Moreover, their normal operating mode is the mode with the shutdown. Typical representatives of this group are smartphones, which most of the time operate independently of the network using a built-in battery. These devices will be called standalone.

We are interested in the electricity that is consumed from the power grid. To determine the energy consumed by non-autonomous devices, it is sufficient to use known measuring devices, for example, wattmeter. Averaging their performance over quite large periods of time, such as an hour or day, we can conclude on which of the devices of the computer technology or the Internet of things is more energy-efficient.

Unfortunately, we cannot use this method for stand-alone devices. Because the devices are connected to the electrical network at short intervals when the batteries are charging. And all the rest of the time they are in our pockets disconnected from the network. Do they consume electricity currently? Yes, no doubt. How to measure this energy? There is a problem. Even as an experiment, we cannot carry a watt meter in our pocket. We can judge the level of power consumption by an indirect indicator—the rate of discharge of the battery. This indicator is very unreliable. First, it depends on what our device is currently doing. Secondly, it depends on the state of its battery. It is known that over time all the batteries lose their ability to store electrical energy. This leads to the fact that even when we know how fast the batteries of two nearby smartphones are discharged, we cannot judge exactly which one is more energy-efficient.

There is an indirect method to determine the level of electrical power consumed by the network when charging batteries for mobile devices such as laptops, tablets, and smartphones. Here is an example of calculation for charging the smartphone's battery. As indicated in [10], the most widespread battery capacity of the smartphone today is around 3000 mAh (or 3 Ah). The smartphone is charged when the battery holds about 20% charge. That is, when charging from the electrical network consumes 0.8 times the full capacity of the battery. If you accept the voltage of the battery at 5 V, then the power consumed during charging is $P = 0.8 \times 3 \times 5 = 12$ W. Approximately the same value is suggested by the authors of the paper [11] (10.6 W). The same calculation made for a tablet (average battery capacity of 5000 mAh) gives a result of 20 W. For a laptop, we have an average of 30 W.

The question of reducing energy consumption by non-autonomous computer devices, such as a desktop PC, is being studied quite a long. In early works just studied computer consumption level in general, without allotment its individual elements. The paper [12], published in 1994, contains an interesting recommendation to reduce energy costs at operation personal computer: «Switch off when not in use. Save energy!». In this work, there is no mention of the fact that not only the hardware of the computer but also its software can be responsible for the consumption of energy.

Subsequently, the object of research on energy consumption became individual computer nodes. Thus, in particular [13] shows the power consumed by individual nodes depends on the computer model. Data on the energy consumption of individual nodes, according to [13], are summarized in Table 2. In this table, the power consumed is shown as a percentage of the total power of the entire computer. Data is provided on a computer with a Xeon 5310 processor in two RAM modifications: 4 DIMMs at maximum processor load, and 8 DIMMs with a total capacity of 32 GB.

As you can see from this table, RAM ranks second in terms of power consumption among all computer nodes. Total energy consumption and computer was about 400 W. Moreover, the authors [13] note that the addition of each new RAM chipset increases the power consumption by 20 W. These data once again indicate that the minimization of energy consumed by RAM and CPU is a topical issue. The importance of solving this problem increases with each passing year. After all, every year the number of devices of the Internet increases, in which the presence of both components is mandatory [14].

2.2 *Ways to Reduce RAM Power Consumption*

RAM is one of the resources used by the computer. For her applicable all laws that apply to the generalized concept of "resource". These laws are detail described in [15]. According to the resource approach described in this paper, the power consumed by the RAM depends on the intensity of its use. This applies to any electrical device. And the software determines the intensity of using RAM. Therefore, we can say that the efficiency of the RAM energy using depends on the efficiency (degree of "green") of the software. For today, the science of green software goes through the

Table 2 Power consumed by individual nodes of the computer (in percent)

Devices	Xeon 5310, 4 DIMM	Xeon 5310, 8 DIMM
CPU	41	34
RAM	20	33
HDD	3	2
Power supply system	21	18
Cooler	7	6
Ethernet adapter	2	2
Other devices	6	5

stage interpretation of the accumulation of results in its development. We understand that the software should be "green". But we do not have a final understanding of how to measure the degree of "greenness". Thus, in [16], it is emphasized that now there are only initial (primary) attempts to create green software metrics: «There are various primary studies which uses different software measurement metrics in green and sustainable software development». At the same time, in this article it is said that «Poorly designed and developed software's makes it difficult to maintain and increases the power consumption of hardware's». This thesis once again emphasizes the importance and relevance of optimizing the program code, not only in terms of its speed, but also in terms of providing them with minimal energy consumption. A special necessity to minimize energy consumption arises in creating devices of Internet of things and FPGA-based devices [17, 18]. Interesting application of the principles of green software is given in the article [19]. It shows that by the level of energy consumption we can conclude that the program code contains malicious software. At the same time, there is an understanding of the necessity of using the green software. Thus, in [20, 21], the authors show the efficacy of green software in industrial systems and Safety-Critical I&C Systems.

The main disadvantage of all existing approaches to creating green software is the absence of a single metric for evaluation and comparison. It is about this in [16], referred above. However, in our opinion, there is no need to look for any new exotic metric. Such metric has been around for several hundred years. As we have already said, software cannot consume electricity by itself. Electric power is consumed by the hardware of the computing devices. And the software only manages their work. Therefore, the absolute, measured in units of power, the level of electricity consumption, can serve as a universal metric for estimating and comparing the levels of software greenness. It is clear that the program can be executed by different types of computer systems. That's why the same for functional purpose but different by type devices will consume different power. There is currently no evidence that the same software will provide effective energy consumption for different types of devices. Therefore, we cannot consider program code separately without its connection to the hardware of the computer. That's why all mathematical models that are created to evaluate the energy efficiency of a software code should consider the environment in which this code is executed.

Based on the analysis, we can be done the following conclusions:

- an absolute value of the power, which is consumed by the computing device when the related code is executed, is the only reliable indicator on basis of which it is possible to compare the energy efficiency of software code.
- an indicator of energy efficiency depends not only on the code, but also on the hardware of the computing device.
- an information technology for performance of the energy efficiency of program code, should be focused on software developers who are not specialists in the sphere of electrical engineering. Such information technology is not available at present time.

Proceeding from this, the goal of this work is the development and practical implementation of information technology, which is oriented to the end user and allows to perform the estimation and comparison of electric energy that is consumed by the computing device during of the program code.

To achieve this goal, following tasks have been solved in this work:

- Information technology has been developed to evaluate and compare the electrical energy that consumed by the computing device during performing of the program code;
- The developed information technology is implemented in the form of a software tool that allows to perform such an evaluation and comparison to the end user—the software developer.

3 Information Technology for Evaluating the Computer Energy Consumption

At the core of any information technology are the relevant mathematical models and methods of using them. The methodological basis of the described technology are models and methods, the justification of which is described in detail in [2]. To provide a commonality of the statement, we briefly describe these models and methods.

3.1 Methodological Foundations of the Information Technology

For energy consumption estimation the following mathematical models are used:

- a mathematical model for determination of the absolute level of power, which is necessary to transfer 1 Bit of RAM from “0” into “1”;
- a mathematical model determining the expected value of the number of bits, which transfer from “0” into “1” in recording the random data of the known length.

This model is based on the calculation transient in equivalent circuit of the basic RAM cell (1T1C RAM cell).

Mathematical model for determining of the complex power can be represented by following formula:

$$S = V_{DD} \cdot \sqrt{\frac{1}{T}\left[\frac{A_1^2}{2\alpha_1}\left(e^{2\alpha_1 T} - 1\right) + \frac{2A_1A_2}{\alpha_1 + \alpha_2}\left(e^{\alpha_{12} T} - 1\right) + \frac{A_2^2}{2\alpha_2}\left(e^{2\alpha_2 T} - 1\right)\right]}. \quad (1)$$

where:

$$\alpha_1 = \frac{-R_1C_2 - R_2C_2 - R_1C_1 + \sqrt{(R_1C_2 + R_2C_2 + R_1C_1)^2 - 4R_1R_2C_1C_2}}{2R_1R_2C_1C_2},$$
$$\alpha_2 = \frac{-R_1C_2 - R_2C_2 - R_1C_1 - \sqrt{(R_1C_2 + R_2C_2 + R_1C_1)^2 - 4R_1R_2C_1C_2}}{2R_1R_2C_1C_2},$$

and $\alpha_{12} = \alpha_1 + \alpha_2$.

The following notations are used in the formulas: V_{IH}—logical “1” voltage; V_{REF}—voltage in the information bus; V_{IL}—logical “0” voltage; R_1–information bus resistance; R_2—Source-Drain resistance of field-effect transistor when it turns on; C_1—information bus capacitance; C_2—capacitance of the capacitor in basic RAM cell.

Electrical parameters of RAM equivalent circuit can be found in the special technical guides, for example in the standard [22]. For example, for DDR4 type RAM, in accordance with [22] it possesses the following electrical parameters: R_1 = 1 Ω, R_2 = 2 Ω, C_1 = 300 fF, C_2 = 30 fF, V_{REF} = 1.225 V, V_{IL} = 1 V, V_{IH} = 1.45 V, VDD = 1.5 V. If we know electrical RAM parameters, formula (1) allows to indicate RAM consumed complex power for the move from “0” to “1” of just 1 RAM bit. For DD4 type RAM the calculated power value is 1.89×10^{-4} VA. At first glance the power value seems to be very small. But we should remember that 1.89×10^{-4} VA is consumed as we change the 1 RAM bit state only once. Certainly, the move from “0” to “1” for two, three and more bits requires appropriately more power consumption than 1 bit.

The number of moves from the “0” state to the “1” state is only determined by the data recorded in RAM. The problem of expected value calculation of the amount of RAM bits transiting to “1” in random change of RAM area state is a rather complicated combinatorial task. For practical use the mathematical model of dependence expected value M(k) on the RAM state amount changes on the RAM part length (k) is represented with formula (3).

$$M(k) = -0.00011791 \cdot k^3 + 0.00444567 \cdot k^2 + 0.19515112 \cdot k + 0.22247721 \quad (2)$$

On the base of these models, the method of evaluation of energy consumption in executing the programs based on their source code was created. This method consists of the following steps.

- To get Assembler code of the researched program.

- To carry out the analysis of each of the instructions of Assembler code and determine the number of clock signals, which are necessary for the instruction execution and RAM area length, which is changed by this instruction. The number of clock signals can be learnt in CPU specifications, but the RAM area length corresponds to instruction operands length.
- To calculate the total number of CPU clock signals necessary for a single program execution based on analysis described in previous paragraph. We use *Nt* to denote the calculated number of clock signals.
- To calculate the expected value of the bit amount changed by this instruction based on RAM area length changed by each of the instructions and with the help formula (2).
- To calculate the complex power spent in order to change the RAM state by the instruction according to the formula.

$$S_k = M(k) \cdot S_1,$$

where: k—RAM area length changed by the instruction; $M(k)$—expected value of the number of changed from “0” to “1” bits depending on the RAM area length calculated according to the formula (3); S_1—the complex power used in order to change 1 RAM bit state according to formula (1).
- To calculate the complex power used to change all RAM area states during a single program execution as a sum of complex powers S_k for every separate instruction. We use S_Σ to denote this complex power.
- To calculate a specific power W by the computer for a one second of program execution according to the formula:

$$W = N_t \cdot f_t \cdot S_\Sigma. \tag{3}$$

where: W—consumed energy (*VA* per second), N_t—the number of CPU clock signals necessary for a single program execution, f_t—frequency of CPU clock signal in Hertz;
Value W is a result of the energy consumption evaluation method in executing the programs based on their source code.

3.2 *Description of the Information Technology*

The developed information technology helps to quantify (in units of power measurement) the level of power consumption of RAM and CPU that run under the control of the user program.

For its use, the technology requires the following input:

- program code in Assembler or high-level algorithmic language;
- type of RAM or its electrical parameters;
- CPU type or its clock speed.

The proposed information technology includes the sequential implementation of the following steps.

Step 1 Convert program code to Assembler code. The step is executed only if the program is written in a high-level language. Execution is possible using the compiler of this language. All compilers of the most common programming languages that create the exe module of program (for example C and C ++) have options for constructing an assembly code [23].

Step 2 Preparation of input: assembler program code, type of RAM and CPU.

Step 3 Construction of a mathematical model of RAM power consumption of a given type. To construct the model, Eqs. (1)–(3) are used.

Step 4 Interpretation of Assembler code. In this step, the mathematical expectation of the number of RAM transitions from the state of the logical "0" to the state of the logical "1" according to the formula (8) is performed.

Step 5 Finding the absolute value of the energy consumed by RAM when executing each Assembler code command per unit time (1 s) of continuous execution. The sum of these values gives an estimate of the amount of energy consumed by RAM when executing the entire code.

The sequence of processes of the proposed information technology is conveniently represented using the IDEF0-diagram. The IDEF0 notation consists of blocks, each of which is a "black box" with inputs and outputs, controls and mechanisms that are detailed (decomposed) to the required level. Based on IDEF0-diagrams it is possible to quickly define and describe the key processes of any technology, to reveal their correct sequence. IDEF0 is a very simple and at the same time intuitive methodology of functional modeling.

With this methodology, information can be transferred between developers, consultants and users. The methodology was very carefully developed, it is universal and convenient for designing. Below is the IDEF0 diagram (Fig. 1) of the information technology for determining the energy consumed by the program code.

Logically, the information technology for determining the energy consumed in performing of the program code can be represented in the form of four blocks. Each unit describes a separate information technology. Let's consider these processes in more detail.

The first process "Defining RAM parameters" is performed by user. Based on the user-defined input data—the type of RAM in this process, its electrical parameters V_{IH}, V_{REF}, V_{IL}, R_1, R_2, C_1 and C_2 are determined. The meaning of these parameters was explained earlier in Sect. 2.1. As control mechanisms, with the help of which the process will be successful, standards or technical manuals on RAM are advocated. The listed RAM parameters are the output data of the first process.

The second process "Defining the CPU parameters" is also performed by user. In this process, the type of processor is the input data. With the help of the control mechanism, which is the documentation for the CPU, the user, based on the type of

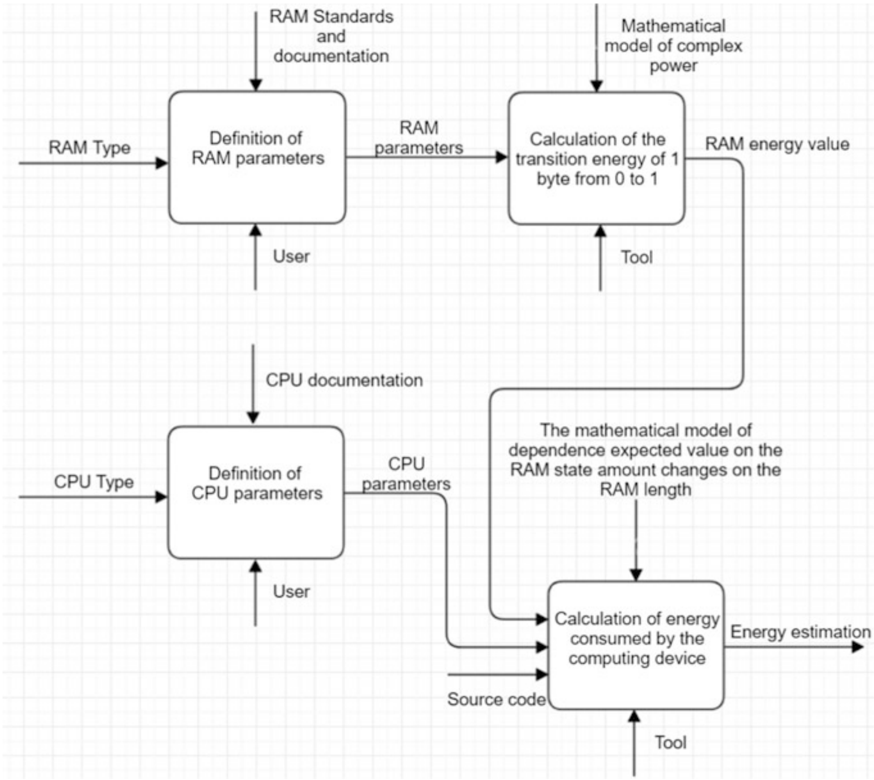


Fig. 1 IDEF0 diagram of the information technology for evaluating the computer energy consumption

processor, determines its parameters—frequency of CPU clock signal and the number of CPU clock signals necessary for an assembler instruction execution. These parameters are the output data of the second process.

The third process is responsible for calculating the transition energy of the 1-bit RAM from the logical “0” state to the logical “1” state. This energy is calculated by the formula (1) developed by the software. The initial parameters are the RAM parameters defined during the first process. The controlling mechanism in this process is the mathematical model of complex power. The output of the process is the energy value that expends the 1-bit transition RAM from the logical “0” state to the logical “1” state.

The fourth, final, process “Calculation of energy consumed by the computing device” is also performed using the developed software. In this process, the results of previous processes are used as input parameters, namely the parameters of the central processor and the energy value that spends the 1-bit transition RAM from the logical “0” state to the logical “1” state. In addition, the program code is used to calculate the energy consumed. This code must be previously submitted to the

Assembler commands. The control mechanism of this process is the mathematical model of the dependence of the expected value on the RAM length. The output of the process is the calculated energy consumed by the RAM of the computing device when the specified program code is executed.

Calculations of power consumption are labor intensive. Particularly laborious is the process of calculating the power consumed by RAM when executing each assembler command of the source code. To simplify these calculations, a program tool for evaluating the computer energy consumption has been developed.

4 Program Tool for Evaluating the Computer Energy Consumption

The object-oriented Java programming language and the Eclipse development environment were selected to write this software product. Java is the foundation almost for any type of network application and is the world standard for WEB-programming and the creation of corporate software. Java programs are transmitted to the bytecode executed by the Java virtual machine (JVM), a program that handles the byte code of each instruction as an interpreter [24].

The advantage of this type of program execution is the full independence of the bytecode from the operating system and hardware. It's make possible to perform Java applications on any device for which exist appropriate virtual machine. Another important feature of Java technology is the flexible security system due to that the execution of the program is fully controlled by the virtual machine. Any operations that exceed the program's authority (for example, attempting unauthorized access to data or connecting to another computer) cause an immediate interruption of the computing process.

The program tool "ESTimation Energy Tool" (ESTET) includes an interface part that allows the user to select the required file with the code of the program that is being explored. Ibid you can choose the type of processor and RAM of the computer on which the program is launched and see the result of the calculations. Except interface part the program includes a logical part in which all mathematical calculations are performed to determine energy consumption.

4.1 ESTET Classes Description

The main class of the created program tool is the ***MainWindow.java*** class. This class organizes the interconnection of all components of the program. Let's consider main methods of this class.

Method *MainFrame()*—the initialization of the program takes place in this class. Also, the UserDir variable sets the path to the default working folder. This folder contains data and results directories.

Method *showWindow()*—thanks to this method it is possible to open an additional window and independently select the necessary file of initial data.

CompParamWindow.java is a class that is used to work with lists of CPUs and types of RAM.

Method *showWindow()*—this method is implemented to view and select the required characteristics of the central processor and the RAM of the computer on which the program is running.

Method *countLinesProc()*—a file with all the necessary characteristics of all known processors is currently being uploaded as a list.

Method *countLinesMem()*—a file with all the necessary characteristics of the RAM is unloaded as a list.

NewMemoryWindow.java is the class that is required to add new RAM to the shared list.

Method *showWindow()*—this method is implemented to create a new memory record that the user cannot find in the list. This is possible, since at the time of writing this program it is impossible to find out what kinds of RAM will come in the future. For this, the user will be given fields to fill the required characteristics of the new RAM.

Method *addNewMemInFile()*—this method implements a new type of memory record in a file that contains a list of all existing types of RAM.

NewProcWindow.java is the class that is required to add new CPU to the shared list.

Method *showWindow()*—this method is implemented to create a new central processor record that the user cannot find in the list. This is possible, since at the time of writing this program it is impossible to find out what kinds of CPUs will come in the future. For this, the user will be given fields to fill the required characteristics of a new type of CPU.

Method *addNewProcInFile()*—this method implements a record of a new type of central processing unit in a file that contains a list of all existing types of the CPU.

AsmFileWindow.java is a method implemented to work with Assembler code.

Method *showWindow()*—this method is implemented for viewing and selecting a file in the *.asm extension to further analyze the input file in the other stages of the program algorithm and break the commands of each line into its components.

Method *openFile()*—using this method, the user can open a window for viewing all the folders and files on his computer and select the desired file in the *.asm extension.

Logic.java is a class that is necessary for creating program logic, namely for creating algorithms and programming mathematical formulas.

Method *deleteSpace()*—in this method, an algorithm that analyzes the assembly line source and removes further unnecessary labels, spaces, and more.

Method *StringWork()*—this method analyzes every input file in its assembler file, determines which type of operation is described in the line, and which types of operands it contains. Also, method writes all the information about a string into a new array, which is further required for mathematical calculations for power consumption.

Method *ComplexPower()*—this method, based on the input data, implements the calculation of the energy consumption of each operation of each input line, then converts the resulting results into a new array.

ResultTableWindow.java—this class is needed to work with the results of calculations.

Method *showWindow()*—after parsing each line of the source file that the user selected, the results are recorded in an array and displayed as a table in a new window. This window is the final point in implementing the algorithm of the work of this software product.

4.2 Working with ESTET Program Tool

You need a version of Windows with no lower than Windows XP to use this program on your computer. The operating memory should be at least 4 GB, free hard disk space of at least 2 GB. The procedure for installing/uninstalling the program is done using the standard Windows installation/uninstall tool. When installed, the program is stored in a separate directory assigned by the user. At the same time, at the user's request, a shortcut is created on the desktop. No changes to the Windows system registry are made.

When the program is launched, the main window of the program appears on the screen. In the main window is located the logo of the program and the choice of language. In total, three languages are available in the interface: English, Ukrainian, or Russian.

Next window of the ESTET Program Tool is the window of the previous settings and selection of very important data for the program, namely the type of RAM and the CPU. A window with two drop down list of types of RAM and CPU respectively appears before the user.

The program already describes the most common types of RAM and CPUs that are equipped with modern computers. This window is shown in Fig. 2.

But if the desired type of RAM or CPU is not found, it is possible to add them to the list. You need to click on the "ADD" button. It will also need to specify the electrical parameters of RAM and data on the CPU speed. Figure 3 shows the window for RAM parameters description.

The window, in which you can specify the CPU data, is shown in Fig. 4. The processor speed is set by its clock speed.

The next step is to specify a text file that contains the assembler code of the program. To select a code file, you need to click the "Select" button. This will open

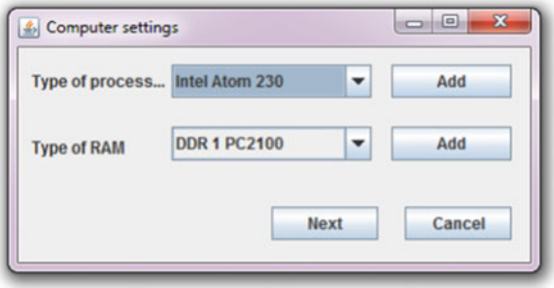


Fig. 2 Window of choice of RAM and CPU

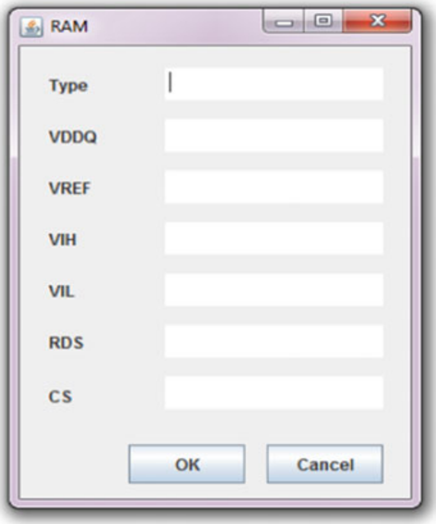


Fig. 3 The window for RAM parameters description

the default file selection dialog. After selecting the file, the program code can be viewed in the window shown in Fig. 5.

To estimate the level of electrical energy that will be consumed by the RAM and CPU when executing this program, you need to click on the "Next" button. The models and methods indicated in Sect. 3 will be used for evaluation. The result of the evaluation can be seen in the window shown in Fig. 6.

An Assembler source code shown in this window separated into instructions and represented in the form of a table. The table shows the number of clock cycles that are consumed to execute the instruction, and the number of RAM bits changed by it. The total power estimation consumed by RAM for performing the instruction is represented in the last column of the table.

The result of program operation is the amount of energy consumed by RAM (WA per second) is represented under the table.

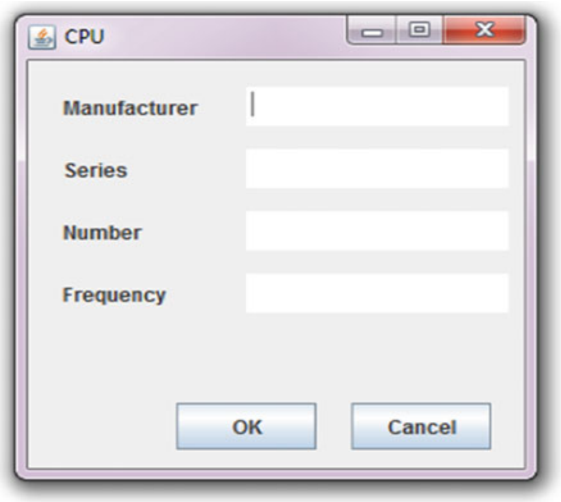


Fig. 4 Window for description of the new CPU type and frequency

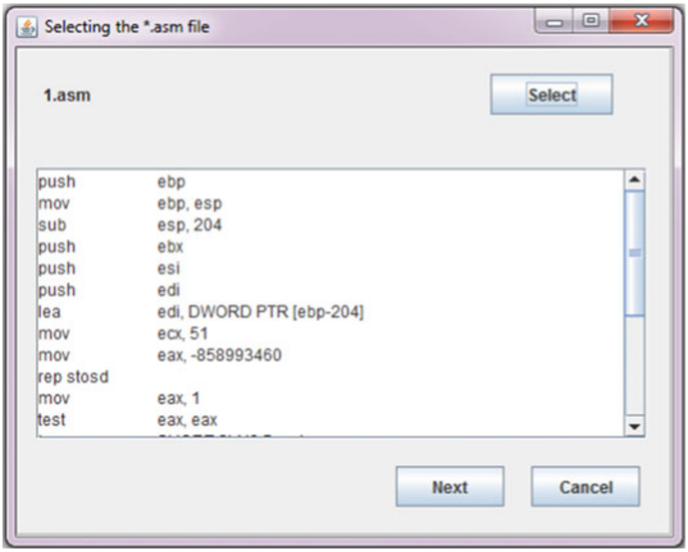
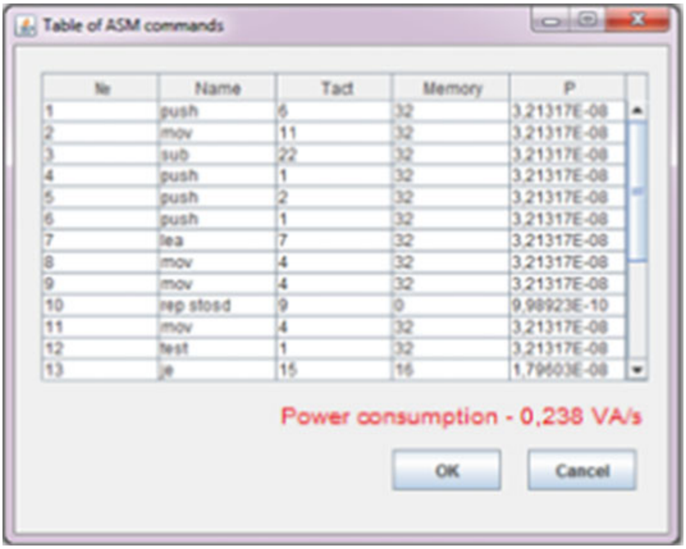


Fig. 5 Window with the source code of the testing program

Thus, with the help of ESTET the software developer's possessing only the program source code can estimate the energy costs on its implementation. It gives the possibility to control and minimize the energy consumption level at the stage of software development. Having several variants of one and the same algorithm the developers can choose the variant, which is the most optimal from the viewpoint of energy consumption.



№	Name	Tact	Memory	P
1	push	6	32	3,21317E-08
2	mov	11	32	3,21317E-08
3	sub	22	32	3,21317E-08
4	push	1	32	3,21317E-08
5	push	2	32	3,21317E-08
6	push	1	32	3,21317E-08
7	lea	7	32	3,21317E-08
8	mov	4	32	3,21317E-08
9	mov	4	32	3,21317E-08
10	rep stosd	9	0	9,98923E-10
11	mov	4	32	3,21317E-08
12	test	1	32	3,21317E-08
13	je	15	16	1,79603E-08

Fig. 6 Presentation of ESTET operation results

5 Conclusion and Future Works

This chapter describes in detail the mathematical models, methods and tools for evaluating the absolute level of power consumption of RAM when executing the program code. Such an assessment can be carried out at any stage of software development. It is known that the basic algorithms laid down in the software are determined at first stages of development [25]. In the future, these algorithms will not change.

But in the previous stages of development, there is still no ready-made software. There are only some of its fragments, often linked to each other. It is not possible to implement this program code to assess its energy efficiency. Therefore, the models, methods, tools and information technology offered in the work are useful. This technology opens the possibility of comparing among themselves the criterion of the level of power consumption of different algorithms. At the same time, these algorithms can be written even in different programming languages.

The proposed method has only one limitation—we need to know an Assembler code of the program. For programs built by compilers and having an exe code, Assembler code can always be obtained. But, if the software is executed in the interpreter mode, then there is no principle for such a code. In this case, the high-level language is transformed into a bytecode executed by the program interpreter. But this interpreter also uses memory resources, which leads to electricity consumption. Separating the energy consumed by the application itself is not possible.

The direction of further research should be the development of similar tools for assessing the level of consumption by the CPU when executing its program code. In this case, it is possible to combine such a system with the description written in this section. Because the input data for both systems is the program code, and the source code - the absolute level of power consumption. The ability to determine the power consumption of RAM along with the CPU by simply adding is the main advantage of the proposed methodology.

References

1. Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.): Green IT Engineering Concepts, Models, Complex Systems Architectures. Studies in Systems, Decision and Control, vol. 74. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-44162-7
2. Maevsky, D.A., Maevskaya, E.J., Stetsuyk, E.D.: Evaluating the RAM energy consumption at the stage of software development. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering Concepts, Models, Complex Systems Architectures. Studies in Systems, Decision and Control, vol. 74, pp. 101–121. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-44162-7_6
3. Miranda, F., Carlos, A.: Handbook of Research on Computational Simulation and Modeling in Engineering. IGI Global, pp. 1–824 (2016). https://doi.org/10.4018/978-1-4666-8823-0
4. Komodakis, N., Tziritas, G., Paragios, N.: Performance vs computational efficiency for optimizing single and dynamic MRFs: setting the state of the art with primal-dual strategies. Comput. Vis. Image Underst. **112**(1), 14–29 (2008). https://doi.org/10.1016/j.cviu.2008.06.007
5. Yoshizumi, T., Goto, T., Isobe, Y., Ohmaki, K., Mori, H., Tsuchida, K.: Parallel algorithm that considers energy efficiency and time efficiency. In: 2016 IEEE International Conference on Industrial Technology (ICIT), Taipei, pp. 1723–1727 (2016). https://doi.org/10.1109/icit.2016.7475023
6. Walker, C.C., Gardner, A.S.: Rapid drawdown of Antarctica's Wordie ice shelf glaciers in response to ENSO/southern annular mode-driven warming in the Southern Ocean. Earth Plan. Sci. Lett. **476**, 100–110 (2017). https://doi.org/10.1016/j.epsl.2017.08.005
7. Shipment Forecast of Laptops, Desktop PCs and Tablets Worldwide from 2010 to 2021 (2017). https://www.statista.com/statistics/272595/global-shipments-forecast-for-tablets-laptops-and-desktop-pcs/
8. Lunden, I.: 6.1B Smartphone Users Globally By 2020, Overtaking Basic Fixed Phone Subscriptions (2015). https://techcrunch.com/2015/06/02/6-1b-smartphone-users-globally-by-2020-overtaking-basic-fixed-phone-subscriptions/
9. Nordrum, A.: Popular Internet of Things Forecast of 50 Billion Devices by 2020 is Outdated (2016). https://spectrum.ieee.org/tech-talk/telecom/internet/popular-internet-of-things-forecast-of-50-billion-devices-by-2020-is-outdated
10. Hoque, M.A., Siekkinen, M., Koo, J., Tarkoma, S.: Full charge capacity and charging diagnosis of smartphone batteries. IEEE Trans. Mob. Comput. **16**(11), 3042–3055 (2017). https://doi.org/10.1109/TMC.2017.2688321
11. Zaman, N., Almusalli, F.A.: Review: smartphones power consumption & energy saving techniques. In: 2017 International Conference on Innovations in Electrical Engineering and Computational Technologies (ICIEECT), Karachi, pp. 1–7 (2017). https://doi.org/10.1109/icieect.2017.7916593

12. Newsham, G.R., Tiller, D.K.: The energy consumption of desktop computers: measurement and savings potential. IEEE Trans. Ind. Appl. **30**(4), 1065–1072 (1994). https://doi.org/10.1109/28.297924
13. Meyer, N., Januszewski, R.: Research on energy-efficient computing solutions. In: PRACE Workshop "New Languages & Future Technology Prototypes" at Leibniz Supercomputing Centrein, Garching near Munich, pp. 217–224 (2010)
14. Chmaj, G., Selvaraj, H.: Energy-efficient computing solutions for internet of things with ZigBee reconfigurable devices. Int. J. Soft. Innov. **4**(1), 31–47 (2016). https://doi.org/10.4018/IJSI.2016010103
15. Drozd, J., Drozd, A., Antoshchuk, S.: Green IT engineering in the view of resource-based approach. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Concepts, Models, Complex Systems Architectures, Studies in Systems, Decision and Control, vol. 74, pp. 43–65. Springer International Publishing, Berlin, Heidelberg (2017). https://doi.org/10.1007/978-3-319-44162-7_3
16. Debbarma, T., Chandrasekaran, K.: Green measurement metrics towards a sustainable software: a systematic literature review. In: 2016 International Conference on Recent Advances and Innovations in Engineering (ICRAIE), Jaipur, pp. 1–7 (2016). https://doi.org/10.1109/icraie.2016.7939521
17. Drozd, A., Antoshchuk, S.: New on-line testing methods for approximate data processing in the computing circuits. In: 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, pp. 291–294, Prague, Czech Republic (2011). https://doi.org/10.1109/idaacs.2011.6072759
18. Drozd, A., et al.: Green experiments with FPGA. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Components, Networks and Systems Implementation. Studies in Systems, Decision and Control, vol. 105, pp. 219–239. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-55595-9_11
19. Maevsky, D.A., Maevskaya, E.J., Stetsuyk, E.D., Shapa, L.N.: Malicious software effect on the mobile devices power consumption. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Components, Networks and Systems Implementation. Studies in Systems, Decision and Control, vol. 105, pp. 155–171. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-55595-9_8
20. Kondratenko, Y., Korobko, V., Korobko, O., Kondratenko, G., Kozlov, O.: Green-IT approach to design and optimization of thermoacoustic waste heat utilization plant based on soft computing. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Components, Networks and Systems Implementation. Studies in Systems, Decision and Control, vol. 105, pp. 287–311. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-55595-9_14
21. Kuchuk, G., Kovalenko, A., Kharchenko, V., Shamraev, A.: Resource-oriented approaches to implementation of traffic control technologies in safety-critical I&C systems. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Components, Networks and Systems Implementation. Studies in Systems, Decision and Control, vol. 105, pp. 313–337. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-55595-9_15
22. Double Data Rate (DDR) SDRAM JESD79F.: In: JEDEC Standard (2008). http://cs.ecs.baylor.edu/~maurer/CSI5338/JEDEC79R2.pdf
23. Oualline, S.: Practical C++ Programming. O'Reilly Media, Inc., p. 549 (2003)
24. Farrel, J.: JAVA Programming. Cengage Learning, p. 976 (2015)
25. Almendro, D., Silveira, I.F.: Quality and tests for open educational resources: a systematization based on software engineering principles. In: 2017 Twelfth Latin American Conference on Learning Technologies (LACLO), La Plata, Argentina, pp. 1–7 (2017). https://doi.org/10.1109/laclo.2017.8120927

Green IT Engineering: Green Wireless Cooperative Networks

Shu Fu and Jinsong Wu

Abstract With the development of IP (Internet Protocol) based next generation of wireless communication, wireless capacity in the physical layer faces unprecedented challenges. Thanks to the coordinated multi-point (CoMP) networks, the performance of wireless transmission can be largely enhanced. To sufficiently explore the performance of CoMP, two key issues of resource allocation, user scheduling and power allocation, should be appropriately determined. In this chapter, we first introduce the heuristic schemes of resource allocation in CoMP based on the Nash non-cooperative game to improve the wireless capacity. Then, we further propose a novel PFM mechanism (power-fractionizing mechanism) to transform the optimization problem of resource allocation of CoMP into a SP (signomial programming) optimization model. Based on GP (geometric programming) approximation, we obtain the local optimization solution of resource allocation in CoMP, empirically, the local optimization solution is also the global optimization solution. By simulation, we verify the effectiveness of our algorithms. In this chapter, we employ Nash non-cooperative game, geometric programming and convex optimization, etc., to improve the performance of the next generation of wireless cooperative networks. All the algorithms fit for the existing framework of network. Moreover, all the algorithms we proposed in this chapter can be also used for other wireless networks for achieving better interference management and energy efficiency.

Keywords Green communication · Cooperative networks · Wireless transmission Nash non-cooperative game

S. Fu
College of Communication Engineering, Chongqing University, Chongqing 400044, People's Republic of China
e-mail: shufu@cqu.edu.cn

J. Wu (✉)
Department of Electrical Engineering, Universidad de Chile, Santiago 8370451, Chile
e-mail: wujs@ieee.org

© Springer Nature Switzerland AG 2019

V. Kharchenko et al. (eds.), *Green IT Engineering: Social, Business and Industrial Applications*, Studies in Systems, Decision and Control 171,
https://doi.org/10.1007/978-3-030-00253-4_3

1 Introduction

In this chapter, to achieve a green networks [1–6], we introduce the cooperative multi-point (CoMP) [7] in multi-cell systems. 3GPP has proposed several transmission scenario of CoMP. In this chapter, we focus on the downlink CoMP with the JP (joint processing) scenario of JT (joint transmission) mode. JP means that all BSs (base stations) in the same cooperative BS set (CBS) share all the user data for wireless transmission. In this chapter, we use CBS and cluster alternately. JT means all BSs in the same CBS can simultaneously transmitted intended information to users in the scheduled user set with the same time-frequency resources. The system model can be described as Fig. 1. BS *A* works in the single-cell mode, and BS *B* and BS *C* simultaneously serve for the user in the CBS containing BS *B* and *C*. This can transform the wireless interference from BS *C* to the user in Cell *B* into intended information and enhance the receiving signal to interference plus noise ratio (SINR) of the user.1 Green Wireless Cooperative Networks. In this chapter, to achieve a green networks [1–6], we introduce the cooperative multi-point (CoMP) [7] in multi-cell systems. 3GPP has proposed several transmission scenario of CoMP. In this chapter, we focus on the downlink CoMP with the JP (joint processing) scenario of JT (joint transmission) mode. JP means that all BSs (base stations) in the same cooperative BS set (CBS) share all the user data for wireless transmission. In this chapter, we use CBS and cluster alternately. JT means all BSs in the same CBS can simultaneously transmitted intended information to users in the scheduled user set with the same time-frequency resources. The system model can be described as Fig. 1. BS *A* works in the single-cell mode, and BS *B* and BS *C* simultaneously serve

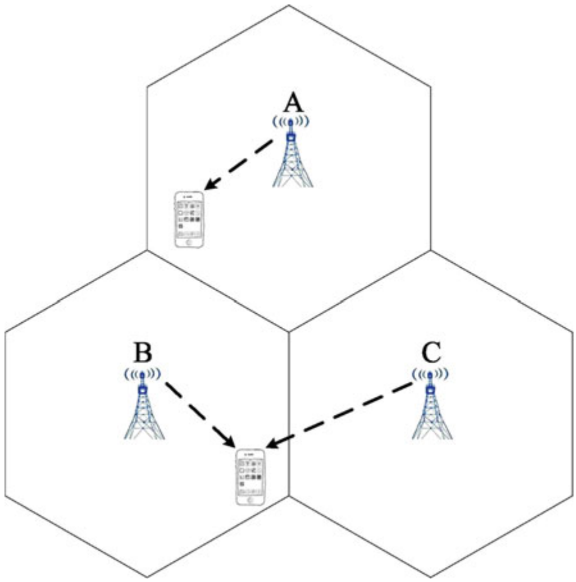


Fig. 1 Downlink CoMP

for the user in the CBS containing BS *B* and *C*. This can transform the wireless interference from BS *C* to the user in Cell *B* into intended information and enhance the receiving signal to interference plus noise ratio (SINR) of the user.

In this chapter, we introduce two categories of system models in CoMP, i.e., single-antenna scenario and multi-antenna scenario. Algorithms in the single-antenna CoMP can be easily followed with low complexity, yet the performance is limited by the available space resources. On the other hand, algorithms in the multi-antenna CoMP possess much better performance than that in the single-antenna CoMP due to the extended available space resources. However, the system complexity appears an important issue effecting the CoMP performance such as delay, etc.

1.1 Single-Antenna Scenario

In this sub-section, we introduce the downlink CoMP in the multi-cell scenario with single-antenna. As discussed above, we focus on the JP of JT working manner between BSs, the system model is shown as in Fig. 1. The cell that the user is located is defined as associated cell, and the cooperative cells are defined as cooperative cells. The set of BSs is denoted as $\boldsymbol{n} = \{1, 2, \ldots, |\boldsymbol{n}|\}$, the maximal transmitting power per BS is denoted as $p_{\max}$. The system frequency bandwidth is assumed as b, and the frequency reuse factor is assumed as 1. The frequency resource is equivalently divided into multiple orthogonal sub-channels, the available number of which is represented as $\boldsymbol{f} = \{1, 2, \ldots, |\boldsymbol{f}|\}$. The wireless transmission and user scheduling on each sub-channel is independent and self-governed. In this chapter, we define the users with the highest receiving reference signal receiving power (RSRP) as central users and that with the lowest RSRP as the edge users. Then, the users per cell can be divided into central users and edge users. Since CoMP aims to improve the performance of edge users, we focus on the edge users and assume that the demands of the central users have been met in other specific timeslots. For a specific edge user, if it is covered by a CBS, we can say that the user belongs to the CBS. BSs in the same CBS simultaneously serve for its edge users with the same time-frequency and space resources. The wireless transmission will lead to wireless interference between CBSs.

The wireless channel quality can be detected by RSRP. If all BSs detect bad wireless channel quality of all the edge users on a specific sub-channel, the sub-channel will be dropped at all the BSs. Based on the detecting RSRP, the channel state information (CSI) of edge users will be fed back to all the BSs in the corresponding CBS.

In the single-antenna scenario, BSs belong to different CBSs across the sub-channels. To further simplify the system model, in this chapter, we ignore the pre-coding technologies, and all BSs in the same CBS serve for only one edge user. On an arbitrary sub-channel i, let $p_{i,j}$ denote the BS transmitting power of BS j on the sub-channel i to the edge user. Let $\boldsymbol{k}_i = \{1, 2, \ldots, |\boldsymbol{k}_i|\}$ denote the set of CBSs on the sub-channel i. The set of BSs in the k-th CBS on the i-th sub-channel is represented as $\boldsymbol{m}_{i,k} = \{1, 2, \ldots, |\boldsymbol{m}_{i,k}|\}, (1 \le k \le |\boldsymbol{k}_i|)$. The observed SINR of the served edge user in the cluster $\boldsymbol{m}_{ik}$ is denoted as γ_{ik}. Let g_{ijk} denote the channel gain of the BS j in the cluster k on the sub-channel i to the edge user. γ_{ik} can be formulated as (1), where σ^2 denote the Gaussian white noise.

$$\gamma_{ik} = \frac{\sum_{j \in \boldsymbol{m}_{ik}} p_{i,j} g_{ijk}}{\sum_{j \notin \boldsymbol{m}_{ik}} p_{i,j} g_{ijk} + \sigma^2}. \tag{1}$$

In this chapter, we consider the wireless transmission with finite-alphabets. Considering (1), the throughput in the cluster k can be written as (2)

$$r_{ik} = \frac{b}{|\boldsymbol{f}|} \log_2\left(1 + \frac{\gamma_{ik}}{\Gamma}\right), \tag{2}$$

where Γ is a constant involved with the aimed bit error rate (BER). Theoretically, γ_{ik} in (2) can be used to calculate the user throughput in the case that both the wireless interference and noise obey the complex Gaussian distribution. In practice, the condition is generally relaxed [8], (2) is then used to approximately calculate the finite-alphabets based user throughput. The approximation error after the relaxation has been studied in the existing references [9]. The system model is formulated as (3)–(4):

$$\text{maximize} \quad \sum_{i \in \boldsymbol{f}} \sum_{1 \le k \le |\boldsymbol{k}_i|} r_{ik}, \tag{3}$$

$$\text{s.t.} \quad \sum_{i \in \boldsymbol{f}} p_{i,j} \le p_{\max}, \quad \forall j \in \boldsymbol{N}. \tag{4}$$

The object (3) is to maximize the user throughput, and (4) is the constraint of the maximal transmitting power of BS.

1.2 Multi-antenna Scenario

In the multi-antenna scenario, BSs per CBS employs block diagonalization (BD) [10] to transmit multiple parallel traffic streams via Multiple-Input Multiple-Output (MIMO). In this sub-section, we virtualize the multi-antenna system as multiple parallel single-antenna system, where each virtualized single-antenna transmitting-receiving pair corresponds to one stream. The wireless interference between the single-antenna transmitting-receiving pairs is nulled due to the BD pre-coding. The specific BD precoding can be achieved by singular value decomposition (SVD) [11]. After the BD pre-coding is determined, the power proportion between BSs in the same CBS can be also determined for serving an arbitrary stream in this CBS.

We assume that each BS and user are configured N_t and N_r antennas, respectively. For simplification, we assume that the number of receiving streams per user is N_r, where $N_t \geq N_r$. For an arbitrary BS $j \in \boldsymbol{n}$, when the allocated power is $p_{\max}$, the power allocation for all streams that the BS j serves for will be terminated. By this, the power proportion between BSs serving for each of the streams can be remained. For an arbitrary BS $m \in \boldsymbol{n}$, let $\boldsymbol{u}_{i,k}$ denote the set of scheduled edge users in the cluster k on the sub-band i. Let $\boldsymbol{s}_{i,k}^{u}$ denote the $n_r \times 1$ vector of transmitting streams in the cluster k on the sub-band I, where $\left|\boldsymbol{s}_{i,k}^{u}\right| = n_r$. Then, we have

$$\boldsymbol{s}_{i,k} = \left[\boldsymbol{s}_{i,k}^{1\ \mathrm{H}}, \boldsymbol{s}_{i,k}^{2\ \mathrm{H}}, \ldots, \boldsymbol{s}_{i,k}^{|\boldsymbol{u}_{i,k}|\,\mathrm{H}}\right]^{\mathrm{H}}. \quad (5)$$

For the cluster k on the sub-channel i, the largest number of scheduled edge users $\left|\boldsymbol{u}_{i,k}\right|_{\text{largest}}$ can be represented as

$$\left|\boldsymbol{u}_{i,k}\right|_{\text{largest}} = \left\lfloor \frac{N_t \times \left|\boldsymbol{m}_{i,k}\right|}{N_r} \right\rfloor. \quad (6)$$

Define $\boldsymbol{H}_{u,m}^{i}$ as the $N_r \times N_t$ channel matrix from the BS m to the user u. On the sub-channel i, BSs in the cluster k to the user u possess the channel matrix as

$$\boldsymbol{H}_{u,k}^{i} = \left[\boldsymbol{H}_{u,\boldsymbol{m}_{i,k}(1)}^{i}, \boldsymbol{H}_{u,\boldsymbol{m}_{i,k}(2)}^{i} \cdots, \boldsymbol{H}_{u,\boldsymbol{m}_{i,k}\left(\left|\boldsymbol{m}_{i,k}\right|\right)}^{i}\right]. \quad (7)$$

Let $\boldsymbol{t}_{i,k}^{\boldsymbol{s}_{i,k}^{u}(s)}$ denote the $\left(\left|\boldsymbol{m}_{i,k}\right| \times N_t\right) \times 1$ transmitting pre-coding vector of the s-th stream in the cluster k on the sub-channel i, where $1 \leq s \leq N_r$. Generally, we assume that $\mathbb{E}\left[\boldsymbol{t}_{i,k}^{\boldsymbol{s}_{i,k}^{u}(s)} \boldsymbol{t}_{i,k}^{\boldsymbol{s}_{i,k}^{u}(s)^{\mathrm{H}}}\right] = \boldsymbol{I}_{\left(\left|\boldsymbol{m}_{i,k}\right| \times N_t\right)}$, where $\boldsymbol{I}_{\left(\left|\boldsymbol{m}_{i,k}\right| \times N_t\right)}$ denote $\left(\left|\boldsymbol{m}_{i,k}\right| \times N_t\right) \times \left(\left|\boldsymbol{m}_{i,k}\right| \times N_t\right)$ identity matrix.

Assume that the user u is covered by the cluster k on the sub-channel i, then, the $\left(\left|\boldsymbol{m}_{i,k}\right| \times N_t\right) \times N_r$ transmitting pre-coding matrix is

$$\boldsymbol{T}_{i,k}^{u} = \left[\boldsymbol{t}_{i,k}^{\boldsymbol{s}_{i,k}^{u}(1)}, \boldsymbol{t}_{i,k}^{\boldsymbol{s}_{i,k}^{u}(2)}, \ldots, \boldsymbol{t}_{i,k}^{\boldsymbol{s}_{i,k}^{u}(N_r)}\right]. \quad (8)$$

According to (8), $\boldsymbol{T}_{i,k}^{\boldsymbol{u}_{i,k}(a)}$ represents the $\left(\left|\boldsymbol{m}_{i,k}\right| \times N_t\right) \times N_r$ transmitting pre-coding matrix from BSs in the cluster k to the α-th user in $\boldsymbol{u}_{i,k}$ on the sub-channel i. Then, the $\left(\left|\boldsymbol{m}_{i,k}\right| \times N_t\right) \times \left(\left|\boldsymbol{u}_{i,k}\right| \times N_r\right)$ transmitting pre-coding matrix from the cluster k to the users in the cluster k on the sub-channel i can be denoted as

$$\boldsymbol{T}_{i,k} = \left[\boldsymbol{T}_{i,k}^{\boldsymbol{u}_{i,k}(1)}, \boldsymbol{T}_{i,k}^{\boldsymbol{u}_{i,k}(2)}, \ldots, \boldsymbol{T}_{i,k}^{\boldsymbol{u}_{i,k}\left(\left|\boldsymbol{u}_{i,k}\right|\right)}\right]. \tag{9}$$

Let $\boldsymbol{t}_{i,k}^{\boldsymbol{s}_{i,k}(s)}$ denote the $\left(\left|\boldsymbol{m}_{i,k}\right| \times N_t\right) \times 1$ transmitting pre-coding matrix of the s-th stream in the cluster k on the sub-channel i. Then, $\boldsymbol{t}_{i,k}^{\boldsymbol{s}_{i,k}(s)}$ corresponds to the s-th column of $\boldsymbol{T}_{i,k}$.

Let $\boldsymbol{Q}_{i,k}^{u} = \mathbb{E}\left[\boldsymbol{T}_{i,k}^{u}\boldsymbol{s}_{i,k}^{u}\boldsymbol{s}_{i,k}^{u}{}^{\mathrm{H}}\boldsymbol{T}_{i,k}^{u}{}^{\mathrm{H}}\right]$ denote the covariance matrix of the user u in the cluster k on the sub-channel i. Likewise, let $\boldsymbol{Q}_{i,k} = \mathbb{E}\left[\boldsymbol{T}_{i,k}\boldsymbol{s}_{i,k}\boldsymbol{s}_{i,k}{}^{\mathrm{H}}\boldsymbol{T}_{i,k}{}^{\mathrm{H}}\right]$ denote the transmitting covariance matrix from the cluster k to the users scheduled in the cluster k on the sub-channel i.

Under the BD precoding, the intra-cluster interference between scheduled users can be eliminated by the receiving pre-coding matrix. For an arbitrary stream s $(1 \le s \le N_r)$ of the user u on the i-th sub-channel in the cluster k, the $1 \times N_r$ receiving pre-coding matrix can be denoted as $\boldsymbol{v}_{i,k}^{\boldsymbol{s}_{i,k}^{u}(s)}$. The receiving pre-coding matrix of the user u from the cluster k on the on the i-th sub-channel is

$$\boldsymbol{V}_{i,k}^{u} = \left[\boldsymbol{v}_{i,k}^{\boldsymbol{s}_{i,k}^{u}(1)^{\mathrm{H}}}, \boldsymbol{v}_{i,k}^{\boldsymbol{s}_{i,k}^{u}(2)^{\mathrm{H}}}, \ldots \boldsymbol{v}_{i,k}^{\boldsymbol{s}_{i,k}^{u}(N_r)^{\mathrm{H}}}\right]^{\mathrm{H}}. \tag{10}$$

Similar to (9), on the sub-channel i, the $\left(\left|\boldsymbol{u}_{i,k}\right| \times N_r\right) \times N_r$ receiving pre-coding matrix from the cluster k to the scheduled users can be represented as

$$\boldsymbol{V}_{i,k} = \left[\boldsymbol{V}_{i,k}^{\boldsymbol{u}_{i,k}(1)^{\mathrm{H}}}, \boldsymbol{V}_{i,k}^{\boldsymbol{u}_{i,k}(2)^{\mathrm{H}}}, \ldots \boldsymbol{V}_{i,k}^{\boldsymbol{u}_{i,k}\left(\left|\boldsymbol{u}_{i,k}\right|\right)^{\mathrm{H}}}\right]^{\mathrm{H}}. \tag{11}$$

Let $\boldsymbol{v}_{i,k}^{\boldsymbol{s}_{i,k}(s)}$ denotes the $1 \times N_r$ receiving precoding matrix for an arbitrary stream s in the cluster k on the sub-channel i. Then, the vector corresponds to the s column in $\boldsymbol{V}_{i,k}$. The specific BD pre-coding scheme between BSs and users can be referenced in [10].

Let $\tilde{n}$ denotes the Additive White Gaussian Noise, for an arbitrary user, it has a zero mean, and the corresponding variance is $(\tilde{n}\tilde{n}^{\mathrm{H}}) = \sigma^2 I_{n_r}$, where I_{n_r} is a $N_r \times N_r$ unit matrix. For an arbitrary user u in the cluster k on the sub-channel i, the interference from users of other sub-channels and users in the cluster k can be nulled. The covariance matrix $R_{i,k}^{u}(-\boldsymbol{Q}_{i,k}^{u})$ of the interference plus noise of user u in the cluster k on the sub-channel i is

$$R_{i,k}^{u}\left(-\boldsymbol{Q}_{i,k}^{u}\right) = \sum_{\substack{k' \in \boldsymbol{k}_i \\ k' \neq k}} H_{u,k'}^{i}\left(\boldsymbol{Q}_{i,k}\right) {H_{u,k'}^{i}}^{\mathrm{H}} + \sigma^2 \mathrm{I}_{N_r}. \tag{12}$$

Therefore, by BD pre-coding, considering finite-alphabets, the user throughput in CoMP can be denoted as

$$R_{\mathrm{CoMP}} = \sum_{i \in \boldsymbol{f}} \sum_{k \in \boldsymbol{k}_i} \sum_{u \in \boldsymbol{u}_{i,k}} \frac{b}{|\boldsymbol{f}|} \log \left| \mathrm{I} + \frac{{H_{u,k}^{i}}^{\mathrm{H}} \left(R_{i,k}^{u}\left(-\boldsymbol{Q}_{i,k}^{u}\right)\right)^{-1} H_{u,k}^{i}}{\Gamma} \right|, \tag{13}$$

where R_{CoMP} is limited by the power proportion of BSs in the same cluster, which is determined by BD precoding, and the maximal transmitting power of BSs. The model needs complex analysis via matrix theorem. This makes the algorithms in the single-antenna scenario difficult to be extended to the multi-antenna scenario. In Ref. [12], orthogonal projection theory is employed in the multi-antenna scenario. However, it is not easy to be followed. In this chapter, we introduce a stream analysis model to extend the algorithms in the single-antenna scenario to the multi-antenna scenario.

We analyze (13) from the viewpoint of streams, in an arbitrary cluster k, $\mathbb{E}(\tilde{n}\tilde{n}^{\mathrm{H}}) = \sigma^2 I_{|\boldsymbol{u}_{i,k}|N_r}$. On the sub-channel i, for the streams s and s' ($1 \le s \le |\boldsymbol{s}_{i,k}|$, $1 \le s' \le |\boldsymbol{s}_{i,k'}|$), we assume that $\boldsymbol{s}_{i,k}(s)$ belongs to the user $\boldsymbol{u}_{i,k}(u)$ in the cluster k, and $\boldsymbol{s}_{i,k'}(s')$ belongs to the user $\boldsymbol{u}_{i,k'}(u')$ in the cluster k'. Then, the stream s in the cluster k will go through the transmitting pre-coding $\boldsymbol{t}_{i,k}^{\boldsymbol{s}_{i,k}(s)}$, channel $\boldsymbol{H}_{\boldsymbol{u}_{i,k'}(u'),k}^{i}$, and the receiving pre-coding $\boldsymbol{v}_{i,k'}^{\boldsymbol{s}_{i,k'}(s')}$. The equivalent channel gain (virtual channel gain) is

$$\tilde{H}_{\boldsymbol{s}_{i,k'}(s'),\boldsymbol{s}_{i,k}(s)}^{i} = \boldsymbol{v}_{i,k'}^{\boldsymbol{s}_{i,k'}(s')} \times \boldsymbol{H}_{\boldsymbol{u}_{i,k'}(u'),k}^{i} \times \boldsymbol{t}_{i,k}^{\boldsymbol{s}_{i,k}(s)}. \tag{14}$$

By BD pre-coding, the intra-interference is eliminated at users, we can further have

$$\tilde{H}_{\boldsymbol{s}_{i,k'}(s'),\boldsymbol{s}_{i,k}(s)}^{i} = 0, \quad k = k', \text{ and } s \neq s'. \tag{15}$$

By (15), the multi-antenna can be equivalent to multiple single-antenna virtual system, where the corresponding channel gain meet (14) and (15). After BD pre-coding, the power allocation scheme in the cluster k on the sub-channel i can be denoted as a $|\boldsymbol{s}_{i,k}| \times |\boldsymbol{s}_{i,k}|$ diagonal matrix, $\boldsymbol{D}_{i,k}$. The interference plus noise of the s-th stream is

$$Z_{i,k}^{\boldsymbol{s}_{i,k}(s)} = \sigma^2 + \sum_{\substack{k' \in \boldsymbol{k}_i \\ k' \neq k}} \sum_{s' \in [1:|\boldsymbol{s}_{ik'}|]} \boldsymbol{D}_{i,k'}(s', s') \times \tilde{H}_{\boldsymbol{s}_{i,k}(s),\boldsymbol{s}_{i,k'}(s')}^{i} \times \left(\tilde{H}_{\boldsymbol{s}_{i,k}(s),\boldsymbol{s}_{i,k'}(s')}^{i}\right)^{\mathrm{H}}, \tag{16}$$

where $Z_{i,k}^{s_{i,k}(s)}$ is the covariance of the interference plus noise for the stream s, which can be estimated by the edge users. Reference [10] introduces several methods of signal estimation such as pilot signal, blind estimation and so on. The receiving signal of stream s is

$$G_{i,k}^{s_{i,k}(s)} = \boldsymbol{D}_{i,k}(s,s) \times \tilde{H}^{i}_{s_{i,k}(s),s_{i,k}(s)} \times \left(\tilde{H}^{i}_{s_{i,k}(s),s_{i,k}(s)}\right)^{\mathrm{H}}. \tag{17}$$

Then, the user throughput of the stream s is

$$r_{i,k}^{s} = \frac{b}{|\boldsymbol{f}|}\log_2\left(1 + \frac{G_{i,k}^{s_{i,k}(s)}}{\Gamma \times Z_{i,k}^{s_{i,k}(s)}}\right). \tag{18}$$

By (18), the matrix theorem based multi-antenna model in (13) can be transformed into the virtual single-antenna system. In (9), for an arbitrary s $\left(1 \le s \le |\boldsymbol{s}_{i,k}|\right)$ in the cluster k on the sub-channel i, the $\left(|\boldsymbol{m}_{i,k}| \times n_t\right) \times 1$ transmitting pre-coding is $\boldsymbol{t}_{i,k}^{s_{i,k}(s)}$, where the transmitting pre-coding vector of the m-th BS to the stream $\boldsymbol{s}_{i,k}(s)$ is

$$\boldsymbol{t}_{i,k}^{s_{i,k}(s),\boldsymbol{m}_{i,k}(m)} = \left[\boldsymbol{t}_{i,k}^{s_{i,k}(s)}((m-1)\times n_t + 1)^{\mathrm{H}}, \boldsymbol{t}_{i,k}^{s_{i,k}(s)}((m-1)\times n_t + 2)^{\mathrm{H}}, \ldots, \boldsymbol{t}_{i,k}^{s_{i,k}(s)}((m-1)\times n_t + n_t)^{\mathrm{H}}\right]^{\mathrm{H}}. \tag{19}$$

Let $\left\|\boldsymbol{t}_{i,k2}^{s_{i,k}(s),\boldsymbol{m}_{i,k}(m)}\right\|_2^2$ denotes the 2-norm of $\boldsymbol{t}_{i,k}^{s_{i,k}(s),\boldsymbol{m}_{i,k}(m)}$, then, the power proportion of the m-th BS to the stream $\boldsymbol{s}_{i,k}(s)$ is

$$F_{i,k}^{s_{i,k}(s),\boldsymbol{m}_{i,k}(m)} = \left\|\boldsymbol{t}_{i,k}^{s_{i,k}(s),\boldsymbol{m}_{i,k}(m)}\right\|_2^2, \tag{20}$$

where $F_{i,k}^{s_{i,k}(s),\boldsymbol{m}_{i,k}(m)}$ denotes the power proportion of the m-th BS in $\boldsymbol{m}_{i,k}$ in $\boldsymbol{D}_{i,k}(s,s)$.

From the viewpoint of virtualization, each BS serves for multiple streams and each stream can be served by multiple BSs with the power proportions determined by BD pre-coding. We can further ignore the subscripts of sub-channel and cluster and virtualize the system as a streams based virtual system, where one stream corresponds to a pair of transmitting-receiving with single-antenna. The channel gain is determined in (14). Let $\boldsymbol{s}_m$ denotes the stream set across all the sub-channels of the BS m ($m \in \boldsymbol{n}$), and $\boldsymbol{m}_s$ denotes the set of BSs serving for the stream s. Then, (20) can be re-written as

$$F_{m'}^{\boldsymbol{s}_{m'}(s')} = F_{i,k}^{\boldsymbol{s}_{i,k}(s),\boldsymbol{m}_{i,k}(m)}, \tag{21}$$

where $m' = \boldsymbol{m}_{i,k}(m)$, and s' denotes the subscript of the stream $\boldsymbol{s}_{i,k}(s)$ in $\boldsymbol{s}_m$. Likewise, the power allocation of the s-th stream on the i-th sub-channel $\boldsymbol{D}_{i,k}(s,s)$ can be re-written as

$$D_{s'} = \boldsymbol{Q}_{i,k}(s,s), \tag{22}$$

for the stream $s \in \boldsymbol{s}_{i,k}$, in the cluster k on the sub-channel i, let $\boldsymbol{s}_{\text{null}}^s$ denotes the streams other than s across the other sub-channels, then, the interference from streams in $\boldsymbol{s}_{\text{null}}^s$ to the stream s can be nulled. We can re-write the virtual channel gain in (14) and (15) as

$$\tilde{H}_{s_0',s_0} = \tilde{H}^i_{\boldsymbol{s}_{i,k'}(s'),\boldsymbol{s}_{i,k}(s)}, \quad \text{where } s_0' = \boldsymbol{s}_{i,k'}(s'), s_0 = \boldsymbol{s}_{i,k}(s) \tag{23}$$

$$s.t. \quad \tilde{H}_{s_0',s_0} = 0, \quad \forall s_0' \in \boldsymbol{s}_{\text{null}}^s. \tag{24}$$

Let $\boldsymbol{s}_{\text{all}}$ denote the set of streams in the system, then, we can re-written (16) and (17) as

$$Z_s = \sigma^2 + \sum_{\substack{s' \in \boldsymbol{s}_{\text{all}} \\ s' \neq s \\ s' \notin \boldsymbol{s}_{\text{null}}^s}} D_s \times \tilde{H}_{s,s'} \times \left(\tilde{H}_{s,s'}\right)^{\text{H}}, \quad \forall s, \tag{25}$$

$$G_s = D_s \times \tilde{H}_{s,s} \times \left(\tilde{H}_{s,s}\right)^{\text{H}}, \quad \forall s. \tag{26}$$

Considering the maximal transmitting power of BS, $p_{\max}$, for an arbitrary stream s, BS m $(m \in \boldsymbol{m}_s)$ serves for the stream with power p_m^s, let r_s denote the throughput of the stream s, we can get the stream model based system model as

$$\text{maximize} \sum_{s \in \boldsymbol{s}_{\text{all}}} r_s = \sum_{s \in \boldsymbol{s}_{\text{all}}} \frac{b}{|\boldsymbol{f}|} \log_2\left(1 + \frac{G_s}{\Gamma \times Z_s}\right), \tag{27}$$

$$s.t. \quad D_s = \sum_{m \in \boldsymbol{m}_s} p_m^s, \quad \forall s \in \boldsymbol{s}_{\text{all}}, \tag{28}$$

$$\frac{p_m^s}{D_s} = F_m^s, \quad \forall s \in \boldsymbol{s}_{\text{all}}, \ \forall m \in \boldsymbol{m}_s, \tag{29}$$

$$\sum_{s \in \boldsymbol{s}_m} p_m^s \leq p_{\max}, \quad \forall m \in \boldsymbol{m}. \tag{30}$$

By the system model, we can easily transformed (13) into multiple transmitting-receiving pairs with single-antenna mode.

2 Heuristic Mechanism in CoMP

2.1 Centralized Cooperative BS Clustering

The existing centralized CoMP cannot commendably cater for the changing environment. In this sub-section, we introduce a centralized CoMP with environment-adaptation (CTS, centralized transmission scheduling) [7]. CTS employs the graph theory, and each BS can get the global information. CTS is executed in iteration, in each of which the transmission and scheduling scheme is updated based on greedy algorithm.

For simplication, in this sub-section, we focus on the single-antenna scenario. We define the wireless link as one transmission between BS and user. In CoMP, one wireless link can denote either one effective wireless connection or one wireless interference connection. Take the features of CoMP into account, the CTS algorithm possesses better flexibility and environment-adaptation than the existing ones.

We first give the graph $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E}, \boldsymbol{C})$ in Fig. 2 to describe the execution of CTS, where $\boldsymbol{E}$ is the set of edges, $\boldsymbol{V}$ is the set of points, $\boldsymbol{C}$ is the set of colors. One point can denote one BS, one user, or both one BS and one user, which is determined by the specific edges. Specifically, edge $V_x \rightarrow V_y$ denotes the wireless link from BS x to a specific user in the cell y, the value of the edge is the corresponding RSRP value.

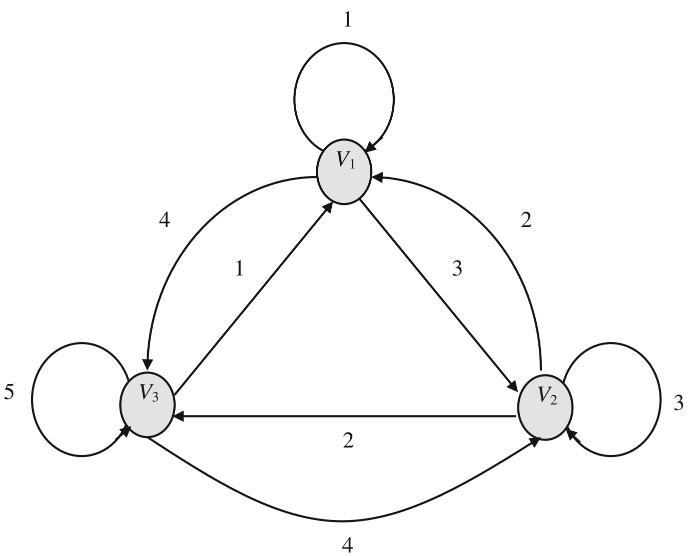


Fig. 2 An example for CTS algorithm

CTS is executed in iteration and independently on each sub-channel. For an arbitrary sub-channel i, the edge $V_1 \rightarrow V_2$ denotes the wireless link from BS 1 to the selected user, t_{12}^i, in Cell 2, and the corresponding RSRP is 3. Edge $V_2 \rightarrow V_2$ denotes the wireless link from BS 2 to the selected user, t_{22}^i, in Cell 2. The observed RSRP value is 3 at the user t_{22}^i. Noticeably, each edge corresponds to one user, where t_{12}^i and t_{22}^i may be not the same user. Non-scheduled users in the same cell can be dealt as the user candidates for scheduling. At most one user candidate can be scheduled finally. At the beginning, each point is colored with one different color and the edges are not colored. Denote $\boldsymbol{V}_y^c$ as the set of points with the same color with Point V_y, if one user is finally scheduled in Cell y, $\boldsymbol{V}_y^c$ denotes the set of BSs serving the user. One key issue in CST is how to determine the selected user for a specific edge $V_x \rightarrow V_y$. For the wireless link from BS x to the user t in Cell y, define the weigh W_{xy}^{it} as

Algorithm 1 CTS (centralized transmission scheduling) algorithm [7].

CTS (centralized transmission scheduling)

Input:

f={1, …, $|\boldsymbol{f}|$}；$\boldsymbol{n}$={1, …, $|\boldsymbol{n}|$}；$\boldsymbol{u}_j$(j ∈ $\boldsymbol{n}$)；$RSRP_0$和 RSRPs。

Output:

Transmission and scheduling on each sub-channel.

for (i=1; $i \le |\boldsymbol{f}|$; i++)

{

1. Initialization:

Let $\boldsymbol{n} \rightarrow \boldsymbol{A}$. For all the wireless links, select the users with (31) and (32). Calculate the credits with (31) and (33). Constitute $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E}, \boldsymbol{C})$. Color the points (Cells) with different colors. Built the $\boldsymbol{n} \times \boldsymbol{n}$ matrix $\boldsymbol{W}$ with RSRPs.

2. Coloring:

Find the maximal value W_{xy}^i in $\boldsymbol{W}$, delete the corresponding rows of BS x and y. If $x \neq y$, delete the corresponding column of the user scheduled in Cell x. Delete BS x and y in $\boldsymbol{A}$, coloring V_x, $V_x \rightarrow V_y$, $V_y \rightarrow V_y$ with the color of V_x. Update the graph. If $x \neq y$, delete all the incoming edges of V_x, user t_{xy}^i is scheduled. Remove Cell y and the users in Cell y. Update RSRPs with (34) and calculate the credits. Select the users in the un-colored points based on (31) and (32). Update $\boldsymbol{W}$. Repeat Step 2 until $|\boldsymbol{A}|$=0 or $\boldsymbol{W}$=**0**

3. Dealing with the un-allocated BSs :

```
if (|A| > 0)
{
        for (each and all BSs j ∈ A)
        {
          BS j ignores the sub-channel i (drop sub-channel i).
        }
}
}
```

$$W_{xy}^{it} = \begin{cases} \dfrac{\mathrm{RSRP}_{xy}^{it}}{\sum_{\forall V_n \notin \boldsymbol{V}_y^c, n \neq x} \mathrm{RSRP}_{ny}^{it} + \sigma^2}, & \text{if } \mathrm{RSRP}_{xy}^{it} > \mathrm{RSRP}_0, \\ 0, & \text{others.} \end{cases} \tag{31}$$

where i denotes the sequence number of sub-channel, RSRP_{xy}^{it} is the corresponding RSRP value. RSRP_0 is the threshold for the BS cooperation, where only when RSRP for one user is larger than it, the user can be served. The selected user t_{xy}^{i} for edge $V_x \to V_y$ is

$$t_{xy}^{i} = \underset{t \in \boldsymbol{u_y}}{\arg\max} \left\{ W_{xy}^{it} \right\}, \tag{32}$$

where $\boldsymbol{u_y}$ is the set of all the edge users in Cell y. When t_{xy}^{i} is determined, the RSRP of edge $V_x \to V_y$ is also determined. Based on the selected user, the graph $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E}, \boldsymbol{C})$ is constituted by all the possible wireless links. In CTS, one point V_y may contain multiple possible selected users, where each user corresponds to one incoming edge of V_y. When the scheduled user in Cell y is determined, RSRP values of all the incoming edges should be updated (34).

Algorithm 1 concludes the process of CTS. For an arbitrary sub-channel i, in Step 1, the graph is initialized, a $|\boldsymbol{n}| \times |\boldsymbol{n}|$ matrix $\boldsymbol{W}$ is constituted, the entry W_{xy}^{i} in $\boldsymbol{W}$ is

$$W_{xy}^{i} = \max_{t \in \boldsymbol{u_y}} \left\{ W_{xy}^{it} \right\} = W_{xy}^{it} \Big|_{t = t_{xy}^{i}}, \tag{33}$$

where W_{xy}^{i} is the credit weigh value of BS x joining CBS serving for Cell y. In Step 2, the maximal value in $\boldsymbol{W}$ is selected, and the cooperation can be determined of the corresponding row and column of the entry. The philosophy behind Step 2 is maximizing the ratio of receiving RSRP and interference RSRP with the selected user (as (31)). After the W_{xy}^{i} has been selected, the corresponding rows of BSs x and y will be deleted in $\boldsymbol{W}$. If $x \neq y$, the corresponding column of Cell x will be deleted, which denotes that BS x and y cooperate to serve for user t_{xy}^{i} in Cell y. Then, all users in Cell x will be not served, and thus all the incoming edges of V_x will be deleted. The other users $\left\{ t_{ny}^{i} \middle| \forall V_n \notin \boldsymbol{V_y^c}, n \neq x \right\}$ in Cell y will be removed. Points V_x, V_y, and edges $V_x \to V_y$, $V_y \to V_y$ will be colored by V_y, the corresponding RSRP value will be updated by (34)

$$\mathrm{RSRP}_{ny}^{it} = \mathrm{RSRP}_{ny}^{it} \Big|_{t = t_{xy}^{i}}, \quad \forall V_n \notin \boldsymbol{V_y^c}, \ \forall t. \tag{34}$$

Based on the refined graph, CTS will find a new set of selected users in the un-colored points and update the remaining entries in $\boldsymbol{W}$. Then, Step 2 will be repeated. When either all the cooperations of BSs have been determined or $\boldsymbol{W}$ = 0, the algorithm stops. When $\boldsymbol{W}$ = 0, BSs of the un-colored points do not serve for any users. For an arbitrary sub-channel i, after the coloring, each color corresponds to one CBS.

Take Fig. 2 as an example, we assume that $\text{RSRP}_0 = 0$, and $\sigma^2 = 1$, then, the initial $\boldsymbol{W}$ is

$$\boldsymbol{W} = \begin{bmatrix} \frac{1}{3+\sigma^2} & \frac{3}{4+\sigma^2} & \frac{4}{2+\sigma^2} \\ \frac{2}{1+\sigma^2} & \frac{3}{7+\sigma^2} & \frac{2}{4+\sigma^2} \\ \frac{1}{2+\sigma^2} & \frac{4}{3+\sigma^2} & \frac{5}{6+\sigma^2} \end{bmatrix} = \begin{bmatrix} 0.25 & 0.6 & 1.33 \\ 0.67 & 0.375 & 0.4 \\ 0.33 & 0.8 & 0.71 \end{bmatrix}, \tag{35}$$

where W_{13} is the maximal entry, then BS 1 and BS 3 cooperate to serve for the user s^i_{13}. Thereafter, we update the matrix $\boldsymbol{W}$ as $\boldsymbol{W} = \begin{bmatrix} 3/(7+\sigma^2) & 2/(\sigma^2) \end{bmatrix} = \begin{bmatrix} 0.375 & 2 \end{bmatrix}$. W_{12} is selected as the maximal value. Then, all the three BSs constitutes one CBS serving for the scheduled user in Cell 3.

2.2 *Distributed Cooperative BS Clustering*

Algorithm 2 gives us a distributed method to determine the BS transmission and user scheduling (DTS, distributed transmission scheduling) in CoMP. Like CTS, DTS is executed independently on each sub-channel. DTS is constituted by three steps. Step 1 is initialization. Step 2 is the main body of DTS, and in Step 3, BSs without transmission relationship will serve for no users on the corresponding sub-channel.

Algorithm 2 DTS (distributed transmission scheduling) algorithm [7].

Distributed Transmission Scheduling (DTS)

Input:

$\boldsymbol{f}$={1, …, $|\boldsymbol{f}|$} ; $\boldsymbol{n}$={1, …, $|\boldsymbol{n}|$} ; $\boldsymbol{u}_j (j \in \boldsymbol{n})$; $m_{\max}$; RSRP_0 and RSRPs.

Output:

Transmission and scheduling on each sub-channel.

for (i=1; $i \le |\boldsymbol{f}|$; i++)

{

1. Initialization:

Let $\boldsymbol{n} \to \boldsymbol{A}$, $m_{\max} \to m$.

2. Transmission and scheduling:

2.1) for ($\forall j \in \boldsymbol{A}$)

{

2.1.1) Arbitrary edge user t in Cell j find $x \le m-1$ RSRPs that meet $\text{RSRP} \ge \text{RSRP}_0$ from the neighboring BSs other than BS j. If either $x \ge 1$ or the RSRP receiving from BS j is larger than RSRP_0, then one set with size $m_t(j) = x+1$ containing the x RSRPs and the RSRP from BS j will be constituted. User t calculates the mean value of the $m_t(j)$ RSRPs as $\overline{\text{RSRP}_m}$, and reports $\overline{\text{RSRP}_m}$ to BS j. The $m_t(j)$ BSs together with t will be also recorded at the user t.

2.1.2) BS j find the $\overline{\text{RSRP}_m}$ with the maximal value, and send a requirement to the corresponding user t' for the BSs list of the $\overline{\text{RSRP}_m}$.

2.1.3) BS j sends $m_{t'}(j)$ cooperation requirements to the BSs on the list.

}

(continued)

Algorithm 2 (continued)

2.2) Each BS checks the signaling of cooperation requirement, and feeds back an ACK signaling to the BS with the maximal $\overline{\text{RSRP}}_m$.
2.3) Each BS counts the number of ACKs, if the BS j receives $m_{t'}(j)$ ACK signaling, it will send cooperation success signaling to all the BSs in $\boldsymbol{A}$. BSs in the list constitutes one CBS and the user in the list is scheduled.
2.4) BSs with transmission relationship will be sent to all the BSs in $\boldsymbol{A}$ and deleted from $\boldsymbol{A}$. If $\boldsymbol{A}$ changes and $|\boldsymbol{A}| \geq m$, repeat Step 2.1). Otherwise, if $m > 0$, then, $\min\{m-1, |\boldsymbol{A}|\} \rightarrow m$, repeat Step 2.1).

3. Dealing with the un-allocated BSs :

```
if (|A| > 0)
{
  BSs in A do not serve for any users on the sub-channel i (drop the sub-channel i).
 }
}
```

Specifically, In Step 1, set $\boldsymbol{A}$ stores the un-allocated BSs which is initialized as $\boldsymbol{n}$ in each BS, the maximal size of CBS $m_{\max} \rightarrow m$, and m will be changed in iteration in Step 2. In Step 2, $m_t(j)$ ($m_t(j) \leq m$) RSRPs will be averaged and sent to the serving BS by each user. In Step 2.1), the maximal value of $\overline{\text{RSRP}}_m$ selected by a BS is the potential candidate of CBS and the corresponding scheduled users. Since BS only exchange with user with the maximal $\overline{\text{RSRP}}_m$, the feedback can be largely cut down. In Step 2.2) to 2.4), a negotiate mechanism has been employed by BSs to determine the CBSs. If no CBSs are determined, the maximal size of CBS $m = m - 1$ in Step 2.4), the algorithm goes to the next iteration. This can guarantee the convergence of the algorithm.

Although Algorithm 2 can achieve a distributed CoMP mechanism with environment-adaptation, the single-antenna mode limits its applicable scenarios. Algorithm 3 further extends such a distributed mechanism into the multi-antenna scenario. Define the channel energy (CE) [13] as the square of the Frobenius norm of the channel matrix between BS and users. In this chapter, BD pre-coding is employed and Chordal distance [14] is used to measure the space spanned by the channel matrix. A larger Chordal distance indicates a space more verging to orthogonality. This can decrease the power waste in BD for spanning an orthogonality space with the channel matrix.

Algorithm 3 shows the transmission and user scheduling in CoMP under the multi-antenna scenario (DTSMA, distributed transmission and scheduling with multiple-antennas). Like DTS algorithm, DTSMA is executed independently on each sub-channel. Step 1 is initialization, and Steps 2–4 are the main body of the algorithm. Thereafter, BSs without transmission will not serve for any users. Specifically, in Step 1, the set of BSs $\boldsymbol{a}$ is initialized as $\boldsymbol{n}$ and stored in each BS, the maximal size of CBS is $m' = m_{\max}$. E_0 is the cooperative threshold of channel energy, and D_0 is the cooperative threshold of Chordal distance. In Step 2.1), edge

Algorithm 3 DTSMA (distributed transmission and scheduling with multiple-antennas) algorithm [15].

<u>Distributed transmission and scheduling with multiple-antennas (DTSMA)</u>

<u>Input:</u>
$\boldsymbol{f}$={1, …, $|\boldsymbol{f}|$}; $\boldsymbol{n}$={1, …, $|\boldsymbol{n}|$}; set of edge users in each cell; N_t; N_r ; $m_{\max}$; E_0; D_0 and RSRPs.

<u>Output:</u>
Transmission and scheduling in CoMP.

for (i=1; $i\leq|\boldsymbol{f}|$; i++)
{

<u>1. Initialization:</u>
Let $\boldsymbol{n} \rightarrow \boldsymbol{a}$, $m_{\max} \rightarrow m'$ in each BS.

<u>2. Determine the CBS candidates :</u>
for (arbitrary $n \in \boldsymbol{a}$)
{
2.1) Arbitrary user u in n measures $|\boldsymbol{\alpha}|$ channel matrices based on RSRP and calculates the corresponding CE.
2.2) Arbitrary edge user u finds the maximal $x \leq m' - 1$ BSs other than BS n with CE larger than E_0. If either $x \geq 1$ or the CE from Cell n is larger than E_0, the $x + 1$ BSs are combined as one CBS candidate $\boldsymbol{c}(u)$. $\boldsymbol{c}(u)$, u and the channel matrix between the CBS and the user u will be combined as one content signaling $W(u)$ and sent to the BS n.
2.3) BS n sends $W(u)$ to the BS with the minimal sequence number in $\boldsymbol{c}(u)$.
2.4) BS n combines all the CBS it received with the same $\boldsymbol{c}(u)$, and reports the BSs and users in $\boldsymbol{c}(u)$.
}

<u>3. User scheduling in CBS candidates:</u>
for (arbitrary CBS candidate) {Executing the user scheduling algorithm and returning the SMV value.}

<u>4. Transmission and scheduling:</u>
for (arbitrary BS $n \in \boldsymbol{a}$)
{
4.1) BS n finds the CBS candidate with the maximal SMV value it received. The SMV and the source BS number , n, will be combined as one CRS signaling and sent to the BSs on the BS list of the CBS candidate.
4.2) Each BS selects the CRS with the maximal SMV value and sends one ACK signaling to the source BS.
4.3) If the BS receives x+1 ACK signaling, it sends CSS signaling to all the BSs on the list.
4.4) The BS list of the determined CBS is sent to BSs in $\boldsymbol{a}$ and then deleted from $\boldsymbol{a}$. If $\boldsymbol{a}$ changes or $|\boldsymbol{a}| \geq m'$, repeat Step 2.1).
}

<u>5. Dealing with the un-allocated BSs :</u>
if ($|\boldsymbol{a}| > 0$)
{
for (arbitrary BS $n \in \boldsymbol{a}$) {BS n does not serve for any users on the sub-channel i.}
}
}

users in an arbitrary BS n in $\boldsymbol{a}$ measure the RSRP to obtain the CE. In Step 2.2), based on E_0, arbitrary user u in Cell n can determine a CBS candidate with size x + 1. Each candidate has one user and the BS set of the user u is $\boldsymbol{c}(u)$. If $x = 0$ and the receiving CE of user u in Cell n is less than E_0, user u does not report the CBS candidate. Then, the sequence number of the user u, $\boldsymbol{c}(u)$, and the corresponding channel matrix will be combined as *Content Signaling* $W(u)$ and sent to BS n. In Step 2.3), each BS analyzes the receiving $W(u)$ and sends $W(u)$ to the BS in $\boldsymbol{c}(u)$ with the minimal sequence number via wired connections. In Step 2.4) CBS candidates will be combined and updated as new CBS candidates, where all the CBS candidates with the same $\boldsymbol{c}(u)$ will be combined and reported to the users in covered by the new CBS candidates. In Step 3, each CBS candidate determine the set of scheduled user. Step 3 can be embedded into DTSMA, this makes the algorithm more flexible. The specific algorithm for user scheduling will be introduced latter. After user scheduling, each CBS candidate can be allocated one identification label, which contains CBS label/the set of BSs, the set of scheduled users and the *Sum Metric Value* (SMV). SMV is used to measure the weigh value of CBS candidates. Step 4 is the transmission of CoMP. Like the DTS algorithm, the CBSs can be determined with SMV values. In Step 5, the un-allocated BSs will be dropped on the corresponding sub-channel.

In Step 3, the user scheduling per CBS candidate should be determined by user scheduling algorithm. In this chapter, we introduce TCEMS (total channel energy maximization scheduling) algorithm to achieve it. Let $G(m,\ n)$ denotes the n-Grassmannian sub-space in the Euclidean space with m domains [16]. Define $\boldsymbol{A}, \boldsymbol{B} \in G(m,n)$ and $\boldsymbol{A}_0$, $\boldsymbol{B}_0$ are the orthogonal basis of $\boldsymbol{A}$and $\boldsymbol{B}$. Then, the Chordal distance between $\boldsymbol{A}$and $\boldsymbol{B}$ is

$$d_c(\boldsymbol{A},\boldsymbol{B}) = \frac{1}{\sqrt{2}}\left\|\boldsymbol{A}_0\boldsymbol{A}_0^{\mathrm{H}} - \boldsymbol{B}_0\boldsymbol{B}_0^{\mathrm{H}}\right\|_{\mathrm{F}}. \tag{36}$$

Then, TCEMS algorithm can be proposed based on Chordal distance. As shown in Algorithm 4, the algorithm is executed in each CBS candidate. Step 1 is the initialization. In Step 2, the user with the maximal CE in each CBS candidate will be selected and put in the set $\boldsymbol{b}$. In Step 3, the algorithm is implemented in iteration. In Step 3.1), for an arbitrary CBS candidate, define the scheduling qualification as the users in $\boldsymbol{u}$ with the Chordal distance larger than D_0 between the orthonormalized channel of the user and the orthonormalized channel spanned by users in $\boldsymbol{b}$. Users with the scheduling qualification will be further measured and the user with the maximal CE will be selected as the scheduled user in this iteration. If no user has the scheduled qualification in 3.2), the algorithm goes to Step 3.5). In Step 3.3), the scheduled user is added to set $\boldsymbol{b}$, and removed from $\boldsymbol{u}$. The overall channel energy F of users in $\boldsymbol{b}$ and the channel matrix $\boldsymbol{G}'$ in $\boldsymbol{b}$ will be updated. In Step 3.4), if $\boldsymbol{u}$ is an empty set, the algorithm goes to Step 3.5). Otherwise, the algorithm goes to Step 3.1) for the next iteration. After the algorithm, users in $\boldsymbol{b}$ is the set of scheduled users, and F is defined as the SMV value of the CBS candidate.

Next, we give the complexity analysis of TCEMS, we define φ as the overall number of flops, which denotes the real floating point operation [17]. As in [17], real addition, multiplication are all counted as one flop operation, complex addition is counted as two flops, and complex multiplication is counted as six flops. For an arbitrary CBS candidate with user set $\boldsymbol{u}$, it can be easily proved that the complexity is $\theta^{\text{TCEMS}} = \text{O}(|\boldsymbol{u}|N_t^3)$. TCEMS possesses the same complexity with the algorithm proposed in [17]. In practice, since TCEMS does not need SVD decomposition, and D_0 can decrease the complexity, TCEMS can be more popular than the existing algorithms.

Algorithm 4 TCEMS (Total channel energy maximization scheduling) algorithm [15].

<u>Total channel energy maximization scheduling (TCEMS)</u>

<u>Input:</u>

CBS candidates on each sub-channel; user set for each CBS candidate and the corresponding channel matrix; D_0.

<u>Output:</u>

User set in each CBS candidate.

for (arbitrary CBS candidate)

{

<u>1. Initialization:</u>

Set of users $\boldsymbol{u}$, channel matrix $\boldsymbol{H}_u$ of an arbitrary user $u \in \boldsymbol{u}$ in the CBS. Let $(\cdot)_0$ denote the orthogonal basis of a matrix. The scheduled user set is initialized as $\boldsymbol{b}$. F=0.

<u>2. Selecting the first user in *b*:</u>

Let $u_1 = \arg_{u \in \boldsymbol{u}} \|\boldsymbol{H}_u\|_\text{F}^2$, $\boldsymbol{b}$=[u_1]. Remove u_1 from $\boldsymbol{u}$. Update $\boldsymbol{u}$. Let $\boldsymbol{G}' = (\boldsymbol{H}_{u_1})^\text{T}$, $F = \|\boldsymbol{H}_{u_1}\|_\text{F}^2$.

<u>3. Update *b*:</u>

3.1) While (the users in $\boldsymbol{b}$ meet (6))

{

Let the scheduling set of users $\boldsymbol{x}$ as an empty set, f=0, chordal distance D=0.

for (arbitrary user u' in $\boldsymbol{u}$)

{

if (D < D_0) {continue};

Let $\boldsymbol{P}' = \boldsymbol{H}_{u'}{}^T$; $\boldsymbol{P} = \boldsymbol{P}'_0 \boldsymbol{P}'_0{}^\text{H}$; $\boldsymbol{G} = \boldsymbol{G}'_0 \boldsymbol{G}'_0{}^\text{H}$.

Calculating $D = \frac{1}{\sqrt{2}} \times \|\boldsymbol{P} - \boldsymbol{G}\|_\text{F}$, and $f' = \|\boldsymbol{H}_{u'}\|_\text{F}^2$.

If ($f' > f$) {$f = f'$; $x = u'$};

}

}

3.2) if ($\boldsymbol{x}$ is an empty set) go to step 3.5), else 3.3);

3.3) Remove $\boldsymbol{x}$ from $\boldsymbol{u}$. Update $\boldsymbol{b} = [\boldsymbol{b}\ x]$, $F = F + \|\boldsymbol{H}_x\|_\text{F}^2$, and $\boldsymbol{G}' = [\boldsymbol{G}'\ \boldsymbol{H}_x{}^T]$.

3.4) if ($\boldsymbol{u}$ is an empty set) go to step 3.5, else 3.1);

3.5) The user set in the CBS candidate is $\boldsymbol{b}$. The sum metric value (SMV) of the CBS candidate is defined as the overall channel energy F.

}

2.3 Nash Non-cooperative Power Game

After the transmission and scheduling in CoMP, another key issue in CoMP is to determine the power allocation scheme. In this sub-section, we employs Nash non-cooperative game based power allocation game (PAG) to execute the power allocation in CoMP.

Definition 1 Based on an appropriate pricing mechanism, the power on an arbitrary sub-channel $i \in \boldsymbol{f}$ has $\boldsymbol{G} = \left[\boldsymbol{n_i}, \{\boldsymbol{p}_{i,j}\}, \left\{u^c_{i,j}(\cdot)\right\}\right]$, where $\boldsymbol{n_i}$ denotes the set of BSs involved in the power game, $\boldsymbol{p}^{\mathrm{PAG}}_{i,j} = \left[0, p^{\mathrm{WF}}_{i,j}\right]$ denotes the strategy space for power allocation on the sub-channel i, and $p^{\mathrm{WF}}_{i,j}$ denotes the initial power allocation of BS j on the sub-channel i with water-filling algorithm. $u^c_{i,j}(\cdot)$ denotes the net utility function of BS j on the sub-channel i. Obviously, $\boldsymbol{p}^{\mathrm{PAG}}_{i,j} = \left[0, p^{\mathrm{WF}}_{i,j}\right]$ is a closed and bounded convex set. Let $p_{i,j}$ denote the transmitting power of BS j on the sub-channel i, and $p_{i,-j}$ denote the wireless interference of BS j on the sub-channel i. Then, the net utility function is

$$u^c_{i,j}\left(p_{i,j}, p_{i,-j}\right) = u_{i,j}\left(p_{i,j}, p_{i,-j}\right) - c_{i,j}\left(p_{i,j}, p_{i,-j}\right), \tag{37}$$

where $u_{i,j}\left(p_{i,j}, p_{i,-j}\right)$ denotes the utility function, $c_{i,j}\left(p_{i,j}, p_{i,-j}\right)$ denotes the pricing function as follows:

$$u_{i,j}\left(p_{i,j}, p_{i,-j}\right) = \frac{b}{|\boldsymbol{f}|}\log_2\left(1 + \frac{\gamma_{ijk}}{\Gamma}\right)\Big|_{j\in \boldsymbol{m_{ik}}}, \tag{38}$$

$$c_{i,j}\left(p_{i,j}, p_{i,-j}\right) = \lambda_{i,j} p_{i,j}, \tag{39}$$

$$u^c_{i,j}\left(p_{i,j}, p_{i,-j}\right) = \frac{b}{|\boldsymbol{f}|}\log_2\left(1 + \frac{\gamma_{ijk}}{\Gamma}\right)\Big|_{j\in \boldsymbol{m_{ik}}} - \lambda_{i,j} p_{i,j}, \tag{40}$$

where $\lambda_{i,j} > 0$ denotes the pricing factor, and γ_{ijk} is

$$\gamma_{ijk} = \frac{p_{i,j} g_{ijk}}{\sum_{j' \notin \boldsymbol{m_{ik}}} p_{i,j'} g_{ij'k} + \sigma^2}\Bigg|_{j\in \boldsymbol{m_{ik}}}. \tag{41}$$

Considering the limitation $0 \le p_{i,j} \le p^{\mathrm{WF}}_{i,j}$, in this sub-section, we design the pricing factor $\lambda_{i,j}$ to guarantee this limitation. By (40)–(41), we have

$$\frac{\partial u_{i,j}^{c}\left(p_{i,j},p_{i,-j}\right)}{\partial p_{i,j}}=-\lambda_{i,j}+\frac{b}{\ln 2|\boldsymbol{f}|}\times\frac{g_{ijk}}{p_{i,j}g_{ijk}+\Gamma\left(\sum_{j'\notin \boldsymbol{m}_{ik}}p_{i,j'}g_{ij'k}+\sigma^{2}\right)}\Bigg|_{j\in \boldsymbol{m}_{ik}}. \tag{42}$$

By $\frac{\partial u_{i,j}^{c}\left(p_{i,j},p_{i,-j}\right)}{\partial p_{i,j}}=0$, we can further get

$$p_{i,j}=\frac{b}{\lambda_{i,j}\ln 2|\boldsymbol{f}|}-\frac{\Gamma(I_{ik}+\sigma^{2})}{g_{ijk}}\bigg|_{j\in \boldsymbol{m}_{ik}}, \tag{43}$$

where the interference I_{ik} is

$$I_{ik}=\sum_{j'\notin \boldsymbol{m}_{ik}}p_{i,j'}g_{ij'k}. \tag{44}$$

To guarantee $p_{i,j}$ meets $0\le p_{i,j}\le p_{i,j}^{\mathrm{WF}}$, by (43), we have the upper and lower bounds of $\lambda_{i,j}$ as

$$\lambda_{i,j}\le\lambda_{i,j}^{\max}=\frac{bg_{ijk}}{\ln 2|\boldsymbol{f}|\Gamma(I_{ik}+\sigma^{2})}\bigg|_{j\in \boldsymbol{m}_{ik}}, \tag{45}$$

$$\lambda_{i,j}\ge\lambda_{i,j}^{\min}=\frac{bg_{ijk}}{\ln 2|\boldsymbol{f}|\left[\Gamma(I_{ik}+\sigma^{2})+p_{i,j}^{\mathrm{WF}}g_{ijk}\right]}\Bigg|_{j\in \boldsymbol{m}_{ik}}. \tag{46}$$

Let $\alpha_{i,j}=e^{-\left(\frac{g_{ijk}}{\beta}\right)}\Big|_{j\in \boldsymbol{m}_{ik}}=\frac{\lambda_{i,j}-\lambda_{i,j}^{\min}}{\lambda_{i,j}^{\max}-\lambda_{i,j}^{\min}}$, where β is a positive constant. Then, we have

$$\begin{cases}\lim\limits_{\beta\to 0}\alpha_{i,j}=0,\\ \lim\limits_{\beta\to+\infty}\alpha_{i,j}=1\,,\\ 0<\alpha_{i,j}<1.\end{cases} \tag{47}$$

Define $\lambda_{i,j}=\alpha_{i,j}\times\lambda_{i,j}^{\max}+\left(1-\alpha_{i,j}\right)\times\lambda_{i,j}^{\min}$, then $\lambda_{i,j}$ meets $\lambda_{i,j}^{\min}<\lambda_{i,j}<\lambda_{i,j}^{\max}$. Therefore, such a pricing mechanism can guarantee the limitation of power allocation, and contain the features of wireless channel quality. These make the algorithm environment-adaptation.

Algorithm 5 WFFD (WF with frequency diversity) algorithm [18].

Water-filling with frequency diversity (WFFD) algorithm

Input:

BS set $\boldsymbol{n}$, $\boldsymbol{m}_s$, $\boldsymbol{s}_m$, $F_m^s(m \in \boldsymbol{n})$, s $(s \in \boldsymbol{s}_m)$, $p_{\max}$, $\widetilde{H}_{s_1,s_2}$ $(\forall s_1,\ s_2 \in \boldsymbol{s}_{\text{all}})$.

Output:

Q_s $(\forall s \in \boldsymbol{s}_{\text{all}})$.

1. Initialization:

$p_m = p_{\max}$, $\boldsymbol{n}' = \boldsymbol{n}$, $\boldsymbol{s}'_m = \boldsymbol{s}_m$, the initial power of arbitrary BS m $(m \in \boldsymbol{n}')$ is $p_m^s = 0$.

2. WF power allocation:

for (m=1; $m \in \boldsymbol{m}'$; m++)

{

BS executes water-filling in $\boldsymbol{s}'_m$ with p_m. Then, the power allocation value p'^s_m from BS m to the stream s after water-filling can be determined.

}

3. BD adjustment:

3.1) for (arbitrary BS m in $\boldsymbol{m}'$)

{

3.1.1) Let $x_m^s = \frac{p'^s_m}{F_m^s}$, $\forall s \in s'_m$. BS m sends x_m^s back to BSs in $\boldsymbol{m}_s$ $(s \in s'_m)$.

3.1.2) BS m calculates $x_0^s = \min_{m \in m_s} x_m^s$ $(\forall s \in \boldsymbol{s}'_m)$. Then BS m update the transmitting power as $p_m^s = p_m^s + F_m^s \times x_0^s$.

}

3.2) for (arbitrary BS m in $\boldsymbol{m}'$)

{

Update $p_m = p_{\max} - \sum_{s \in s'_m} p_m^s$.

if ($p_m = 0$) { for (arbitrary $s \in \boldsymbol{s}'_m$) $\{\boldsymbol{s}'_{m'} = \boldsymbol{s}'_{m'}/\{s\},\ \forall m' \in \boldsymbol{m}_s.\}\}$

}

3.3) for (arbitrary BS m in $\boldsymbol{m}'$)

{

If ($p_m = 0$, or $\boldsymbol{s}'_m$=null set) $\{\boldsymbol{m}' = \boldsymbol{m}'/\{m\}.\}$

}

3.4) if ($\boldsymbol{m}'$ is an empty set) the algorithm terminates. $Q_s = \sum_{m \in \boldsymbol{m}_s} p_m^s$, $\forall s \in \boldsymbol{s}_{\text{all}}$. Otherwise, the algorithm goes to Step 2.

Moreover, we can further derive the power updating function as

$$p_{i,j} = \left. \frac{\left(1 - e^{-\left(\frac{g_{ijk}}{\beta}\right)}\right) \times \Gamma(I_{ik} + \sigma^2) \times p_{i,j}^{\text{WF}}}{\Gamma(I_{ik} + \sigma^2) + e^{-\left(\frac{g_{ijk}}{\beta}\right)} \times p_{i,j}^{\text{WF}} \times g_{ijk}} \right|_{j \in \boldsymbol{m}_{ik}}. \tag{48}$$

One key issue in (48) is to determine the $p_{i,j}^{\text{WF}}$ for each user. In the scenario of single-antenna mode, $p_{i,j}^{\text{WF}}$ can be directly obtained by water-filling algorithm. However, in the scenario of multi-antenna scenario, streams based model is needed and the power proportion between BSs serving one stream should be met under BD pre-coding.

For the multi-antenna scenario, let $\boldsymbol{s}_m$ denote the list of streams served by BS m, let $\boldsymbol{m}_s$ denote the BS list serving the stream s. The power proportion of BS m serving the stream s is F_m^s. The algorithm is shown in Algorithm 5. In Step 1, the algorithm is initialized, the set of BS is $\boldsymbol{n}'$. The algorithm is executed in iteration until $\boldsymbol{n}'$ changes to an empty set in Step 3.4). When any stream in $\boldsymbol{s}'_m$ meets Step 3.2), $\boldsymbol{s}'_{\mathrm{m}}$ will be updated. The maximal initial power of BS m is $p_m = p_{\max}$. BS m $(\mathrm{m} \in \boldsymbol{n}\prime)$ serves for the stream s $\left(\mathrm{s} \in \mathrm{s}'_{\mathrm{m}}\right)$ with initial power $p_m^s = 0$. In Step 2, arbitrary BS m in $\boldsymbol{n}'$ executes water-filling algorithm to streams in $\boldsymbol{s}'_m$ with the overall power p_m. The virtual channel gain of stream s is $\tilde{H}_{s,s}$. After water-filling power allocation, BS m $(\mathrm{m} \in \boldsymbol{n}')$ serves for the stream s $\left(s \in s'_m\right)$ with p'^s_m. In Step 3.1.1), for an arbitrary s, BS m calculate the unit power x_m^s with p'^s_m. In Step 3.1.2), define the unit power of stream s is x_0^s, i.e. the minimal power in $\{x_m^s\}_{m \in \boldsymbol{m}_s}$. Based on x_0^s, BS m can adjust its power as $p_m^s = p_m^s + F_m^s \times x_0^s$, where $F_m^s \times x_0^s$ is the added power in this iteration. In Step 3.2), BS m in $\boldsymbol{n}'$ will calculate and update the overall power p_m. If $p_m = 0$, streams in $\boldsymbol{s}'_m$ will be deleted from $\boldsymbol{s}'_{m'}$ $(m' \in \boldsymbol{m}_s)$. In Step 3.3), arbitrary BS m with $p_m = 0$ will be deleted from $\boldsymbol{n}'$. Likewise, arbitrary BS m with empty $\boldsymbol{s}'_m$ will be deleted from $\boldsymbol{n}'$. In Step 3.4), if $\boldsymbol{n}'$ is an empty set, the algorithm terminates. Otherwise, the algorithm goes to Step 2 for the next iteration.

Notably, the Step 2 in Algorithm 5 can be embedded. When any BS m in $\boldsymbol{m}\prime$ implements equal power allocation (EPA) in $\boldsymbol{s}'_m$. The algorithm can become EPAFD (frequency diversity based EPA algorithm)

Definition 2 For power allocation vector $\boldsymbol{p} = \left[p_{i,1}, p_{i,2}, \ldots, p_{i,|\boldsymbol{n}|}\right]$ after game, if arbitrary power $p'_{i,j} \in \boldsymbol{p}_{i,j}^{\mathrm{PAG}}$ has $u_{i,j}^c\left(p_{i,j}, p_{i,-j}\right) \geq u_{i,j}^c\left(p'_{i,j}, p'_{i,-j}\right)$, then $\boldsymbol{p} = \left[p_{i,1}, p_{i,2}, \ldots, p_{i,|\boldsymbol{n}|}\right]$ is an Nash equilibrium (NE).

Theorem 1 The proposed Nash non-cooperative power game possesses at least one NE.

Proof According to Nash theorem, when the following two conditions hold, NE exists [19]: (1) $\boldsymbol{p}_{i,j}^{\mathrm{PAG}}$ is a non-empty closed and bounded convex set in the real Euclidean space $\boldsymbol{R}^{\boldsymbol{N}}$; (2) $u_{i,j}^c\left(p_{i,j}, p_{i,-j}\right)$ is a continuous and quasi-concave function in $\boldsymbol{p}_{i,j}^{\mathrm{PAG}}$. Obviously, the condition (1) holds for $\boldsymbol{p}_{i,j}^{\mathrm{PAG}} = \left[0, p_{i,j}^{\mathrm{WF}}\right]$. By $\frac{\partial^2 u_{i,j}^c\left(p_{i,j}, p_{i,-j}\right)}{\partial p_{i,j}}$, the condition (2) can be easily proved. Then, Theorem 1 holds. ■

Theorem 2 The NE of the proposed power game possesses unique NE.

Proof Reference [20] proposes a framework for power game, i.e. standard interference function. Let $p_{i,j}(I_{ik})$ denote the updated power $p_{i,j}$, if $p_{i,j}(I_{ik})$ meets:

(1) positive: $p_{i,j}(I_{ik}) > 0$; (2) monotonicity: if $I'_{ik} \geq I_{ik}$, then $p_{i,j}(I'_{ik}) \geq p_{i,j}(I_{ik})$; (3) scalable: for an arbitrary $\tau > 1$, $\tau p_{i,j}(I_{ik}) > p_{i,j}(\tau I_{ik})$ always holds, then, $p_{i,j}(I_{ik})$ is a standard interference function. It can be easily proved the proposed power game in (48) meet the three conditions, and thus it is a standard interference function. As in [20], there exists an unique NE in the standard interference function. Therefore, Theorem 2 holds. ■

Assume that users in each CBS report g_{ijk} and $I_{ik} + \sigma^2$ to the BS, in PAG, the algorithm can be executed in distributed manner. Algorithm 6 gives a simple searching algorithm to achieve the NE, the proof of the convergence can be referenced in [20].

For the multi-antenna scenario, the PAG algorithm can be easily extended from the single-antenna scenario and implemented under the streams based model.

2.4 Case Study

Algorithm 6 NE searching in an arbitrary BS j [7].

NE searching algorithm in BS *j*

Input:

f={1, …, $|f|$}；β；$p_{i,j}^{\mathrm{WF}}$；g_{ijk} $(\forall i, k)$；Error control constant ε.

Output:

$p_{i,j}$， $\forall\, i \in \boldsymbol{s}$。

1. Initialization:

令 $p_{i,j}^{\mathrm{WF}} \rightarrow p_{i,j}$， $\forall i$.

2. NE searching:

for (i=1; i≤$|\boldsymbol{f}|$; i++)

{

2.1) BS j transmits with $p_{i,j}$, and receives the $I_{ik} + \sigma^2$ from the users in its coverage.

2.2) BS j updates $p_{i,j}$ with (48), if the absolute difference value between power and the updated power is larger than ε. Go to Step 2.1), otherwise the algorithm terminates.

}

In Figs. 3 and 4, we give the simulation results of the algorithms under the fourth-generation of wireless networks. We employ the traditional wireless scenario as in [7]. 19 cells are assumed in the system. The fading model is $\mathcal{CN}(0, 8\ \mathrm{dB})$, the path loss model is $148 + 37.6 \log_{10} d$, where d is the distance between the BS and the user. The number of users per cell is 100, for simplication, the user with the minimal SINR in each cell is taken as the edge user. The cell radius is 500 m,

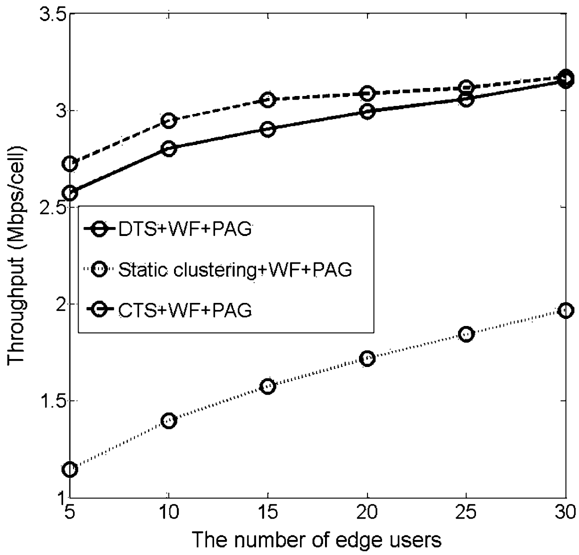


Fig. 3 The performance of CoMP in the single-antenna scenario

maximal BS transmitting power is 43 dBm for the single-antenna scenario and 46 dBm for the multi-antenna scenario, respectively. The wireless bandwidth is 20 MHz. Antenna power gain is 5 dB, the noise power spectral density is −174 dBm/Hz. The number of users per cell is 100 and the number of edge users is 10 per cell. Figure 3 gives the performance of the single-antenna based algorithms. For the single-antenna scenario, we assume that $m_{\max} = 3$, and $\mathrm{RSRP}_0 = 2.22 \times 10^{12}$ mW. As shown in Fig. 3, the introduced dynamic algorithms of CoMP indeed outperform the traditional static off-line algorithm, where the CBS is pre-determined. This indicates that the dynamic BS clustering can better adapt to the changing wireless environment due to the more feasible and real-time resource allocation in the space domain. In other simulations not reported here, we find that CTS based and DTS based algorithms possess nearly the same performance due to the heuristic philosophy behind them. For the multi-antenna scenario, we assume that $m_{\max} = 3$, $E_0 = 1.89 \times 10^{-13}$ J, and $D_0 = 1.14$. In Fig. 4, the performance of the introduced algorithms in the multi-antenna scenario is shown. The traditional wireless scenario in [18] is referenced for the multi-antenna model. We assume the number of receiving antennas per user is 2. As the number of transmitting antennas per BS increases, the performance largely increases in CoMP. This indicates the effectiveness of multi-antenna configuration effecting the performance of CoMP which verify the large potential of space resource for improving the wireless performance. Since the more flexible power allocation achieved by water-filling, the WFFD based algorithm outperforms the EPAFD based algorithm as in Fig. 4.

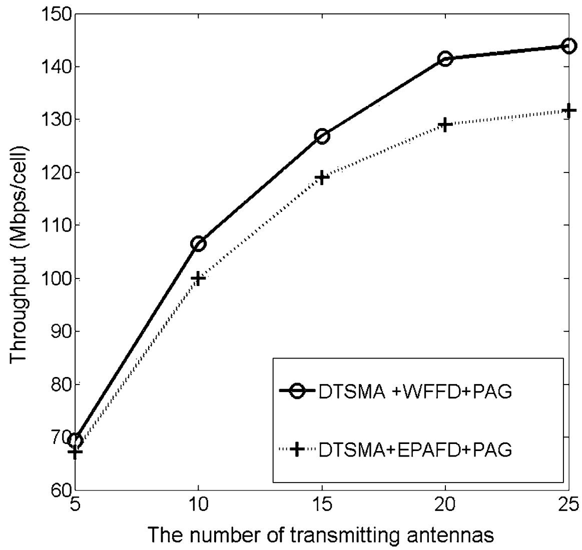


Fig. 4 The performance of CoMP in the multi-antenna scenario

3 Optimization in Wireless Cooperative Networks

3.1 Optimization Model

Geometric programming (GP) [21, 22] provides an effective method to achieve optimization in resource allocation. For simplication, in this section, we focus on the single-antenna scenario without frequency diversity. We introduce a novel power-fractionizing mechanism (PFM) [21] to execute the joint CBS, user scheduling, and power allocation in CoMP via GP.

In CoMP, for an arbitrary BS j, we sort the users in the cell with RSRP above RSRP_0 in ascending order. Assume the number of edge users is x, then the number 1 to x after sorting is dealt as the set of edge users, $\boldsymbol{u}_j$. All the sets of edge users of BSs are further combined as the set of edge users in the system, $\boldsymbol{u}^{\mathrm{e}}$. BSs serving for the users in $\boldsymbol{u}^{\mathrm{e}}$ constitutes a set $\boldsymbol{m}_u$. The signal to interference plus noise ratio (SINR) of user u, γ_u^{e}, is $\gamma_u^{\mathrm{e}} = \frac{\sum_{j \in \mathbf{m}_u} p_i g_j^u}{\sum_{j \notin \mathbf{m}_u} p_i g_j^u + \sigma^2}$. The throughput of user u is $r_u^{\mathrm{e}} = b \log_2\left(1 + \gamma_u^{\mathrm{e}}\right)$. Define a binary variable s_j^u, where $s_j^u = 1$ denotes user u is served by the BS j, and $s_j^u = 0$ denotes user u is not served by the BS j. Then, the system model can be formulated as

$$\text{maximize} \quad \sum_{u \in \boldsymbol{u}^{\mathrm{e}}} r_u^{\mathrm{e}}, \tag{49a}$$

$$s.t. \quad \boldsymbol{m}_u \cap \boldsymbol{m}_{u'} = \emptyset, \quad \forall u, u' \in \boldsymbol{u_i}, \tag{49b}$$

$$\sum_{u \notin \boldsymbol{u}_j} s_j^u = 0, \quad \forall j \in \boldsymbol{n}, \tag{49c}$$

$$0 \leq p_j \leq p_{\max}, \quad \forall j \in \boldsymbol{n}. \tag{49d}$$

Equation (49a) is to maximizing the system throughput, (49b) denotes that CBSs cannot work in overlapping manner, (49c) denotes that BSs only serve for users in $\boldsymbol{u}^{\mathrm{e}}$. (49d) is the limitation of the BS transmitting power.

3.2 Optimization via Geometric Programming

Power-fractionizing mechanism (PFM) is such a mechanism that each BS can serve for all the users with RSRP $\geq$ RSRP_0 in the system at the same frequency. Obviously, PFM is a logical other than a practical method. With such an assumption, the wireless networks can be described as a biconnected manner between BSs and users. Define the power of BS serving on e user as the power split block. To further transform the PFM based model into the practical model, let the product of arbitrary two power split block equivalent to a very small positive constant δ, then the transformed PFM based model can approximately equal to the original model.

For an arbitrary user u, define $\boldsymbol{J}_u$ as the set of BSs with RSRP $\geq$ RSRP_0 received by the user u. Let p_j^u be the power split block from BS $j \in \boldsymbol{J}_u$ to the user u. Define δ as a very small positive constant δ, T as an arbitrary variance, then the original model in (49a)–(49d) can be transformed as a special form of GP model, i.e. signomial programming (SP):

$$\text{minimize} \quad T \tag{50a}$$

$$s.t. \quad \prod_{u \in \boldsymbol{u}^{\mathrm{e}}} \frac{\sum_{u' \in \boldsymbol{u}^{\mathrm{e}}/u} \sum_{j \in \boldsymbol{J}_{u'}} p_j^{u'} g_j^u + \sigma^2}{\sum_{u' \in \boldsymbol{u}^{\mathrm{e}}} \sum_{j \in \boldsymbol{J}_{u'}} p_j^{u'} g_j^u + \sigma^2} \leq T \tag{50b}$$

$$p_j^u \times p_j^{u'} \leq \delta, \quad \forall j \in \boldsymbol{n}, \ \ u \neq u', \tag{50c}$$

$$\sum_{u \in \boldsymbol{U}^{\mathrm{e}}} p_j^u \leq p_{\max}, \quad \forall j \in \boldsymbol{n}, \ \forall u \in \boldsymbol{U}^{\mathrm{e}}, \tag{50d}$$

The "maximize" in (49a) has been transformed as "minimize" in (50a) to meet the standard form of SP model. (50b) leads to the objective of the model as maximizing system throughput. (50c) guarantees at most one user scheduled on each sub-channel. (50d) is the maximal limitation of BS transmitting power. Since in GP, all the variables are positive, the non-negative feature of power can be directly guaranteed.

Notably, the PFM method proposed in this section can be easily extended to the scenario of frequency diversity, where the condition (50c) should be changed as the product of arbitrary two power split blocks less than a very small positive constant.

For the SP model in (50a)–(50d), the denominator in (50b) can be approximated by a monomial as

$$\sum_{u' \in \boldsymbol{u}^{\mathrm{e}}} \sum_{j \in \boldsymbol{J}_{u'}} p_j^{u'} g_j^u + \sigma^2 \cong \left(\frac{\sigma^2}{\tilde{\lambda}_0^u}\right)^{\tilde{\lambda}_0^u} \times \prod_{u' \in \boldsymbol{u}^{\mathrm{e}}} \prod_{j \in \boldsymbol{J}_{u'}} \left(\frac{p_j^{u'} g_j^u}{\lambda_j^{uu'}}\right)^{\lambda_j^{uu'}}, \tag{51a}$$

$$\tilde{\lambda}_0^u = \frac{\sigma^2}{\sum_{u' \in \boldsymbol{u}^{\mathrm{e}}} \sum_{j \in \boldsymbol{J}_{u'}} p_j^{u'} g_j^u + \sigma^2}, \tag{51b}$$

$$\lambda_j^{uu'} = \frac{p_j^{u'} g_j^u}{\sum_{u' \in \boldsymbol{u}^{\mathrm{e}}} \sum_{j \in \boldsymbol{J}_{u'}} p_j^{u'} g_j^u + \sigma^2}. \tag{51c}$$

Algorithm 7 Solution of the SP model in CoMP [21]

The joint transmission scheduling and power allocation in CoMP

Input:

$\boldsymbol{n}=\{1, \ldots, |\boldsymbol{n}|\}$, $\boldsymbol{u}^{\mathrm{e}}$, $\{\boldsymbol{J}_u\}$, σ^2, δ, φ, $p_{\max}$, and the error precision control constant ε.

Output:

The matrix of power allocation $\bar{\boldsymbol{P}} = \{\bar{p}_j^u\}$.

Step1:

Let $0 \to k$, initialize $\bar{\boldsymbol{P}}$ and $\boldsymbol{P}^*$ as zeros matrices.

Let $\boldsymbol{p}^k = \{p_j^u = \varphi \mid \forall u \in \boldsymbol{U}^{\mathrm{e}}, \ \forall j \in \boldsymbol{J}_u\}$.

Step 2:

Based on $\boldsymbol{p}^k$, calculate $\tilde{\lambda}_0^u$ and $\lambda_j^{uu'}$ in (51b) and (51c), respectively. By (51a), we can obtain the approximate GP model of the SP model.

Step3:

Using the standard interior point method to solve the GP model in Step 2, then, we can get the global optimal solution of the GP model according to the GP theorem as $\boldsymbol{p}^* = \{p_j^{u*}\}$.

Step 4:

If $\|\boldsymbol{p}^* - \boldsymbol{p}^k\| \le \varepsilon$, goes to 5. Otherwise, let $k+1 \to k$, $\boldsymbol{p}^* \to \boldsymbol{p}^k$, goes to Step 2.

Step 5:

For $\forall j \in \boldsymbol{n}$, let $s = \arg\max_u (p_j^{u*})$, and $p_j^{s*} \to \bar{p}_j^{s}$.

Such an approximation can transform the SP model into a GP model. By the GP theorem, the SP model can be transformed into a series of GP models meeting KKT (Karush-Kuhn-Tucker) conditions [21] and converged to at the local optimization of the original model in (49a)–(49d).

The specific solution of the SP model can be referenced in Algorithm 7. In Step 1, the power allocation is initialized as $\boldsymbol{p}^k = \left\{p_j^u\right\}$, then, $\tilde{\lambda}_0^u$ and $\lambda_j^{uu'}$ will be calculated by (51b) and (51c) respectively. In Step 2, the SP model can be transformed into a GP model by approximating the denominator in (51a) with a monomial. In Step 3, the GP is solved and $\boldsymbol{p}^k$ in Step 4 will be updated. If $\boldsymbol{p}^k$ between the two iterations has an absolute difference larger than ε, the algorithm goes to Step 2 for the next iteration. If $\left\|\boldsymbol{p}^* - \boldsymbol{p}^k\right\| \leq \varepsilon$ is met, where $\boldsymbol{p}^*$ is the power allocation in the last iteration, the algorithm terminates. In Step 5, the power split block with the maximal value for each BS will be selected and one transmission relationship will be determined. The other power split block on the sub-channel will be nulled. According to the GP theorem, SP model can converge to at least the local optimal solution of the original model, and empirically, this solution is also the global solution of the original model.

3.3 Case Study

In Fig. 5, we employ the model of the traditional wireless scenario for simulation as in [21]. 7 cells are assumed in the system. The fading model is $\mathcal{CN}(0, 5\ \text{dB})$, the path loss model is $148 + 37.6\log_{10} d$, where d is the distance between the BS and the user. the number of users per cell is 100, for simplication, the user with the minimal SINR in each cell is taken as the edge user. The cell radius is 500 m, maximal BS transmitting power is 43 dBm, and the wireless bandwidth is 3 MHz. Antenna power gain is 5 dB, the noise power spectral density is −174 dBm/Hz. For comparison, $m_{\max} = 3$ for the DTS algorithm. As shown in Fig. 5, the joint optimization by GP indeed outperforms the two-step based heuristic method. As RSRP_0 increases, the performance in CoMP decreases with the joint optimization method. This is because that the larger RSRP_0 limits the flexibility in the optimization philosophy based algorithm.

Another observation in Fig. 5 is that as the value of RSRP_0 further increases, the total edge user throughput achieved by the joint optimization decreases monotonically, and the gap to the mechanism of DTS + WF + game gradually narrows. This is because a larger RSRP_0 imposes a more stringent constraint on the candidate user set with $\text{RSRP} \geq \text{RSRP}_0$, as well as the flexibility of dynamic CBS clustering.

As discussed above, δ can guarantee at most one user scheduled by one BS. Since the power split block with the maximal value will be selected, the power pumped into the other power split blocks will be nulled and wasted. When δ is small enough, the remaining power on the other power split blocks is very small, and such an approximation does not lead to deteriorative performance in CoMP. However, as δ enlarges, more power may effuse to the other power split blocks. However, the optimal solution cannot cater for the original model with such a power nulling. Therefore, in practice, δ should be chosen as small as possible. On the other hand, when δ is very small, the precision of the computing for optimization will be also another problem in engineering. How to achieve such a tradeoff between system performance and precision will be significant for the application of PFM mechanism. Moreover, when δ is ignored in such an architecture, the model can be dealt as non-othogonal multiple access (NOMA) system. More in-depth discussion in such a scenario will be left for the future work.

The complexity of Algorithm 7 meet that of SP model. According to GP theorem, GP possesses linear complexity with the number of variables, yet SP possesses geometry complexity with the number of variables. Therefore, the

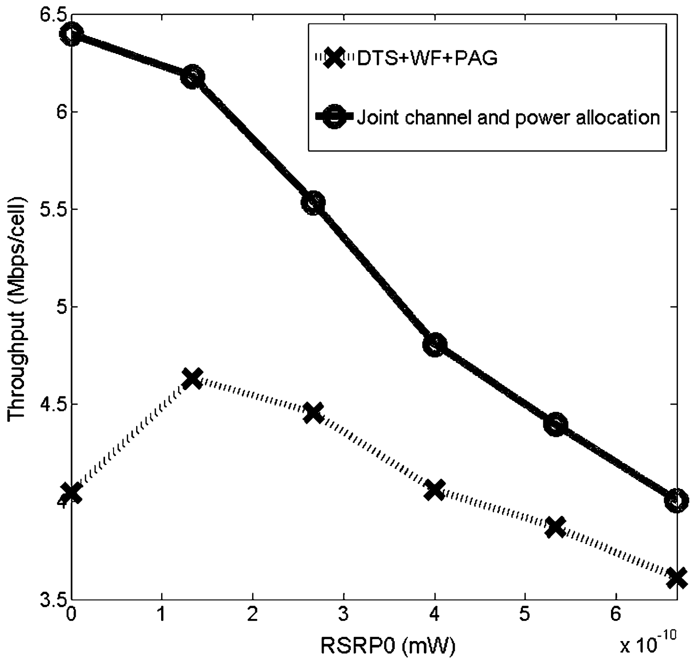


Fig. 5 The impact of $RSRP_0$ on the edge user throughput in CoMP

complexity of Algorithm 7 increases largely when the system-scale is enlarged, which makes the SP model stays in theory research stage, especially in the ultra-dense wireless cells based next generation of wireless networks. However, as the capacity of computing in the networks enhanced, such an algorithm with global viewpoint will possess better environment-adaptation and thus promising in the future.

4 Conclusion

The next generation of wireless networks employs CoMP technology to achieve a more flexible resource allocation with enlarged space resource. To design a distributed and environment-adaptive CoMP, in this chapter, we have first introduced a series of algorithms to achieve such a goal with better performance than the existing ones. Especially, with the stream analysis model, the single-antenna based algorithms can be easily extended to the multi-antenna scenario. In this chapter, all the algorithms are based on two-steps heuristic method, which limits the performance and the adaptation of algorithms. Although the two-step heuristic algorithm for resource allocation in CoMP can improve the performance largely, the heuristic mechanism makes the performance of the algorithm lacks stability with the changed wireless environment. In this chapter, we have also further introduced the joint optimization mechanism of transmission scheduling and power allocation in CoMP via geometric programming. We have proved that the joint optimization can achieve a better performance than the heuristic method. Some remaining problems, such as complexity and the limitation of single-antenna scenario for the algorithm will be left for the future research.

References

1. Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.): Green IT Engineering: Concepts, Models, Complex Systems Architectures. Studies in Systems, Decision and Control, vol. 74. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-44162-7
2. Kharchenko, V., Illiashenko, O.: Concepts of Green IT Engineering: taxonomy, principles and implementation. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Concepts, Models, Complex Systems Architectures. Studies in Systems, Decision and Control, vol. 74, pp. 3–19. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-44162-7_1
3. Kondratenko, Y., Korobko, O.V., Kozlov, O.V.: PLC-based systems for data acquisition and supervisory control of environment-friendly energy-saving technologies. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Concepts, Models, Complex

Systems Architectures. Studies in Systems, Decision and Control, vol. 74, pp. 247–267. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-44162-7_13
4. Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.): Green IT Engineering: Components, Networks and Systems Implementation. Studies in Systems, Decision and Control, Vol. 105. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-55595-9
5. Kondratenko, Y., Korobko, V., Korobko, O., Kondratenko, G., Kozlov, O.: Green-IT approach to design and optimization of thermoacoustic waste heat utilization plant based on soft computing. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Components, Networks and Systems Implementation. Studies in Systems, Decision and Control, vol. 105, pp. 287–311. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-55595-9_14
6. Kuchuk, G., Kovalenko, A., Kharchenko, V., Shamraev, A.: Resource-oriented approaches to implementation of traffic control technologies in safety-critical I&C systems. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Components, Networks and Systems Implementation. Studies in Systems, Decision and Control, vol. 105, pp. 313–337. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-55595-9_15
7. Fu, S., Wu, B., Wen, H., Ho, P.-H., Feng, G.: Transmission scheduling and game theoretical power allocation for interference coordination in CoMP. IEEE Trans. Wirel. Commun. **13**(1), 112–123 (2014)
8. Han, Z., Ji, Z., Liu, K.J.R.: Non-cooperative resource competition game by virtual referee in multicell OFDMA networks. IEEE J. Sel. Areas Commun. **25**(6), 1079–1090 (2007)
9. Xiao, C., Zheng, Y.R., Ding, Z.: Globally optimal linear precoders for finite alphabet signals over complex vector Gaussian channels. IEEE Trans. Signal Process. **59**(7), 3301–3314 (2011)
10. Zhang, J., Chen, R., Andrews, J.G., et al.: Networked MIMO with clustered linear precoding. IEEE Trans. Wirel. Commun. **8**(4), 1910–1921 (2009)
11. Shifrin, T.: Multivariable Mathematics: Linear Algebra, Multiva Riable Calculus, and Manifolds, pp. 1–492. Wiley (2004)
12. Nguyen, D.H.N., Le, H.N., Ngoc, T.L.: Block-diagonalization precoding in a multiuser multicell MIMO system: competition and coordination. IEEE Trans. Wirel. Commun. **13**(2), 968–981 (2014)
13. Zhou, B., Bai, B., Li, Y., et al.: Chordal distance-based user selection algorithm for the multiuser MIMO downlink with perfect or partial CSIT. In: Proceedings of IEEE AINA, pp. 77–82 (2011)
14. Ko, K., Lee, J.: Multiuser MIMO user selection based on chordal distance. IEEE Trans. Commun. **60**(3), 649–654 (2012)
15. Fu, S., Zhou, H., Qiao, J., et. al.: Distributed transmission scheduling and power allocation in CoMP. IEEE Syst. J. (2017)
16. Bandeira, C.I.R., Moreira, D.C., Normando, P.G., et al.: Performance of joint antenna and user selection schemes for interference alignment in MIMO interference channels. In: Proceedings of IEEE VTC, pp. 1–5 (2013)
17. Shen, Z., Chen, R., Andrews, J., et al.: Low complexity user selection algorithms for multiuser MIMO systems with block diagonalization. IEEE Trans. Signal Process. **54**(9), 3658–3663 (2006)
18. Fu, S., Wen, H., Wu, J., Wu, B.: Energy-efficient precoded coordinated multi-point transmission with pricing power game mechanism. IEEE Syst. J. **11**(2), 578–587 (2017)
19. Tanimoto, J.: Fundamentals of Evolutionary Game Theory and Its Applications. Springer, pp. 1–214 (2015)
20. Yates, R.: A framework for uplink power control in cellular radio system. IEEE J. Sel. Areas Commun. **13**(7), 1341–1347 (1995)

21. Fu, S., et. al.: Power-fractionizing mechanism: achieving joint user scheduling and power allocation via geometric programming. IEEE Trans. Veh. Technol. (2015)
22. Chiang, M., Tan, C.W., Palomar, D.P., et al.: Power control by geometric programming. IEEE Trans. Wirel. Commun. **6**(7), 2640–2651 (2007)

Checkable FPGA Design: Energy Consumption, Throughput and Trustworthiness

Alex Drozd, Svetlana Antoshchuk, Julia Drozd, Konstantin Zashcholkin, Miroslav Drozd, Nikolay Kuznietsov, Mohammed Al-Dhabi and Valery Nikul

Abstract Green FPGA design represented in the directions of energy efficiency and safety which are tightly connected in the areas of critical application is considered. The array structures that are traditionally used in digital components of safety-related systems, reduce a checkability of circuits, creating a problem of the hidden faults which can be accumulated in a normal mode and reduce the fault tolerance of the circuit and safety of system in the emergency mode. Soft and cardinal ways of array structure reduction are offered. The soft way consists in development of the truncated arithmetical operations implementing into reduced array structures. The cardinal way consists in paralleling of calculations in serial codes with the use of bitwise pipelines. The comparative analysis in complexity, throughput and energy consumption of the iterative array and bitwise pipeline multiplier is executed experimentally with use of Altera Quartus II. Methods of on-line testing in checking of mantissas by inequalities are developed for the

A. Drozd (✉) · S. Antoshchuk · J. Drozd · K. Zashcholkin · M. Drozd · N. Kuznietsov · M. Al-Dhabi · V. Nikul
Odessa National Polytechnic University, 1, Shevchenko Avenue, Odessa, Ukraine
e-mail: drozd@ukr.net

S. Antoshchuk
e-mail: asgonpu@gmail.com

J. Drozd
e-mail: dea_lucis@ukr.net

K. Zashcholkin
e-mail: const-z@te.net.ua

M. Drozd
e-mail: miroslav_dr@mail.ru

N. Kuznietsov
e-mail: koliaodessa@mail.ru

M. Al-Dhabi
e-mail: aldhabi@mail.ru

V. Nikul
e-mail: leraniku@gmail.com

© Springer Nature Switzerland AG 2019

V. Kharchenko et al. (eds.), *Green IT Engineering: Social, Business and Industrial Applications*, Studies in Systems, Decision and Control 171,
https://doi.org/10.1007/978-3-030-00253-4_4

truncated operations. A method of increase in safety of FPGA circuits in opposition to accumulation of the hidden faults and a method of monitoring in integrity of FPGA project are suggested on the basis the program code diversity use.

Keywords Green FPGA design · Energy efficiency · Safety-related system
Digital component · Checkability of circuit · Hidden faults · Array structure
Truncated operation · Bitwise pipeline · On-line testing
Diversity of a program code · Integrity monitoring

1 Introduction

1.1 Motivation

FPGA (Field Programmable Gate Array) design is an important area in Industrial Application of Green IT-Engineering, which is positioned in two main directions: energy efficiency and safety [1, 2]. Energy efficiency characterizes a degree energy quantum use [3, 4], and safety determines the direction in which energy would be expended: for the creation or the corruption [5, 6]. These directions are tightly connected in the areas of critical application. The objects of the increased risk which provide production, transmission and consumption of the great amount of electric power, have already become a natural human environment. These objects are becoming more complex and more powerful. Their number is growing. The equilibrating task of the increasing risks is defined for Green IT, which is implemented in instrumentation and control safety-related systems aimed at providing of the functional safety of both the system and the object for accident prevention and of their consequences reducing [7]. The cascade power networks disconnections and accidents at the power stations lead to the considerable electric power losses as a result of stoppings in its delivery and production. The important feature of the modern stage of Green IT development is creation of safety-related systems on FPGA [8]. This fact represents the high evaluation of FPGA opportunities design.

FPGA design supports all types of circuit parallelism: matrix, pipeline and preparation of results which are selected with the use of the resource-based approach at the levels of resource development: replications, diversifications and autonomy, accordingly [9]. However, the matrix parallelism, reflecting the bottom level of resource development is the dominant one [10, 11]. It is represented not only in the matrix organization of chips, but, first of all, in orientation of design to processing of numbers in parallel codes.

The matrix parallelism demonstrates a lot of drawbacks inherent in the bottom level of resources development. Array structures occupy big spaces, but do not provide the appropriate temporal parallelism. The iterative array multiplier of n-bit binary codes executes the operation for one clock period and for this purpose contains n^2 operational elements, $2n - 2$ of which are connected consistently.

Therefore, each operational element is used in a clock period for a short period of time with the coefficient of $K_M = (2n - 2)^{-1}$ [12]. For $n = 32$ and $n = 64$, operational elements are used only for 1.6% and for 0.8% of time, accordingly. The iterative array divider contains $(n + 1)^2$ sequentially connected operational elements. For each of them it defines utilization coefficient $K_D = (n + 1)^{-2,}$ i.e. for $n = 32$ and $n = 64$ operational elements are used for 0.1 and 0.02%, accordingly. The rest of the time in a clock cycle it is used for the distribution of waves of parasitic transitions, the amount of which repeatedly exceeds the number of the functional ones. This glitch problem leads to the significant increase in a dynamic component of energy consumption [13, 14]. Input, processing and output of data in parallel codes causes a large number of connections in array circuits, which makes their topology more complicated, reduces technological effectiveness in production and reliability in maintenance.

Therefore, we suggest to develop Green IT on the basis of array structures reduction, suggesting two ways: The Soft Way and Cardinal one.

The Soft Way consists in execution of the truncated arithmetical operations which almost twice simplify array circuits and reduce time of computation with maintaining single precision when the result inherits the size of an operand [15]. Therefore, the truncated operations executed with mantissas are the main method of processing the approximate data in floating-point formats [16]. The resource approach represents a tendency of the amplifying domination of the approximate calculations over the exact data processing [9, 17]. In the safety-related systems, the initial data arriving from sensors is a result of measurements, i.e. also approximate data.

The Cardinal Way presupposes the transition to the following level of the resources development: from replication of array structures to diversification on the basis of pipeline enhancement. Modern computer systems and their digital components, as a rule, are built like a pipeline which reflects the diversification level. However, pipeline sections are single-cycle nodes with array structures and their problems. We suggest to reduce array structures of the pipeline sections. The maximum reduction of array structure to one operational element will transform the system into the bitwise pipeline for data processing in consecutive codes. To what extent such progressive decision can be competitive in the world of array structures?

Decades of matrix parallelism domination promoted the considerable achievements in its enhancement that substantially strengthened the position of matrix decisions. FPGA design uses these achievements effectively. Preparation of results is used for an increase in array structures efficiency: circuits of the accelerated distribution of add carries of parallel codes are prepared; iterative array multipliers are prepared with high indexes in throughput and energy saving; libraries suggest the wide range of nodes with array structure [18].

Can a cardinal solution compete with array structures in these conditions?

The affirmative answer is based on one indisputable advantage of bitwise pipelines in comparison with array structures. This advantage consists in the high checkability which is very important for the safety-related systems.

These systems are designed for operation in two modes: normal mode and emergency one. The functional safety is ensured by the use of fault-tolerant decisions which together with dual-mode operation of systems lead to the considerable structural redundancy of digital circuits and, as a result, to a low checkability of their array structures. It creates a problem of the hidden faults which in a normal mode are not detected as there are no input data necessary for this purpose. The continuous normal mode adds to the accumulation of the hidden faults. In the most responsible emergency mode such faults are represented, they reduce fault tolerance of components and the functional safety of systems. Points of a circuit in which this problem is represented belong to potentially hazardous ones [19, 20].

The Problem of the Hidden Faults is known by unsuccessful attempts to solve it with the use of the imitation modes. The accident simulation executed in a case of a switch-off emergency protection repeatedly led to the alert conditions. The Chernobyl catastrophe took place in the case of the turned-off emergency protection.

Register structures of bitwise pipelines are the elements of testable design (scan registers). The serial codes propagated in circuit by scan registers that do not leave place for hidden faults. Therefore, bitwise pipelines deserve an attention directed to the receiving of their estimates in comparison with array structures and development of models, methods and means for their enhancement.

FPGA design represents additional opportunities of an increase in the functional safety of safety-related systems on the basis of reconfigurable components creation, thanks to high version redundancy of circuits in LUT-oriented architecture (LUT—Look-Up Table) [21, 22]. The version redundancy allows to select the most effective configuration of a component at a stage of its design, and also in the process of FPGA project program code forming [23].

1.2 Related Works

The known works related to the problems of green FPGA design can be subdivided into three groups which show:

- The possibilities of FPGA design that allow to carry out comparative assessment of matrix circuits and bitwise pipelines;
- The development of on-line testing methods for circuits with the reduced array structures;
- The use of version redundancy of circuits for increase in safety of the digital components implemented in FPGA projects.

The possibilities of FPGA design are considered on the example of CAD Altera [24]. For the comparative analysis of matrix and pipeline circuits in energy consumption, utilities of Power Play Early Power Estimators Spreadsheets and Power Play Power Analyzer described in [25, 26] are used. An assessment of temporal parameters is executed by temporal simulation of circuits with the use of Timing Analyzer utility [27].

The review of the on-line testing methods which increase their trustworthiness in the approximate results checking allows to select the residue checking methods and methods checking results by inequalities [28]. The important feature of approximate data consists in their non-uniform structure which contains most and least significant bits [29]. Faults in circuits cause the errors in these bits, they are accordingly essential and inessential for the result trustworthiness.

The residue checking methods developed for the truncated operations executed in iterative array multipliers and dividers of mantissas are effective in case of high probability of an essential error [30]. The operations of a denormalization of operands and normalization of results considerably reduce this probability for results of all prior and all following operations with mantissas, accordingly. In these conditions residue checking methods reduce the trustworthiness as a result of ignoring of approximate data structure as they do not distinguish between an error in most and least significant bits [31].

The methods of checking results by inequalities show high trustworthiness, distinguishing between essential and inessential errors and they do not depend on a method of obtaining result. It allows to use them directly for checking of the truncated operations. However, these methods are not developed for the main formats regulated by the standards IEEE Standard 754-1985 and IEEE Standard 754™-2008 for floating-point arithmetic [32, 33].

The version redundancy inherent to FPGA is represented in an opportunity to have a set of versions of a program code for the same hardware implementation of the circuit decision in the LUT-oriented architecture. The versions are formed in the case of the data transfer from one LUT to another in a direct or inverse value. Use of such version redundancy became a basis for an increase in the functional safety of FPGA projects by an increase in a checkability of the circuit in a normal mode and trustworthiness of results in the emergency one [23].

The version redundancy of the program code allows to execute monitoring of FPGA project integrity, too. The integrity of the information object (IO) is a part of security which can be considered as a necessary condition in the functional safety support.

The simplest scheme for the monitoring integrity determines a digest, which presupposes a standard hash sum with hash function usage, for IO [34]. The IO integrity is considered to be confirmed if a calculated hash sum coincides with a standard one. The disadvantages of this scheme are the following: (1) A standard hash sum is exposed; (2) The hash sum is physically separated from IO; (3) The fact of IO integrity is being controlled cannot be actually concealed from the outside surveillance. All these factors make it possible for the integrity violation to be concealed by substituting the hash sum.

The drawbacks 1 and 2 can be changed by encrypting of the hash sum and by attaching this procedure result to the initial IO [35]. To monitor the integrity, the controlling information H is separated from verified IO A. As a result, the information object acquires the state A_0, which it had before attaching the controlling part. The hash sum for object A_0 is compared to the hash sum H earlier subjected to the decryption. The decision of the object integrity is made according to the results of this comparison. The disadvantages of this decision are the following:

(4) Complexity or even impossibility of the direct insertion of the encrypted hash sum into the active IO body; (5) Disclosure of the fact of the controlling information existence which gives the possibility of breaking the encrypted data (brute force attack, complex attack, social engineering).

The approach allows to attach the controlling information to the information object not with the help of mere concatenation, but is known in the form of digital watermark (DWM). This method is able to conceal the fact of controlling information existence in the object. Methods and approaches of this type [36, 37] are thoroughly studied for passive IOs. The experience of the passive IO integrity provision can be used in developing of active IO integrity control methods and exactly the program code of FPGA.

1.3 Goal and Structure

The purpose of the researches is to develop Green Technologies on the basis of array structures reduction and the use of version redundancy of FPGA projects for an increase in the safety of the digital components in the areas of critical application.

Section 2 represents the comparative analysis of matrix circuits and bitwise pipelines in complexity, throughput and energy consumption for FPGA projects. The received estimates allow to define ways of increasing the bitwise pipelines efficiency. The method of throughput increase and lowering in energy consumption for the bitwise pipeline multiplier by processing two bits of operands in a clock cycle is offered.

Section 3 suggest on-line testing methods for digital components safety increase in normal and emergency modes. These methods are oriented on checking of the truncated arithmetical operations, implemented in the reduced array structures with the raised checkability which is important for normal mode. Methods execute checking of mantissas by inequalities which increase trustworthiness of on-line testing in emergency mode by distinguishing of essential and inessential errors. Methods are based on features of standards IEEE Standard 754-1985 and IEEE Standard 754™-2008.

Section 4 represents version redundancy of the program code in the FPGA project for a solution of problem of the hidden faults.

Section 5 describes the method of integrity monitoring of the FPGA project on the basis of version redundancy of its program code.

2 Bitwise Pipelines Versus Array Structures

2.1 Comparative Analysis of Matrix Circuits and Bitwise Pipelines

The high checkability of bitwise pipelines which cardinally solves a problem of the hidden faults, raises questions about other factors of efficiency in comparison with traditionally used array structures. To the following factors belong the complexity, throughput and energy consumption. The question of the relative efficiency of bitwise pipelines is raised in connection with unequal conditions of array and bitwise structures implementation in FPGA projects. The modern CAD (Computer-Aided Design) first of all is oriented on the array structures support.

Comparison between an array structure and a bitwise pipeline is executed experimentally on the example of multiplication operation with the use of a CAD Altera Quartus II v13 64 bit and FPGA Altera of the Cyclone II EP2C35F672C6 family [38]. Throughput is evaluated in the temporal parameters received by means of the utility TimeQuest Timing Analyzer [27], and energy consumption is calculated depending on the power determined by the utility of Power Play [25, 26]. Throughput and energy consumption are compared in a case of identical circuits complexity: one iterative array multiplier is compared to several pipelines of the same summary complexity. The complexity is estimated by the number of the LE (Logic Elements).

Table 1 represents the results of experiments for the size of operands $n = 8$ and $n = 16$.

Results of experiments show advantage of bitwise pipelines in throughput of 1.19 and 2.67 times and loss in energy consumption in 4 and 3 times (the relative efficiency —0.26 and 0.33) for the size of operands $n = 8$ and $n = 16$, accordingly. The relative efficiency of bitwise pipelines is increased with the growth of the operands size.

Table 1 Results of comparison between iterative array and bitwise pipeline multipliers

Index	*n*	Array structure	Bitwise pipeline	Effectiveness
Complexity, amount of LE	8	179	159 (5 pipelines)	–
	16	739	626 (10 pipelines)	–
Delay of clock cycle, ns	8	8.069	2.38	–
	16	17.238	2.38	–
Throughput, amount of operations/s	8	124×10^6	148×10^6	1.19
	16	58×10^6	155×10^6	2.67
Power, mW	8	6.37	26.11	–
	16	11.24	77.51	–
Energy consumption, mW ns	8	51.4	198.9	0.26
	16	193.8	590.3	0.33

Specific energy consumption, i.e. in the terms of one 8-bit multiplication, for iterative array multipliers is defined as $K_M = PT/N$, where P—power consumption, T_M—runtime of one operation in the iterative array multiplier, $N = (n/8)^2$—the equivalent number of 8-bit operations [39].

Specific energy consumption for bitwise pipeline multipliers is defined as

$$K_P = P_P T_P/(NC),$$

where

P_P	the power consumed by C pipelines which in the amount have complexity of the iterative array multiplier;
$P_P = PT/\tau$, τ	pipeline clock cycle;
$C = C_M/C_P$, C_M and C_P	complexity of array and bitwise pipeline multipliers;
$T_P = Z\tau$	runtime of one operation by the pipeline;
Z	amount of clock cycles in pipeline operation, $Z = 2n\tau$.

Then $K_P = (PT/\tau)\ (Z\tau)/(NC) = PT/NZ/C) = K_M Z/C$ and this equality allows to evaluate relative efficiency $K_{P/M} = K_M/K_P$ of the bitwise pipeline in energy consumption as $K_{P/M} = C/Z$. Thus, the relative energy efficiency $K_{P/M}$ of the pipeline multiplier depends only on relative complexity of array structure and the amount of clock cycles in a pipeline operation.

2.2 A Method of Increase in Energy Efficiency of the Bitwise Pipeline Multiplier

Based on the received assessment, we suggest a method of energy efficiency increase of the bitwise pipeline multiplier by reduction of Z, processing in each clock cycle of two operands bits. Multiplier calculates 4 groups of conjunctions: 1.1–1.4, 2.1–2.4, 3.1–3.4 and 4.1–4.4 shown in Table 2 for $n = 8$.

Conjunctions are designated by the codes that consist of the operands numbers. Conjunctions 3.1–3.4 of the third group after computation are delayed on one clock cycle as it is represented in Table 2.

We compare bitwise pipeline multipliers 1 and 2, processing 1 and 2 bits in clock cycle, accordingly. Table 3 represents the results of the experiments for the size of operands $n = 8$ and $n = 16$.

Results of experiments show advantage of bitwise pipeline processing 2 bits in clock cycle by suggested method in throughput in 1.24 and 1.02 times and in energy consumption in 1.60 and 1.64 times for the size of operands $n = 8$ and $n = 16$, accordingly.

Table 2 Computation of conjunctions on clock cycles and weights of product bits

Clock cycle	1		2		3		4		5		6		7		8	
Weights of bits	2^0	2^1	2^2	2^3	2^4	2^5	2^6	2^7	2^8	2^9	2^{10}	2^{11}	2^{12}	2^{13}	2^{14}	2^{15}
1.1	11		13		15		17									
1.2			31		33		35		37							
1.3					51		53		55		57					
1.4							71		73		75		77			
2.1		21		23		25		27								
2.2				41		43		45		47						
2.3						61		63		65		67				
2.4								81		83		85		87		
3.1			22		24		26		28							
3.2					42		44		46		48					
3.3							62		64		66		68			
3.4									82		84		86		88	
4.1		12		14		16		18								
4.2				32		34		36		38						
4.3						52		54		56		58				
4.4								72		74		76		78		

Table 3 Results of comparing the pipeline multipliers 1 and 2

Index	*n*	Bitwise pipeline 1	Bitwise pipeline 2	Effectiveness
Complexity, amount of LE	8	159 (5 pipelines)	182 (4 pipelines)	–
	16	626 (10 pipelines)	625 (6 pipelines)	–
Delay of clock cycle, ns	8	2.38	2.38	–
	16	2.38	2.38	–
Throughput, amount of operations/s	8	148×10^6	184×10^6	1.24
	16	155×10^6	158×10^6	1.02
Power, mW	8	26.11	26.06	–
	16	77.51	56.80	–
Energy consumption, mW ns	8	198.9	124.0	1.60
	16	590.3	360.5	1.64

3 On-Line Testing Methods for Digital Components Safety Increase

3.1 Requirements to On-Line Testing of Safety-Related Systems

On-line testing of digital components of safety-related systems inherits their features which determine requirements to checking of result trustworthiness. Diversification of an operating mode of safety-related systems on normal and emergency modes leads to diversification of these requirements determining a directivity of on-line testing in these modes. In a normal mode the safety-related system is waiting for an emergency mode and is directed on maintenance of a high level of readiness to anticipate accident and to reduce its consequences. Therefore, on-line testing performs functions of the testing that are not absolutely peculiar to it and aimed at fault detection. However, testing is executed with the use of tests. On-line testing is using uncontrollable input sequences of operating words and completely depends in the trustworthiness on these words in their ability to become tests.

For on-line testing the checkability of circuits becomes the functional, i.e. represents dependence on input data, and for safety-related systems—dual-mode [19, 20].

At the same time, the high checkability of circuits leads to more frequent manifestation of the errors caused by the transient faults in the emergency mode. These errors, as a rule, are inessential for trustworthiness of the approximate results peculiar to safety-related systems, however they reduce trustworthiness of on-line testing, distracting its resources in emergency mode on their parrying. The three-channel majority system will correctly continue to function in a case of a permanent fault of one channel, but transient faults against the background of this permanent fault will lead to incorrect results.

The problem of different requirements to on-line testing in the normal and emergency mode needs to be solved in a complex, raising a checkability of circuits and weakening influence of inessential errors. Such decision shall be based on the development of on-line testing methods which at the same time meet two conditions: they allow to service circuits with the reduced array structures and distinguish essential and inessential errors in approximate results.

Methods checking mantissas by inequalities completely meet these requirements. They are oriented on checking in trustworthiness of n-bit results of homogeneous arithmetical operations with n-bit operands and do not depend on a method of obtaining result which can be calculated by the circuits with the reduced array structure. At the same time, these methods evaluate the size of an error and detect essential errors more likely in comparison with inessential errors.

3.2 Checking Mantissas by Inequalities in Truncated Operations

The method of mantissas checking by inequalities determines the lower and upper boundaries of a result calculated on operands and compares the result with these boundaries. The result is considered to be non-authentic if it exceeds an upper bound or if it appears below the lower bound [28].

The boundaries are created of the restrictions which are superimposed on mantissas of the normalized floating-point numbers in the formats regulated by IEEE Standard 754-1985 and IEEE Standard 754™-2008 [32, 33].

For mantissas multiplication operation $A \cdot B = V$, the operands are restricted to inequalities $1 \leq A < 2$ and $1 \leq B < 2$, which are integrated in the following conditions: $(A - 2)(B - 2) > 0$ and $(A - 1)(B - 1) \geq 0$, defining lower bounds of the product: $V_{\mathrm{LM.1}} = 2(A + B) - 2)$, $V_{\mathrm{LM.2}} = A + B - 1$ and $V_{\mathrm{LM.1\text{-}2}} = \max(V_{\mathrm{LM.1}}, V_{\mathrm{LM.2}})$, where $V_{\mathrm{LM.1}} < V$, $V_{\mathrm{LM.2}} \leq V$ and $V_{\mathrm{LM.1\text{-}2}} \leq V$. Similarly, conditions $(A - 1)(B - 2) \leq 0$ and $(A - 2)(B - 1) \leq 0$ define upper bounds of the product: $V_{\mathrm{HM.1}} = 2A + B - 2$, $V_{\mathrm{HM.2}} = A + 2B - 2$ and $V_{\mathrm{HM.1\text{-}2}} = \min(V_{\mathrm{HM.1}}, V_{\mathrm{HM.2}})$, where $V_{\mathrm{HM.1}} \leq V$, $V_{\mathrm{HM.2}} \leq V$ and $V_{\mathrm{HM.1\text{-}2}} \leq V$. Lower bounds $V_{\mathrm{LM.1}}$ and $V_{\mathrm{LM.2}}$ serve as a basic for combined lower bound $V_{\mathrm{LM.1\text{-}2}}$. The upper bounds of $V_{\mathrm{HM.1}}$ and $V_{\mathrm{HM.2}}$ are basic for combined upper bound $V_{\mathrm{HM.1\text{-}2}}$ [40].

The described 3 upper and 3 lower bounds allow to make 9 methods for checking mantissas by inequalities. The trustworthiness of methods is defined by an ability of boundaries to correctly distinguish essential and inessential errors.

The methods are evaluated in the trustworthiness by experiments with the use of program model of the iterative array multiplier. The program model is developed on a demo version of Embarcadero Delphi 10.1 [41] for an assessment in trustworthiness of mantissas checking methods by inequalities. Simulation is executed for two $\mathrm{M_B}$ and $\mathrm{M_C}$ methods which use pairs of basic boundaries $V_{\mathrm{LM.1}}$, $V_{\mathrm{HM.1}}$ and pair of combined boundaries $V_{\mathrm{LM.1\text{-}2}}$, $V_{\mathrm{HM.1\text{-}2}}$.

Experiments are made on random sequences of input data—mantissas of the normalized operands. Operation of truncated multiplication is executed under the influence of single faults like shorts between two points of the circuit within one operational element. Points of the circuit and an operational element are selected randomly.

Experiments separate errors into essential and inessential specifying quantity n_{MSB} of most significant bits in the product of mantissas.

Simulation allows to determine probability P_E of an essential error, detection probabilities P_{DE} and P_{DI} of essential and inessential errors, trustworthiness $D_{B.n}$ of the $\mathrm{M_B}$ method or trustworthiness $D_{C.n}$ of the $\mathrm{M_C}$ method, where n is the mantissa size. Calculated trustworthiness are compared to trustworthiness D_R of residue checking method which in an experiment showed the detection of all errors, i.e. its probability of error detection $P_D = 1$, and trustworthiness $D_R = P_E$. The result of comparing is determined for the $\mathrm{M_B}$ and $\mathrm{M_C}$ methods as $\delta_{B.n} = D_{B.n}/D_R$ and $\delta_{C.n} = D_{C.n}/D_R$.

Results of simulation of the M_B method for the size of mantissas $n = 8$ are shown in Table 4.

The received results show an increase in probability P_E with growth of the number n_{MSB}. Lowering of n_{MSB} leads to growth of P_{DE}, P_{DI}, $D_{B.8}$ and $\delta_{B.8}$ indexes.

Results of simulation of the M_B method for the size of mantissas $n = 15$ are shown in Table 5.

The high trustworthiness of the method is reached due to the low detection probability P_{DI} of an inessential error and increases with growth of detection probability P_{DE} of an essential error.

Results of simulation of the M_C method for the size $n = 8$ are shown in Table 6.

The trustworthiness of the method grows together with the probability P_E of an essential error, thanks to the high detection probability P_{DE} of an essential error and the decreasing detection probability P_{DI} of an inessential error.

Results of simulation of the M_C method for the size of mantissas $n = 15$ are shown in Table 7.

The trustworthiness of the method grows together with the probability P_E of an essential error in a case of insignificant lowering in detection probabilities P_{DE} and P_{DI} of an essential and inessential error.

Table 4 Results of simulation of the M_B method for $n = 8$

n_{MSB}	3	4	5	6	7	8
P_E, %	6.7	12.9	20.0	31.4	44.5	58.5
P_{DE}, %	100	81.8	64.9	50.0	39.0	31.7
P_{DI}, %	12.7	10.7	7.3	4.5	3.2	1.4
$D_{B.8}$, %	87.3	88.1	86.4	80.4	70.5	59.0
$\delta_{B.8/R}$	12.9	6.8	4.3	2.6	1.6	1.0

Table 5 Results of simulation of the M_B method for $n = 15$

n_{MSB}	5	7	9	11	13	15
P_E, %	6.7	12.3	21.6	34.1	48.6	66.1
P_{DE}, %	66.7	42.1	24.2	17.3	13.3	9.7
P_{DI}, %	2.0	0.1	0	0	0	0
$D_{B.8}$, %	95.4	91.5	82.8	71.3	57.9	40.2
$\delta_{B.8/R}$	14.2	7.4	32.8	2.1	1.2	0.6

Table 6 Results of simulation of the M_C method for $n = 8$

n_{MSB}	3	4	5	6	7	8
P_E, %	6.7	12.9	20.0	31.4	44.5	58.5
P_{DE}, %	100	100	95.1	90.9	84.9	82.2
P_{DI}, %	74.5	72.9	71.5	70.4	68.9	67.1
$D_{B.8}$, %	30.5	36.5	42.8	48.9	55.5	61.5
$\delta_{B.8/R}$	4.6	2.8	2.1	1.6	1.2	1.0

Table 7 Results of simulation of the method M_C for n = 15

n_{MSB}	5	7	9	11	13	15
P_E, %	6.7	12.3	21.6	34.1	48.6	66.1
P_{DE}, %	100	88.9	80.4	76.4	73.5	71.9
P_{DI}, %	68.2	67.5	67.0	66.7	66.2	66.0
$D_{B.8}$, %	36.9	40.7	43.6	48.5	53.3	58.4
$\delta_{B.8/R}$	5.5	3.3	2.0	1.4	1.1	0.9

The M_B and M_C methods are alternative, they increase trustworthiness with reduction and growth in probability P_E of an essential error, accordingly.

Comparison of trustworthiness of the M_B and M_C methods with the residue checking represents an advantage of checking mantissas by inequalities and this advantage grows with reduction in probability P_E of an essential error.

4 The Hidden Faults in FPGA Projects: Problem and Decision

The problem of the hidden faults has a continuation in connection with the development of digital components of the safety-related systems on FPGA with LUT-oriented architecture [18, 19]. It is necessary to mention that this problem is peculiar for the safety-related systems as they belong to the dual-mode type of systems. Systems with one operating mode do not face such problem as the hidden faults remain hidden throughout all operation and are not dangerous. Therefore, use of FPGA in the safety-related systems requires carrying out the analysis on their tolerance in relation to the hidden faults.

LUT is the generator of Boolean functions. In case of four inputs A, B, C and D, LUT generates a Boolean function of four variables. Values of this function are stored in 16-bit memory of LUT. Functioning of LUT is characterized by sets M_N and M_E of memory bits (their addresses at inputs A, B, C and D) to which addressing in normal and emergency mode is executed, accordingly.

In case of $M_E \subseteq M_N$ all bits of memory addressed in emergency mode are used in the normal one. It interferes with the accumulation of the hidden faults in the normal mode.

The set $M_{PH} = M_N \backslash M_E$, $M_{PH} \neq \varnothing$ of the memory bits used in LUT only in emergency mode is a basis for a set of potentially hazardous points of the circuit. The hidden faults can be accumulated in these bits in a normal mode and reduce fault tolerance of digital components and the functional safety of safety-related systems in an emergency mode.

As an example, the digital component calculating the Boolean function $F(X) = 1$ for $X \bmod 3 = 0$ and $F(X) = 0$ for remaining values of X where $X = \{x_5, x_4, x_3, x_2, x_1\}$, $X = 0 \div 31$ is taken. Input data change in the ranges R_N and R_E respectively in normal and emergency mode. Design of this digital component on FPGA ALTERA Quartus II defines the circuit, represented in Fig. 1.

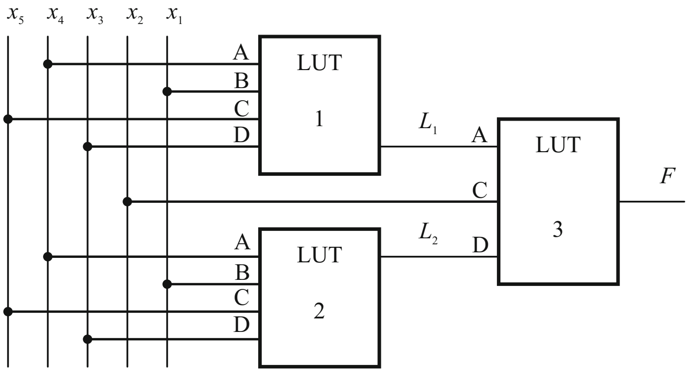


Fig. 1 The circuit of a digital component on FPGA with the LUT-oriented architecture

The circuit consists of three LUT: LUT 1, LUT 2 and LUT 3 which realize the functions $L_1(x_3, x_5, x_1, x_4)$, $L_2(x_3, x_5, x_1, x_4)$ and $F(L_2, x_2, L_1)$ described by codes $6BBD_{16}$, $97E9_{16}$ and $0AA0_{16}$, accordingly.

Values of the variables x_3, x_5, x_1, x_4 at the inputs of LUT 1 and LUT 2 and also corresponding to them values from the range X of the input data are shown in Table 8.

Each number # of the variables x_3, x_5, x_1, x_4 is created for two values of the range X in case of $x_2 = 0$ and $x_2 = 1$ accordingly.

We suggest three examples of partition of a set X of the input data on the ranges R_N and R_E of normal and emergency mode: (a) $R_N = 0 \div 15$ and $R_E = 16 \div 31$; (b) $R_N = 7 \div 26$ and $R_E = 0 \div 6$, $27 \div 31$; (c) $R_N = 0 \div 26$ and $R_E = 27 \div 31$.

The use of LUT memory (LUT 1 and LUT 2) in the normal and emergency modes is represented for the considered examples in Fig. 2a–c, accordingly.

Memory of LUT contains the bits N, E and O, which are used only in normal, emergency or in the both modes, accordingly. The example of **a** shows only bits like E and N, i.e. $M_N \cap M_E = \varnothing$ and $M_{PH} = M_E$. The example of **b** contains also bits of O type. The hidden faults cannot be accumulated in bits of N and O types. The example of **c** represents an example of a wide range of values of the M_N set where only the first three bits are used. Then bits of O type can also accumulate the hidden faults as well as bits like E.

Table 8 Values at the inputs of LUT 1, LUT 2 and corresponding to them values of X

#	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x_4	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
x_1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
x_5	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
x_3	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
X	0	8	1	9	16	24	17	25	4	12	5	13	20	28	21	29
	2	10	3	11	18	26	19	27	6	14	7	15	22	30	23	31

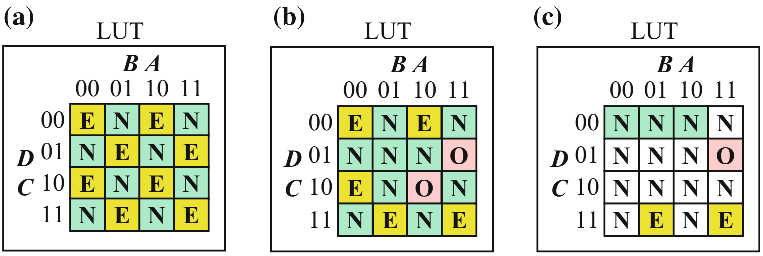


Fig. 2 The use of LUT memory in the normal and emergency modes

An example of **c** represents changes of input data in the insignificant range. In such cases the manual regulation of input data is used for the detection of the hidden faults. It allows to check bits of O type and to exclude them from a set of potentially hazardous points of the circuit. However manual regulation is powerless before a problem of the hidden faults for bits like E.

We suggest a method for the solving of this problem, using the version redundancy of the LUT-oriented architecture described in [23]. Two sequentially connected LUT allow to transfer between direct or inverse value without influence on a remaining part. The inverse value is created by an inverse of all bits in memory of the first LUT. Inverse on an input of the second LUT is compensated by the appropriate change of places of the bits in the memory. Such change concerns only the LUT program code and proves a number of its versions.

Inverting of A input or C input in examples **a** and **b** replaces all bits like E with bits of N type. Continued use of both versions (initial and changed) excludes a possibility of accumulation of the hidden faults in all LUT bits. Inverting of all LUT inputs interchanges the position of the first and last three bits of memory, solving a problem in the example of **c**. Manual regulation can be completely excluded by the continuous use of several program codes providing exchange of places for all bits like E with bits of N type.

5 The Integrity Monitoring of LUT-Oriented IO

The given paper suggests a method to monitoring the integrity of FPGA program code, which is an active IO with the LUT-oriented architecture (hereinafter—LUT-object) [42]. The suggested method proves that the monitoring information is embedded in IO and is represented in the form of DWM. The integrity monitoring scheme within the framework of this method is represented in Fig. 3.

The suggested method is based on the application of the following compositions: (a) Fridrich-Goljan-Du’s method (hereinafter the Fridrich’s method) [43, 44] in the point that concern the provision of the object state recovery; (b) method of the equivalent conversion of codes in LUT unit pairs [23].

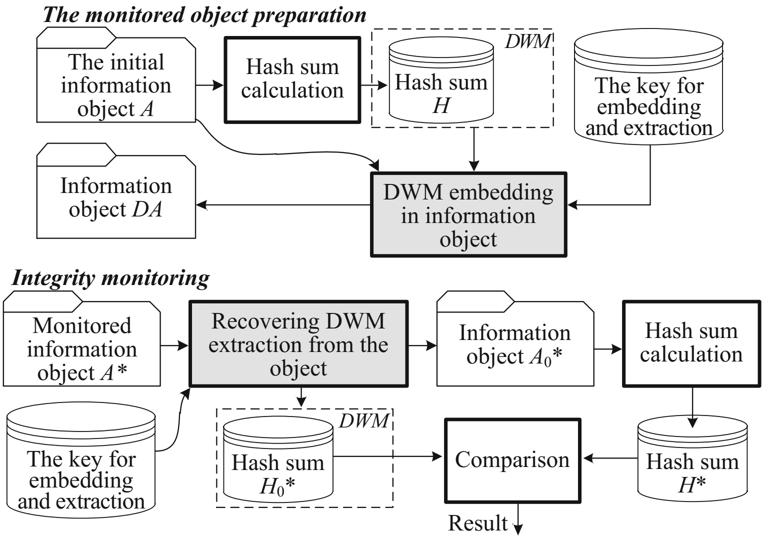


Fig. 3 Integrity monitoring scheme used in the suggested method

The Fridrich's method is only used for the passive IOs and is based on the application of two functions: (a) function of discrimination employed for classification of IO groups elementary parts by including them in one of three classes: regular *R*, singular *S* and unused *U*. The indication of the function value in each of the groups requires the complex calculation procedure; (b) flipping-function used for the conversion of groups of IO elementary parts in accordance with some rule.

Unlike the Fridrich's method the approach suggested in the given paper proves that IO element classification is performed without discrimination function [45]. The classification is based on correlation of zero values and one values in the codes of LUT-units: if more zero values are present in the code then the unit is classified as a regular one (*R*-unit); if there are more one values in the code the unit is classified as singular (*S*-unit); if the amounts of zero and one values of the code coincide then the unit is classified as an unused one (*U*-unit).

Within the framework of our method the Flipping-function calculation, unlike the Fridrich's method, includes only the equivalent inversion of LUT unit values in accordance with the procedure suggested in [23]. As a result of the unit program, code inversion of the *R*-unit is converted into *S*-unit, *S*-unit is converted into *R*-unit, and *U*-unit does not change its class. The sequence of actions in the point concerning the insertion of DWM with monitoring information into LUT-objects, which is suggested by the described method, may be represented in the following way.

Stage 1. The sequential examination of the LUT-object units in the order determined by the embedding path, which is an organized LUT unit subset obtained as a result of employing the preset rule of unit examination, is carried out. Wherein the classifying of units is performed. The result is represented as a binary *RS*-vector: value 0 is put for *R*-units, value 1 is put for *S*-units, and no values are put for *U*-units.

Figure 4 shows an example of the embedding path containing 20 LUT_1–LUT_{20} units. The hexadecimal value of program code, the number of zero values n_0 and one values n_1 in this code, the result of classifying the unit and the *RS*-vector value is demonstrated for each of the units. Vector *RS* = 100011001000001000 consists of 18 bits since the information of the two *U*-units is not included in the *RS*-vector contents.

Stage 2. The compression of the obtained *RS*-vector with the help of any lossless compression method is performed. As a result, a compressed vector RS_{com} is formed. E.g. applying the Huffman's algorithm [46] for the considered *RS* vector we obtain the following system of uneven prefix codes: 000 → 1, 001 → 00, 011 → 010, 100 → 011. Substituting the obtained codes in the *RS*-vector we get the 12-bit vector RS_{com} = 011010001001. The length differences of the vectors *RS* and RS_{com} is Δ = 6.

Stage 3. The DWM with monitoring information (binary succession, whose length is not more than Δ) is attached to the vector RS_{com}. In Fig. 5 the values of bits of the initial *RS*-vector, compressed vector RS_{com} and the example of attaching the *DWM* = 011100 are shown. As a result of concatenation of the vector RS_{com} and DWM, we obtain the vector *RS** for the resultant IO.

Stage 4. While examining the embedding path units the *U*-units are ignored, and *R* and *S* units are modified to introduce them in the vector *RS**. In Fig. 5 the values of vector *RS**, which are different from the corresponding values of vector *RS,* are distinguished. The program codes of LUT-units 1, 2, 3, 6, 13, 15, 18 of the initial object are to be inverted in accordance with the principles of equivalent conversions [23] to transfer to the vector *RS** values.

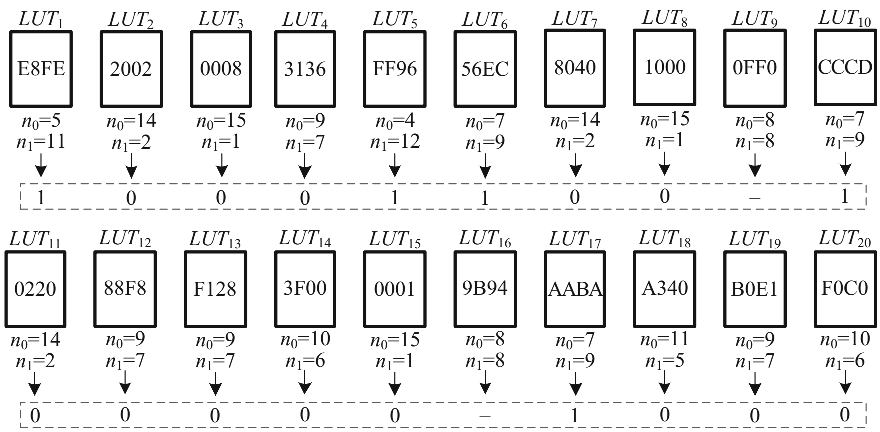


Fig. 4 The example of *RS*-vector value calculation

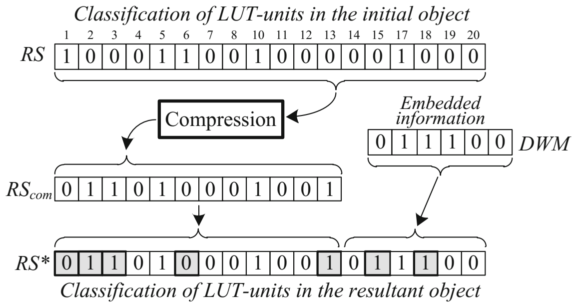


Fig. 5 The binary vector *RS** creation

The monitoring information embedded in IO can be obtained by applying the hashing methods, which are differently cryptographically strong and have different amount of bits. Hash sums obtained with the help of various methods can give the different number of LUT-units, which require their code inversion. On this basis before *Stage 3* the authors offer to carry out the process of hashing not with a single method but with the several ones taken from the previously stipulated list in order to choose the best variant of hash sum for the given IO. Similarly the RS vector compression can be performed with some alternative methods to choose the RS_{com} vector variant, which gives the minimum amount of LUT-units requiring their code inversions (Fig. 6). In this case the chosen hashing method for IO and compression ones for RS-vector are to become components of the key to extract the monitoring information out of IO.

The procedure of DWM extraction and the initial form recovery of LUT-object for the suggested method is the following.

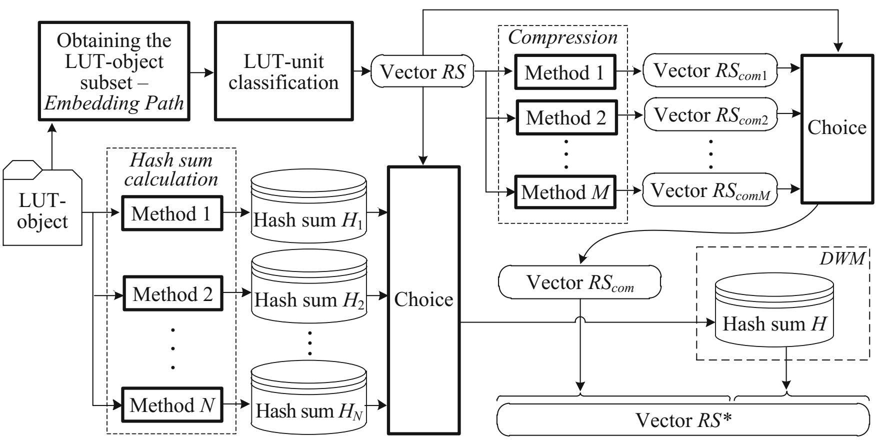


Fig. 6 Search procedure of vector *RS** minimizing the amount of LUT-unit code inversions

Stage 1. The sequential examination of the LUT-object units is performed in the order determined by the embedding path. Wherein the binary vector RS' is formed on the basis of the LUT-unit classification.

Stage 2. The last Δ bits, which contain DWM with monitoring information are read from the vector RS'.

Stage 3. The rest bits of the vector RS' are subjected to the decompression procedure, which is reverse to the compression performed at the stage of the information embedding. As a result the binary vector RS_{decom} is created.

Stage 4. According to the vector RS_{decom} values, the initial IO form recovery is produced. For this purpose, the LUT-unit classes are modified in accordance with the vector RS_{decom} values with the help of equivalent inversion.

In order to confirm the FPGA-project, functioning invariability before and after the monitoring information being embedded in them an experiment was made. The 40 projects for FPGA Altera Cyclone II were used in the experiment. In the course of the project implementation the processing speed changes, energy consumption and thermal dissipation were researched. These characteristic measurements were carried out with the help of modules CAD Altera Quartus II Timing Analyzer and Power Play. The changes of processing speed were 0.18%, energy consumption and thermal dissipation—0.22%, on the average. As one can see the change values are present within the limits of error value of measurement means. So the results of the experiment have confirmed that LUT-objects, in which the monitoring information was embedded with the help of the proposed method, took their original forms after this information extraction.

6 Conclusions

Development of Green IT Engineering on the way of increase in energy efficiency and safety of FPGA projects is restricted to domination of array structures in circuitry decisions. The main shortcoming of array structures consists in their low checkability creating a problem of the hidden faults in safety-related systems. These faults can be accumulated in matrix circuits in the normal mode and reduce fault tolerance of circuits and safety of systems in the most responsible emergency mode. Increase in a checkability of circuits requires reduction of array structures.

The way of array structure reduction is perspective, but requires overcoming traditions in use of matrix decisions which passed a long way of enhancement and show high rates in throughput and energy efficiency.

We suggest soft and cardinal ways of array structures reduction. The soft way is based on execution of the truncated arithmetical operations implemented in the reduced array structures. The truncated operations are the main method in processing of approximate data—mantissas of floating-point numbers. The cardinal way consists in transition from array structures to bitwise pipelines which parallelize calculations in serial codes.

Experiments on FPGA showed that bitwise pipelines do not interfere with matrix circuits in throughput, but they consume more energy. The suggested method of pipeline multiplication with processing of two bits in a clock cycle allows to increase throughput and to reduce energy consumption.

Development of on-line testing methods in checking of mantissas by inequalities showed increase in trustworthiness of methods for the truncated operations in comparison with residue checking.

Advantage of FPGA design in diversity of a program code is used in the suggested method for an increase in the functional safety of FPGA projects in opposition to the hidden faults. Continuous use of several program codes for the circuits with LUT-oriented architecture allows to exclude accumulation of the hidden faults in a normal mode of safety-related systems.

We propose to use diversity of a program code also in suggested method for monitoring in integrity of the FPGA project in support of security and functional safety of digital components in structure of safe systems.

References

1. Murugesan, S., Gangadharan, G.R.: Harnessing Green IT Principles and Practices. Wiley, UK (2012)
2. Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.): Green IT engineering: concepts, models, complex systems architectures, studies in systems, decision and control, vol. 74. Springer, Berlin (2017). https://doi.org/10.1007/978-3-319-44162-7
3. Yakovlev, A.: Energy-modulated computing. In: Proceedings of Design, Automation and Test in Europe Conference and Exhibition 2011 (DATE 2011), Grenoble, France, 1–6 (2011)
4. Maevsky, D.A., Maevskaya, E.J., Stetsuyk, E.D.: Evaluating the RAM energy consumption at the stage of software development. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds) Green IT Engineering Concepts, Models, Complex Systems Architectures. Studies in Systems, Decision and Control, vol. 74, pp. 101–121. Springer, Cham. (2017) https://doi.org/10.1007/978-3-319-44162-7_6
5. Kharchenko, V., Gorbenko, A., Sklyar, V., Phillips, C.: Green computing and communications in critical application domains: challenges and solutions. In: 9th International Conference on Digital Technologies (DT'2013), pp. 191–197. Zhilina, Slovak Republic (2013)
6. Maevsky, D.A., Maevskaya, E.J., Stetsuyk, E.D., Shapa L.N.: Malicious software effect on the mobile devices power consumption. In: Kharchenko V., Kondratenko Y., Kacprzyk J. (eds) Green IT Engineering: Components, Networks and Systems Implementation. Studies in Systems, Decision and Control, vol. 105, pp. 155–171. Springer, Cham. (2017) https://doi.org/10.1007/978-3-319-55595-9_8
7. IEC 61508-1:2010. Functional Safety of Electrical/Electronic/Programmable Electronic Safety Related Systems—Part 1: General requirements. Geneva: International Electrotechnical Commission (2010)
8. Bakhmach, E., Kharchenko, V., Siora, A., Sklyar, V., Tokarev, V.: Design and qualification of I&C systems on the basis of FPGA technologies. In: Proceedings of 7th International Topical Meeting on Nuclear Plant Instrumentation, Control, and Human-Machine Interface Technologies (NPIC&HMIT 2010), pp. 916–924. Las Vegas, Nevada (2010)
9. Drozd, J., Drozd, A., Antoshchuk, S.: Green IT engineering in the view of resource-based approach. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering:

Concepts, Models, Complex Systems Architectures, Studies in Systems, Decision and Control, vol. 74, pp. 43–65. Springer, Heidelberg (2017). https://doi.org/10.1007/978-3-319-44162-7_3
10. Drozd, J., Drozd, A., Antoshchuk, S., Kushnerov, A., Nikul, V.: Effectiveness of matrix and pipeline FPGA-based arithmetic components of safety-related systems. In: 8th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, pp. 785–789. Warsaw, Poland (2015). https://doi.org/10.1109/idaacs.2015.7341410
11. Palagin, A.V., Opanasenko, V.N., Kryvyi, S.L.: Resource and energy optimization oriented development of FPGA-based adaptive logical networks for classification problem. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Components, Networks and Systems Implementation, vol. 105, pp. 195–218. Springer, Heidelberg (2017)
12. Melnik, A.O.: Architecture of Computer. Volinska oblasna drukarnja, Lutsk, Ukraine (2008) (Ukrainian)
13. Shum, W., Anderson, J.H.: FPGA glitch power analysis and reduction. In: International Symposium on Low power electronics and design (ISLPED), pp. 27–32 (2011)
14. Vikas, D.: A review on glitch reduction techniques. Int. J. Res. Eng. Technol. **3**(2), 145–148 (2014)
15. Garofalo, V.: Truncated binary multipliers with minimum mean square error: analytical characterization, circuit implementation and applications. Ph.D. Dissertation, University of Studies of Naples “Federico II”, Naples, Italy (2008)
16. Kahan, W.: IEEE standard 754 for binary floating-point arithmetic. Lecture Notes on the Status of IEEE 754, Elect. Eng. and Computer Science University of California, Berkeley CA 94720-1776 (1996)
17. Kondratenko, Y., Kondratenko, V.: Soft computing algorithm for arithmetic multiplication of fuzzy sets based on universal analytic models. In: Ermolayev, V. et al. (eds) Information and Communication Technologies in Education, Research, and Industrial Application. Communications in Computer and Information Science: ICTERI’2014, pp. 49–77. Springer, Switzerland (2014). https://doi.org/10.1007/978-3-319-13206-8_3
18. Cyclone II Architecture. Cyclone II Device Handbook Version 3.1. Altera Corporation (2007). http://www.altera.com/literature/hb/cyc2/cyc2_cii51002.pdf
19. Drozd, A., Kharchenko, V., Antoshchuk, S., Sulima, J., Drozd, M.: Checkability of the digital components in safety-critical systems: problems and solutions. In: IEEE East-West Design & Test Symposium, Sevastopol, Ukraine, pp. 411–416 (2011). https://doi.org/10.1109/ewdts.2011.6116606
20. Drozd, M., Drozd, A.: Safety-related instrumentation and control systems and a problem of the hidden faults. In: The 10th International Conference on Digital Technologies 2014, Zhilina, Slovak Republic, pp. 137–140 (2014). https://doi.org/10.1109/dt.2014.6868692
21. Palagin, A.V., Opanasenko, V.N.: Design and application of the PLD-based reconfigurable devices. In: Adamski, M., Barkalov, A., Wegrzyn, M. (eds.) Design of Digital Systems and Devices. Lecture Notes in Electrical Engineering, vol. 79, pp. 59–91. Springer Verlag, Berlin Heidelberg (2011)
22. Kondratenko, Y., Gordienko E.: Implementation of the neural networks for adaptive control system on FPGA. In: Katalinic, B. (ed.) Annals of DAAAM for 2012 and Proceeding of the 23th Int. DAAAM Symp “Intelligent Manufacturing and Automation”, vol. 23, No. 1. Published by DAAAM International, pp. 389–392, Vienna, Austria, EU (2012)
23. Drozd, A., Drozd, M., Kuznietsov, M.: Use of natural LUT redundancy to improve trustworthiness of FPGA design. CEUR Workshop Proceedings **1614**, 322–331 (2016)
24. Design Optimization for Altera Devices. Qii52005-2.0. Quartus II Handbook, vol. 2. Altera Corporation (2004)
25. PowerPlay Early Power Estimator. User Guide. Altera Corporation (2013)
26. PowerPlay Power Analysis. Quartus II Handbook Version 13.1.0. Altera Corporation (2013). http://www.altera.com/literature/hb/qts/qts_qii53013.pdf

27. Using TimeQuest Timing Analyzer. Altera Corporation—University Program (2013). http://altera.com/up/pub/Intel_Material/13.0/Tutorials/Timequest.pdf
28. Drozd, A., Antoshchuk, S.: New on-line testing methods for approximate data processing in the computing circuits. In: 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, Prague, Czech Republic, pp. 291–294 (2011). https://doi.org/10.1109/idaacs.2011.6072759
29. Kekre, H.B., Mishra, D., Khanna, R., Khanna, S., Hussaini, A.: Comparison between the basic LSB replacement technique and increased capacity of information hiding in LSB's method for images. Int. J. Comput. Appl. **45**(1), 33–38 (2012)
30. Drozd, A.V., Lobachev, M.V., Hassonah, W.: Hardware check of arithmetic devices with abridged execution of operations. In: the European Design and Test Conference (ED & TC 96), Paris, France, p. 611 (1996)
31. Yatskiv, V., Tsavolyk, T., Zhengbing, H.: Multiple error detection and correction based on modular arithmetic correcting codes. In: 8-th 2015 IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems, IDAACS'2015, Warszawa, Poland, pp. 850–854 (2015)
32. ANSI/IEEE Std 754-1985, IEEE Standard for Binary Floating-Point Arithmetic (1985)
33. IEEE Std 754™-2008 (Revision of IEEE Std 754-1985) IEEE Standard for Floating-Point Arithmetic. IEEE 3 Park Avenue New York, NY 10016–5997, USA (2008)
34. Vacca, J.: Computer and Information Security Handbook, 2nd edn. Morgan Kaufmann, Waltham, Mass (2013)
35. Schneier, B., Kohno, T., Ferguson, N.: Cryptography Engineering. Wiley, Hoboken, N. J. (2013)
36. Cox, I., Miller, M., Bloom, J., Fridrich, J.: Digital Watermarking and Steganography. Morgan Kaufmann Publishers, Amsterdam (2008)
37. Fridrich, J.: Steganography in Digital Media. Cambridge University Press, New York (2010)
38. Cyclone FPGA Family Data Sheet. Altera Corporation (2003). http://www.altera.com
39. Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (Eds.): Green IT Engineering: Components, Networks and Systems Implementation, Studies in Systems, Decision and Control, vol. 105. Springer, Berlin, Heidelberg (2017). https://doi.org/10.1007/978-3-319-55595-9_11
40. Drozd, O., Al-dhabi, M., Antoshchuk, S., Martinyuk, A., Drozd, M.: Models and methods checking mantissas by inequalities for on-line testing of digital circuits in critical. In: Proceedings of IEEE East–West Design and Test Symposium, Novi Sad, Serbia, pp. 79–83 (2017)
41. Delphi 10 Seattle: Embarcadero (2015). https://www.embarcadero.com/ru/products/delphi
42. Huffmire, T.: Handbook of FPGA Design Security. Springer, Dordrecht (2013)
43. Fridrich, J., Goljan, M., Du, R.: Lossless data embedding—new paradigm in digital watermarking. EURASIP J. Adv. Sig. Process 185–196 (2002)
44. Goljan, M., Fridrich, J., Du, R.: Distortion-free data embedding for images. In: Proceedings of the 4th International Workshop on Information Hiding (IHW-01), USA, Pittsburg, pp. 27–41 (2001)
45. Zashcholkin, K., Ivanova, O.: The control technology of integrity and legitimacy of LUT-oriented information object usage by self-recovering digital watermark. CEUR Workshop Proc **1356**, 498–506 (2015)
46. Salomon, D., Motta, G.: Handbook of Data Compression. Springer, London (2010)

A Prospective Lightweight Block Cipher for Green IT Engineering

Alina Andrushkevych, Yurii Gorbenko, Olexandr Kuznetsov, Roman Oliynykov and Mariia Rodinko

Abstract This chapter provides general requirements to modern block ciphers required for implementation at lightweight cryptographic transformations for critical distributed environment applications with Green IT conformance. It is given an overview of well-known block ciphers and lightweight primitives PRESENT and CLEFIA, defined at ISO/IEC 29192-2. It is given a specification of the lightweight block cipher Cypress that was recently developed and presented in Ukraine. Cypress does not use heavy computation operations, nor require any precomputed tables that allows efficient hardware implementation. The Cypress performance in software is approximately three times higher than AES one on Windows, Linux and Android platforms.

1 Introduction

Nowadays, the interest to Green IT is grown rapidly: modern development trends of advanced mobile technologies, Internet of Things (IoT) require significant limitation on energy consumption of electronic devices and number of GE on chip [1, 2].

A. Andrushkevych
Kharkiv National University of Radioelectronics, Kharkiv, Ukraine
e-mail: alvasamoilova@gmail.com

Y. Gorbenko · O. Kuznetsov · R. Oliynykov (✉) · M. Rodinko
Information Systems and Technologies Department, Computer Science Faculty,
V. N. Karazin Kharkiv National University, Kharkiv, Ukraine
e-mail: roliynykov@gmail.com

Y. Gorbenko
e-mail: gorbenkou@iit.kharkov.ua

O. Kuznetsov
e-mail: kuznetsov@karazin.ua

M. Rodinko
e-mail: m.rodinko@gmail.com

© Springer Nature Switzerland AG 2019

V. Kharchenko et al. (eds.), *Green IT Engineering: Social, Business and Industrial Applications*, Studies in Systems, Decision and Control 171,
https://doi.org/10.1007/978-3-030-00253-4_5

Such applications must provide secure services for end users which is implemented through cryptography.

Conventional crypto widely applied for securing remote distributed communication, like block ciphers Kalyna [3], AES [4] and others do not satisfy requirements for devices with highly constrained power consumption that raises importance of lightweight crypto primitives. Lightweight cryptography is cryptographic algorithms or protocols that are designed for implementation in confined environments such as RFID tags, sensors, contactless smart cards, medical devices [5], and so on [6]. Nevertheless, most known lightweight block ciphers do not provide the security level sufficient for cipher application in post quantum period or enough good Green IT properties. Creating a block cipher that is fast, lightweight and provides high security, is a hard and important task.

Since the selection of AES, many block ciphers with lightweight cryptography properties have been considered, among them CLEFIA and PRESENT, which are well known for their safety and implementation. Both ciphers are presented in ISO/IES 29192-2 [7] and are ready for use in practical systems. However, PRESENT has the following disadvantages: small block size and legacy key length, combined with quite inefficiency in software implementation; moreover, characteristics of hardware implementation of GOST 28147-89 [8] that had been adapted almost 30 years ago, was significantly more efficient than the PRESENT ones that was developed many years later especially for hardware solutions. CLEFIA has enough good external parameters like block size and key length, but in implementation characteristics it is close to AES and does not satisfy strict requirements of energy consumption and number of GE on chip for highly constrained devices. In software implementation CLEFIA is slower than AES that makes CLEFIA even less preferable solution comparing to already existing ones.

Thus, Green IT requires application of more secure and fast lightweight ciphers. One of such primitives called Cypress [9] was recently developed and presented in Ukraine. Cypress supports 256- and 512-bit block size and key length correspondingly. The new cryptographic solution is based on ARX-transformation integrated into the Feistel network as a round function. The cipher key schedule is non-injective and repeat the main principles of Kalyna one. Cypress does not use heavy computation operations, nor require any precomputed tables that allows efficient hardware implementation. Round transformation of the cipher is optimized for efficient software implementation on both 32- and 64-bit platforms. The Cypress performance in software is approximately three times higher than AES one on Windows, Linux and Android platforms. The analysis of Cypress shown that statistical properties of the encryption algorithm and the key schedule satisfy the requirements of NIST STS; the required avalanche effect is reached after the fourth round of encryption both for 256- and 512-bit cipher.

The **motivation** of this chapter is to provide more efficient cryptographic solution for Green IT environment.

The **goals** of this chapter are the following:

- to make comparative analysis of lightweight ciphers to conventional ones;
- to show the disadvantages of known lightweight ciphers and the need of development a new efficient solution;
- to provide a specification and properties overview of lightweight block cipher Cypress that was recently developed and presented in Ukraine.

The **structure** of the chapter includes:

- a comparison between conventional block ciphers and lightweight block ciphers CLEFIA and PRESENT;
- specification and description of basic properties of the lightweight block cipher Cypress.

2 Characteristics and Parameters of Block Symmetric Ciphers

Let's consider basic characteristics and parameters of the most known traditional block ciphers.

"AES" is a symmetric block cipher that processes 128-bit blocks of data using encryption keys 128, 192 and 256 bits. AES algorithm is also known as an algorithm Rijndael. Both for encryption and for decryption algorithm AES uses cyclic function that consists of four different byte-oriented transformations: byte substitution using a table lookup (S-block), shift of the line of array of state (current state), mixed data in each column of the array of state, adding round key to the state. Block symmetric cipher AES standardized internationally as ISO/IEC 18033-3 [10].

"Camellia" is an encryption algorithm with a block length of 128 bits and a key length of 128, 192, 256 bits. One of the finalists of the European competition NESSIE (along with AES and Shacal-2), the development of Japanese companies Nippon Telegraph and Telephone Corporation and Mitsubishi Electric Corporation. Standardized internationally in ISO/IEC 18033-3 [10]. Cycle function uses non-linear transformation, linear transformation and byte permutation.

"CAST" is a Canadian algorithm. It was developed by Carlisle Adams and Stafford Tavares. It uses 64-bit message block and 128-bit key, has 16 rounds of transformation. In ISO/IEC 18033-3 [10] was standardized an algorithm CAST-128, which uses eight S-blocks that implement 8-bit input vectors in 32-bit output vectors. Four S-blocks used in round function algorithm, the rest in the scheme of deploying key. The algorithm based on the Feistel chain, uses nonlinear transformation, XOR and modular arithmetic.

"SEED" is a symmetric algorithm, developed by the Korean Information Security Agency (Korean Information Security Agency, KISA) in 1998. Also internationally standardized in ISO/IEC 18033-3 [10]. The algorithm is widely used by financial and banking institutions, industrial enterprises and government agencies in South Korea and in the TLS and S/MIME. The algorithm is a Feistel cipher with 16 rounds of 128-bit blocks and 128-bit key. Cycle function uses nonlinear transformation, linear transformation and byte permutation.

"Blowfish" is a symmetric block cipher that has a block length of 64 bits and a key length of 448 bits [11]. It is a Feistel cipher which performed on simple and quick operations: XOR 32-bit words, substitution, addition. Data encryption is a simple function that repeats 16 times. Blowfish has established itself as a reliable algorithm used to hash passwords, email security and file (GnuPG), security protocols at the network and transport levels (SSH—transport layer, OpenVPN—creating encrypted channels).

"GOST 28147-89" is Soviet standard of symmetric encryption, introduced in 1990, belongs to the category of block ciphers which working on the architecture balanced Feistel cipher [8]. Supports block length of 64 bits and a key length of 256 bits, using 32 cycles. The main operations of the encryption algorithm are key addition by modulo 2^{32}, non-linear transformation and shift. Currently not used.

"DES" is an algorithm for symmetric encryption, developed by IBM and approved by USA government in 1977 as an official standard (FIPS 46-3) [12]. Block size for DES is 64 bits. At the core of the algorithm is the Feistel cipher with 16 cycles and a key length of 56 bits. The algorithm uses a combination nonlinear (S-blocks) and linear transformations. Currently not used.

"TDES" is a symmetric block cipher created by Whitfield Diffie, Martin Hellman and Walt Tachman in 1978 [12]. TDES algorithm is based on DES. Implemented in PGP and S/MIME, is used in the management of keys in a standard ANSI X9.17 and ISO 8732 in PEM (Privacy Enhanced Mail). At the core of the algorithm is the Feistel chain a combination of linear and nonlinear transformations. Supports block length of 64 bits and a key length of 56, 112, 168 bits, using 48 cycles.

"Kalyna" is the national standard of Ukraine as DSTU 7624:2014 [3], designed by Ukrainian company "IIT". Algorithm has block and key length of 128, 256, 512 bits (the length of the key is equal to the block length, or twice as long), it provides normal, high and ultrahigh security level. Cycle transformation use the round key addition, non-linear substitution (four substitutions table byte for byte), cyclic shift of the bits of state, matrix multiplication and round key addition by modulo 2.

"Kuznechik" is an encryption algorithm that standardized in Russia as GOST 34.12-2015 [13]. The length of the input block is 128 bit, key length—256 bit. The process of encryption based on the sequential application of multiple identical rounds, each of which includes the round key addition, nonlinear (substitutions table byte for byte) and linear transformations.

"BelT" is a national standard of symmetric encryption and integrity control of the Republic of Belarus. Standardized as STB 34.101.31-2011 [14]. This is the algorithm with a 256 bit key and 128-bit blocks of message. The algorithm use a Feistel chain with 8 rounds, each round operates 32-bit blocks. Cycle function uses nonlinear transformation, linear transformation and byte permutation.

The results of the analysis of the main properties of modern BSCs are summarized in Table 1. Table 1 lists the name of the cipher, the country in which the

Table 1 The main properties and transformations of modern BSCS

Cipher	Country, standard	N_b	N_k	K	Basic transformations
"AES"	USA, FIPS-197, ISO/IEC 18033-3	128	128	10	Square-like SPN-structure. The cyclic transformation is based on the multiplication substitution tables on the MDR matrix over a finite field
			192	12	
			256	14	
"CAST"	Canada	64	128	16	The basis of the scheme is the Feistel chain using nonlinear operations, XOR and modular arithmetic
"Camellia"	Japan, ISO/IEC 18033-3	128	128	18	The classic Feistel chain with the previous and final whitening uses a nonlinear transform, a linear scattering unit, and a byte reshuffle
			192	24	
			256		
"SEED"	Korea, ISO/IEC 18033-3	128	128	16	The basis of the scheme is the Feistel chain The cyclic transformation is based on the substitution tables, the linear scattering block
"Blowfish"	USA, Not patented	64	448	16	The scheme is based on the Feistel chain, which operates with 32-bit blocks. The cyclic transformation is based on a substitution table, a linear scattering block
"GOST 28147-89"	USSR, 1990, GOST 28147-89	64	256	32	The scheme is based on the Feistel chain, which operates with 32-bit blocks. The cyclic transformation is based on a substitution table, a linear scattering block
"DES"	USA, FIPS 46-3, Not used	64	56	16	The basis of the algorithm is a Feistel network with a combination of nonlinear and linear transformations
"TDES"	USA, ISO/IEC 18033-3	64	168	48	The basis of the algorithm is a Feistel network with a combination of nonlinear and linear transformations
"Kalyna"	Ukraine, DSTU 7624:2014	128	128	10	Square-like SPN-structure. The cyclic transformation is based on the multiplication substitution tables on the MDR matrix over a finite field
		128	256	14	
		256			
		256	512	18	
		512			
"Kuznechik"	Russia, GOST 34.12-2015	128	256	10	Square-like SPN-structure. The cyclic transformation is based on the multiplication substitution tables on the MDR matrix over a finite field
"BelT"	Republic of Belarus, STB 34.101.31-2011	128	256	8	The scheme is based on the Feistel chain, which operates with 32-bit blocks. The cyclic transformation is based on a substitution table, a linear scattering block

cipher were developed, the standard (national or international). The size of the cipher block is denoted as N_b bit, the key size is N_k bit, the number of conversion cycles is K.

The lightweight block ciphers PRESENT and CLEFIA have the following characteristics.

"PRESENT"—a lightweight block cipher with 64 bits of the data block and 80 (or 128) bits of the key block [7]. It is based on the permutation substitution pattern and consists of 31 rounds. For both key lengths of the PRESENT cipher, the user key is first entered in the key register (for 80 bits—80 bits, and for 128 bits—128 bits).

"CLEFIA"—a lightweight block cipher CLEFIA has a length of data blocks of 128 bits and a key block length of 128, 192, 256 bits [7]. The cipher algorithm is based on the generalized Feistel structure. The transformation has 18, 22, 26 rounds of encryption and decryption for 128, 192, 256 key block lengths respectively.

We summarize the analyzed data on the characteristics of ciphers and the levels of security that are provided in Table 2.

So, as a conclusion, it can be noted that for constructing block symmetric ciphers, methodological approaches such as SPN structures, IDEA-like structures based on the Lai-Massey scheme, and structures based on Feistel network can be used. Regarding the level of security, it should be noted that only the Ukrainian cipher "Kalyna" provides ultra-high level of security, the specified parameters allow to use this cipher, even taking into account the latest trends in the development of quantum computing systems. We can assume that the cipher "Kalyna" can be applied to the post-quantum period. In particular, the level of security of symmetric ciphers against quantum cryptanalysis in the attack on the key and on the message block are presented in Table 3.

The lightweight CLEFIA and PRESENT ciphers provide a satisfactory, normal, and even high level of security which, combined with high performance, is extremely useful when used in limited environments such as RFID tags, sensors, contactless smart cards, medical devices and so on.

Thus, the appearance of the technology of lightweight cryptography is already in competition with proven international standards of encryption due to increased performance, and ensuring an adequate level of security. The applying of lightweight cryptography in communication channels is significantly increase their speed, but there are doubts about the stability of ciphers of lightweight cryptography, which need to be solved by conducting cryptanalysis and further studies of such ciphers. In addition, should take into account the possible risks and threats of the appearance of quantum computing systems, that is, should focus on the post-quantum period of the use of cryptographic information security.

The next section presents a description of the new post quantum lightweight block cipher with both high security level and performance which is suitable for solution of Green IT security issues.

Table 2 Characteristics of modern BSCS

Cipher	Length of the block (bit)	Length of the key (bit)	Type of structure	Security level
AES	128	128	SPN-structure	Normal level, high level
		192		
		256		
CAST	64	128	Feistel chain	Satisfactory level
Camellia	128	128	Feistel chain	Normal level, high level
		192		
		256		
SEED	128	128	Feistel network	Normal level, high level
		192		
		256		
Blowfish	64	448	Feistel network	Satisfactory level
GOST 28147-89	64	256	Feistel network	Satisfactory level, normal level
TDES	64	168	Feistel network	Satisfactory level
Kalyna	128	128	SP-network	Normal level, high level, ultra high level
	128	256		
	256			
	256	512		
	512			
Kuznechik	128	256	LSX-scheme and Feistel network	Normal level, high level
BelT	128	256	Feistel network	Normal level, high level
PRESENT	64	80	SP- network	Satisfactory level
		128		
CLEFIA	128	128	The generalized Feistel structure	Normal level, high level
		192		
		256		

3 A Post Quantum Lightweight Block Cipher Cypress

Along with the already known approaches to block ciphers design, such as SPN and Feistel network, an ARX—(Add-Rotate-XOR) transformation is becoming widespread in the area of lightweight primitives design. This is due to the fact that modulo addition and rotation operations are fast and simple to implement.

This section contains a specification and properties overview of post quantum lightweight block cipher Cypress [9] developed in Ukraine. The cipher based on Feistel network, an ARX-transformation is used as a round function. Cypress is optimized both for 32- and 64-bit platforms and provides high performance (several times more than AES).

Table 3 Resistance of symmetric ciphers against an attack on a key block and the message using quantum algorithm

Cipher	Length block/ key, bits	Needed space of memory for attack on message block/key, q-bits	Resistance to the attack in case of using quantum computer	
			On the block of message, quantum operations	*On the key, quantum operations*
AES-128	128/128	128/128	$2^{64}(10^{19,2})$	$2^{64}(10^{19,2})$
AES-256	128/256	128/256	$2^{64}(10^{19,2})$	$2^{128}(10^{38,4})$
CAST-128	64/128	64/128	$2^{32}(10^{9,6})$	$2^{64}(10^{19,2})$
Camellia-128	128/128	128/128	$2^{64}(10^{19,2})$	$2^{64}(10^{19,2})$
Camellia-192	128/192	128/192	$2^{64}(10^{19,2})$	$2^{96}(10^{28,8})$
Camellia-256	128/256	128/256	$2^{64}(10^{19,2})$	$2^{128}(10^{38,4})$
SEED	128/128	128/128	$2^{64}(10^{19,2})$	$2^{64}(10^{19,2})$
Blowfish	64/448	64/448	$2^{32}(10^{9,6})$	$2^{224}(10^{67,2})$
GOST 28147-89	64/256	64/256	$2^{32}(10^{9,6})$	$2^{128}(10^{38,4})$
DES	64/56	64/56	$2^{32}(10^{9,6})$	$2^{28}(10^{8,4})$
TDES	64/168	64/168	$2^{32}(10^{9,6})$	$2^{134}(10^{40,2})$
Kalyna -128	128/128	128/128	$2^{64}(10^{19,2})$	$2^{64}(10^{19,2})$
Kalyna -256	256/256	256/256	$2^{128}(10^{38,4})$	$2^{128}(10^{38,4})$
Kalyna -512	512/512	512/512	$2^{256}(10^{76,8})$	$2^{256}(10^{76,8})$
Kuznechik	128/256	128/256	$2^{64}(10^{19,2})$	$2^{128}(10^{38,4})$
BelT	128/256	128/256	$2^{64}(10^{19,2})$	$2^{128}(10^{38,4})$
PRESENT-128	64/128	64/128	$2^{32}(10^{9,6})$	$2^{64}(10^{19,2})$
CLEFIA-256	128/256	128/256	$2^{64}(10^{19,2})$	$2^{128}(10^{38,4})$

3.1 The Requirements to the Prospective Lightweight Block Cipher

The cipher was developed according to the following requirements [9]:

- High level of cryptographic strength against exhaustive search attacks that requires the necessity to support 256- and 512-bit keys;
- Application of strong key schedule to protect against attacks on key schedules;
- High performance on different software and hardware platforms (including mobile ones);
- A compact software implementation both for desktop systems (workstations, servers) and mobile platforms (smartphones, tablets);
- Versions of algorithm optimized both for 32- and 64-bit systems;
- The minimal memory size needed for a high-speed implementation (a compact code and no precomputation tables);
- A constant encryption time on modern CPUs regardless parameters processed (to protect against side-channel attacks).

3.2 The General Parameters of the Block Cipher Cypress

The encryption algorithm operates on l-bit blocks using k-bit key, where $l, k \in \{256, 512\}$, $l = k$. Operations are performed on s-bit words, where $s \in \{32, 64\}$. The general parameters of the cipher are presented in Table 4 [9].

It should also be noted that Cypress-256 is oriented to be used on 32-bit platforms, while Cypress-512—on 64-bit ones (including the requirements to compact implementation and reduced power consumption).

3.3 The Encryption Algorithm

The general scheme of the encryption algorithm is given in Fig. 1.

The inputs of the encryption algorithm are a plaintext $P = (P_0, P_1, \ldots, P_7)$ and round keys $RK^{(0)}, RK^{(1)}, \ldots, RK^{(t-1)}$. Each key $RK^{(i)} = (RK_0^{(i)}, RK_1^{(i)}, RK_2^{(i)}, RK_3^{(i)})$ consists of four s-bit words. Round keys are formed using the key schedule based on the encryption key $K = (K_0, K_1, \ldots, K_7)$.

According to the Feistel network, a plaintext P is divided into two sub blocks $L_0 = (P_0, P_1, P_2, P_3)$ and $R_0 = (P_4, P_5, P_6, P_7)$. The output of ith iteration is obtained as:

$$L_i = R_{i-1} \oplus F(L_{i-1}, RK^{(i-1)}),$$
$$R_i = L_{i-1.}$$

The round function F consists of addition modulo 2 of a subblock L_{i-1} with a key $RK^{(i)}$ and a double application of the function $h(P_0', P_1', P_2', P_3')$ which takes four s-bit words as an input. The output value of the function h is calculated as:

$$P'_0 = ADD(P'_0, P'_1), \quad P'_3 = XOR(P'_3, P'_0), \quad P'_3 = ROTL(P'_3, r1),$$
$$P'_2 = ADD(P'_2, P'_3), \quad P'_1 = XOR(P'_1, P'_2), \quad P'_1 = ROTL(P'_1, r2),$$
$$P'_0 = ADD(P'_0, P'_1), \quad P'_3 = XOR(P'_3, P'_0), \quad P'_3 = ROTL(P'_3, r3),$$
$$P'_2 = ADD(P'_2, P'_3), \quad P'_1 = XOR(P'_1, P'_2), \quad P'_1 = ROTL(P'_1, r4),$$

Table 4 The general parameters of the cipher Cypress

	Cypress-256	Cypress-512
Block size (l), bits	256	512
Key length (k), bits	256	512
Word length (s), bits	32	64
Number of rounds (t)	10	14

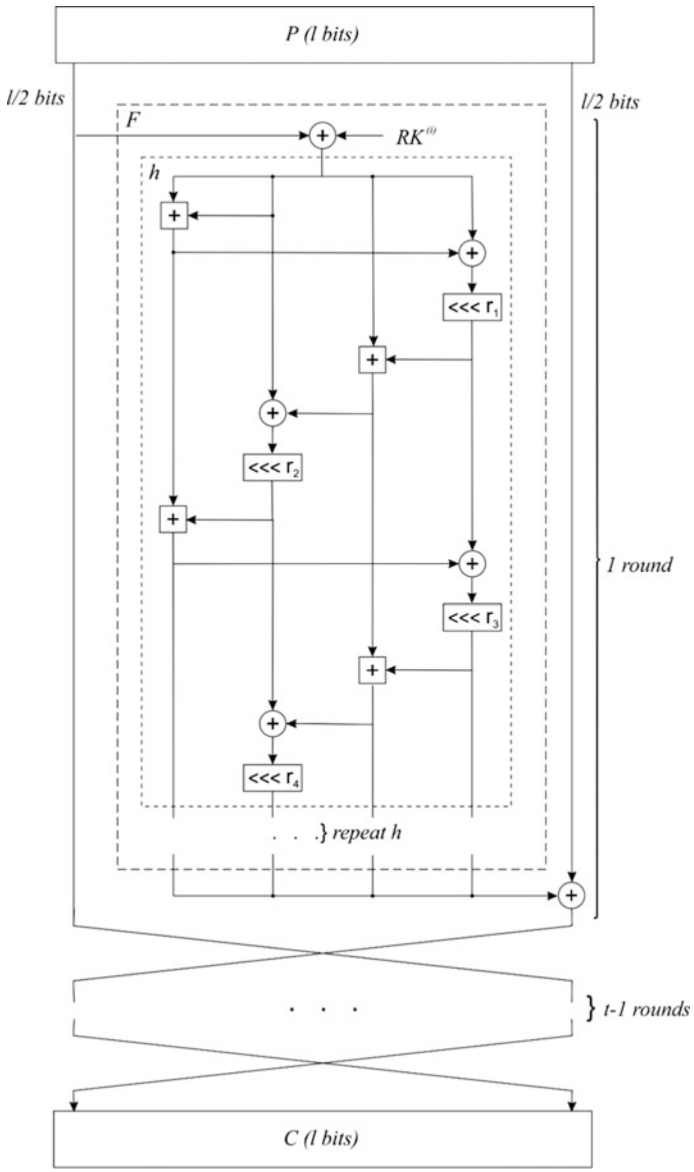


Fig. 1 The encryption procedure of Cypress

where

$ADD(x, y)$ an addition modulo s of two s-bit words;
$XOR(x, y)$ an addition modulo 2 of two s-bit words;
$ROTL(x, r)$ an r-bit left rotation of an s-bit word

The rotation values (r_0, r_1, r_2, r_3) depend on the block size and equal:

- for Cypress-256 $(r_0, r_1, r_2, r_3) = (16,12,8,7)$;
- for Cypress-512 $(r_0, r_1, r_2, r_3) = (32,24,16,15)$.

The decryption algorithm is identical to the encryption one, but round keys are applied in the reverse order.

3.4 The Key Schedule

Cipher's round keys are formed using a non-injective key schedule based on Kalyna one [3]. It is shown that the complexity of exhaustive search attacks against non-injective key schedules does not decrease in comparison with injective ones. Moreover, non-injective key scheduled provide additional protection against side channel attacks and some others.

The encryption uses t round keys with a length of $4 \times s$ bit. Firstly, the intermediate key K_σ with a length of $4 \times s$ bit is produced. The key formation is implemented using an encryption key $K = (K_0, K_1, \ldots, K_7)$ in the following way:

$$
\begin{aligned}
&K_l = (K_0, K_1, K_2, K_3), \quad K_r = (K_4, K_5, K_6, K_7),\\
&st = XOR(0x1, K_l), \quad st = h(h(st)),\\
&st = ADD(st, K_r), \quad st = h(h(st)),\\
&st = XOR(st, K_l), \quad K_\sigma = st.
\end{aligned}
$$

Round keys are formed basing on the encryption key K, the intermediate key K_σ and a constant tmv = (0x000F000F, 0x000F000F, 0x000F000F, 0x000F000F) in the following way:

$$
\begin{aligned}
&st = (K_0, K_1, K_2, K_3), \quad K_t = ADD(K_\sigma, tmv),\\
&st = ADD(st, K_t), \quad st = h(h(st)),\\
&st = XOR(st, K_t), \quad st = h(h(st)),\\
&st = ADD(st, K_t), \quad RK_{2i} = st,\\
&tmv = ShiftLeft(tmv, 0x1),\\
&st = (K_4, K_5, K_6, K_7), \quad K_t = ADD(K_\sigma, tmv),\\
&st = ADD(st, K_t), \quad st = h(h(st)),\\
&st = XOR(st, K_t), \quad st = h(h(st)),\\
&st = ADD(st, K_t), \quad RK_{2i+1} = st,
\end{aligned}
$$

where the function *ShiftLeft* includes the 1-bit left shift of each s-bit word.

After the formation of every pair "even/odd key", the values of *tmv* та *K* are modified in the following way:

$$tmv = ShiftLeft(tmv, 1),$$
$$K = ROTLKey(K, 1),$$

where

ROTLKey(*K*,1) a left rotation of an array of *s*-bit words

3.5 Cypress Statistical Properties

The statistical properties of the encrypting transformation and the key schedule were evaluated according to NIST STS methods [15]. To evaluate the statistical properties, input sequences with high redundancy were chosen.

To test the encrypting transformation it was chosen the sequence of plaintexts $m_0 = 0$, $m_1 = 1$, …, $m_i = i$, $m_n = 100{,}000$ (with high redundancy). The output sequence of ciphertexts (that was obtained in ECB mode) was tested according to NIST STS.

The statistical profiles of the output sequences of encrypting transformation for 256- and 512-bit block size are presented in Figs. 2 and 3.

To test the key schedule it was chosen the sequence of encryption keys $K_0 = 0$, $K_1 = 1$, …, $K_i = i$, $K_n = 100{,}000$ (with maximum redundancy). The obtained round keys formed the output sequence that was tested according to NIST STS.

The statistical profiles of the output sequences of round keys for 256- and 512-bit key length are presented in Figs. 4 and 5.

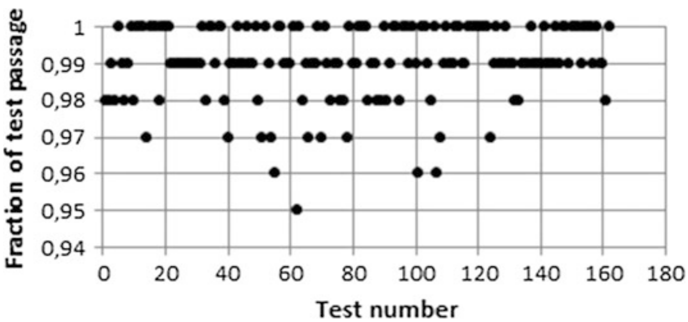


Fig. 2 The statistical profile of the encrypting transformation with 256-bit block size

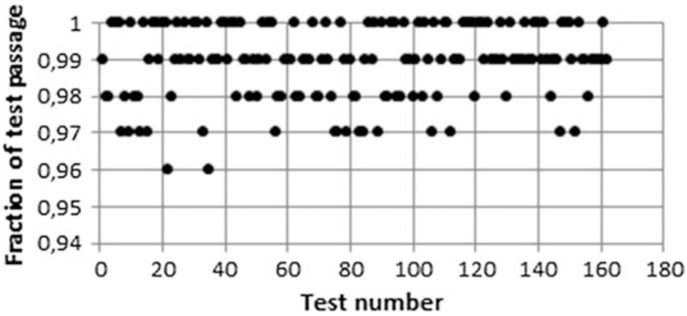


Fig. 3 The statistical profile of the encrypting transformation with 512-bit block size

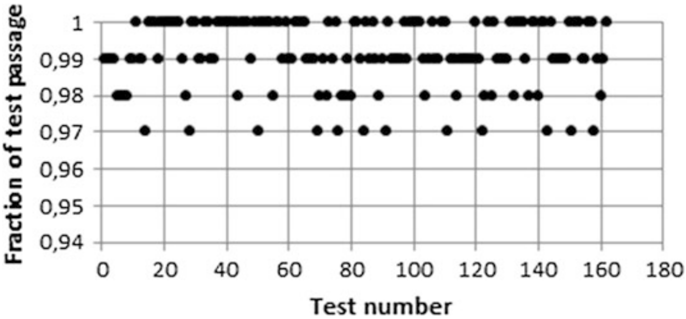


Fig. 4 The statistical profile of the key schedule with 256-bit key length

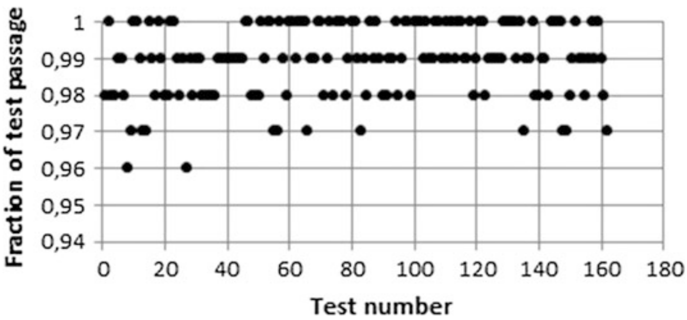


Fig. 5 The statistical profile of the key schedule with 512-bit key length

3.6 Cypress Avalanche Effect Properties

The avalanche effect [16] is a cipher property which means that changing a small number of bits in a plaintext leads to "avalanche" change of bits in a ciphertext. If a block cipher does not have enough avalanche effect, an attacker can make an assumption about input information relying on output information, so the avalanche effect achievement is an important goal while designing a cipher.

It is considered that an algorithm satisfies the avalanche effect if changing one bit of a plaintext leads to a change not less than a half bits in a ciphertext.

To evaluate the avalanche effect of Cypress the following indicators were calculated:

- the expected value (EV) of a number of output bits that were changed when changing one input bit, minimum among all input bits, for N data blocks;
- the expected value of a number of output bits that were changed when changing one input bit, maximum among all input bits, for N data blocks;
- the standard deviation (SD) of a number of output bits that were changed when changing one input bit, minimum among all input bits, for N data blocks;
- the standard deviation of a number of output bits that were changed when changing one input bit, maximum among all input bits, for N data blocks.

The results of the avalanche indicators calculation for Cypress-256 and Cypress-512 are presented in Tables 5 and 6.

As can be seen from Tables 5 and 6, both Cypress-256 and Cypress-512 satisfy the avalanche effect after the 4th round.

3.7 Cypress Performance Analysis

Using different operation systems and hardware, the performance of Cypress-256 and Cypress-512 was evaluated and compared against AES performance.

The speed evaluation was made on the following platforms:

- Intel Core i3/Windows 7 x32 with Visual C++ 2010 compiler;
- Intel Core i3/Windows 7 x64 with Visual C++ 2010 compiler;
- Intel Core i5/Linux (64 bit) with g++ v.4.8 compiler;
- ARM Cortex-A7/Android 4.2.2 Jelly Bean (32 bits).

The results of ciphers speed evaluation on different platforms are presented in Table 7.

As can be seen from Table 7, the block cipher Cypress is faster than AES on all of the selected platforms. On the x86 platform with 32-bit architecture Cypress-256 is 2.5 times faster than AES-256. On the x86 platform with 64-bit architecture Cypress-512 is approximately 3 times faster than AES-256. On the ARM Cortex-A7 platform Cypress-256 and Cypress-512 is approximately 3 times faster than AES-256.

Table 5 The avalanche indicators of Cypress-256

Number of rounds of encryption	Indicator	Value	Number of rounds of encryption	Indicator	Value
1	EV minimum	1	6	EV minimum	127,941
	EV maximum	65,0254		EV maximum	128,078
	SD minimum	0		SD minimum	63,2596
	SD maximum	49,8347		SD maximum	64,7305
2	EV minimum	62,3417	7	EV minimum	127,94
	EV maximum	128,016		EV maximum	128,066
	SD minimum	32,1742		SD minimum	63,1499
	SD maximum	81,6093		SD maximum	64,7323
3	EV minimum	125,375	8	EV minimum	127,927
	EV maximum	128,06		EV maximum	128,064
	SD minimum	63,3573		SD minimum	63,2817
	SD maximum	82,1095		SD maximum	64,7861
4	EV minimum	127,929	9	EV minimum	127,924
	EV maximum	128,079		EV maximum	128,072
	SD minimum	63,2875		SD minimum	63,2531
	SD maximum	64,6699		SD maximum	64,7662
5	EV minimum	127,926	10	EV minimum	127,941
	EV maximum	128,09		EV maximum	128,075
	SD minimum	63,2186		SD minimum	63,2797
	SD maximum	64,8311		SD maximum	64,778

Table 6 The avalanche indicators of Cypress-512

Number of rounds of encryption	Indicator	Value	Number of rounds of encryption	Indicator	Value
1	EV minimum	1	8	EV minimum	255,889
	EV maximum	161,034		EV maximum	256,096
	SD minimum	0		SD minimum	126,625
	SD maximum	417,279		SD maximum	129,672
2	EV minimum	98,0896	9	EV minimum	255,9
	EV maximum	256,052		EV maximum	256,078
	SD minimum	63,6531		SD minimum	126,646
	SD maximum	481,193		SD maximum	129,51
3	EV minimum	225,032	10	EV minimum	255,911
	EV maximum	256,081		EV maximum	256,109
	SD minimum	126,813		SD minimum	126,321
	SD maximum	482,083		SD maximum	129,6

(continued)

Table 6 (continued)

Number of rounds of encryption	Indicator	Value	Number of rounds of encryption	Indicator	Value
4	EV minimum	255,911	11	EV minimum	255,912
	EV maximum	256,084		EV maximum	256,1
	SD minimum	126,4		SD minimum	126,691
	SD maximum	129,707		SD maximum	129,297
5	EV minimum	255,915	12	EV minimum	255,909
	EV maximum	256,1		EV maximum	256,138
	SD minimum	125,956		SD minimum	126,219
	SD maximum	129,338		SD maximum	129,697
6	EV minimum	255,862	13	EV minimum	255,895
	EV maximum	256,096		EV maximum	256,103
	SD minimum	126,708		SD minimum	126,53
	SD maximum	129,394		SD maximum	129,524
7	EV minimum	255,915	14	EV minimum	255,89
	EV maximum	256,102		EV maximum	256,108
	SD minimum	126,838		SD minimum	126,607
	SD maximum	129,639		SD maximum	129,597

Table 7 The performance of Cypress and AES, Mbit/s

Platform	Cypress-256	Cypress-512	AES-256
Intel Core i3/Windows 7 x32	1796,86	786,24	711,13
Intel Core i3/Windows 7 x64	1878,5	2617,74	858,77
Intel Core i5/Linux (64 bit)	3954,55	5395,81	1969,65
ARM Cortex-A7/Android 4.2.2 (32 bit)	122	136	43

4 Conclusions

Modern block symmetric ciphers play an important role in ensuring the security of information systems and technologies, and they are applied in various cryptographic applications standalone as well as a part of other cryptographic primitives. Modern block ciphers must satisfy to different requirements that characterize their ability to perform functional tasks, for example: security against various cryptanalytic attacks; good statistical properties of the encryption algorithm; reliability of mathematical base; practical security of the encryption algorithm; absence of weak and equivalent keys; small complexity of direct and inverse transformations; transparency and openness of the design and structure of the cipher; versatility and flexibility of the cryptographic algorithm, etc. For designing block ciphers today are used various methodological approaches which based on the use of SPN structures, IDEA-like schemes(based on the Lai-Massey scheme), Feistel chains structure.

For devices with constrained resources, it is important to use cryptographic operations with limited energy consumption. Application of a lightweight symmetric key algorithm reduces energy consumption for end-use devices.

The appearance of lightweight cryptography technology became a competition to proven international standards of encryption through higher speeds and at the same time providing an adequate level of security. The use of lightweight cryptography in communication channels is likely to significantly increase their speed, however, there are doubts about strength of ciphers of lightweight cryptography, which need to be resolved by conducting cryptanalysis and further studies of such ciphers for use in low-power hardware. In addition to this, should be taken account of the possible risks and threats of the appearance of quantum computing systems, that is, to focus on the post-quantum period of the use of cryptographic information security.

Taking into account the modern requirements to block ciphers, a post quantum high-The cipher is based on Feistel network, the round function is an ARX-transformation, and the key schedule is non-injective (uses the design principles of Kalyna key schedule). The algorithm supports 256- and 512-bit block size and key length that provide the required security level. Cypress-256 is oriented to be used on 32-bit platforms and Cypress-512—on 64-bit ones.

The research of the statistical properties has shown that Cypress and its key schedule satisfy the requirements of random sequences statistical testing of NIST STS.

The research of the avalanche indicators has shown that both Cypress-256 (the total number of rounds is 10) and Cypress-512 (the total number of rounds is 14) satisfy the requirements of the avalanche effect after the 4th round.

The performance evaluation was carried out on x86 (x86_64) platform with Windows and Linux operation systems (OS) and ARM-v7 platform with Android OS. On the different platforms Cypress has the following performance:

- on x86 platform with 32-bit architecture Cypress-256 is 2.5 times faster than AES-256;
- on x86 platform with 64-bit architecture Cypress-512 is approximately 3 times faster than AES-256;
- on ARM-v7 platform with 32-bit architecture Cypress-256 is approximately 3 times faster than AES-256.

Thus, compared with Kalyna and AES, the block cipher Cypress has the following advantages:

- a high performance regardless of the platform used;
- an existence of two variants of the algorithm optimized both for 32- and 64-bit platforms;
- a compact implementation both for desktop systems and mobile platforms;
- a high-speed implementation does not require the usage of precomputation tables that provides the minimal memory usage;

- a possibility to organize effective secure high-speed communication channels between mobile systems and servers including those that use hardware accelerators;
- constant block encryption time on modern CPUs regardless specific values of plaintexts and encryption keys being processed.

References

1. Doukas, N.: Technologies for greener internet of things systems. Green IT Engineering: Components, Networks and Systems Implementation, pp. 23–42. Springer, Cham (2017)
2. Chemeris, A., Dmitri, L.: Sergey S (2017) Influence of software optimization on energy consumption of embedded systems. Green IT Engineering: Components Networks and Systems Implementation, pp. 111–133. Springer, Cham (2017)
3. DSTU 7624:2014, Information technology, Cryptography protection of information, The block symmetric conversion algorithm, Kiev: Ministry of Economic Development of Ukraine, 2015, 238 p
4. Pub, N.I.S.T.F.I.P.S.: 197: Advanced encryption standard (AES). Fed. Inf. Process. Stand. Publ. **197**(441), 0311 (2001)
5. Rani, D.J, Roslin, S.E.: Light weight cryptographic algorithms for medical internet of things (IoT)-a review. In Green Engineering and Technologies (IC-GET), 2016 Online International Conference on IEEE (2016)
6. Katagi, M., Moriai, S.: Lightweight Cryptography for the Internet of Things. Available: https://www.researchgate.net/publication/267246530_Lightweight_Cryptography_for_the_Internet_of_Things [May 2012]
7. Information Technology, Security Techniques, Lightweight Cryptography, Part 2: Block ciphers (ISO/IEC 29192-2), 41 p
8. GOST 28147-89, Systems of information processing, cryptography security, cryptography conversion algorithm, Moscow: Publishing standards, 1989, 20 p
9. Rodinko M.Yu., Oliynykov R.V. A post quantum lightweight symmetric block cipher "Cypress" Radiotechnics: All-Urk. Sci. Interdep. Mag. 2017. N. **189**: 100–107
10. Information technology, Security techniques, Encryption algorithms, Part 3: Block ciphers (ISO/IEC 18033-3), 80 p
11. Schneier, B.: Description of a new variable-length key, 64-bit block cipher (Blowfish). In Fast Software Encryption—Proceedings of the Cambridge Security Workshop, Cambridge, United Kingdom, Lectures Notes in Computer Science 809, Springer (1994)
12. Gorbenko, I.D., Yu, I.: Gorbenko, Prykladna kryptolohiya. Teoriya. Praktyka. Zastosuvannya: monohrafiya [Applied cryptology. Theory. Practice. Application: monograph.]. Kharkiv, Ukraine: Fort (2012) (in Ukrainian)
13. GOST R 34.12-2015, Information technology, Cryptography protection of information, Block ciphers, Moscow: Standartinform, 25 p (2015)
14. STB 34.101.31-2011, Information technology and security, Information security, Cryptography encryption and integrity control algorithms, Minsk: Gosstandart 35 p (2011)
15. A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications. NIST Special Publication 800-22/[Rukhin A. and others]—Revision 1a, April (2010)
16. Feistel, H.: Cryptography and computer privacy. Sci. Am **228**(5), 15 (1973)

Lightweight Stream Ciphers for Green IT Engineering

Olexandr Kuznetsov, Olexandr Potii, Artem Perepelitsyn, Dmytro Ivanenko and Nikolay Poluyanenko

Abstract At the moment the most widely used are symmetric cryptographic transformations, in particular, stream ciphers. The development of an efficient synchronous stream cipher is reduced to the construction of a pseudo-random sequence generator with defined cryptographic properties. It should be noted that in devices with limited computing power, low volume and low power consumption the implementation of reliable cryptographic methods is extremely complicated. Limited physical parameters, low power consumption, low computing power and other characteristic attributes of “green” IT engineering forces the use of new approaches for designing cryptographic protection tools. The main cryptographic transformations are considered and experimental studies of performance and statistical security are conducted. We propose new methods and hardware and software tools for lightweight stream encryption that meet the current requirements of “green” IT engineering. It is proposed synthesis method for the construction of nonlinear-feedback shift register, which allows creating nonlinear registers with design features that correspond to the certain predefined criteria.

Keywords Green IT engineering · Internet of things · Stream cipher

O. Kuznetsov (✉) · O. Potii · D. Ivanenko · N. Poluyanenko
Information Systems and Technologies Department, Computer Science Faculty,
V.N. Karazin, Kharkiv National University, Kharkiv, Ukraine
e-mail: kuznetsov@karazin.ua

O. Potii
e-mail: potav@ua.fm

D. Ivanenko
e-mail: i8o.dima@gmail.com

N. Poluyanenko
e-mail: nlfsr01@gmail.com

A. Perepelitsyn
Computer Systems Networks and Cybersecurity Department,
National Aerospace University «Kharkov Aviation Institute», Kharkiv, Ukraine
e-mail: a.perepelitsyn@khai.edu

© Springer Nature Switzerland AG 2019

V. Kharchenko et al. (eds.), *Green IT Engineering: Social, Business and Industrial Applications*, Studies in Systems, Decision and Control 171,
https://doi.org/10.1007/978-3-030-00253-4_6

1 Introduction

In the first two parts of Green IT Engineering [1, 2], conceptual issues, models and architectures of building environmentally clean information technologies, as well as sectoral and public aspects of environmental engineering were considered. In particular, special attention was paid to the issues of building green Internet computing, cloud and IoT [1], Green IT for industry and intelligent networks [2]. At the same time, issues of building systems and means of cyber security remain open and unexplored, to which, in addition to the traditional requirements of durability and reliability, conditions for low power consumption, restrictions on physical parameters and processing power are also advanced.

In the field of information security, the sphere of Green IT Engineering is green cryptography. In green cryptography two main paradigms are distinguished.

The first concept is the concept of recycling cryptographic design. The second concept is Lightweight cryptography.

In the framework of the first concept, green cryptography is a school of thought that observes a natural progression of cryptographic primitives, from the moment they're designed to the moment they're deployed; it is realized through recycling, where design strategies, components, and primitives, are recycled, supported by a twofold argument: Firstly, if we are comfortable, cryptanalytically-speaking, with a design strategy, component, or primitive, we should attempt to recycle it. This makes sense, because the most compelling reason to use a design strategy, component, or primitive is that it has "earned its bones," cryptanalytically. Secondly, to ensure the simplicity of implementations, we should recycle primitives, whenever and wherever possible. This also makes sense, because complexity is the culprit behind nearly all instances of cryptography failing in practice. Concisely, recycling-based green cryptography is about maximizing confidence in cryptographic primitives while minimizing complexity in their implementation [3].

Lightweight cryptography is characterized by constraints on the implementation characteristics of the cryptography, including the resource cost, the performance characteristics, and the security. Among the factors currently considered in a profile are gate count (GE) for hardware implementations, memory footprint for software implementations, latency and throughput, power dissipation, security level, attack models, and implementation attack resistance. Lightweight cryptography is an active research direction, as witnessed by the number of algorithms aiming at "low-cost" implementations designed over the last years [4, 5].

2 Formalization Stream Crypto Transformation

The fourth part of the international standard ISO/IEC 18033 describes the algorithms of stream symmetric encryption [6]. According to the definition given in the standard, under the stream cipher it is a cryptographic mechanism that uses a keystream to encrypt the plaintext by bit or by block. There are two types of streaming conversions: a synchronous stream cipher, in which the keystream is generated only from the secret key (and the initialization vector), and the asynchronous stream cipher (self-synchronizing), in which the keystream is generated from the secret key and some past encryption texts (and the initialization vector). The Keystream according to ISO/IEC 18033-4 is a pseudorandom sequence kept secret and is used by encryption and decryption algorithms of stream ciphers. If a part of the key thread is known to the attacker, then it must be computationally impracticable for the attacker to receive any information about the remainder of the key thread. Consider the generic models for stream ciphers, as defined in ISO/IEC 18033-4, consisting of a combination of a key generator and a source function.

2.1 Model of Synchronous Stream Symmetric Cipher

Let X_i, C_i, Z_i, i, S_i denotes the characters of the open text, ciphertext, the keystream and the internal state at the time i, respectively. The secret key K, specifies the specific kind of cipher transformation and is selected according to a certain probabilistic distribution. Typically, the key is selected according to a uniform distribution, but in some cases, it may not be possible to select the key at random.

In general, the synchronous flow symmetric cipher can be described by equations

$$Z_i = f(S_i, K, IV),\ C_i = \varphi(X_i, S_i, K, IV)$$

where

f key generation function,
φ output function,
IV initialization vector.

This condition $C_i = X_i + f(S_i, K, IV)$ allows you to implement streaming encryption-decryption without delay. In this case, streaming encryption is served according to the scheme shown in Fig. 1. The sequence $\{Z_i = f(S_i, K, IV)\colon i \geq 1\}$ is called the keystream. To ensure stable streaming encryption, the keystream should be random.

2.2 The Model of the Synchronous Keystream Generator in Accordance with ISO/IEC 18033-4

The synchronous keystream generator is a finite automaton, which is defined in the following way (Fig. 2).

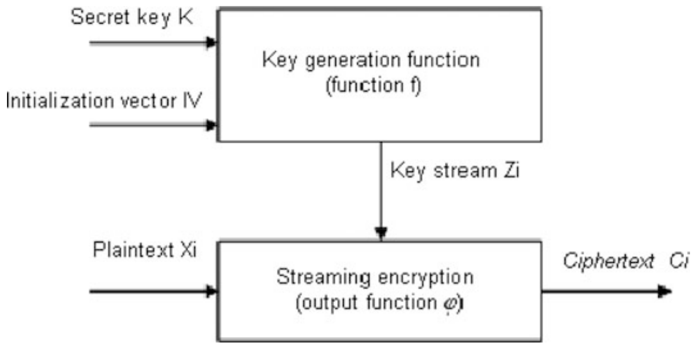


Fig. 1 The general scheme of synchronous stream ciphering

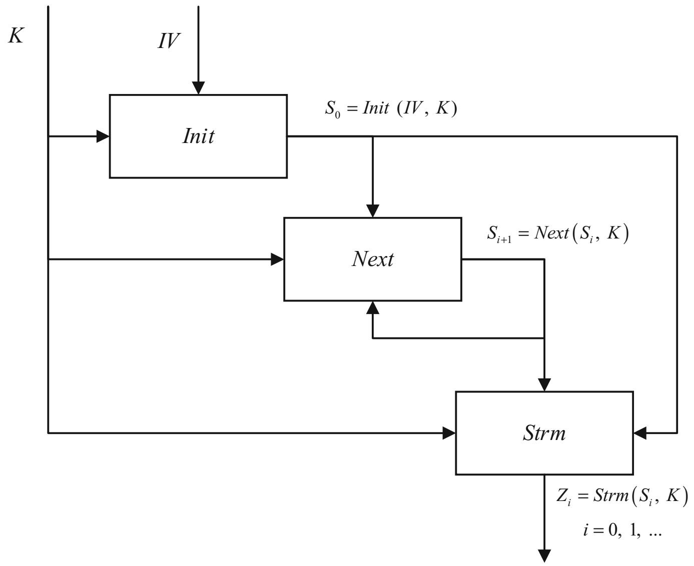


Fig. 2 General scheme of generation of the keystream in the synchronous cipher

1. *The initialization function*, *Init*. Input: key K and initialization vector IV. Oputput: the initial state S_0 for the keystream generator.
2. *The the next state function, Next*. Input: Current state of the keystream generator S_i. Output: next state of the keystream generator S_{i+1}.
3. *The keystream function, Strm*. Input: state of the keystream generator S_i. Output: block of keystream Z_i.

When a synchronous key generator is initiated for the first time, it will enter the initial state S_0, defined as

$$S_0 = Init(IV,\ K).$$

On request, the synchronous key generator for $i = 0,\ 1,\ \ldots$ will perform:

1. The output of keystream block $Z_i = Strm(S_i,\ K)$;
2. Update the state of the finite automaton $S_{i+1} = Next(S_i,\ K)$.

Therefore, to define the synchronous key generator, it is necessary to describe the functions *Init*, *Next* i *Strm*.

The main advantage of synchronous ciphers is the absence of a multiplication error. Only the distorted data transmission bit (or block) will be decrypted incorrectly. The transmission and receiving side synchronization is provided by the insertion of an additional sequence of data (markers). This allows for enhanced imitation protection since the insertion of false information will necessarily lead to a violation of synchronization and will be detected on the receiving side. The disadvantages of synchronous ciphers include some vulnerability to attacks with the selected open source text. If an attacker know some symbols (or blocks) of plaintext and corresponding ciphertext, he can change the ciphertext so that it will be correctly encryption.

2.3 The Model of Self-synchronizing Keystream Generators

In general, the asynchronous symmetric stream cipher can be described by the equations:

$Z_i = f(C_{i-1},\ C_{i-2},\ \ldots,\ C_{i-r}, S_i, K, IV),\ C_i = \varphi(X_i, C_{i-1},\ C_{i-2},\ \ldots,\ C_{i-r}, S_i, K),$
here $C_{i-1},\ C_{i-2},\ \ldots,\ C_{i-r}$—sequence of r ciphertext blocks at times $i-1,$ $i-2, \ldots, i-r$.

Therefore, the value of the keystream at a certain point in time is determined by the values of the key, the initialization vector, the internal state, and the values of the r ciphertext blocks at the previous moments of time.

In this case, stream encryption $C_i = X_i + f(C_{i-1},\ C_{i-2},\ \ldots,\ C_{i-r}, S_i, K, IV)$ submit the scheme, which is shown in Fig. 3.

2.4 The Model of the Self-synchronizing Keystream Generator ISO/IEC 18033-4

Generating a keystream for asynchronous stream ciphers depends on the previous encrypted texts, the key, the initialization vector. ISO/IEC 18033-4 defines a common model for the asynchronous stream cipher keystream generator (Fig. 4).

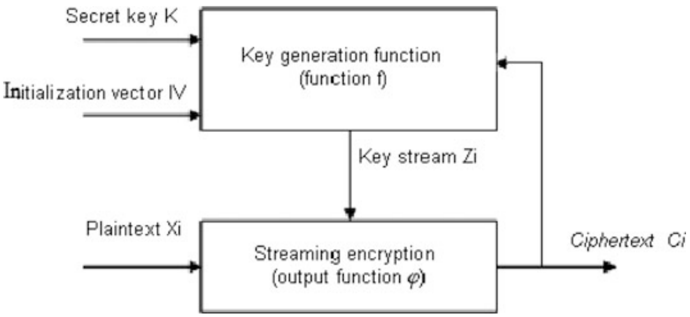
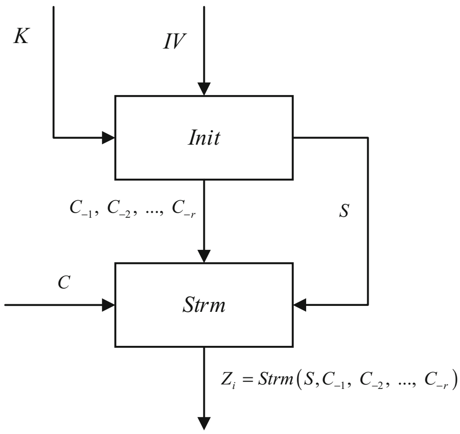


Fig. 3 Asynchronous streaming encryption scheme

Fig. 4 General scheme of generating a keystream in the asynchronous cipher



1. *The initialization function*, *Init*, which takes as input a key K and an initialization vector IV and outputs an internal input for the keystream generator S and r dummy ciphertext blocks $C_{-1}, C_{-2}, \ldots, C_{-r}$.
2. *The keystream function*, *Strm*, that takes as input S and r ciphertext blocks $C_{-1}, C_{-2}, \ldots, C_{-r}$, and outputs a keystream block Z_i.

To define a self-synchronizing keystream generator it is only necessary to specify the number of feedback blocks r and the functions *Init* and *Strm*.

The advantages of asynchronous ciphers include the dissemination of statistics on all plaintext ciphertext. Indeed, since each bit (or block) plaintext affects the value of all the bits (blocks) ciphertext, the statistical properties of the plaintext apply to all ciphertext. So it becomes difficult to realize the attack, based on the use of redundancy plaintext. Lack of asynchronous ciphers is an expansion of mistakes. Truly, if an error occurs when sending a bit or ciphertext block, this error spreads to the next r blocks that will be decrypted incorrectly. If there are many errors, then the use of asynchronous ciphers is inappropriate.

The most widely used in modern information and telecommunication systems are synchronous stream ciphers. To construct them it is enough to determine the PRNG, whose cryptographic properties are based on the stability of streaming encryption. In this paper, we consider the most well-known generators and analyze their efficiency, in particular, investigate the statistical properties of moulded PRNG and the complexity of practical implementation.

3 Review of Stream Ciphers

In order to conduct comparative studies of the efficiency of stream ciphers, world-known cryptographic algorithms are selected. These algorithms are the most trusted and distributed. In particular, keystreams of stream ciphers HC-128, HC-256, Grain, MICKEY, MUGI, Rabbit, Salsa20, SNOW 2.0, Sosemanuk and block symmetric were tested. AES encryption with a key length of 128 and 256 bits. The purpose of the testing was to study the statistical indicators of security and performance, compared with corresponding indicators lightweight stream cipher «Enocoro» and «Trivium», as defined in the third part of the international standard ISO/IEC 29192.

Stream symmetric cipher "HC-256" was developed in 2004 [7]. HC-256 is a simple, secure, software-oriented cipher with effective implementation and can be used freely. The simplified version of HC-128 was presented to eSTREAM. 256-bit key and 256-bit initialization vector are used for initialization. Recommended maximum length of the key sequence is 2^{128}.

Stream symmetric cipher "Grain" presented by Martin Hall, Thomas Johansson and Willie Meyer in 2004 at the international eSTREAM competition for the second profile (hardware-oriented ciphers) [7]. Symmetric algorithm for synchronous current encryption, which is aimed at using computers with a limited number of gate, low power and memory. Depending on the hardware implementation, Grain can be bit-oriented or word-oriented. At the entrance to the Gran v1 is a 80-bit key and a 64-bit initialization vector. At the heart of the design of the algorithm are 2 offset registers—with linear and nonlinear feedback and the output function. Recommended length of the keystream, which can be generated in one key/vector pairs—2^{44} bit.

***Stream symmetric cipher "Mickey* 2.0**" was presented in 2005 by Steve Babbage and Matthew Dodd [7]. The cipher is intended for hardware platforms with limited resources, that is, the streaming code MICKEY was developed as a hardware-oriented cipher. To initialize the initial state, a key length of 80 bits and an initialization vector of up to 80 bits are used. The maximum possible length of the keystream is 2^{40} bits per key, but with the use of different initialization vectors of one length. The MICKEY encryption algorithm has a simple hardware implementation, but it provides a high level of security. Due to the use of irregular movement of registers of displacement, as well as new methods, high resistance to certain cryptanalytic attacks is ensured.

Stream symmetric cipher «MUGI» is a key thread generator, which was recommended by the CRYPTREC project for use by the Government of Japan in 2003 [8]. The algorithm has been standardized in ISO/IEC 18033-4. As the initial data, MUGI uses a 128-bit secret key, a 128-bit initialization vector. MUGI uses non-linear substitution blocks and linear transformations using the MDS matrix of the AES algorithm. As the initial data, MUGI uses a 128-bit secret key, a 128-bit initialization vector. MUGI uses non-linear substitution blocks and linear transformations using the MDS matrix of the AES algorithm. Basic cipher designs are similar to Panama cipher designs. The MUGI is a word-oriented cipher.

Stream symmetric cipher «Rabbit». The developers of the algorithm are Martin Boesgaard, Matthew Vesterager, Thomas Pedersen, Jesper Christian and Ove Skavinius [8]. In May 2005, this cipher was presented at eStream as a software-oriented algorithm. The algorithm uses a 128-bit key and a 64-bit initialization vector. On a single key/vector pair, up to 267 bits of the key thread can be generated.

Stream symmetric cipher «Salsa20» was developed by Daniel Bernstein [7]. The algorithm became the winner of the eSTREAM contest in the class of software-oriented algorithms. To initialize the internal state using a key length of 256 bits, 64-bit nonce and a 64-bit keystream block position. The maximum length of the pseudo-random key sequence is 2^{70} bits.

Stream symmetric cipher «SNOW 2.0» is a keystream generator [7], which uses 128 or 256-bit secret key K as input and 128-bit initialization IV as input. Cipher is a word-oriented. The authors of the algorithm are Thomas Johanson and Patrick Ekdal. The algorithm has been standardized in ISO/IEC 18033-4. For SNOW 2.0, the maximum recommended number of bits of the keystream, produced in one pair (K, IV), is $23 * 2^{50}$ bits. This restriction is justified in terms of ensuring the security of the algorithm against cryptanalytic attacks.

Stream symmetric cipher «Sosemanuk»—it is a synchronous software-oriented stream cipher [7]. The length of the key is 128 and 256 bits. Cipher works with 128 bit initial value, is provided with 128-bit security. Sosemanuk algorithm uses some of the basic principles of the streaming cipher «SNOW 2.0» and some transformations of the block cipher SERPENT [6].

Block symmetric cipher «AES». The cipher is the national standard of the USA (FIPS-197) [9] and the international standard ISO/IEC 18033-3 [10]. Cipher uses a key length of 128, 192, or 256 bits. There are 10, 12 or 14 rounds of encryption depending on the length of the key. AES has a quick hardware and software implementation. The cipher is used as a stream cipher in the CFB mode.

Stream symmetric cipher «Enocoro». This is a hardware-oriented cryptographic algorithm [11]. Standardized in the third part of ISO/IEC 29192-3 [12]. This is a byte-oriented cipher with a 128-bit key length and a 64-bit initialization vector. Despite the fact that Enocoro is a hardware-oriented cipher, it also has an effective software implementation. To achieve different requirements, byte operations are used.

The structural scheme of the Enocoro key generator is shown in Fig. 5. *The intrinsic station* consists of two bytes a_0, a_1. Also used buffer b, which is stored in nb bytes b_0, b_1,…, b_{nb-1}. The function of state update use $b_{k_1}, \ldots, b_{k_4}$ as input parameters. It consists of appropriate S-box s8, linear transformation L over the field GF (28) and operations of bitwise addition (XOR)., At the beginning, the initialization function sets the registers of the key value K, the initialization vector IV, and the internal state C (certain constants).

Stream symmetric cipher "Trivium" is a symmetric hardware-oriented stream cipher. The authors of the cipher are Christoph de Cannierre and Bart Prenel. Standardized in ISO/IEC 29192-3 [12]. Trivium is the easiest eSTREAM project cipher (second profile) which demonstrates excellent cryptoscope results. According to the specification, the Trivium algorithm is a stream cipher designed to generate 2^{64} bits of a keystream with 80 bits of a secret key and 80 bits of initialization vector. The cipher is bit-oriented.

The encryption scheme contains 288 bits of the internal state and is indicated by (s_1, ..., s_{288}). The process of generating a keystream is iterative and uses the value of 15 special bits of states to update the 3 bits of the state and count 1 bits of the keystream z_i. The algorithm is initiated by installing an 80-bit key and an 80-bit vector of initialization in a 288-bit internal state. The graphical representation of the process of generating a keystream is depicted in Fig. 6.

The "Enocoro" and "Trivium" algorithms are hardware-oriented stream ciphers whose construction focuses on flexibility. The main goal is compactness in a medium with limited input parameters, energy efficiency on small-energy platforms, and speed in applications that require high-speed encryption.

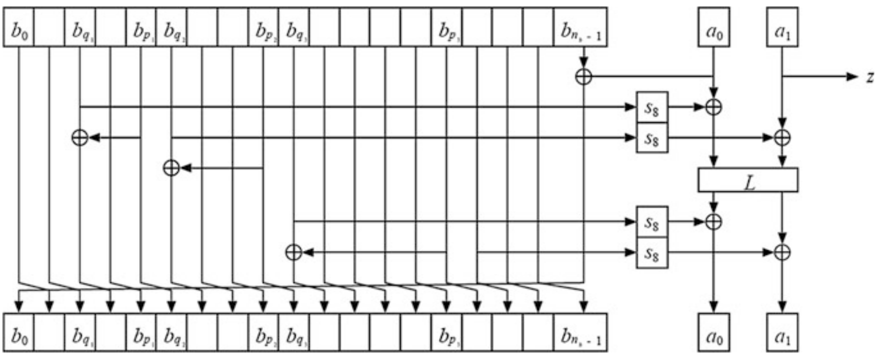


Fig. 5 The structural scheme diagram of the key generator "Enocoro"

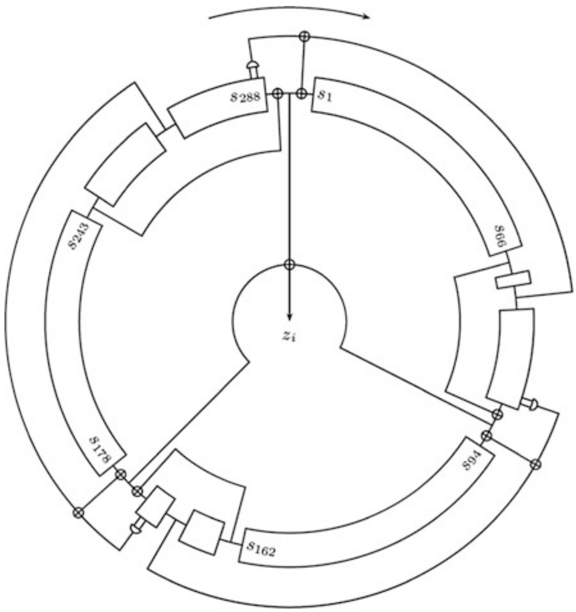


Fig. 6 The scheme of calculation of the function of generation of the keystream "Trivium"

4 Test Results

For conducting experimental studies of cryptographic properties, statistical testing of source sequences (keystream) was performed using the method of NIST STS [13]. The random sequence hypothesis was tested by 15 independent statistical tests (188 tests are performed with different input parameters), each of which calculates the probability of passing the test as: $P_j, j = 1, \ldots, 188$. This probability is used by the rule of judgments (criterion of consent) about the truth or falsity of the hypothesis, that means, if the value of the probability P_j of passing some j test is not below a certain threshold value $\alpha \in [0.96.0.99]$, for example, if $P_j \geq 0.99$, then the hypothesis H_0 is accepted (on some j-test).

To improve the reliability of the test, the mathematical expectation of the number of tests passed by each generator was estimated, with each of the 100 independent tests considered as one observation. The sequences were formed by a pair of random key K/random vector of initialization IV. The results obtained are presented in Table 1:

– "M096" and "M099"—estimates of mathematical expectation (sample average) of the number of passed statistical tests by criterion $P_j \geq 0.96$ and $P_j \geq 0.99$ respectively;
– "D096" and "D099" ("S096" and "S099")—estimates of dispersions (mean square deviations) of the results of testing the number of passed statistical tests according to the criteria $P_j \geq 0.96$ and $P_j \geq 0.99$, accordingly;

Table 1 Results of statistical testing of modern stream ciphers

	M099	D099	S099	P099	M096	D096	S096	P096	MIN
HC-256	133.8	32.2	6.1	1.0	186.7	2.0	1.4	1.00	182
Grain	132.4	57.8	7.6	1.0	186.9	1.4	1.2	1.00	182
Mickey	133.5	62.4	7.9	1.0	186.6	2.3	1.5	1.00	179
MUGI	132.2	53.3	7.3	1.0	186.5	1.0	1.0	1.00	185
Rabbit	132.7	16.0	4.0	1.0	187.2	0.5	0.7	1.00	185
Salsa20	134.2	28.1	5.3	1.0	187.0	1.0	1.0	1.00	183
Snow 2.0	132.8	24.0	4.9	1.0	186.8	0.5	0.7	1.00	183
Sosemanuk	131.7	49.0	7.0	1.0	186.8	2.3	1.5	1.00	184
AES-128	127.1	19.4	4.4	1.0	186.6	0.3	0.6	1.00	185
Enocoro	132.9	51.8	7.2	1.0	187.2	0.8	0.9	1.00	185
Trivium	130.2	98.0	9.9	1.0	187.2	1.4	1.2	1.00	182

- "P099"—the value of the fiducial probability for the number of passed statistical tests by criterion $P_j \geq 0.99$ and for accuracy $\varepsilon = 2$;
- "P096"—the value of the fiducial probability for the number of passed statistical tests by criterion $P_j \geq 0.96$ and for accuracy $\varepsilon = 1$;
- "Min096" is the minimum value of the number of passed statistical tests by criterion $P_j \geq 0.96$.

The high cryptographic indexes of various stream cryptographic transformations are confirmed by the given results of testing. Investigated ciphers (including low-resource cryptography algorithms) showed a high number of successfully passed tests: 130–133 by criterion $P_j \geq 0.99$ and 186–187 by criterion $P_j \geq 0.96$. These estimates are obtained with high reliability ($P_\partial = 0.99$ for $P_j \geq 0.99$ and $P_\partial \approx 1$ for $P_j \geq 0.96$).

To conduct research on the speed of cryptanalysis, we decided to take the already proven methodology for the analysis of stream ciphers, the methodology of the competition eSTREAM. This technique provides three criteria for the analysis of stream ciphers. Though more detailed tests might be added in the future. We will experiment with each criterion anddetermine the best stream cipher by the obtained results.

1. **Criterion—Encryption Rate for Long Streams**. This is where stream ciphers have the biggest potential advantage over block ciphers, and hence this figure is likely to be the most important criterion in many applications. The testing framework measures the encryption rate by encrypting a long stream(1 Gb). The encryption speed, in cycles/byte, is calculated by measuring the number of bytes encrypted in 1 μsec. Note that the time to set up the key or the IV is not considered in this test. The results of the experiment are shown in Fig. 7.
2. **Criterion—Packet Encryption Rate**. This criterion allows you to determine at what length of the packet the stream cipher starts to show its real performance in speed. Since stream ciphers do not have to significantly degrade their outlookers

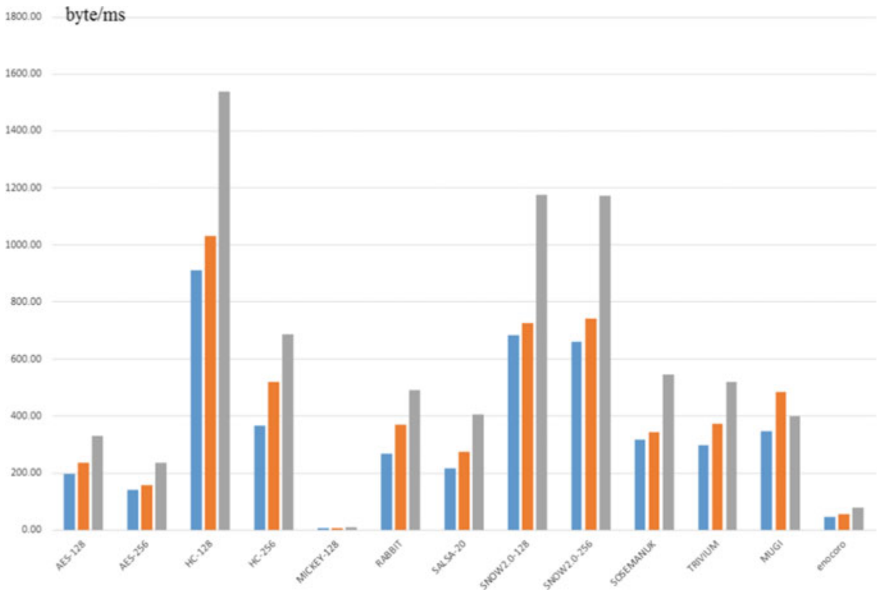


Fig. 7 Encryption rate for long streams

on small packets, we will see the pattern/characteristic of the algorithm structure of each cipher. The packet encryption rate is measured by applying the encrypt_packet function to packets of different lengths. Each call to encrypt_-packet includes a separate IV configuration and, if supported by authenticated encryption, the finalization phase of the MAC. The packet lengths (40, 576 and 1500 bytes) were selected to display traffic on the Internet. The results of the experiment are collected in Tables 2, 3, and 4.

3. **Criterion—Key and IV Setup**. The last test measures the efficiency of the key setup and the IV setup. This is probably the least critical of the three criterions, considering that the efficiency of the IV setup is already reflected in the packet encryption rate and that the time for the key setup will typically be negligible compared to the work needed to generate and exchange the key. The results of the experiment are shown in Fig. 1.

Experimental Setup. All ciphers were implemented using a computer with Intel (R) Pentium(R) CPU P6200 @ 2.13 GHz, 2128, RAM: 2–2 GB (with a frequency of 1333 MHz), Cache: Level 1 (128 KB); Level 2 (512 KB); Level 3 (3072 KB) in the Windows 8.1 operating system Professional 64bit, on the Microsoft Visual Studio 2012 Compiler 32bit version 11.00.50727.1. After the implementation of ciphers was launched on three platforms: Intel Core i7-6820HQ 2.7Gh; Intel Core i7-5500u 2.4Gh and Intel Pentium P6200 2.13Gh (Fig. 8).

The test results showed that according to the criterion of encryption of long messages the best indicators were “HC-128” and “SNOW 2.0”.

Table 2 The experimental results (Intel Pentium P6200 2.13Gh)

Algorithm	Encryption packets								
	50 packets/1500 byte			120 packets/576 byte			350 packets/40 byte		
	Time (µs)	Speed (bytes/µs)	Speed (packets/µs)	Time (µs)	Speed (bytes/µs)	Speed (packets/µs)	Time (µs)	Speed (bytes/µs)	Speed (packets/µs)
HC-128	657	2.282	0.076	1363	0.423	0.088	3673	0.011	0.095
HC-256	4210	0.356	0.012	9658	0.060	0.012	27350	0.001	0.013
Grain	135271	0.011089	0.00037	128127	0.004496	0.000937	44250	0.000904	0.00791
Mickey-128	19010	0.079	0.003	18628	0.031	0.006	8423	0.005	0.042
MUGI	253	5.928854	0.197628	296	1.94792	0.405817	317	0.126143	1.103753
Rabbit	409	3.670	0.122	409	1.408	0.293	211	0.190	1.662
Salsa-20	601	2.496	0.083	541	1.065	0.222	170	0.236	2.062
Snow 2.0-128	220	6.812	0.227	220	2.618	0.545	177	0.226	1.974
Snow 2.0-256	221	6.790	0.226	223	2.582	0.538	177	0.226	1.976
Sosemanuk	285	5.258	0.175	297	1.938	0.404	186	0.215	1.883
AES-128	691	2.170	0.072	621	0.927	0.193	179	0.224	1.957
AES-256	983	1.526	0.051	911	0.632	0.132	245	0.163	1.429
Enocoro	109120	0.998554964	0.033282324	107602.9	0.388853462	0.081011138	50481.743	0.057555835	0.503641259
Trivium	376	3.986	0.133	404	1.427	0.297	313	0.128	1.119

Table 3 The experimental results (Intel Core i7-5500u 2.4Gh)

Algorithm	Encryption packets								
	50 packets/1500 byte			120 packets/576 byte			350 packets/40 byte		
	Time (µs)	Speed (bytes/µs)	Speed (packets/µs)	Time (µs)	Speed (bytes/µs)	Speed (packets/µs)	Time (µs)	Speed (bytes/µs)	Speed (packets/µs)
HC-128	412	3.637	0.121	844	0.682	0.142	2241	0.018	0.156
HC-256	2148	0.698	0.023	4916	0.117	0.024	14225	0.003	0.025
Grain	110360	0.013592	0.000453	85329	0.00675	0.001406	29499	0.001356	0.011865
Mickey-128	10651	0.141	0.005	10462	0.055	0.011470	4963	0.008	0.071
MUGI	177	8.479	0.282646	205	2.809	0.585366	212	0.188	1.65
Rabbit	210	7.156	0.239	209	2.763	0.576	105	0.381	3.333
Salsa-20	262	5.723	0.191	233	2.471	0.515	77	0.517	4.528
Snow 2.0-128	118	12.722	0.424	117	4.906	1.022	101	0.397	3.476
Snow 2.0-256	110	13.599	0.459	120	4.79	0.999	103	0.388	3.395
Sosemanuk	150	10	0.333	171	3.376	0.703	122	0.329	2.876
AES-128	313	4.8	0.16	291	1.983	0.413	91	0.439	3.837
AES-256	432	3.474	0.116	406	1.417	0.295	113	0.354	3.095
Enocoro	90778.49879	1.200311477	0.040006973	89167.65	0.469242421	0.097758838	42299.661	0.0686954	0.601059177
Trivium	211	7.105	0.237	238	2.419	0.504	175	0.228	1.998

Table 4 The experimental results (Intel Core i7-6820HQ 2.7Gh)

Algorithm	Encryption packets								
	50 packets/1500 byte			120 packets/576 byte			350 packets/40 byte		
	Time (µs)	Speed (bytes/µs)	Speed (packets/µs)	Time (µs)	Speed (bytes/µs)	Speed (packets/µs)	Time (µs)	Speed (bytes/µs)	Speed (packets/µs)
HC-128	257	5.846	0.195	546	1.055	0.22	1492	0.027	0.235
HC-256	1420	1.056	0.035	3413	0.169	0.035	9605	0.004	0.036
Grain	108503	0.013824	0.000461	1000586	0.005726	0.001193	34241	0.001168	0.010222
Mickey-128	7819	0.192	0.006	7542	0.076	0.016	3532	0.011	0.099
MUGI	225	6.663705	0.22124	255	2.259	0.470773	248	0.16116	1.410153
Rabbit	154	9.772	0.326	158	3.655	0.761	88	0.453	3.964
Salsa-20	190	7.886	0.263	175	3.299	0.687	58	0.687	6.014
Snow 2.0-128	70	21.337	0.711	71	8.09	1.685	64	0.624	5.46
Snow 2.0-256	71	21.246	0.708	71	8.09	1.685	64	0.628	5.494
Sosemanuk	91	16.575	0.552	102	5.658	1.179	79	0.506	4.4259
AES-128	230	6.525	0.217	211	2.736	0.57	60	0.668	5.843
AES-256	323	4.64	0.155	300	1.917	0.399	81	0.491	4.3
Enocoro	66641.35593	1.635052785	0.0545046	93.75	64280.48	0.650918741	0.135609492	36.00	30103.245
Trivium	153	9.791	0.326	168	3.429	0.714	128	0.312	2.726

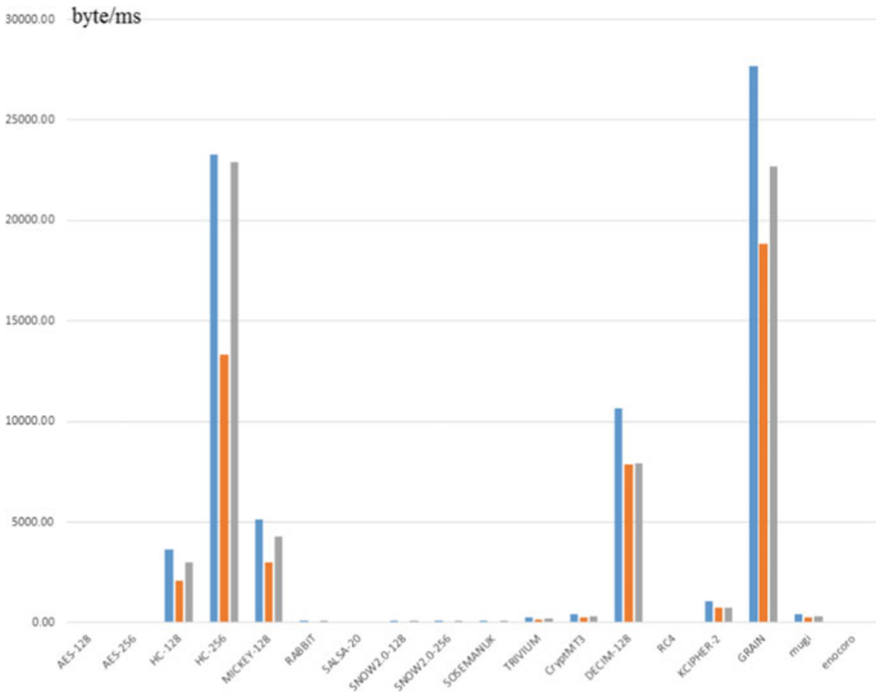


Fig. 8 Research results

The experimental results for the short message encryption criterion have shown that the best performance is in SNOW 2.0 and SOSEMANUK, respectively. SNOW 2.0 and SOSEMANUK have better performance when there are many, but small, packages. According to the criteria of the KEY and IV setup shown in Fig. 2, we can conclude that the worst indicators in the NS-128, HC-256, MICKEY-128. This explains the poor performance of the NS-128 and HC-256 by the criterion for encrypting short messages—the algorithm takes a lot of time to update the new key parameters. The best indicators are AES-128, AES-256 and SALSA-20.

Algorithms of lightweight cryptography are almost not inferior to other investigated stream ciphers in all indicators. However, they are intended for use in low-resource information systems, the requirements for their implementation are minimal.

A promising direction in the development of lightweight stream ciphers is the use of a nonlinear-feedback shift register (NLFSR). In particular, stream ciphers “Trivium” and “Grain”, which showed high levels of statistical safety and fast-acting, use the NLFSR technique.

In this paper, based on the use of nonlinear-feedback shift register, we propose new methods and hardware and software for streaming low-resource encryption which meet modern requirements of “green” IT engineering.

5 NLFSR as the Basic Element of Promising Stream Ciphers in Lightweight Cryptography

Generalized structure of second-order NLFSR for register which composed of the $L = 4$ cells is shown in Fig. 9. In the above construction is used multiplication ofonly two cells, such NLFSR called Second-order NLFSR in $GF(2)$.

In Fig. 9 $a_{ij} \in \{0, 1\}$ corresponds to the presence or absence of feedbacks, $q_i(t) \in \{0, 1\}$—is i-th register value in the moment of time t, Q—is generated bit sequence. The symbol ⊗ means nonlinear multiplication function (AND), the symbol ⊕ means linear addition function (XOR).

Task of pseudorandom sequence (PSS) generator synthesis on the base of NLFSR represents the possibility of determining generator structure rely on preassigned statistical and structural requirements for the synthetic sequence. In general, a non-linear function has to satisfy the following requirements: maximum linear complexity; meet the balance probability criterion; ensure minimum ultimate autocorrelation criterion deviation; feedback combinational circuit should have minimum acceptable complexity etc. Synthesis algorithm realization of optimal non-linear register, even on conditions that the stage of register number is comparatively small, is so labour-intensive that its workability is practically excluded.

Permanent improvement of computing technologies, new mathematical methods for cryptanalysis appearance—all of these facts set up new cryptographic security objectives for the PSS generation systems. In connection with this, when developing new encryption algorithms, two main issues are particularly acute: justification of the criteria for the PSS generation structure and issues, associated with generator which meets the chosen requirements synthesis itself.

Nowadays the issue of quite big (for $L > 20$) registers with specified constructional characteristics which satisfy the introduced cryptographic security objectives synthesis isn't solved. Current methods are based on De Bruijn sequence construction and feedbacks which form ready-built sequence structure restoring.

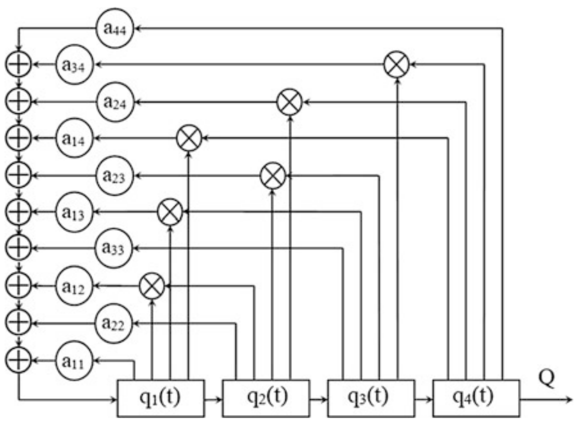


Fig. 9 Generalized construction of the second-order NLFSR which consists of $L = 4$ cells

None of existing methods doesn't allow to form NLFSR with specified constructional characteristics. In addition, existing methods are very exacting to time costs. But this doesn't permit to search for quite big registers.

In current paper is offered partial solution of the mentioned problem.

5.1 The Method of NLFSR Synthesis with the Given Characteristics

The developed method of synthesizing NLFSR uses set of formulated requirements and limitations relative to the feedbacks coefficients. Noncompliance of these requirements points on register inability to generate sequence of maximal period.

Requirement 1: For the ensuring of maximal period the last register linear feedback coefficient should always be present: $a_{LL} = 1$.

Requirement 2: The sum of all non-zero feedback coefficients a_{ij} must be an even number, that is: $\sum_{ij} a_{ij}$—even number.

Requirement 3: There should be no nonlinear feedback received from the output register and any other, that is: $\forall a_{iL} = 0,\ i = \overline{\{1, L-1\}}$.

Requirement 4: Polynomial based on the second-order NLFSR which generate sequence of maximal period shouldn't be symmetrical to itself, that is: $a_{ij} \neq a_{(L-j)(L-i)}$.

Requirement 5: Polynomial based on the second-order NLFSR which can generate M- sequence should has more than one ($a_{LL} = 1$ from the requirement 1) linear feedback coefficient, that is: $\sum_{i-1}^{L-1} a_{ii} \geq 1$.

Requirement 6: A prerequisite for second-order M-NLFSR existence is the simultaneous non-compliance with the next pair of conditions:

$$\sum_{i,j \in A} a_{ij} \begin{cases} \text{even number, where } A \text{ is the whole set of odd numbers from 1 to } L; \\ \text{odd number, where } A \text{ is the whole set of even numbers from 1 to } L. \end{cases}$$

Requirement 7: A prerequisite for second-order M-NLFSR such as that the number of all $a_{ij} = 1$ is even number existence is parity of at least one sum for elements $a_{ij} = 1$ in the i-th rows and columns for any i from 1 to L.

Requirement 8: Let T is any number in the interval from 1 to L inclusive. The condition for the formation of a ring with the period T of the sequence which is generated by NLFSR will be a fulfillment of condition:

$$\sum_{i=0}^{k+i\cdot T \leq L} \sum_{j=i}^{k+j\cdot T \leq L} a_{(k+i\cdot T)(k+j\cdot T)} \begin{cases} \text{an even number, for all } k \in [1, T-1] \\ \text{an odd number, for } k = T \end{cases}$$

Requirement 9: In general, in consideration of the introduced notation, feedback for NLFSR in the moment of time t could be specified as follows:

$$q_1(t+1) = \sum_{i=1}^{L} a_{ii} q_i(t) + \sum_{i=1}^{L-1} \sum_{j=i+1}^{L} a_{ij} q_i(t) q_j(t).$$

A condition under which any of the periods will be repeated is the generation by the register a value of $q(t) = q(t+T)$ for all $t = 1, \ldots, T$.

To find the set of combinations a_{ij} which generate M-sequence is necessary and sufficient to determine and exclude the whole set of combinations a_{ij} which generate the sequence $Q = \{q_1(t),\ q_1(t+1),\ q_1(t+2),\ \ldots,\ q_1(t+T-1)\}$, with the periods $T \in (1,\ 2, \ldots, T_{\max} - 1)$.

The developed NLFSR synthesis method consists of 3 stages:

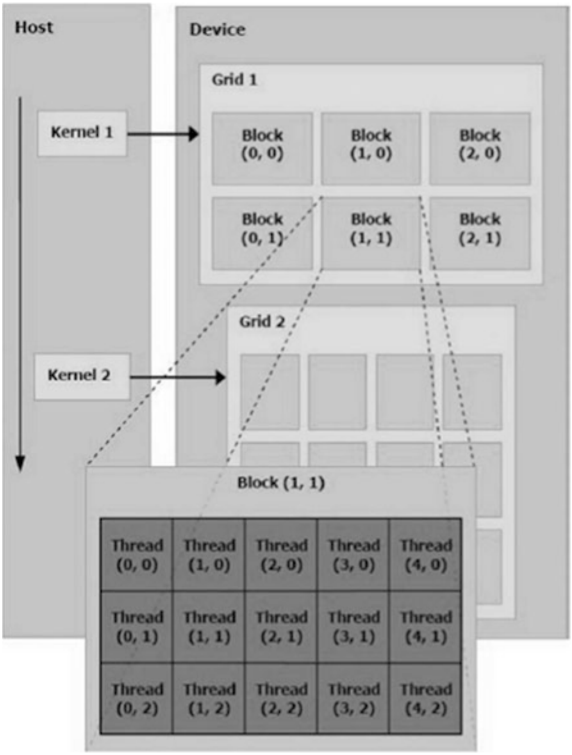
Stage 1—based on selected criteria (generator of a polynomial density, possibility of generating multiple bits per cycle, complexity of calculations of future M-NLFSR in software or hardware implementation, correlation immunity, etc.) some feedback coefficients are fixed;
Stage 2—preliminary calculations are carried out, throught which the overwhelming majority (90–95%) of polynomials that do not comply with Requirements 1–9 and, consequently, do not unambiguously form the M-sequence, are eliminated;
Stage 3—using the hardware device the verification of NLFSR which passed Stage 2 is carried out.

To increase the speed of NLFSR formation, it is suggested to transfer part of the computational process to the graphics processors (GPU—Graphics processing units). The architecture of NVIDIA graphics cards, which allows to use them for non-graphical calculations, was called CUDA (Common Unified Device Architecture), and the specification which was developed for it facilitated [14] the algorithms working on the GPU creating process.

Since the graphics card is a separate computational module with its own computing node and RAM, in the program for CUDA the code is divided into two parts: a host is executed on the CPU; a device is executed on the GPU. Parallelism in CUDA is provided in the following way: when you run the code for execution, you create not one but a group of threads (Fig. 10). The whole group of threads which was created when the code was run is called a grid. Elements of the grid are blocks. They form one- or two-dimensional structure. Each block consists of a group of threads, united in a one-dimensional, two-dimensional or three-dimensional array. The number of blocks in the grid and threads in the block is set at code startup.

Parallel program in the messaging model is a collection of conventional sequential programs that are processed simultaneously. Usually, each of these sequential programs runs on its processor and has access to its own local memory. The clear advantage of such calculations organization is the ability to write and debug a program on a uniprocessor system. In the paper [15], the possibilities of

Fig. 10 The flows on the GPU organization structure



using GPU for streaming encryption systems were analyzed, in particular, the possibilities of transferring some of the computation such as generating a pseudo-random sequence from the CPU to the GPU and comparing the performance of the obtained solutions were studied. By performance we mean the number of bits generated per unit of time. In the course of computational experiments, the performance increment was up to 27 times on the Achterbahn-128 stream cipher polynomial (the feedback function was performed in accordance with the optimized implementation given in [16]) and up to 41 times on the M-NLFSR which has simpler feedback structure. This result is clearly inferior to the increase in performance of the NLFS operation with the use of PLD calculations. It should be noted that the computing capabilities of the GPU can be applied at the stage of preliminary calculations, which is characterized by the use of large amount of RAM, which makes the use of PLD technologies ineffective. Performed calculations showed more than fivefold time reduction on the preliminary calculation stage. Thus, with the use of firmware platform application for parallel computing on the NVIDIA CUDA 8.0 graphics processor (CPU—Intel Core i5-3210M 2.5 GHz, GPU—NVIDIA GeForce GT 630M), we have managed to achieve a significant reduction in the time spent while reducing by 95% sets of researched RFOS at the second stage. In addition, to increase the performance were used 6 FPGA Altera boards Cyclone II family EP2C35F672C6 for 33,216 LEs each. One of the important advantages of implementing cryptographic algorithms on FPGAs is the possibility of building parallel and asynchronous architecture, which excels solutions based on microprocessors, GPUs or CPUs in performance [17].

The project for the FPGA is written in Verilog HDL, for compilation and simulation, CAD Altera Quartus Prime Version 16.0.0 was used. An example of a structural scheme which is implemented in the project presented in Fig. 11 and consists of the following main modules:

- NLFSR—module in which NLFSR is implemented. The number of Les which are used in this module depends on the dimension of the NLFSR;
- UART_USB1 (UART_USB2)—module that performs received packets via the UART channel decoding functions, packet forming, and sending information back to the UART channel. The module uses 62 LEs.
- AltPLL—module for the assignment of clock rate which is different from quartz oscillator, uses built-in PLL block, does not use LEs.
- a module for the task of clock frequency other than a quartz oscillator, uses a built-in PLL block, does not use LEs.
- Aij_modul—the main module, performs the initialization of the feedback coefficients for NLFSR. After the initialization of all NLFSR it launchs their work (NLFSR modules) and control the formation of their M-sequence. At the end of the cycle, it sets the system to its original state and informs the UART module to send the result. The module contains 38 LEs of constant quantity for maintenance of the module, and also the variable number of LEs used for storage and the process of initialization of feedback coefficients, as well as NLFSR modules. The number of LEs depends on the number of NLFSR modules and variable feedback factors. Using the time calculation formula (in hours) t_0^{FPGA} given in [17], which must be spent on searching for the M- NLFSR on the stage 3:

$$t_0^{FPGA} = \frac{1}{n \cdot 3600\, cek} \cdot \frac{k_0^{FPGA}}{k_{NLFSR}^{FPGA}} \cdot t_{cycle}^{FPGA}$$

where

n	the number of used FPGAs
k_0^{FPGA}	the number of polynomials to be analyzed
k_{NLFSR}^{FPGA}	the number of simultaneously tested NLFSR on each FPGA
t_{cycle}^{FPGA}	time spent per cycle in the NLFSR module

We have calculated the time taken to find NLFSR of length $L = 31$. The results are shown in Fig. 12.

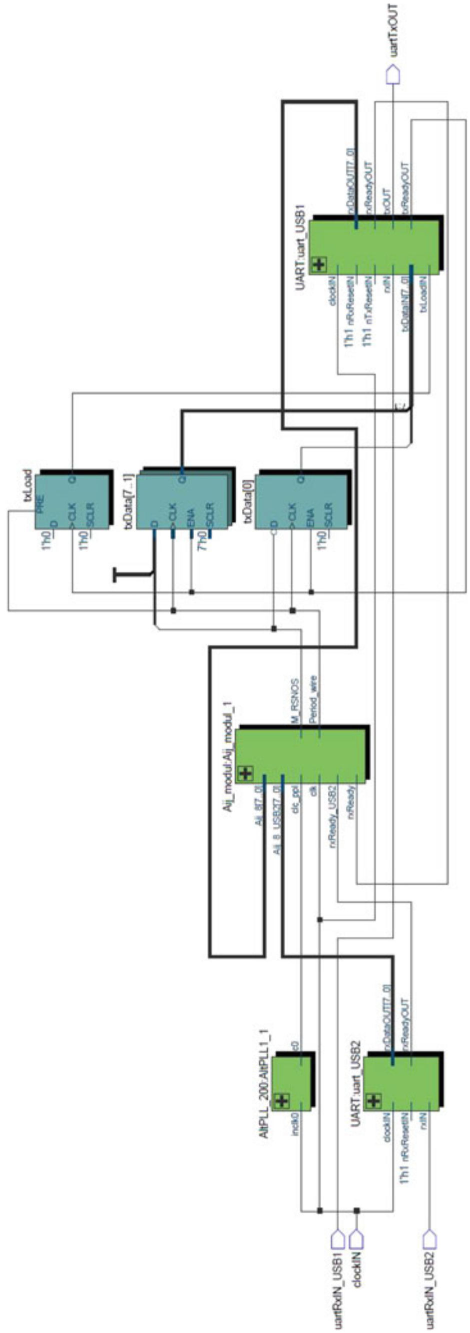


Fig. 11 The block diagram of the project implementation

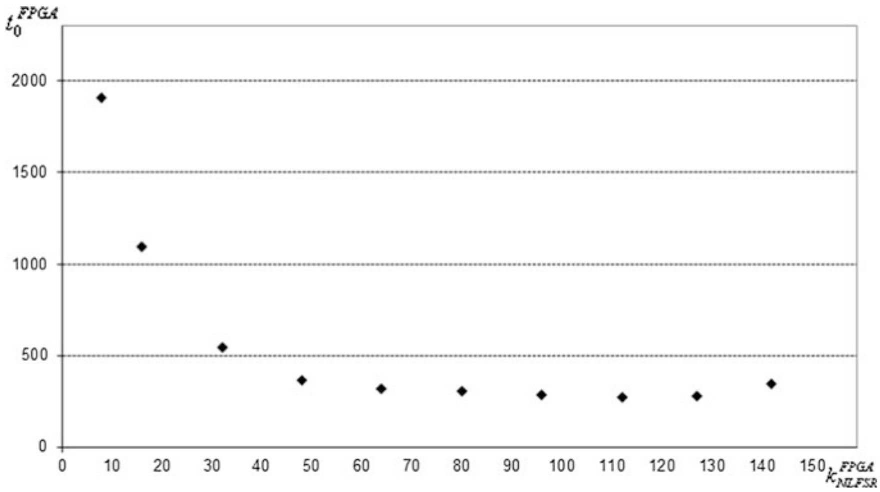


Fig. 12 Relation of t_0^{FPGA} (in hours) from k_{NLFSR}^{FPGA} for $L = 31$

As we can see from the diagram above, the optimal value is achieved with 127 simultaneously tested registers on each board and corresponds to approximately 11.5 days for testing NLFSR set equal to 245 553 141. This approximately reduces by 15 times the time spent on the check in the hardware device, compared to the one given in the paper [18].

5.2 M-NLFSR Synthesis Results

The M-NLFSR which corresponds to predefined set of criteria was searched using the proposed method. In the paper, only the criterion of realization simplicity and the specified size of the register L were taken as an example.

Using the method and the complex described above, an M-NLFSR of dimension $L = 31$ were obtained:

$$x^{31} + x^{27} + x^{24} + x^{22} + x^{21} + x^{17} + x^{15} + x^{12} + x^{11} + x^{10} + x^{20} \cdot x^{22} + x^{20} \cdot x^{23} + x^{20} \cdot x^{25} + x^{20} \cdot x^{27} + x^{20} \cdot x^{29} + x^{20} \cdot x^{30}$$

$$x^{31} + x^{30} + x^{27} + x^{26} + x^{25} + x^{23} + x^{19} + x^{17} + x^{15} + x^{14} + x^{13} + x^{11} + x^{10} + x^{20} \cdot x^{21} + x^{20} \cdot x^{23} + x^{20} \cdot x^{27} + x^{20} \cdot x^{28} + x^{20} \cdot x^{29}$$

$$x^{31} + x^{29} + x^{28} + x^{26} + x^{24} + x^{23} + x^{21} + x^{18} + x^{17} + x^{16} + x^{15} + x^{13} + x^{10} + x^{20} \cdot x^{21} + x^{20} \cdot x^{22} + x^{20} \cdot x^{23} + x^{20} \cdot x^{25} + x^{20} \cdot x^{26} + x^{20} \cdot x^{27} + x^{20} \cdot x^{30}$$

As we can see, in the all found M-NLFSR there are no feedbacks from the first 9 cells. So that on their basis it is possible to generate up to 10 output bits simultaneously due to multiple realizations of the feedback functions, which will allow increasing the speed of the PRS generation up to 10 times.

In addition, the inherent second-order nonlinearity allows to reduce the number of operations for generating one bit only from the twentieth cell. Therefore to get the output bit for the first of the given M-PCHOC it is necessary to perform 15 XOR operations and 1 AND operation, for which it is necessary to use only 5 LEs in the used FPGA models.

It is worth paying attention that these polynomials have a small size both in hardware and software implementation, high generation speed and the possibility of its increas due to generation several bits per cycle, introduced nonlinearity into feedbacks. All of this make them ideal for the role of basic elements in PSS generators in the lightweight cryptography.

It should be noted that the search did not include specific criteria for cryptographic security, such as polynomial density, correlation immunity, and so on. These polynomials are given as an example of the feasibility of implementing the proposed method. Its direct application for efficient schemes of stream encryption that meets all modern requirements of “green” IT engineering designing is a promising direction for further research.

6 Conclusions

Research in the field of green cryptography is developed in two directions: recycling cryptographic design and lightweight cryptography.

Modern cryptographic information security algorithms should not only ensure the reliability of the operation and the security levels set, but also take into account the severe constraints on the complexity of practical implementation, energy consumption and the cost of hardware resources, especially in limited information environments. The requirements of green engineering are becoming more and more relevant, therefore the design of promising means of protection should take into account the current trend of light weight cryptography.

The conducted experimental researches of modern stream ciphers have shown that according to the ratio of stability/complexity of light weight cryptography algorithms, by all indicators, are almost inferior to other investigated stream ciphers. They are intended for use in low-resource information systems, the requirements for their implementation are minimal and this direction of research should be considered the most promising in the future.

References

1. Kharchenko, V., Kondratenko, Y., Kacprzyk, J.: Green IT engineering: concepts, models, complex systems architectures. In: Studies in Systems, Decision and Control, vol. 74, 305p. Springer (2017)
2. Kharchenko, V., Kondratenko, Y., Kacprzyk, J.: Green IT engineering: components, networks and systems implementation. In: Studies in Systems, Decision and Control, vol. 105, 355p. Springer (2017)
3. Troutman, J., Rijmen, V.: Green cryptography: cleaner engineering through recycling, part 2. Secur. Priv. IEEE **7**, 64–65 (2009). https://doi.org/10.1109/MSP.2009.120
4. Bol, D., Durvaux, F., Hocquet, C., Kerckhof, S., Standaert, F.: Towards green cryptography: a comparison of lightweight ciphers from the energy viewpoint. CHES (2012)
5. Biryukov, A., Perrin, L.: State of the art in lightweight symmetric cryptography. https://eprint.iacr.org/2017/511.pdf
6. ISO/IEC 18033-4: Information technology—security techniques—encryption algorithms—part 4: stream ciphers (2011). http://www.iso.org/iso//home/store/catalogue_ics/catalogue_detail_ics.htm?csnumber = 54532
7. eSTREAM: the ECRYPT stream cipher project. http://www.ecrypt.eu.org/stream/
8. CRYPTREC: cryptography research and evaluation committees. https://www.cryptrec.go.jp/Serpent. http://www.cl.cam.ac.uk/rja14/serpent.html
9. FIPS-197: advanced encryption standard (AES). National Institute of Standards and Technology (2001). http://csrc.nist.gov/publications/fips/fips197/fips-197.pdf
10. ISO/IEC 18033-3. Information technology—security techniques—encryption algorithms—part 3: block ciphers. (2010) https://www.iso.org/standard/54531.html
11. Pseudorandom number generator Enocoro. http://www.hitachi.com/rd/yrl/crypto/enocoro/
12. ISO/IEC 29192-3: Information technology—security techniques—lightweight cryptography—part 3: stream ciphers (2012). https://www.iso.org/standard/56426.html
13. Special Publication 800-22: A statistical test suite for random and pseudorandom number generators for cryptographic applications. http://csrc.nist.gov/publications/nistpubs/800-22-rev1a/SP800-22rev1a.pdf
14. CUDA, what is it? [electronic resource]. Access mode: http://www.nvidia.ru/object/what_is_cuda_new_ru.html
15. Poluyanenko, N.A., Potii, A.V.: Using parallel computing technologies in graphic processors for stream encoders generators. Radiotechnics, **189**, 108–114 (2017)
16. Achterbahn stream cipher site [electronic resource]. Access mode: http://www.matpack.de/achterbahn/achterbahn-128–80.zip
17. Perepelitsyn, A.: FPGA technologies in medical equipment: electrical impedance tomography. In: Perepelitsyn, A., Shulga, D. (eds.) Proceedings of IEEE East-West Design & Test Symposium (EWDTS 2012), pp. 437–440 (2012)
18. Poluyanenko, N.: Development of the search method for non-linear shift registers using hardware, implemented on field programmable gate arrays. EUREKA Phys. Eng., 1, 53–60 (2017). https://doi.org/10.21303/2461-4262.2017.00271

Part II
Modelling and Experiments with Green IT Systems

Semi-Markov Availability Model for Infrastructure as a Service Cloud Considering Energy Performance

Oleg Ivanchenko, Vyacheslav Kharchenko, Borys Moroz, Leonid Kabak, Ivan Blindyuk and Kyrylo Smoktii

Abstract This paper addresses a stochastic approach based on Semi-Markov Process modeling in order to determine availability level and energy performance for Infrastructure as a Service (IaaS) Cloud. It is generally agreed today that Cloud Service Providers try to offer diverse IT services, using large number of physical machines (PMs). It really is created serious energy performance issue for cloud computing systems. In this situation we also propose to use information of Technical and Information State Control System in order to improve monitoring process for IaaS Cloud. Therefore, main part of the efforts has been devoted to the development of a monolithic Semi-Markov Availability model for IaaS Cloud and the evaluation power consumption for non-failed, that is available PMs. Semi-Markov Availability model for the cloud infrastructure considering sudden failures of physical and virtual machines are considered, too. Mathematical modeling results will be present in order to give certain clarifications on the possible use of approach suggested.

Keywords Infrastructure as a service cloud · Semi-Markov Availability model Energy performance model · Green aspects

O. Ivanchenko (✉) · B. Moroz · L. Kabak · I. Blindyuk
University of Customs and Finance, Volodimira Vernadskogo Street, Dnipro, Ukraine
e-mail: vmsu12@gmail.com

B. Moroz
e-mail: moroz.boris.1948@gmail.com

L. Kabak
e-mail: kabak.leo@gmail.com

I. Blindyuk
e-mail: ivanblindyuk@mail.ru

V. Kharchenko
National Aerospace University "KhAI", 17, Chkalova Street, Kharkiv, Ukraine
e-mail: v.kharchenko@csn.khai.edu

K. Smoktii
Vasyl' Stus Donetsk National University, 21, 600-Richchya Street, Vinnytsia, Ukraine
e-mail: smoktii@gmail.com

© Springer Nature Switzerland AG 2019

V. Kharchenko et al. (eds.), *Green IT Engineering: Social, Business and Industrial Applications*, Studies in Systems, Decision and Control 171,
https://doi.org/10.1007/978-3-030-00253-4_7

1 Introduction

1.1 Motivation

Nowadays migration of computer and information resources for an enterprise environment into cloud is one of the most widely used and perspective direction for development of IT services [1]. In this situation different companies and organizations try to transform their traditional IT services into cloud services. Therefore, Cloud Service Providers (CSPs) aim to ensure high availability level and quality of service (QoS) using large number of physical machines (PMs) [2]. Furthermore, we can describe situation, relating to the public services for different components of society as pervasive dependence on technical, computer and cloud systems [3, 4].

Hence, hosting complex IT services based on using large number of the PMs creates serious energy performance issue for cloud computing systems [5]. The solution of this issue depends on multiple factors, covering diverse aspects in order to provide necessary value of QoS metrics. Based on this understanding, green aspects are important component to design and choose the best cloud infrastructure architecture. Obviously, green aspects are actual issues for Infrastructure as a Service (IaaS) Cloud, too. One argument in support of this conclusion is concrete incentive the increasing number of physical machines, that relate to large CSPs, such as Amazon, Microsoft, Google [6]. This issue is complex task, which need to solution based on implementation and development Green IT for different types and models of the cloud computing systems.

In spite of the fact, that several aspects of Green IT for different computer and information systems, including programmable logic controllers and systems based on soft computing have been extensively described in works [7–9], we consider that situation regarding green aspects for cloud infrastructures is clearly not sustainable. It should be noted that fundamentals of Green IT Engineering and green aspects within different domains of decisions and control of the complex systems are detailed in [10]. Therefore, challenge of assessment the energy performance of IaaS Cloud for the purpose of ensuring normal work of different CSPs is crucial today. However, in order to quite accurate solve this task, vendors and users of cloud services must consider real availability and reliability level for all cloud infrastructure components. In this respect it is clear, that assessments of availability and energy performance of the IaaS Cloud should be determined jointly based on appropriate monolithic models. Therefore, new types of Semi-Markov Availability models for IaaS Cloud we will be proposed by us. Doubtless, next step is the joint use of both modeling resources in order to solve overall task.

1.2 State-of-the-Arts

On the other hand, we can observe situation, when CSPs constantly try to provide more range of users' services by increasing number of physical and virtual machines (VMs). No doubt, it is also an additional source of the growing power consumption, cooling and operational costs for IaaS Cloud. It is the cause of the increasing of carbon dioxide emissions. Hence, the problem must be solved on the basis of usage of the achievements in Green IT Engineering domain [9, 10]. As the basic provisions for energy performance analysis of IaaS Cloud have been proposed are the following: (a) development of stochastic scalable availability models; (b) development of Markov availability sub-models with closed form solution; (c) development of scalable performance models (d) development of overall power consumption and cooling cost model.

The scientific provisions outlined in [2, 6] have been considered as the basis of original technique for stochastic evaluating the characteristics of scalable availability level of the IaaS Cloud. In accordance with the proposed technique, authors offered to use stochastic reward net [11] in order to determine availability characteristics of physical machines, which are integral parts of IaaS Cloud pools.

The principles of mathematical modeling outlined in [12] form the basis of the approach using Continuous Time Markov Chains (CTMCs) for determining of availability characteristics of the cloud infrastructure components. In accordance with [13], we can use CTMCs, when Markov process implies Markov property. It is noticeable that the use of this approach has become increasingly difficult, when cloud service providers try to maintenance and repair cloud computing system, considering full information range and set of technical states for different system's components. In this case Markov property can't be explained based on trivial reasoning and inferences.

On the other hand, we have to notice that the issue of ensuring performance of the cloud infrastructure can hardly be ignored. In order to solve this task several researchers offered to use additional subsystem for the allocation of resources, called Resource Provisioning Decision Engine (RPDE) [14]. Continued application of RPDE have been allowed to improve performance of cloud infrastructure and obtained quantitative estimates, using CTMCs as the basis for modeling process [15].

Note that some scientists usually choose Markov models in preference to Semi-Markov models, when they try to obtain availability assessments for PMs of IaaS Cloud [16]. Although in real situation at least for Cloud Infrastructure Markov property is not always supposed. Where does that lead us? It means, that with solving concrete situation task we can't get accurate result as availability level for IaaS Cloud will be higher than in real situation. Therefore, thus researches ought to be considerate of definite states, that to consider for modeling of certain situations for different technical and computer systems, including cloud infrastructures. In particular, analytical and stochastic [15, 16] models regarding to cloud infrastructures need to describe different events, including sudden and hidden failures, repairs and monitoring services for PMs. We can't afford to ignore issues, that to relate to the control and monitoring of technical states of different components for cloud infrastructures. Results of researches show, that due to combination of deterministic and stochastic durations into working cycle of IaaS Cloud, this approach for cloud infrastructure availability assessment with application technical state control system can be presented by us.

Turning to overall power consumption and cooling cost model, it should be noted, that availability quantitative assessments of different components are crucial part of the stochastic modeling process for cloud infrastructure. In [2, 6] the stochastic link among availability characteristics for cloud infrastructures subsystems, power consumption and cooling cost parameters was found by authors.

However, in our view, the overall parameter estimates of power consumption and availability level of the different cloud infrastructures components can be determined by different approaches, including the use of Semi-Markov models [17, 18]. Since, the behavior of the cloud infrastructure in the future depends from current states of the infrastructure and from previous history relating to different events, including hackers' attacks, deliberate malicious impacts, control and monitoring operations for physical machines, data bases, networks etc. In this paper we propose to obtain the overall estimates of energy performance by using Semi-Markov Availability models for IaaS Cloud with multiple pools of PMs [19].

1.3 Goal and Structure

A goal of this paper is to propose an approach to determining of availability level for IaaS Cloud, considering energy performance of PMs and different negative events, such as sudden, hidden failures of cloud infrastructures components and deliberate malicious interferences. First of all, we try to build and use a novel Semi-Markov Process (SMP) model with not-special states in order to obtain availability assessments for IaaS Cloud with three pools of PMs and VMs. Turning to third model, we will use famous researches results based on energy performance assessments [2, 6] and availability assessments for IaaS Cloud.

It is a well-known fact, that issues related to failures, faults and hacker attacks lead to resource exhaustion of systems and infrastructures. It is undeniable, that vulnerability analysis plays a critical role in the prevention and mitigation of software security attacks [20], and ensuring of availability of the IaaS Cloud. Furthermore, we have considered the assumption, that different types of failures of the PMs and VMs, information components may be influenced by overall availability level of IaaS Cloud, but the activation of those negative events depends on possibilities of implementation the control system of technical and information states for cloud infrastructure.

The paper is organized as follows. Key points of the approach for availability analysis of IaaS Cloud is considered in Sect. 2. New Semi-Markov model for IaaS

Cloud with three pools of PMs and VMs is described is Sect. 3. In fact, we propose to use the novel monolithic model in order to determine overall availability level of cloud infrastructure, which has a smooth outline architecture with multiple pools of physical and virtual machines. In Sect. 4, the results of stochastic modeling based on power consumptions assessments are presented, too. The numerical results of the Semi-Markov Availability modeling and stochastic power consumptions assessments for IaaS Cloud are illustrated in Sect. 5. Conclusions and the future researches are discussed in Sect. 6.

2 Approach for Availability Analysis of IaaS Cloud

Before performing of availability modeling of the cloud infrastructures we will try to analyze famous Semi-Markov models which are worth considering in order to solve overall task. Approach based on Non-Markovian Models [18] is basis of theoretical provisions for different types of models of the technical and computer systems. According to this approach, SMP contains two branches of analytical models. First branch really is the module SMP models based on Embedded Markov Chains (EMCs) [19]. Solving $v = vP$, where $P = K(\infty)$ is the transition probability matrix for the EMCs [18], therefore we can obtain steady-state availability A_∞ for our system. Second branch implements as module SMP models based on embedded Discrete Time Markov Chains (DTMCs). Similarly solving system of equations $v = vK(t)$, where $K(t) = K_{ij}(t)$ is global kernel [19], therefore we also get steady-state availability $A(t)$ as function of time [21, 22]. Next we will try to use SMP modeling in order to determine steady-state availability $A(t)$ for IaaS Cloud with three pools of PMs and VMs. Indeed, we propose to employ Semi-Markov Modeling Approach on which the availability researches for cloud infrastructures are based. One cannot deny that nowadays number of PMs and VMs is climbing rapid, hence large providers tries to manage their services on based of using different approaches and techniques. To model the failure growth of IaaS Cloud, we can use SMP modeling based on considering information and technical states of virtual and physical resources. At the same time researches of famous scientists [2, 6, 12, 14, 15] show that an effective route to reduction of power consumption for Infrastructure as a Service (IaaS) Cloud is to use three pools of PMs. Figure 1 shows the taxonomy model for availability analysis of IaaS Cloud with multiple pools of PMs and VMs and with application Technical State Control System (TSCS).

In accordance with Fig. 1 IaaS Cloud contains hot, warm and cold pools of PMs. Turning to concrete SMP availability models of IaaS Cloud, we assume that hot pool contains only fully operational PMs, warm pool also includes normal working machines, but these PMs are not ready and cold pool contains only turn off PMs.

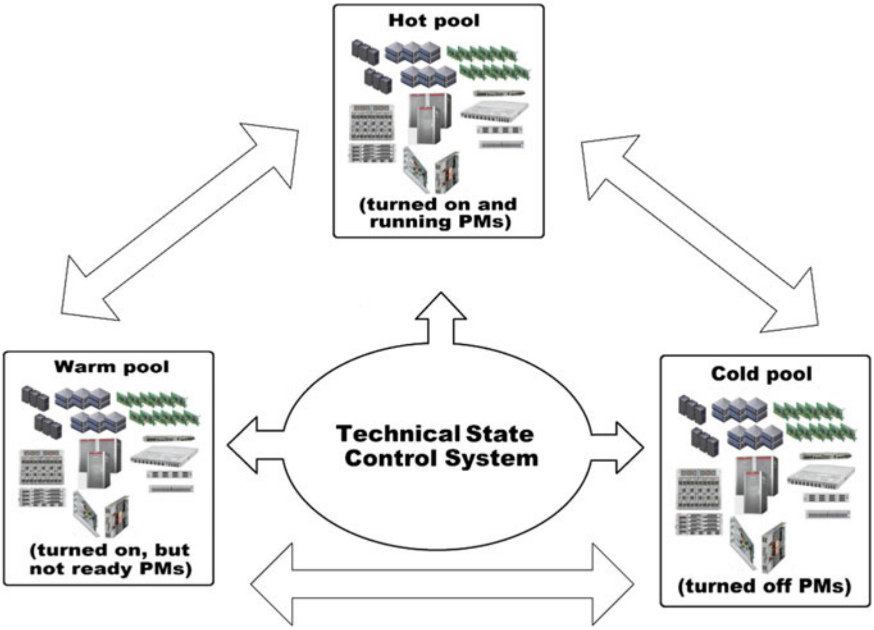


Fig. 1 Taxonomy model for availability analysis of IaaS Cloud with multiple pools of PMs and VMs and with application TSCS [19]

Deployment of VMs are performed by PMs, using information of TSCS [19]. In other words, system TSCS is worked so that the cloud infrastructure can repair, remove or replace a failed or not available PMs. Various signs of illegal of harmful impact on different software and hardware components of IaaS Cloud is also detected by TSCS. Possibility of migrating available PM from one pool to another is important advantage relating to the TSCS.

We cannot ignore the fact that use of the Semi-Markov models, that best fit for describing cloud infrastructures behavior give us an opportunity to demonstrate trends of their availability change based on considering of physical and virtual distribution of the resources among pools, involving power consumption and cooling cost. It seems to confirm the idea, that building and solution of monolithic SMP availability model for IaaS Cloud with three pools of PMs and VMs based on embedded DTMCs with global kernel $K(t) = K_{ij}(t)$ is significant task for next phase of analytical modeling.

3 Monolithic Semi-Markov Availability Model for IaaS Cloud with Three Pools of Physical and Virtual Machines

To improve availability modeling process is described behavior of IaaS Cloud considering sudden failures for hot and warm PMs and VMs. Figure 2 shows finite graph for monolithic SMP model for determining and analysis availability level of IaaS Cloud with three pools of PMs and VMs. For monolithic model overall denotes of the available and unavailable states are displayed in a similar way, that is user uses yellow and red color, respectively.

According to Fig. 2, the cloud computing system consist of three PMs and six VMs. The difference between monolithic model (MNM) (Fig. 2) and analytical and stochastic models (ASMs) is that MNM considers physical and virtual machines states, while ASMs consider states for only PMs, but not VMs states. In Fig. 2, if three PMs and six VMs fail, cloud infrastructure becomes unavailable. State S_{17} is unavailable state for IaaS Cloud. At the same time, states $S_{A_{IaaS}} = \{S_0, S_1, S_5, S_6, S_{10}, S_{11}, S_{15}, S_{16}\}$ are operational states for cloud infrastructure, which set by vendors or users in accordance with service level agreement (SLA) on technical and finance levels. In fact, duration of downtime intervals is written in SLA. In order to solve this task, CSPs try to perform quite large part of the work using diverse techniques and models. Investigators in this domain are often use mathematical models, including stochastic combinatorial, Markov and Semi-Markov models and techniques considering set of the object states or without considering set of the object states etc. As a result of this analysis, we cannot approve that IaaS Cloud is object

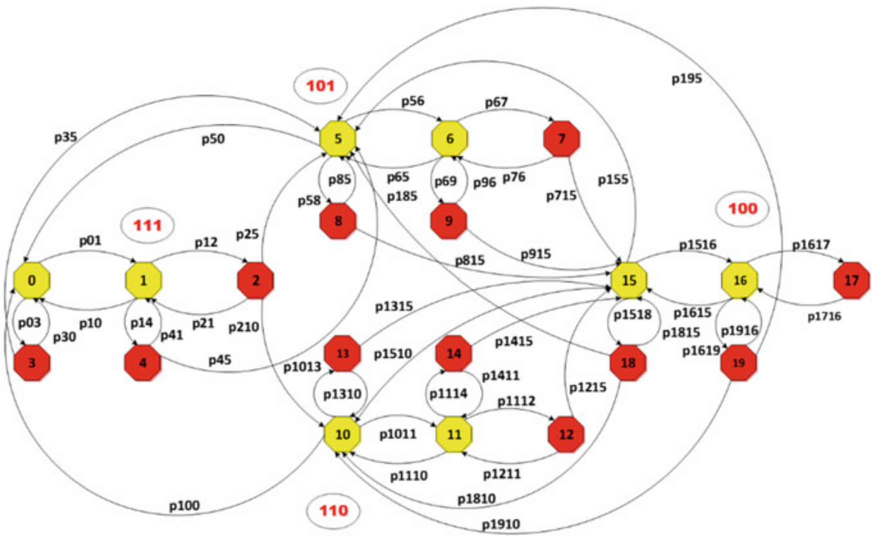


Fig. 2 Finite graph for monolithic SMP availability model with three pools of PMs and VMs

Markov process, which has the Markov property [13]. Indeed, since several important properties of the IaaS Cloud do not depend on historical background, therefore we apply Semi-Markov models in preference to Markov models.

In our modeling process we use the following assumptions for IaaS Cloud with three pools of PMs and VMs.

- The IaaS Cloud is available when available at least one PM and one VM.
- When a hot PM fails, the failed PM can be replaced by available warm or cold PM.
- Failed hot PM, when warm and cold pools empty is put in a queue for repair [12].
- We assume that all migration delays for PMs and VMs are not exponentially distributed.
- Failed VMs also can be replaced or repaired with repair rate.
- We also consider that all times to failures for all PMs and VMs are exponentially distributed. At the same time, we suppose that mean time to failure (MTTF) of $1/\lambda_i$, $i = 1, \ldots, 7$, where λ_1, λ_2—sudden failure rates of hot PM and warm PM, respectively; λ_3, λ_4—sudden failure rates of hot VM and warm VM, respectively; $\lambda_5 = \lambda_1 + \lambda_2$—overall sudden failure rates of hot and warm PMs; $\lambda_6 = \lambda_3 + \lambda_4$—overall sudden failure rates of hot and warm VMs; $\lambda_7 = 0.5\lambda_3$—sudden failure rate of last VM. We also assume, that MTTFs of PMs is lager then MTTFs of VMs by a factor of ten.
- All times to repair are not exponentially distributed random values, too. In this case we prefer to use Erlang-k distribution, where $k = 2$ [13, 22]. Repair of PMs and VMs is to be performed with mean time to repair (MTTR) of $2/\mu_j$, $i = 1, \ldots, 4$, where μ_1, μ_2—repair rates of hot PM and warm PM, respectively; μ_3, μ_4—repair rates of hot and warm VMs, respectively. We consider, that MTTRs of PMs is greater than MTTFs of VMs by a factor of two.

Invoking total probability relation, $\sum_{i=0}^{N} \pi_i = 1$, where $N = 19$, gives [13] steady-state probability vector $\pi_{ss} = \{\pi_0, \pi_1, \pi_5, \pi_6, \pi_{10}, \pi_{11}, \pi_{15}, \pi_{16}\}$. Components of the vector π_{ss} are steady-states probabilities for subset of available states $S_{A_{IaaS}} = \{S_0, S_1, S_5, S_6, S_{10}, S_{11}, S_{15}, S_{16}\}$ of the IaaS Cloud with three pools of PMs and VMs. The arguments we have presented for SMP modeling, suggested that cumulative distribution function (CDF) of failures for transitions from state S_i to state S_j are given by:

$$Q_{ij}(t) = 1 - e^{-\lambda_i t} = Exp(\lambda_i). \tag{1}$$

In addition, overall equation for CDF of repair and migration operations for transitions from S_j state to S_i states can be written as:

$$Q_{ij}(t) = 1 - \left(1 + \mu_j t\right)e^{-\mu_j t} = Erlang\left(2, \mu_j\right). \tag{2}$$

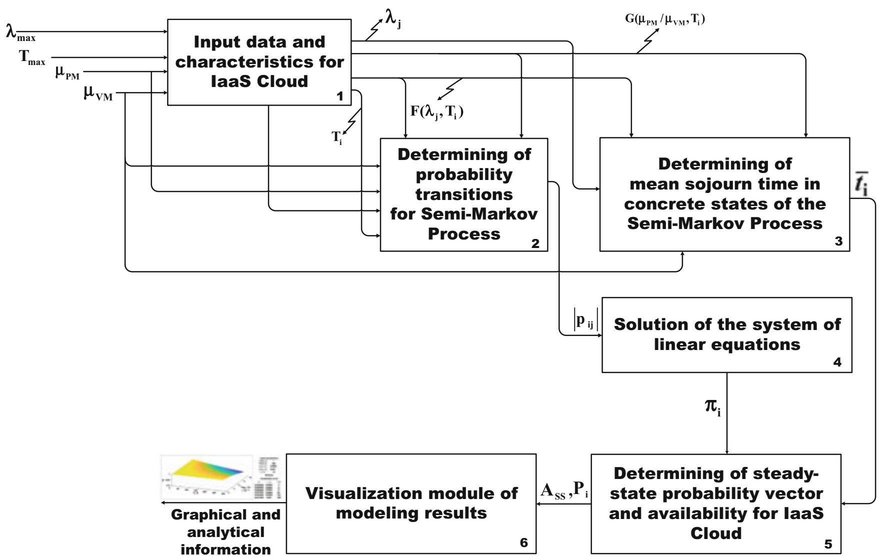


Fig. 3 Functional block diagram for implementation of SMP modeling process of the IaaS Cloud with three pools of PMs and VMS

As a result, the overall equation for steady-state availability of IaaS Cloud with three pools of PMs and VMs are evaluated as follows:

$$A_{IaaS} = \pi_0 + \pi_1 + \pi_5 + \pi_6 + \pi_{10} + \pi_{11} + \pi_{15} + \pi_{16}. \quad (3)$$

Functional block diagram is built in order to consider dynamic modeling process for cloud infrastructure. Figure 3 shows the functional block diagram for implementation of this modeling process.

Thus according to mentioned functional diagram, all non-zero elements [22] of the transition probability matrix $P = |p_{ij}| = K(\infty)$ [17, 18] can be determined as:

$$P_{ij}(\infty) = p_{ij} = \int_0^{\infty} \prod_{\ell \neq i} [1 - Q_{i\ell}(\theta)]\, dQ_{ij}(\theta). \quad (4)$$

List of transitions for the SMP availability model of IaaS Cloud obtained, using the Eqs. (1) and (2) are shown by Tables 1 and 2.

Next, we determine the vector v of steady-state probabilities v_i, where $i = 0, 1, 2, \ldots, 19$ based on solution [17] of the system of equations $v = vP$, $\sum_{i=0}^{19} v_i = 1$. After that, we can also determine steady-state probabilities for SMP availability modeling of IaaS Cloud, applying the equation [17, 18, 22]:

Table 1 List of transitions for the Semi-Markov Availability model of IaaS Cloud

Transitions for IaaS cloud	Transitions for PMs and VMs	CDFs for time transitions
$0 \to 1$	$(2,1) \to (1,1)$	$Exp(\lambda_3)$
$\to 3$	$\to (2,0)$	$Exp(\lambda_1)$
$\to 0$	$\to (2,1)$	$Erlang(2,\mu_3)$
$1 \to 2$	$(1,1) \to (0,1)$	$Exp(\lambda_3)$
$\to 4$	$\to (1,0)$	$Exp(\lambda_1)$
$2 \to 1$	$(0,1) \to (1,1)$	$Erlang(2,\mu_3)$
$\to 5(10)$	$\to (2,1)$	$Erlang(2,\mu_1)$
$3 \to 0(5)$	$(2,0) \to (2,1)$	$Erlang(2,\mu_1)$
$4 \to 1$	$(1,0) \to (1,1)$	$Erlang(2,\mu_1)$
$\to 5$	$\to (2,1)$	$Erlang(2,\mu_1)$
$\to 0$	$\to (2,1)$	$Exp(\lambda_3)$
$5 \to 6$	$(2,1) \to (1,1)$	$Exp(\lambda_1)$
$\to 8$	$\to (2,0)$	$Erlang(2,\mu_3)$
$\to 5$	$\to (2,1)$	$Exp(\lambda_3)$
$6 \to 7$	$(1,1) \to (0,1)$	$Exp(\lambda_1)$
$\to 9$	$\to (1,0)$	$Erlang(2,\mu_3)$
$7 \to 6$	$(0,1) \to (1,1)$	$Erlang(2,\mu_1)$
$\to 15$	$\to (2,1)$	$Erlang(2,\mu_1)$
$8 \to 5(15)$	$(2,0) \to (2,1)$	$Erlang(2,\mu_1)$
$9 \to 6$	$(1,0) \to (1,1)$	$Erlang(2,\mu_2)$
$\to 15$	$\to (2,1)$	$Exp(\lambda_6)$
$\to 0$	$\to ((2,1))$	$Exp(\lambda_5)$
$10 \to 11$	$(2,1) \to (1,1)$	$Erlang(2,\mu_4)$
$\to 13$	$\to (2,0)$	$Exp(\lambda_6)$
$\to 10$	$\to (2,1)$	$Exp(\lambda_5)$
$11 \to 12$	$(1,1) \to (0,1)$	$Erlang(2,\mu_4)$
$\to 14$	$\to (1,0)$	$Erlang(2,\mu_2)$
$12 \to 11$	$(0,1) \to (1,1)$	$Erlang(2,\mu_2)$
$\to 15$	$\to (2,1)$	$Erlang(2,\mu_2)$
$13 \to 10(15)$	$(2,0) \to (2,1)$	$Erlang(2,\mu_2)$
$14 \to 11$	$(1,0) \to (1,1)$	$Exp(\lambda_3)$
$\to 15$	$\to (2,1)$	$Exp(\lambda_1)$
$\to 5(10)$	$\to (2,1)$	
$15 \to 16$	$(2,1) \to (1,1)$	
$\to 18$	$\to (2,0)$	

Table 2 List of transitions for the Semi-Markov Availability model of IaaS Cloud

Transitions for IaaS cloud	Transitions for PMs and VMs	CDFs for time transitions
$\to 15$	$\to (2,1)$	$Erlang(2,\mu_3)$
$16 \to 17$	$(1,1) \to (0,1)$	$Exp(\lambda_7)$
$\to 19$	$\to (1,0)$	$Exp(\lambda_1)$
$17 \to 16$	$(0,1) \to (1,1)$	$Erlang(2,\mu_3)$
$18 \to 5(10)$	$(2,0) \to (2,1)$	$Erlang(2,\mu_2)$
$\to 15$		$Erlang(2,\mu_1)$
$19 \to 5(10)$	$(1,0) \to (2,1)$	$Erlang(2,\mu_2)$
$\to 16$	$\to (1,1)$	$Erlang(2,\mu_1)$

$$\pi_i = \frac{v_i h_i}{\sum_{i=0}^{N} v_i h_i}, \tag{5}$$

where h_i—mean sojourn times for all states of SMP modeling process.

Based on the above approach, we can present an algorithm to calculate and visualization values of steady-state availability A_{IaaS} of the IaaS Cloud with multiple pools of PMs and VMs as illustrated in Algorithm 1.

Algorithm 1: AVAILABILITY OF AN IaaS Cloud $\left(A_{ss}(T)=\{\pi_1,\pi_2,\ldots,\pi_m\}\right)$

Calculate the time modeling interval of $A_{ss}(T)$ as $T=\sum_{i=1}^{n} t_i$
Entry of initial parameters λ_{max}, λ_{step}, t_{min}, t_{step}, μ_j
for $i=1$ *to* n **do**
  $\lambda_i=\lambda_{max}-i\cdot\lambda_{step}$
  for $j=1$ *to* m **do**
    $T_j=t_{min}+j\cdot t_{step}$;
    $h_j=F(\lambda_i,\mu_j)$;
    $p_{ij}=G(\lambda_i,\mu_j,T_j)$;
    $v_i=L(h_j,p_{ij})$;
    $\xi_i=v_i\cdot p_{ij}$;
    $\eta_i=\sum_{i\in S} v_i\cdot p_{ij}$;
    $\pi_i=\xi_i/\eta_i$;
    $\pi_{i(\lambda)}(t)=\sum_{\lambda\in S_A}\pi_\lambda$;
  end
end
return $A_{ss}(T)=\pi_{i(\lambda)}(t)$;
mesh $(t,\lambda_i,A_{ss}(T))$;
visualization $A_{ss}(T)$

In accordance with Fig. 3, functional diagram exploited for determining of overall availability level for IaaS Cloud contains six modules, namely one module of input data and characteristics, four computing modules and one visualization module of modeling results.

One must admit that vendors and users of cloud services can also apply this algorithm in order to develop toolkit for control and analyze energy performance of IaaS Cloud.

4 Stochastic Energy Performance Model of the IaaS Cloud

Energy performance is one of the most important green factors of cloud infrastructures development. Therefore, we propose to leverage stochastic performance energy model of the IaaS Cloud based on joint usage of power consumption assessments and monolithic Semi-Markov Availability model for IaaS Cloud with three pools of PMs and VMs.

We have slightly modified famous equation [2] and proposed to determine total power consumption (TPC) of the cloud infrastructure W_{IaaS}, applying value of the overall steady-state availability for IaaS Cloud A_{IaaS} instead of the separate steady-state probabilities for hot p_h, warm p_w and cold p_c pool respectively. The modified equation is given by:

$$W_{IaaS} = A_{IaaS}\left(\sum_{x=0}^{n_h}\sum_{y=0}^{n_w}\sum_{z=0}^{n_c} W(x, y, z)\right), \tag{6}$$

where $W_{(x,y,z)}$ (Watt) is the total power consumption with x, y and z non-failed PMs in hot, warm and cold pool respectively [2], as computed from the monolithic Semi-Markov Availability model for IaaS Cloud with three pools of PMs and VMs developed in Sect. 3. Determining the number of PMs for ensuring of energy performance requires significant and sustained attention in availability assessment operation of the IaaS Cloud. Indeed, according to Fig. 2 we employ three PMs with index (111) for subset of the states S_0, S_1, S_2, S_3, S_4, that is for one hot PM, one warm PM and one cold PM. It means, hot and warm PMs consume electricity, but cold PM doesn't consume electricity. Table 3 shows overall states of the model (Fig. 2), considering power supply of the PMs.

We can also determine number of PMs, that is value of parameter N_{PMs}, using information from Table 3. This parameter is given by:

$$N_{PMs} = (n_h + 1) + (n_w + 1) + (n_c + 1) - 1. \tag{7}$$

Table 3 States of power supply of the PMs

Index	Power supply of the PMs	Subset states of SMP availability model
(111)	Hot, warm PMs consume electricity; cold PM does not consume electricity	S_0, S_1, S_2, S_3, S_4
(101)	Hot PM consumes electricity; cold, warm PMs do not consume electricity	S_5, S_6, S_7, S_8, S_9
(110)	Hot, warm PMs consume electricity; cold PM does not consume electricity	$S_{10}, S_{11}, S_{12}, S_{13}, S_{14}$
(100)	Hot PM consumes electricity; old, warm PMs do not consume electricity	$S_{15}, S_{16}, S_{17}, S_{18}, S_{19}$

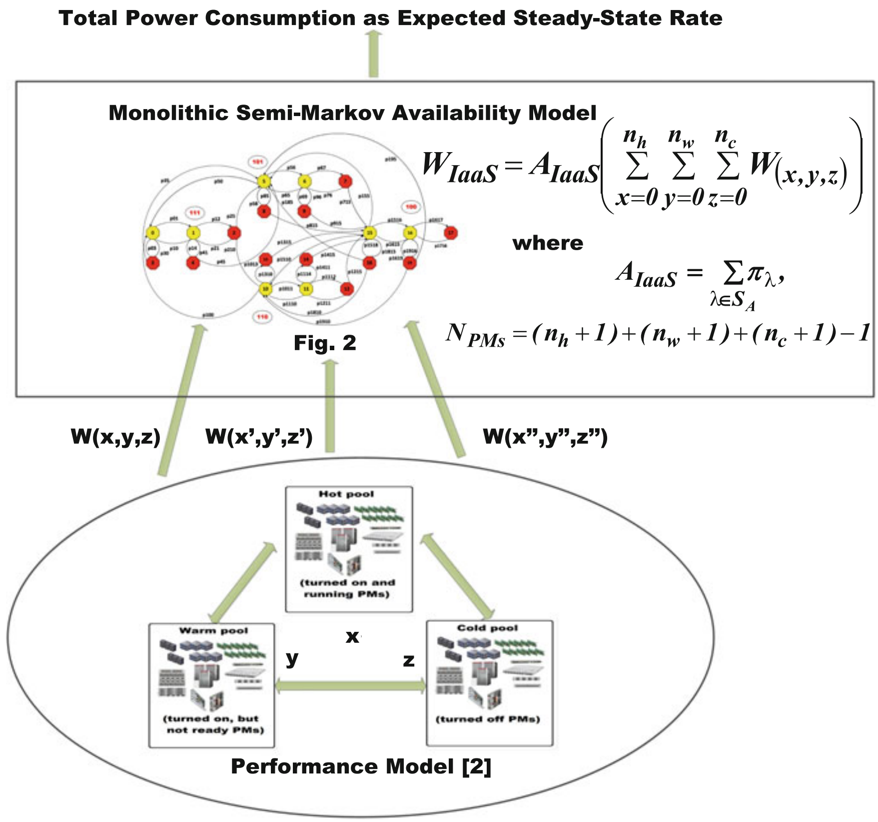


Fig. 4 Import graph to compute the total power consumption of the IaaS Cloud with three pools of PMs and VMs

Figure 4 shows the import graph to determine total power consumption of the IaaS Cloud as expected steady-state rate. According to Fig. 4, procedure of determining of TPC parameter for the IaaS Cloud is planned to carry out in three phases: (a) preliminary assess performance of cloud infrastructure with multiple pools of PMs; (b) determining of IaaS Cloud overall availability level; (c) determining of TPC parameter for IaaS Cloud.

In order to get preliminary, assess performance of cloud infrastructure some scientists prefer to use multiple pools architecture contained different number of PMs. This means that according to Fig. 4, before a concrete researcher begins to benefit Semi-Markov Availability assessments, he will try to obtain information about quantity of PMs and VMs by applying the performance model [2]. The second phase illustrates the possibilities of Semi-Markov modeling to obtain assessments of availability for IaaS Cloud with multiple pools of physical and virtual machines in accordance with Fig. 2. The Semi-Markov Availability model for IaaS Cloud with special states was built on the basis of monolithic principle, unlike the Markov sub-models closed form solution for every pool of PMs [6].

Third phase implements stochastic approach in order to determine TPC parameter for cloud infrastructure with three pools of different types of PMs and VMs, using information from Table 3 and Eqs. (6), (7).

5 Numerical Modeling Illustration

First of all, let us try to get numerical modeling results for SMP availability model based on changing of operational parameters of the IaaS Cloud and reliability characteristics for PMs. The values of this parameters are presented in Table 4. The values are submitted as the basis for mathematical modeling of the IaaS Cloud behavior for an operational period of time T hours and for concrete types of servers (PMs).

In Fig. 5 we change failure rate of hot physical machines over time $T \in [0; 2160]$ hours, that is throughout quarter. According to Fig. 5, with the increase in MTTR $(2/\mu_1)$ of hot PMs the overall availability level of cloud infrastructure is as well grown. Further, we performed SMP availability modeling for different variants of target load (TL) of concrete types servers in accordance with [23]. Figures 5, 6 and 7 show dynamics of changing of the TPC of IaaS Cloud considering availability and reliability characteristics, including power consumption parameters for usage variants of the servers.

In Fig. 5 we change failure rate of hot PMs over time $T \in [0; 2160]$ hours, that is throughout quarter.

According to Fig. 5, the overall availability level of IaaS Cloud can be increasing, if mean time to repair of hot physical machines $2/\mu_1$ is reducing. However, the overall availability level of cloud infrastructure can be reducing, if failure rate of hot physical machines λ_1 is increasing. Further, we performed SMP availability modeling for different variants of target load (TL) of concrete types servers in accordance with [23]. Figures 6 and 7 shows dynamics of changing of the TPC of IaaS Cloud considering availability and reliability characteristics, including power consumption parameters for usage variants of the servers.

Comparative assessments of energy performance modeling results considering overall availability level of the IaaS Cloud, workload operation (WO), TL, average active power (AAP) and Performance to Power Ratio (PPR) [23] for different types of servers are presented in Figs. 8, 9, 10, 11 and Table 5.

Table 4 Numerical values for SMP availability model

Failure rate of hot PMs	Repair rate of hot PMs	Failure rate of hot VMs	Repair rate of hot VMs
from $\lambda_{1_{\min}} = 2 \times 10^{-5}$ to $\lambda_{1_{\max}} = 1 \times 10^{-4}$ 1/h	from $\mu_{1_{\min}} = 4; 5; 8; 10$ to $\mu_{1_{\max}} = 20; 30; 40; 50$ 1/h	λ_3 lower than failure rate of PMs by a factor ten	λ_3 higher than failure rate of PMs by a factor two

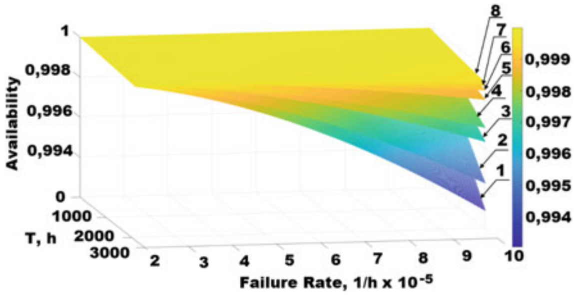


Fig. 5 Availability of the IaaS Cloud for: 1—$\mu_1 = 4$ 1/h; 2—$\mu_1 = 5$ 1/h; 3—$\mu_1 = 8$ 1/h; 4—$\mu_1 = 10$ 1/h; 5—$\mu_1 = 20$ 1/h; 6—$\mu_1 = 30$ 1/h; 7—$\mu_1 = 40$ 1/h; 8—$\mu_1 = 50$ 1/h

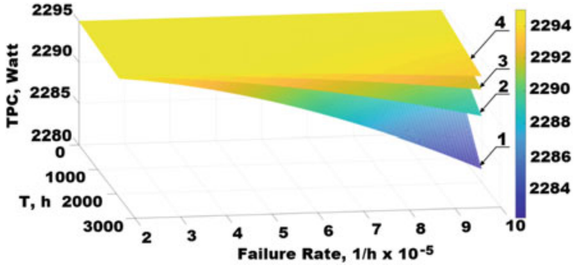


Fig. 6 TPC of IaaS Cloud with five servers Hewlett Packard Enterprise ProLiant ML350 Gen10 for TL 100%, when: 1—$\mu_1 = 5$ 1/h; 2—$\mu_1 = 10$ 1/h; 3—$\mu_1 = 20$ 1/h; 4—$\mu_1 = 40$ 1/h

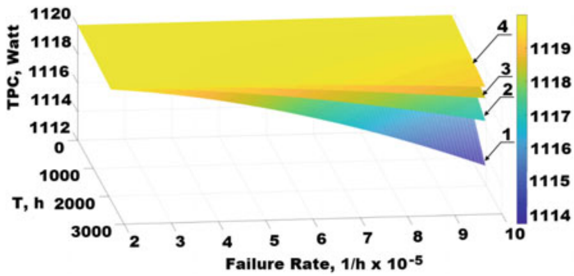


Fig. 7 TPC of IaaS Cloud with five servers Hewlett Packard Enterprise ProLiant ML350 Gen10 for TL 50%, when: 1—$\mu_1 = 5$ 1/h; 2—$\mu_1 = 10$ 1/h; 3—$\mu_1 = 20$ 1/h; 4—$\mu_1 = 40$ 1/h

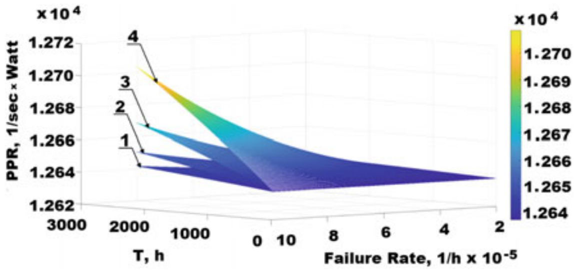


Fig. 8 PPR of IaaS Cloud considering infrastructure availability level with five servers Hewlett Packard Enterprise ProLiant ML350 Gen10 for TL 100%, when: 1—$\mu_1 = 5$ 1/h; 2—$\mu_1 = 10$ 1/h; 3—$\mu_1 = 20$ 1/h; 4—$\mu_1 = 40$ 1/h

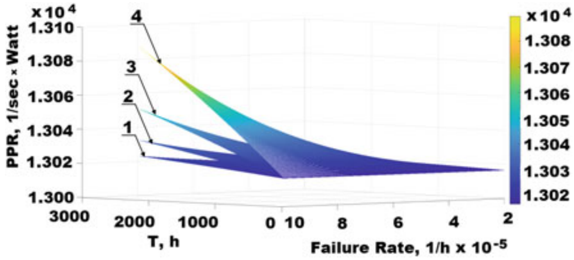


Fig. 9 PPR of IaaS Cloud considering infrastructure availability level with five servers Hewlett Packard Enterprise ProLiant ML350 Gen10 for TL 50%, when: 1—$\mu_1 = 5$ 1/h; 2—$\mu_1 = 10$ 1/h; 3—$\mu_1 = 20$ 1/h; 4—$\mu_1 = 40$ 1/h

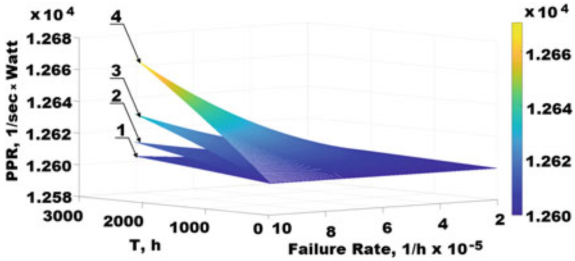


Fig. 10 PPR of IaaS Cloud considering infrastructure availability level with five servers Dell Inc. PowerEdge R940 for TL 100%, when: 1—$\mu_1 = 5$ 1/h; 2—$\mu_1 = 10$ 1/h; 3—$\mu_1 = 20$ 1/h; 4—$\mu_1 = 40$ 1/h

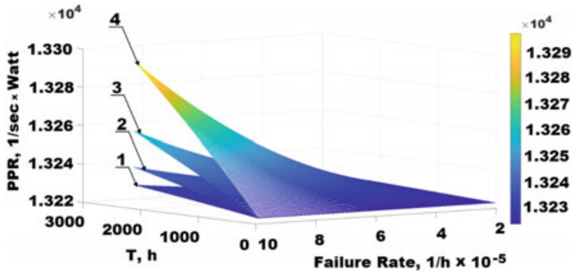


Fig. 11 PPR of IaaS Cloud considering infrastructure availability level with five servers Dell Inc. PowerEdge R940 for TL 50%, when: 1—$\mu_1 = 5$ 1/h; 2—$\mu_1 = 10$ 1/h; 3—$\mu_1 = 20$ 1/h; 4—$\mu_1 = 40$ 1/h

Table 5 Benchmark results summary [23]

Performance		Power		PPR without the considering availability level
TL	WO	AAP (Watt) without the considering availability level	AAP (Watt) considering availability level	
For five servers Hewlett Packard Enterprise ProLiant ML350 Gen10				
100%	29,003,230	2295	2289	12,638
50%	14,578,560	1120	1117	13,017
For five servers Dell Inc. PowerEdge R940				
100%	57,644,020	4575	4562	12,600
50%	28,892,060	2185	2179	13,223

At Figs. 8 and 9 the modelled results confirm that PPR considering overall availability level of IaaS Cloud with five servers Hewlett Packard Enterprise ProLiant ML350 Gen10 lower than without giving due consideration the availability value. Let's us also consider what are advantages of analytical and stochastic approach based on Semi-Markov Modeling Process relating to the assessments of different power consumption parameters for IaaS Cloud in another variant of PMs configuration. For instance, Figs. 10 and 11 show modeled results for five servers Dell Inc. PowerEdge R940.

In accordance with Figs. 8, 9, 10, 11 and Table 5 we illustrate to a possibility as improve energy performance of IaaS Cloud using availability, reliability and total power consumption models of infrastructure components. Usage of these models allows to support the requirements of ensuring availability of the cloud infrastructure with multiple pools of physical and virtual machines and as well to improve energy performance characteristics.

Note that we obtained the same results for both modeling variants. In particular, average active power for cloud infrastructure with five servers Hewlett Packard Enterprise ProLiant ML350 Gen10 has lowered by 0.265%. At the same time analogical parameter for cloud infrastructure with five servers Dell Inc. PowerEdge R940 has also lowered by 0.285%.

6 Conclusions and Future Researches

This paper explores the availability and energy performance of cloud computing system oriented diverse services of an IaaS Cloud. Changes of behavior of the IaaS Cloud, considering reliability of the PMs and overall availability level of the infrastructure are modeled with the help a Semi-Markov Process. Since the mathematical model allows to consider combinations between deterministic and stochastic intervals of time and use different distributions for time to failure and time to repair and time to migration, too. A new approach for building and implementation of the monolithic SMP availability model for IaaS Cloud with three pools of PMs and VMs are proposed and characterized. Unlike closed form solution models [2, 6] the overall monolithic SMP availability model considers interconnection among three pools of PMs and VMs. Results of the SMP modeling can be used in order to build optimal IaaS Cloud architectures.

Mathematical modeling results considering power consumption of PMs and VMs of the IaaS Cloud based on joint usage of stochastic assessments and monolithic Semi-Markov Availability model are obtained and analyzed, too. Figures 7, 8, 9, 10 and 11 show numerical values of modeling results. The approach can be used to obtain power consumptions assessments and applied to implement optimal energy performance regimes for IaaS Cloud. Furthermore, proposed model may be employed by vendors and users of cloud services to create new web applications, stipulating employing of the availability and energy performance actual levels for different architectures of cloud infrastructure. Undoubtedly, the proposed models will also be improved and extended in order to perform additional analytical features and reviews for cloud computing systems and cloud service providers in future.

References

1. Joshi, P.R., Islam, S., Islam, S.: A framework for cloud based E-government from the perspective of developing countries. Future Internet **9**(4), 80 (2017)
2. Ghosh, R., Longo, F., Xia, R., Naik, V.K., Trivedi, K.S.: Stochastic model driven capacity planning for an infrastructure-as-a-service cloud. IEEE Trans. Serv. Comput. **7**(4), 667–680 (2014)
3. Bashir, R.N., Qadri, S., Saleem, R.M., Naeem, M., Ghafoor, Y.: Human computer interaction (HCI) in ubiquitous computing. Int. J. Innov. Appl. Stud. **9**(2), 534 (2014)
4. Nadeem, F.: Ranking grid-sites based on their reliability for successfully executing jobs of given durations. Int. J. Comput. Netw. Inf. Secur. **7**(5), 9 (2015)
5. Kharchenko, V., Illiashenko, O.: Concepts of Green IT Engineering: Taxonomy, Principles and Implementation. In: Kharchenko, V., Kondratenko, Y., Kasprzyk J. (eds.) Green IT Engineering: Concepts, Models, Complex Systems Architectures. Studies in Systems, Decision and Control, vol. 74, pp. 3–19. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-44162-7_1
6. Ghosh, R.: Scalable stochastic models for Cloud Services. Dissertation, Duke of University (2012)
7. Kondratenko, Y., Korobko, O., Kozlov, O.: PLC-based systems for data acquisition and supervisory control of environment-friendly energy-saving technologies. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Concepts, Models, Complex Systems Architectures. Studies in Systems, Decision and Control, vol. 74, pp. 247–267. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-44162-7_13
8. Kondratenko, Y., Korobko, V. Korobko, O., Kondratenko, G., Kozlov, O.: Green-IT approach to design and optimization of thermoacoustic waste heat utilization plant based on soft computing. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Components, Networks and Systems Implementation. Studies in Systems, Decision and Control, vol. 105, pp. 287–311. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-55595-9_14
9. Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Concepts, Models, Complex Systems Architectures. Studies in Systems, Decision and Control, vol. 74. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-44162-7
10. Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Components, Networks and Systems Implementation. Studies in Systems, Decision and Control, vol. 105. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-55595-9
11. Ghosh, R., Longo, F., Naik, V.K., Trivedi, K.S.: Quantifying resiliency of IaaS cloud. In: 2010 29th IEEE Symposium on Reliable Distributed Systems, pp. 343–347, Oct 2010
12. Ghosh, R., Longo, F., Frattini, F., Russo, S., Trivedi, K.S.: Scalable analytics for IaaS cloud availability. IEEE Trans. Cloud Comput. **2**(1), 57–70 (2014)
13. Bolch, G., Greiner, S., de Meer, H., Trivedi, K.S.: Queuing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications, 878 p. Wiley, New York (2006)
14. Ghosh, R., Trivedi, K.S., Naik, V.K., Kim, D.S.: End-to-end performability analysis for infrastructure-as-a-service cloud: An interacting stochastic models approach. In: 2010 IEEE 16th Pacific Rim International Symposium on Dependable Computing (PRDC), pp. 125–132, Dec 2010

15. Ghosh, R., Longo, F., Naik, V.K., Trivedi, K.S.: Modeling and performance analysis of large scale IaaS clouds. Future Gener. Comput. Syst. **29**(5), 1216–1234 (2013)
16. Matos, R., Araujo, J., Oliveira, D., Maciel, P., Trivedi, K.S.: Sensitivity analysis of a hierarchical model of mobile cloud computing. Simul. Model. Pract. Theory **50**, 151–164 (2015)
17. Vinayk, R., Dharmaraja, S.: Semi-Markov modeling approach for deteriorating systems with preventive maintenance. Int. J. Perform. Eng. **8**(5), 515–526 (2012)
18. Distefano, S., Trivedi, K.S.: Non-Markovian state-space models in dependability evaluation. Qual. Reliab. Eng. Int. **29**(2), 225–239 (2013)
19. Ivanchenko, O., Kharchenko, V., Ponochovny, Y., Blindyuk, I., Smoktii, O.: Semi-Markov availability model for an infrastructure as a service cloud considering hidden failures of physical machines. ICTERI **1844**, 634–644 (2017)
20. Li, X., Chang, X., Board, J.A., Trivedi, K.S.: A novel approach for software vulnerability classification. In: 2017 Annual of Reliability and Maintainability Symposium (RAMS), pp. 1–7, Jan 2017
21. Xie, W., Hong, Y., Trivedi, K.S.: Analysis of a two-level software rejuvenation policy. Reliab. Eng. Syst. Saf. **87**(1), 13–22 (2005)
22. Volkov, L.: Control of Air Complexes Operation, 368 p. High Education, Moscow (1981)
23. Fourth Quarter 2017 SPECpower_ssj2008 Results, Dec 2017. Available: http://www.spec.org/power_ssj2008/results/res2017q4/

Improving Big Data Centers Energy Efficiency: Traffic Based Model and Method

Georgiy Kuchuk, Andriy Kovalenko, Iraj Elyasi Komari, Artem Svyrydov and Vyacheslav Kharchenko

Abstract Green aspects of data centers (DCs) which provide the appropriate hosting services and data storage are considered. Various approaches to transmission control of integrated data streams (IDS), which are the result of traffic aggregation, being transmitted in a DC related to Big Data processing, are analyzed. It is proposed a complex, which includes several models and a method; it uses traffic samples of several sessions as input data to perform short-term traffic prediction. For the case of elastic traffic, the application of the developed complex will reduce the DC's resources usage, which are required to process data, down by 3%. As a result of the resources reallocation of the DC on the basis of the short-term prediction, the gain on the effectiveness of their use with such type of traffic will be 4–8%. Application of the proposed complex allows using the resources of the DC in more efficient way (reduces transmission time and, correspondingly, increases energy efficiency of the whole system).

G. Kuchuk · V. Kharchenko
National Aerospace University "KhAI" named after N.E. Zhukovskiy,
17 Chkalova str., Kharkiv 61070, Ukraine
e-mail: kuchuk56@ukr.net

V. Kharchenko
e-mail: v.kharchenko@csn.khai.edu; research@radiy.com

A. Kovalenko (✉) · A. Svyrydov
Kharkiv National University of Radio Electronics, 14 Nauki ave.,
Kharkiv 61166, Ukraine
e-mail: andriy_kovalenko@yahoo.com; research@radiy.com

A. Svyrydov
e-mail: SvyrydovArtem@gmail.com

A. Kovalenko · V. Kharchenko
Research and Production Corporation Radiy, 29 Academic Tamm str.,
Kropyvnytskyi 25009, Ukraine

I. E. Komari
Andimeshk Branch, Islamic Azad University, Andimeshk, Iran
e-mail: irajelyasi@gmail.com

© Springer Nature Switzerland AG 2019

V. Kharchenko et al. (eds.), *Green IT Engineering: Social, Business and Industrial Applications*, Studies in Systems, Decision and Control 171,
https://doi.org/10.1007/978-3-030-00253-4_8

Keywords Big Data · Service · Green · Technology · Processing
Traffic

1 Introduction

1.1 Motivation

Information systems and information technologies are of great importance now. They are fundamental inputs to any industry's productive activity. However, information systems' and technologies' production, use and retirement demand costs and have a large impact on the environment. The adoption of green technologies reduces these costs through the use of relevant materials and products, reduction of energy consumption and damaging environmental effects [1]. For green information systems and technologies development and use should be guided by three next principles: eco-efficiency, eco-equity and eco-effectiveness [2]. Such principles involve the use of minimal number of methods and tools that are damaging for the nature and inefficient in applying of non-regenerative resources. In particular, this is achieved by improvement of technologies of data transmission and processing in order to use the resources in more efficient way.

At the present time DCs, which provide webhosting and data storage services, are actively developing. Unfortunately, the energy consumption is rising at the same time. Reduction of energy consumption can be achieved as due to energy-efficient equipment usage or by software "greening". In this way, Software Improvement Group research shows that only 25% of algorithms in software technologies provide optimal computing effectiveness [3]. The same research shows that only 35% of technologies' potential provide problems solving, the rest 65% provide optional functionality.

Efficiency of the applied technology depends on the quality of integrated data streams (IDS) control within DCs. At the same time, the quality of such control depends heavily on accuracy of short-range prediction of traffic behavior. Traffic models, created by IDS, are used in Big Data transmission and processing.

Improving the traffic behavior accuracy through modeling based on its samples, especially during Big Data transmission and processing, will allow reducing the time of data transmission and therefore increasing system energy efficiency.

1.2 Analysis of Related Works

During the development of information technologies, considerable attention is given to the green aspects. In many works there are considered various approaches in solving this problem. Finding solutions for complex problems, related with certain aspects optimization, including traffic control, is a highly complex modern problem for such networks.

Kuchuk et al. [4] proposed mathematical statement for several discrete optimization problems, applicable to resources reallocation in computer networks. To solve the problems, appropriate specific algorithms are required.

Such issues as traffic control with the purpose of reducing time of data transmission were discussed by many authors, for example, in [5–9]. In this way, in [5] the issues of optimal TCP traffic control under the condition of its local elasticity are considered. Routing algorithm for the multiprotocol label switching (MPLS) are considered in depth in [6]. For the near-elastic traffic, they allow achieving the near-optimal result, but they are absolutely not applicable with δ-spikes during data transmission process. Similar problems are also present in applying control algorithms, described in [7–9].

In [10] the attention has been given to reducing the time of relational system transactions transmission. Similar problems were considered by the authors together with the assessment of its security in [11]. In [12, 13], in assessing the effectiveness of WiMax technology, authors analyzed traffic features in wireless networks. Also, the wireless traffic effectiveness was studied in [14] for ad hoc networks, and in [15] for sensor networks.

However, works [5–15] failed to take into account the Big Data centers traffic features. In these works, authors have been focused on elastic and poorly-aggregated traffic. At the same time, without taking in account the fact, that in congested scenarios or close to them, traffic gains fractal character and, as a result, the number of spikes increases. Moreover, if a number of active sources increases, as in a case of Big Data centers, then traffic becomes highly-aggregated [16].

In [4] authors consider greening issues for safety-critical I&C systems. It is clearly shown that consideration of Big Data centers traffic features allows to improve accuracy of short-range traffic behavior prediction through the usage of models, which are based on traffic samples related to several consequent sessions.

Considering the analysis results presented above, it is critical for Big Data centers traffic control to quickly determine the traffic character based on small number of its samples. Moreover, it is required to predict the traffic behavior in terms of dynamic mathematical models, based on traffic samples related to several consequent sessions. Thus, Big Data centers traffic control requires the development of radically new methods that could actively and quickly respond to described above problems.

1.3 Goals and Structure

One of the important problems in Big Data centers traffic control is increasing of accuracy of short-range prediction of traffic behavior based on its behavior models. The goal of this chapter is to describe appropriate approaches intended for traffic control problem solution. In this way, it is necessary to determine the character of traffic in order to choose appropriate model. The model shall be chosen from the complex of traffic models that corresponds to real traffic of the Big Data center. This complex uses measurements of several sessions as input data and it contains the following main components: static model of single source traffic, dynamic two-dimensional model of traffic, method of aggregated traffic spikes characterization. Based on the results, obtained using the chosen model, characteristics and volume of aggregated traffic's spikes are being predicted, as well as amount of traffic, which exceeds pre-defined threshold, and maximal data transmission rate are being determined. This allows reallocating the Big Data center resource and increase its efficiency that results in decrease of system energy consumption.

The chapter is structured in the following way. Chapter 2 describes traffic models for Big Data centers. Traffic source statistical characteristics are defined. Spline interpolation model of single traffic source and dynamic two-dimensional Spline interpolation model of aggregated traffic are proposed. Chapter 3 focuses on solving the problem of energy efficiency of Big Data centers through the short-range prediction of traffic behavior using the proposed complex of traffic models.

2 Traffic Models in Big Data Centers

This section proposes a complex of traffic models applicable to Big Data centers. Such complex allows to perform short-term prediction of traffic behavior. The prediction results can be used to achieve load balance in physical lines of data transmission in order to increase energy efficiency of Big Data centers.

2.1 Static Model of Single Source Traffic

The first stage of study is focused on the determination of statistical characteristics of first and second order of traffic point source (TPS).

Let us consider Big Data center, consisting of I TPSs. Let i-th source ($i = \overline{1, I}$) characterized by stochastic process $V_i(t)$, which determines a set of stochastic functions for data transmission rate of certain sessions. Such stochastic process, analyzed during certain time interval $[0, T_i]$, may be defined in such key parameters:

$V_i^{(\max)} = \max\limits_{t \in [0, T_i]} V_i(t)$ is a peak rate of i-th TPS;

$V_i^{(av)} = \int_0^{T_i} V_i(t)dt$ is an average rate of i-th TPS;

$k_i^{(p)} = V_i^{(\max)} / V_i^{(av)}$ is a burstiness coefficient, defining relation between peak and average rates;

$T_i^{(\max)}$ is a duration time of maximal rate.

Let data volume, transferred in a time $t \in [0, T_i]$, is characterized by stochastic value $W_i(t)$ with maximum possible value of $W_i^{(\max)} = V_i^{(\max)} \cdot T_i^{(\max)}$, i.e. $V_i(t) = \frac{\mathrm{d}W_i(t)}{\mathrm{d}t}$. Let us consider the probability for scenario when i-th TPS reaches its peak transfer rate for the concerned time interval of $[0, T_i]$: $p_i = P\left(V_i(t) = V_i^{(\max)} \right)$. This probability gives also a measure for transferred volume:

$$V_i(t) = V_i^{(\max)} \Rightarrow \left(V_i(t) \cdot T_i^{(\max)} = V_i^{(\max)} \cdot T_i^{(\max)} \right) \equiv \left(W_i(t) = W_i^{(\max)} \right)$$
$$\Rightarrow P_i = P\left(W_i(t) = W_i^{(\max)} \right).$$

Let us divide the time interval of $[0, T_i]$ into two subsets of embedded intervals $\Im_i^{(1)}$ and $\Im_i^{(2)}$ in the following way, where $V_i^{(min)}$ denotes minimal transfer rate (Fig. 1):

$$\Im_i^{(1)} \cup \Im_i^{(2)} = [0, T_i]; \quad \Im_i^{(1)} \cap \Im_i^{(2)} = \varnothing; \tag{1}$$

$$\forall t^{(1)} \in \left[t_i^{\prime(1)}; t_i^{\prime\prime(1)} \right] \subset \Im_i^{(1)} \Rightarrow V_i\left(t^{(1)} \right) = V_i^{(\max)}; \tag{2}$$

$$\forall t^{(2)} \in \left[t_i^{\prime(2)}; t_i^{\prime\prime(2)} \right] \subset \Im_i^{(2)} \Rightarrow V_i\left(t^{(2)} \right) < V_i^{(\max)}. \tag{3}$$

This partition allows defining the peak load duration time for the traffic on given interval, as:

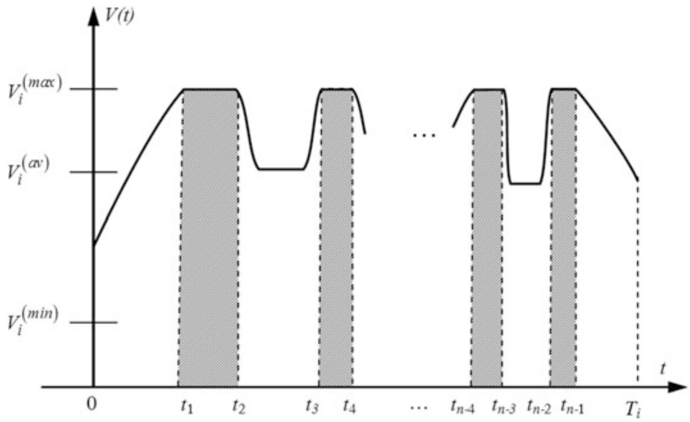


Fig. 1 Diagram for $[0, T_i]$ interval partitioning

$$T_i^{(\max)} = \sum \left[t_i^{\prime(1)}; t_i^{\prime\prime(1)} \right] \subset \Im_i^{(1)} \left| t_i^{\prime\prime(1)} - t_i^{\prime(1)} \right| \tag{4}$$

and determine a probability value for the scenario when i-th TPS reaches its peak transfer rate on the given time interval. For this purpose let us calculate areas, corresponding to the subsets of $\Im_i^{(1)}$ and $\Im_i^{(2)}$:

$$\Theta_1 = \int\limits_{\Im_i^{(1)}} V_i(t)dt = V_i^{(\max)} \cdot T_i^{(\max)}; \Theta_2 = \int\limits_{\Im_i^{(2)}} V_i(t)dt; \tag{5}$$

$$\Theta = \Theta_1 + \Theta_2 = V_i^{(\max)} \cdot T_i^{(\max)} + \Theta_2. \tag{6}$$

In this case, required probability is equal to

$$p_i = \frac{\Theta_1}{\Theta} = \frac{V_i^{(\max)} \cdot T_i^{(\max)}}{V_i^{(\max)} \cdot T_i^{(\max)} + \Theta_2}. \tag{7}$$

Applying the integral theorem of mean value, we obtain:

$$\exists\, \eta \in [0, T_i] \mid \Theta_2 = V_i(\eta) \cdot \left(T_i - T_i^{(\max)} \right). \tag{8}$$

Given that $\Theta = \Theta_1 + \Theta_2 = V_i^{(av)} \cdot T_i$, as well as (5), we can find the value of Θ_2:

$$\Theta_2 = \Theta_2 = V_i^{(av)} \cdot T_i - V_i^{(\max)} \cdot T_i^{(\max)}. \tag{9}$$

We should notice, that from (7–9) it follows:

$$V_i(\eta) = \frac{\Theta_2}{T_i - T_i^{(\max)}} = \frac{V_i^{(cp)} \cdot T_i - V_i^{(\max)} \cdot T_i^{(\max)}}{T_i - T_i^{(\max)}}; \tag{10}$$

$$p_i = \frac{V_i^{(\max)} \cdot T_i^{(\max)}}{V_i^{(av)} \cdot T_i} = k_i^{(p)} \cdot \frac{T_i^{(\max)}}{T_i}. \tag{11}$$

Since $\Theta_2 > 0$ and taking (9) into account, we obtain:

$$V_i^{(av)} \cdot T_i - V_i^{(\max)} \cdot T_i^{(\max)} > 0; \;\; \frac{V_i^{(\max)}}{V_i^{(av)}} \cdot \frac{T_i^{(\max)}}{T_i} < 1 \Rightarrow k_i^{(p)} \cdot \frac{T_i^{(\max)}}{T_i} < 1. \tag{12}$$

If we know a probability value for the scenario when i-th TPS reaches its peak transfer rate, it is possible to calculate the probability that the peak transfer rate would not be reached:

$$q_i = 1 - p_i = P\left(V_i(t) < V_i^{(\max)}\right). \tag{13}$$

For the following analysis we should know the dispersion for possible values of V_i stochastic variable inside the given interval of $[0, T_i]$. Thus, we need to define the density of $f(V_i)$ distribution for given stochastic variable, when $\sup D[V_i]$ is reached. Then the dispersion is calculated using the following equation:

$$D[V_i] = \int\limits_{-\infty}^{\infty} V_i^2 f(V_i)dV_i - M^2[V_i] = \int\limits_0^{V_i} V_i^2 f(V_i)dV_i - \left(\int\limits_0^{V_i} V_i f(V_i)dV_i\right)^2. \tag{14}$$

It should be noted that V_i parameter possesses the value within the range from $V_i^{(\min)}$ to $V_i^{(\max)}$ $\left(0 \le V_i^{(\min)} < V_i^{(\max)}\right)$. Hence we can approximate $D[V_i]$ via division the interval of $\Delta = \left[V_i^{(\min)}; V_i^{(\max)}\right]$ into n equal parts of $\Delta_j = \left[V_i^{(j-1)}; V_i^{(j)}\right]$ in such a way that $\bigcup_{j=1}^n \Delta_j = \Delta$; $\bigcap_{j=1}^n \Delta_j = \varnothing$.

Then, it is possible to choose ξ_j point within each interval Δ_j using integral theorem of mean value and the dispersion could be written, based on (14), as

$$D[V_i] = \sum_{j=1}^m \xi_j^2 \cdot \int\limits_{V_i^{j-1}}^{V_i^j} f(V_i)dV_i - \left(\sum_{j=1}^m \xi_j \int\limits_{V_i^{j-1}}^{V_i^j} f(V_i)dV_i\right)^2. \tag{15}$$

Having noticed, that $\int_{V_i^{j-1}}^{V_i^j} f(V_i)dV_i = p_j$, $\sum_{j=1}^n p_j = 1$ and denote $\vec{p} = (p_1, \ldots p_n)$, it is possible to find dispersion extremum (15), employing Lagrange's method of multipliers for the optimization problem:

$$\sum_{j=1}^n p_\mathrm{j} \cdot \xi_j^2 - \left(\sum_{j=1}^n p_j \cdot \xi_j\right)^2 \to extr \quad \text{where} \sum_{j=1}^n p_j = 1; \ \ \mathrm{p_j} \ge 0. \tag{16}$$

Let's compile Lagrangian for the problem (16):

$$L(\vec{p}, \lambda) = \sum_{j=1}^n p_j \cdot \xi_j^2 - \left(\sum_{j=1}^n p_j \cdot \xi_j\right)^2 - \lambda\left(\sum_{j=1}^n p_j - 1\right)$$

and find its unconditional extremum using the following system of equations:

$$\begin{cases} \frac{\partial L}{\partial p_j} = 0, j = \overline{1, n}; \\ \frac{\partial L}{\partial \lambda} = 0 \end{cases} \Rightarrow \begin{cases} \xi_j^2 - 2\xi_j \left(\sum_{j=1}^{n} p_j \cdot \xi_j \right) - \lambda = 0; \\ \sum_{j=1}^{n} p_j - 1 = 0. \end{cases} \quad (17)$$

From the first equation we can find ξ_j:

$$\xi_j = \tilde{m}_\xi \pm \sqrt{\tilde{m}_\xi^2 + \lambda}\ \text{, where } \tilde{m}_\xi = \sum_{j=1}^{n} p_j \cdot \xi_j\ , \quad (18)$$

that is for the Lagrange multipliers, satisfying a condition of $\lambda > -\tilde{m}_\xi^2$, there exist only two following distinct values of ξ_j: ξ_1 and ξ_2:

$$n = 2;\ p_1 = p_i;\ p_2 = 1 - p_1 = q_i,$$

that is the probability density function has discrete representation, and the density of distribution is $f(V_i) = p_i\delta(V_i - \xi_1) + q_i(V_i - \xi_2)$, where $\delta(\cdot)$ is a delta function.

Now we can calculate mean value and dispersion for stochastic value of V_i:

$$M[V_i] = p_i\xi_1 + q_i\xi_2 = p_i\xi_1 + (1 - p_i)\xi_2; \quad (19)$$

$$\begin{aligned} D[V_i] &= p_i\xi_1^2 + (1 - p_i)\xi_2^2 - M^2[V_i] = p_i\xi_1^2 + (1 - p_i)\xi_2^2 - p_i^2\xi_1^2 \\ &\quad - 2p_i\xi_1\xi_2(1 - p_i) - (1 - p_i)^2\xi_2^2 = (1 - p_i)p_i(\xi_1 - \xi_2)^2. \end{aligned} \quad (20)$$

It is evident, that sup $D[V_i]$ is achieved at $(\xi_1 - \xi_2)^2$ multiplier largest extremum, that is $\xi_1 = V_i^{(\max)}$; $\xi_2 = V_i^{(\min)}$, and, hence,

$$M[V_i] = p_i V_i^{(\max)} + (1 - p_i)\, V_i^{(\min)}. \quad (21)$$

But, $M[V_i] = V_i^{(av)}$, and the $V_i^{(\min)}$ value for the real processes without loss of generality could be equated to zero. Therefore,

$$V_i^{(av)} = p_i V_i^{(\max)} \Rightarrow \ p_i = \frac{V_i^{(av)}}{V_i^{(\max)}} = \left(k_i^{(p)}\right)^{-1}; D[V_i] = (1 - p_i)p_i\left(V_i^{(\max)}\right)^2. \quad (22)$$

Let us find the p_i value, where maximum of V_i dispersion is achieved:

$$\frac{\mathrm{d}D[V_i]}{dp_i} = \left(V_i^{(\max)}\right)^2(-2p_i + 1) \ \Rightarrow\ p_i = 1/2, \quad (23)$$

that is, when $p_i = q_i = 1/2$, there is maximal dispersion of V_i stochastic value. It can also be noted that, in terms of TPS evaluation, the maximum duration of a packet, which is transmitted with peak rate of $T_i^{(\max_d)}$, can be also defined as:

$$k_i^{(p)} \cdot \frac{T_i^{(\max_d)}}{T_i} = \frac{1}{2} \Rightarrow \quad T_i^{(\max_d)} = \frac{T_i}{2 \cdot k_i^{(p)}}. \tag{24}$$

Thus, static model of single source traffic for Big Data centers, proposed in this subsection, allows performing probabilistic assessment of data transmission rate, using approximation of its density function. The model is the basis for further modelling of aggregated traffic related to several sources.

2.2 *Spline Interpolation of Big Data Centers Traffic*

Let us suppose that Big Data centers traffic is processed by K services, moreover k-th service $\left(k = \overline{1, K}\right)$ processes the traffic generated by L_k sources. Let $W^{(k,\ell_k)}(t)$ is a data volume, arriving to k-th service from ℓ_k $\left(\ell_k = \overline{1, L_k}\right)$ source within single session $t \in [t_0, T_c]$. For analysis of given function character let us use local splines technique, which allows considering various behavior patterns of $W^{(k,\ell_k)}(t)$, within different time periods of the session, with least approximation error as regard to a norm in $L_2[t_0, T_c]$ space. For the constriction of local $n^{(k,\ell_k)}$-spline with interpolation nodes of $t_i^{(k,\ell_k)}$, where $i = \overline{0, n^{(k,\ell_k)}}$, $t_0^{(k,\ell_k)} = t_0$, $t_{n^{(k,\ell_k)}}^{(k,\ell_k)} = T_c$, it is necessary at every $\tau_i^{(k,\ell_k)} = \left[t_{i-1}^{(k,\ell_k)}, t_i^{(k,\ell_k)}\right]$ interpolation interval to build the interpolation polynomial of $P_{n^{(k,\ell_k)},i}^{(k,\ell_k)}(t)$. Then spline, which approximates $W^{(k,\ell_k)}(t)$, in a general case can be represented as

$$S^{(k,\ell_k)}(t) = \sum_{i=1}^{n^{(k,\ell_k)}} \alpha_i^{(k,\ell_k)} \cdot P_{n^{(k,\ell_k)},i}^{(k,\ell_k)}(t), \tag{25}$$

where $\alpha_i^{(k,\ell_k)}$ is a set of constant coefficients in $\left\{P_{n^{(k,\ell_k)},i}^{(k,\ell_k)}\right\}$ linear-independent functional basis, which greatly depends on behavioral performance of $W^{(k,\ell_k)}(t)$ on the given interval of $\tau_i^{(k,\ell_k)}$. In that way, if

$$W^{(k,\ell_k)}(t) \in C^{\beta}(\tau_i), \tag{26}$$

then corresponding approximating polynomial can have a degree less than $2\beta + 1$, and if $W^{(k,\ell_k)}(t) \in C^{\infty}(\tau_i)$ then the approximation is possible using trigonometric splines.

Consider the construction of piecewise-polynomial spline inside the τ_i interval when assumption (26) is fulfilled. Let us specify the nodes of $t_{j,i}^{(k,\ell_k)}$ $\left(j = 0, n_i^{(k,\ell_k)}\right)$

and define appropriate values of $W^{(k,\ell_k)}\left(t_{j,i}^{(k,\ell_k)}\right)$ function. As a degree of approximating polynomial let us consider $\gamma = \overline{0,\ \beta - 1}$ number. For each j-th node we build interpolation polynomial of $P_\beta\left(t, W_{j,i}^{(k,\ell_k)}\right)$ based on the values in $t_{j-\gamma,i}^{(k,\ell_k)}, \ldots, t_{j-\gamma+\beta,i}^{(k,\ell_k)}$ nodes. Then relevant piecewise-polynomial spline for $\tau_i^{(k,\ell_k)}$ interval can be defined as

$$S_i^{(k,\ell_k)}(t,\beta) = Q_{2\beta+1}^{(k,\ell_k)}(t,j), \tag{27}$$

where

$$\frac{d^m}{dt^m} Q_{2\beta+1}^{(k,\ell_k)}(t,j)\bigg|_{t=t_{j,i}^{(k,\ell_k)}} = \frac{d^m}{dt^m} P_\beta\left(t, W_{j,i}^{(k,\ell_k)}\right)\bigg|_{t=t_{j,i}^{(k,\ell_k)}};$$

$$\frac{d^m}{dt^m} Q_{2\beta+1}^{(k,\ell_k)}(t,j)\bigg|_{t=t_{j+1,i}^{(k,\ell_k)}} = \frac{d^m}{dt^m} P_\beta\left(t, W_{j,i}^{(k,\ell_k)}\right)\bigg|_{t=t_{j+1,i}^{(k,\ell_k)}}; \quad m = \overline{1,\beta}.$$

In this case $Q_{2\beta+1}^{(k,\ell_k)}(t,j) = P_\beta\left(t, W_{j,i}^{(k,\ell_k)}\right) + R_{2\beta+1}^{(k,\ell_k)}(t,j)$, where $R_{2\beta+1}^{(k,\ell_k)}(t,j)$ is a correction to classical interpolation polynomial inside τ_i interval; it is equal, when (26) is fulfilled, to:

$$R_{2\beta+1}^{(k,\ell_k)}(t,j) = (t_{j+1} - t_j)^{\beta+1} \cdot W_{2\beta+1}^{(k,\ell_k)}(t_{j-\gamma}, t_{j-\gamma+1}, \ldots, t_{j-\gamma+\beta+1}) \cdot q_{2\beta+1}(\tilde{t}, j),$$

where $W_{2\beta+1}^{(k,\ell_k)}\left(t_{j-\gamma}, t_{j-\gamma+1}, \ldots, t_{j-\gamma+\beta+1}\right)$ is the difference quotient of $2\beta+1$ order;

$$\tilde{t} = \frac{t - t_j}{t_{j+1} - t_j};\ \theta_r(\tilde{t}) = \frac{\tilde{t}^{j+1} \cdot (\tilde{t} - 1)^r}{r!\, j!} \cdot \sum_{m=0}^{\beta-r} (-1)^m \frac{(\beta+m)!}{m!} (\tilde{t} - 1)^m;$$

$$q_{2\beta+1}(\tilde{t}, j) = \frac{t_{j+\beta-\gamma+1} - t_{j-\gamma}}{t_{j+1} - t_j} \cdot \sum_{r=0}^{\beta} \left(\left(\frac{d^2}{d\tilde{t}^2} \prod_{\xi=0}^{\beta} \left(\tilde{t} - \frac{t_{j-\gamma+\xi}}{t_{j+1} - t_i} \right) \bigg|_{\tilde{t}=1} \right) \cdot \theta_r(\tilde{t}) \right).$$

It such case, a Lagrange interpolation polynomial of β order can be chosen as interpolation polynomial:

$$P_\beta\left(t, W_{t_{j,i}}^{(k,\ell_k)}\right) = \frac{\sum_{m=0}^{\beta} W^{(k,\ell_k)}\left(t_{m,i}^{(k,\ell_k)}\right) \times \prod_{m_1=0}^{\beta}\left(t - t_{m_1,i}^{(k,\ell_k)}\right)}{\prod_{m_2=0}^{m-1}\left(t_{m,i}^{(k,\ell_k)} - t_{m_2,i}^{(k,\ell_k)}\right) \cdot \prod_{m_3=m+1}^{\beta}\left(t_{m,i}^{(k,\ell_k)} - t_{m_3,i}^{(k,\ell_k)}\right)}. \tag{28}$$

The measure of inaccuracy of (28) can be estimated as a norm of $R_{\beta,i}^{(k,\ell_k)} = \left\| W^{(k,\ell_k)}(t) - S_i^{(k,\ell_k)}(t,\beta) \right\|_{L_2(\tau_i)}$ deficiency value, and it follows that

$$R_{\beta,j,i}^{(k,\ell_k)}(t) \leq \sup_{t \in \tau_{j,i}^{(k,\ell_k)}} \left| \frac{d^{n_i^{(k,\ell_k)}}}{dtn_i^{(k,\ell_k)}} W_{j,i}^{(k,\ell_k)}(t) \right| \cdot \frac{\prod_{m=0}^{\beta} \left(t - t_{m,i}^{(k,\ell_k)}\right)}{(\beta+1)!}, \tag{29}$$

that is, with further increase in β the accuracy of interpolation approximation also increases, however processing complexity is increasing in its turn.

That's why the choice of $\beta^{(i)} \leq \beta$ value for each τ_i interval is a separate task, which depends on the requirements to $W^{(k,\ell_k)}(t)$ approximation and traffic character.

2.3 The Synthesis of Dynamic Two-Dimensional Model of Big Data Center Traffic

Key traffic parameters that allows determining its main characteristics are the volume of transmitted data and transmission rate. The process of transmission rate change is stochastic. However, finding of probabilistic characteristics corresponding to real traffic is problematical in many cases. In a case of frequently changing transmission rate, it is impossible re-planning network resources rapidly. Therefore, it is necessary to synthesize such network traffic model that could allow rapid finding the limits of its main characteristics variation on the basis of cumulative statistics.

Let us consider that characteristics of Big Data centers traffic are unknown, however consider that the time interval of the next session can be divided into the finite number of subintervals. Within each of the subintervals it is possible to determine statistical characteristics of the traffic. However, in contrast to existing models, we will describe the process behavior dependent not only on the time, but also on initial data (in particular, on initial data transmission rate within each session).

Let $\left\{ W^{(k,\ell_k)}(t) \right\}$ is dynamic functions family that characterizes data volume, which incomes to k-th service $(k = \overline{1,K})$ from ℓ_k source with average rate of $V^{(k,\ell_k)}(t) = \frac{dW^{(k,\ell_k)}(t)}{dt}$. Such rate can be described by an a priori unknown dependence of $V^{(k,\ell_k)}(t) = \Psi\left(W^{(k,\ell_k)}, t\right)$. Then

$$V^{(k,\ell_k)}(t_0) = \left. \frac{dW^{(k,\ell_k)}(t)}{dt} \right|_{t=t_0} = V_0^{(k,\ell_k)}; \quad t \in \left[t_0^{(k,\ell_k)}, t_0^{(k,\ell_k)} + T^{(k,\ell_k)} \right], \tag{30}$$

where $t_0^{(k,\ell_k)}$, $T^{(k,\ell_k)}$ are the lower limit and length of the considered time interval, appropriately.

Let us perform R observations within N different sessions, with initial data of $V_{0,n}^{(k,\ell_k)}$ $(n = \overline{0, N-1})$ at $t_{r,n}^{(k,\ell_k)}$, $r = \overline{0, R-1}$ time moments. For the setting of two-dimensional model area within limits of each session, we consider relative

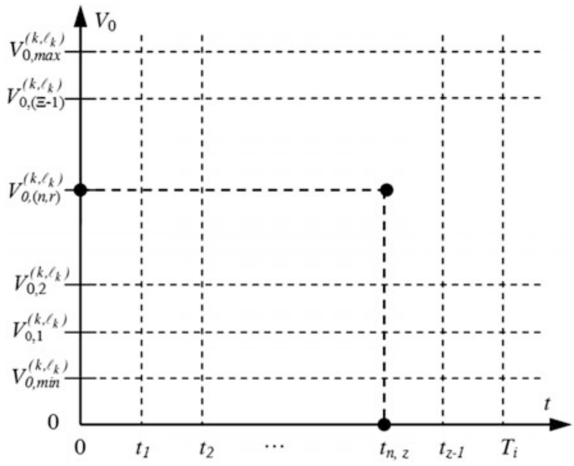


Fig. 2 Diagram of two-dimensional interpolation area

time, shifted for $t_{0,n}^{(k,\ell_k)}$, and we choose $T^{(k,\ell_k)} = \max_n T_n^{(k,\ell_k)}$ as the modelling interval length (Fig. 2).

Obtained measurements fall within permissible region of the model

$$\Im^{(k,\ell_k)} = \left[V_{\min}^{(k,\ell_k)}, V_{\max}^{(k,\ell_k)}\right] \times \left[0; T^{(k,\ell_k)}\right] \subset \Re^2, \tag{31}$$

where $V_{\min}^{(k,\ell_k)}, V_{\max}^{(k,\ell_k)}$ are minimal and maximal variation borders of first-order derivatives for functions of $\left\{ W^{(k,\ell_k)}(t) \right\}$ family within considered sessions. Hence, a problem of dynamic model construction for Big Data center traffic could be formulated in the following way: to build a family of $\left\{ \tilde{W}^{(k,\ell_k)}\left(V_0^{(k,\ell_k)}, t\right) \right\}$ spline-functions, which satisfy to the following condition of model adequacy:

$$\begin{gathered} \rho\left(W^{(k,\ell_k)}\left(V_0^{(k,\ell_k)}, t\right), \tilde{W}^{(k,\ell_k)}\left(V_0^{(k,\ell_k)}, t\right)\right) \le \varepsilon^{(k,\ell_k)}; \\ \left(V_0^{(k,\ell_k)}, t\right) \in \Im^{(k,\ell_k)}, \end{gathered} \tag{32}$$

where $\tilde{W}(\cdot)$ is a function, which approximates $W(\cdot)$ with given initial conditions, and $\left(t, V_0^{(k,\ell_k)}\right)$ argument separates out certain function (for which assumption (30) is fulfilled) from the $\left\{W^{(k,\ell_k)}(t)\right\}$ family.

Since $\Im^{(k,\ell_k)} \subset \Re^2$, then in (32) distance $\rho(\cdot)$ is a standard metric for L^2 space, that is

$$\max_{t,W_0^{(k,\ell_k)}} \left| W^{(k,\ell_k)}\left(V_0^{(k,\ell_k)}, t\right) - \tilde{W}^{(k,\ell_k)}\left(V_0^{(k,\ell_k)}, t\right) \right| \leq \varepsilon^{(k,\ell_k)};$$
$$\left(V_0^{(k,\ell_k)}, t\right) \in \Im^{(k,\ell_k)}. \tag{33}$$

For the further construction of the model, we employ the outcomes of real traffic statistical data gathering:

$$M^{(k,\ell_k)} = \left\{ m_{nr}^{(k,\ell_k)}, n = \overrightarrow{0, N-1}, r = \overrightarrow{0, R-1} \right\}, \tag{34}$$

where each $m_{nr}^{(k,\ell_k)}$ measurement can be defined as the sum of modeled measurement $\tilde{m}_{nr}^{(k,\ell_k)}$ and $\delta\tilde{m}_{nr}^{(k,\ell_k)}$ deficiency, which is related to measurement errors. These errors are uncorrelated and have a zero mean value.

Since traffic elasticity is expected to be in nodes, then for the synthesis of mathematical model we consider generalized interpolation polynomial of two-dimensional spline for $\tilde{W}^{(k,\ell_k)}\left(V_0^{(k,\ell_k)}, t\right)$ function in a basis of L^2 metric space over $\Im^{(k,\ell_k)}$ area:

$$\tilde{W}^{(k,\ell_k)}\left(V_0^{(k,\ell_k)}, t\right) = \sum_{(\xi=\overline{0,R-1})}^{(\zeta=\overline{0,N-1})} \alpha_{\zeta,\xi}^{(k,\ell_k)} \times f_{\zeta,\Xi}^{(k,\ell_k)}\left(V_0^{(k,\ell_k)}\right) \phi_{\xi,\mathrm{Z}}^{(k,\ell_k)}(t), \tag{35}$$

where $f_{\zeta,\Xi}^{(k,\ell_k)}\left(V_0^{(k,\ell_k)}\right) \phi_{\xi,\mathrm{Z}}^{(k,\ell_k)}(t)$ is a basis of interpolation decomposition in L^2, which is defined off $\Im^{(k,\ell_k)}$ wherein Ξ and Z define the spline order along the n and r axes respectively; $\alpha_{\zeta,\xi}^{(k,\ell_k)}$ are unknown coefficients of the decomposition; for the summation over several indices the notation of $\sum_{(j_1=\overline{1,J})}^{(i_1=\overline{1,I_1})}(\cdot)$ is used, which is equivalent to generally accepted of $\sum_{i_1=1}^{I_1}\sum_{i_2=1}^{I_2}(\cdot)$.

For the $\varepsilon^{(k,\ell_k)}$-adequacy (35) it is necessary, during while spline construction, to superpose the following delimitation onto the elements of $\delta\tilde{\mathrm{M}}(\cdot)$ deficiencies set from (34):

$$\max_{t,V_0^{(k,\ell_k)}} \left| \delta\tilde{m}^{(k,\ell_k)}\left(V_0^{(k,\ell_k)}, t\right) \right| \leq \varepsilon^{(k,\ell_k)};$$
$$\delta\tilde{m}^{(k,\ell_k)} \in \delta\tilde{M}^{(k,\ell_k)}, \quad \left(V_0^{(k,\ell_k)}, t\right) \in \Im^{(k,\ell_k)}, \tag{36}$$

which will allow, according to adequacy estimation method, defining values of spline orders.

If (36) is met, then unknown coefficients of decomposition (35) can be found using the following system of algebraic equations:

$$\sum_{(\xi=\overline{0,R-1})}^{(\zeta=\overline{0,N-1})} \alpha_{\zeta,\xi}^{(k,\ell_k)} \times f_{\zeta,\Xi}^{(k,\ell_k)}\left(V_{0,n}^{(k,\ell_k)}\right)\phi_{\xi,Z}^{(k,\ell_k)}(t_r) = m_{nr} \tag{37}$$

Applying Lagrange polynomial as basis functions in interpolation decomposition in (35) and (37) results in the following model of dynamic traffic:

$$\begin{aligned}\tilde{W}^{(k,\ell_k)}\left(V_0^{(k,\ell_k)},t\right) &= \sum_{(\xi=\overline{0,R-1})}^{(\zeta=\overline{0,N-1})} \alpha_{\zeta,\xi}^{(k,\ell_k)} \cdot L_{\zeta,\Xi}^{(k,\ell_k)}\left(V_0^{(k,\ell_k)}\right) \cdot L_{\xi,Z}^{(k,\ell_k)}(t) \\ &= \sum_{(r=\overline{0,R-1})}^{(n=\overline{0,N-1})} m_{n,r}^{(k,\ell_k)} \cdot \sum_{\zeta=0}^{Z} \frac{\Im_\zeta\left(V_0^{(k,\ell_k)}\right)}{\Im_\zeta\left(V_{0,\zeta}^{(k,\ell_k)}\right)} \cdot \sum_{\xi=0}^{\Xi} \frac{\aleph_\xi(t)}{\aleph_\xi\left(t_{n,\xi}\right)},\end{aligned} \tag{38}$$

where $\Im_\zeta\left(V_0^{(k,\ell_k)}\right) = \prod_{i=0,i\neq\zeta}^{Z}\left(V_0^{(k,\ell_k)} - V_{0,i}^{(k,\ell_k)}\right)$; $\aleph_\xi(t) = \prod_{j=0,\,j\neq\xi}^{\Xi}\left(t - t_{n,j}\right)$.

In order to fulfil adequacy condition (38) it is necessary to choose γ decomposition steps of $\Im^{(k,\ell_k)}$ area with $\left(V_{0,n}^{(k,\ell_k)},t\right)$ nodes according to the condition of error estimation minimization for two-dimensional interpolation. Performed analysis revealed that when a number of elasticity disturbance points $W^{(k,\ell_k)}\left(V_0^{(k,\ell_k)},t\right)$ is relatively small, then dimension of γ decomposition is small and allows rapid analysis of network performance quality.

Proposed dynamic two-dimensional model allows to synthesize the Big Data center aggregated traffic (generated by several sources), on the basis of accumulated statistics related to measurements of several sessions data.

3 Improving Energy Efficiency of Big Data Centers

Improving energy efficiency of Big Data centers could be achieved by balancing load in physical lines. In turn, at a level of virtual channels, load balance could be achieved on the basis of traffic behavior short-term prediction. For the prediction we can use set of traffic models using measurements, proposed in previous subsection. However, the presence of big number of traffic spikes can reduce the model adequacy. That is the reason why in this section we propose a method of aggregated traffic spikes characterization.

3.1 Method of Aggregated Traffic Spikes Characterization

In a case of traffic convergence from individual sources and its further aggregation the spikes number is greatly increasing that complicates its spline-approximation.

Thus, during traffic control, it is necessary to know the characteristics of such spikes. We should notice that data transmission rate $V(t)$, available to the data flow, is a stochastic process of $\xi(t)$ and has probabilistic description. By physical reasons, it always exists maximum transmission rate limit:

$$\xi_{\max} = \max\ \xi(t). \tag{39}$$

The average data transmission rate within a session time is equal to:

$$m = M[\xi(t)] = \frac{1}{T}\int_{0}^{T} \xi(t)dt. \tag{40}$$

In Sect. 2.1 the following burstiness coefficient was considered:

$$B = \tau_{\max}/m,\ \tau_{\max} = \max\xi(t). \tag{41}$$

However, the given characteristic gives no way for estimating the rate of $\xi(t)$ stochastic process over time, since it has no relation with its spectral and, hence, correlation properties. In order to consider the stochastic function time scale for describing data transmission rate, it was introduced one more traffic characteristic—the peak rate average time T_p, which is defined by spikes' duration time. To determine such characteristic, it is necessary to choose the measurement level of peak duration, that will allow to determine:

– required volume of buffer memory;
– jitter estimation (frequency of delay time pulsing).

Figure 3 represents stochastic process $\xi(t)$ during time of T, where C is a fixed level; $\tau = \tau_1 + \tau_2$ is a total duration of spikes; $S = S_1 + S_2$ is a total area of spikes; t_{01} and t_{02} are appropriate time moments when the spikes start.

Real stochastic processes are continuous functions of time with bounded spectrum due to inertial properties of a message source and processing units, which could be considered as a low-pass filter. In particular, such notions, as short and long messages and gaps between them, are relative and depend on performance of a

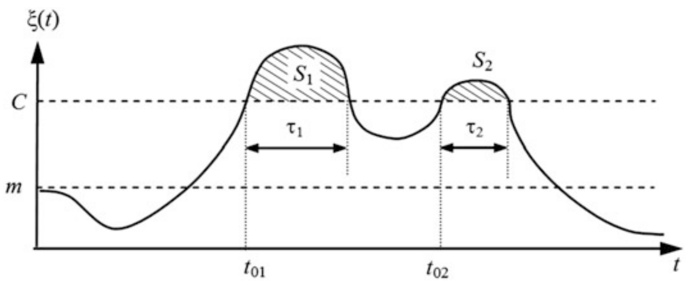


Fig. 3 Example of $\xi(t)$ stochastic process implementation

switching system. Traffic source is considered to be pulsing, if total time of through connection establishment and disconnection is less than interval of next message. It is the basis for using of parabolic approximation during the derivation of main formulas. These functions have finite numbers of their maximums and minimums on bounded time interval of *T*.

Implementation of $\xi(t)$ processes can cross the level bottom up for several times (with positive derivative). The number of such crossings corresponds to the number of positive spikes for a stochastic process above the level of *C*, for which it exists the following very rigorous mathematical expression:

$$N^{+}(C,T) = \int_{0}^{T} dt \int_{0}^{\infty} \xi' \cdot f(C,\xi';t) d\xi', \quad \xi' = \xi'(t), \tag{42}$$

where *f*(·) is a distribution density function for $\xi(t)$ stochastic process.

For stationery stochastic processes, Eq. (42) becomes simpler:

$$N^{+}(C,T) = f(C) \int_{0}^{\infty} \xi' \cdot f(\xi')\, d\xi'. \tag{43}$$

Equation (43) allows defining average number of spikes for any arbitrary distribution law of $\xi(t)$ stochastic process.

If Gaussian process is stationery, then the average value of total number of process crossings at level *C* per unit time is equal to

$$N(C) = \frac{1}{\pi}\sqrt{-\rho_0''}\exp\left[-\frac{1}{2}\left(\frac{c-m}{\sigma}\right)^2\right], \tag{44}$$

where ρ_0'' is a second derivative from correlation coefficient; σ is a mean-square deviation.

In a case of sufficiently great level $C \gg 0$, the overwhelming majority of spikes have short duration and is equal to a number of maximums, which exceed level of *C*. A smoothly varying stochastic process of $\xi(t)$ in a small neighborhood of a maximum set could be approximated using the parabola.

Let t_0 is the time moment when considered positive spike starts, that is $\xi(t_0) = C, \;\; \xi'(t_0) > 0$. Let us expand this function of $\xi(t)$ in a Taylor series in the neighborhood of t_0 point and let us confine to the quadratic term of series:

$$\xi(t) \approx \xi(t_0) + \xi'(t_0)(t-t_0) + \frac{1}{2}\xi''(t_0)(t-t_0)^2. \tag{45}$$

Considering the fact that with τ spike duration the equation of $\xi(t_0) = \xi(t_0+\tau) = C$ is true, we obtain:

$$\tau = -2\xi'(t_0)/\xi''(t_0). \tag{46}$$

Having applied received correlations to the normal stationery process with correlation function of $K(\tau) = \sigma^2 \cdot \rho(\tau)$, we can find the probability density of spikes duration when the value of C is large:

$$f(\tau, C) = \frac{1}{12\sqrt{\pi}} \left(-\rho_0''\right) \frac{C\tau}{\sigma} \exp\left[-\frac{1}{8}\left(-\rho_0''\right)\frac{C^2\tau^2}{\sigma^2}\right], \tag{47}$$

where $\rho_0'' = \rho''(t_0)$.

Parabolic approximation of spikes shape allows finding distribution of the spikes area:

$$S = \int\limits_{t_0}^{t_0+\tau} [\xi(t) - C]\, \mathrm{dt}. \tag{48}$$

Considering Eq. (46) and parabolic approximation of a spike,

$$S = \frac{2 \cdot (\xi'(t_0))^3}{3 \cdot (\xi''(t_0))^2}. \tag{49}$$

Then the probability density of distribution for the spikes exceeding level of C:

$$f(S, C) = \frac{1}{3}\lambda^{2/3} S^{-1/3} \exp\left[-\frac{1}{2}(\lambda S)^{2/3}\right], \tag{50}$$

where

$$\lambda = \frac{2}{3}\frac{C^2}{\sigma^3}\sqrt{-\rho_0''};\, C \gg \sigma. \tag{51}$$

Derived formula (50) is critically important, since spikes area could be considered as a buffer memory volume in switching nodes. Such volume is required to avoid the packet losses when network resources are insufficient (in that case level of C defines highest transmission rate available). Delayed packets could be transmitted at the intervals when traffic intensity is lower than the mentioned level. Moreover, allowed delay time, obviously, will be defined based on network requirements.

Let us carry out a more detailed analysis of formula (50). Probability density function of spikes distribution in terms of area exceeding level of C for normal stationery process, normalized by λ coefficient, is represented in Fig. 4.

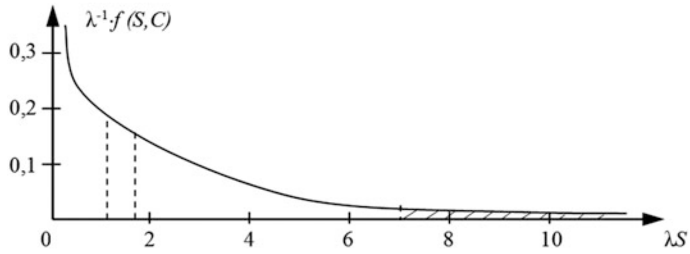


Fig. 4 Spikes area distribution

Let us define the average value of spikes area.

$$\begin{aligned} S_{av} = M[f(C,S)] &= \frac{1}{3}\int_0^\infty S\lambda^{2/3}S^{-1/3}\exp\left[-\frac{1}{2}(\lambda S)^{2/3}\right]ds \\ &= \frac{1}{3}\lambda^{2/3}\frac{3}{2}\left(\frac{1}{2}\lambda^{2/3}\right)^{-5/2}\Gamma\left(\frac{5}{2}\right) = \sqrt{\frac{\pi}{2}}\cdot\frac{3}{\lambda}, \end{aligned} \tag{52}$$

or, taking (51) into account,

$$S_{av} = \sqrt{\frac{\pi}{2}}\frac{9\sigma^3}{2C^2\sqrt{-\rho_0''}}. \tag{53}$$

The average value of spikes duration, that define additional delay due to storing of redundant data in buffer memory, is equal to

$$\tau_{av} = \frac{2\pi\sigma}{C\sqrt{-\rho_0''}}. \tag{54}$$

As an example let us consider media stream traffic with correlation coefficient of $\rho(t) = \exp(-\alpha\tau^2)$. Considering Eqs. (44), (47) and (50):

$$N(C) = A\Delta f_e \cdot \exp\left[-\frac{1}{2}\cdot\frac{(C-m)^2}{\sigma^2}\right]; \tag{55}$$

$$f(\tau, C) = \frac{1}{3\sqrt{\pi}}B\cdot\Delta f_e^2\frac{C\tau}{\sigma}\cdot\exp\left[-\frac{1}{2}B\Delta f_e^2\frac{C^2\tau^2}{\sigma^2}\right]; \tag{56}$$

$$\lambda = D\frac{C^2}{\sigma^3}\Delta f_e, \tag{57}$$

where constants are equal to the following: $A = \sqrt{8/3}$; $B = 2\pi/3$; $D = 4\sqrt{2}\pi/\left(3\sqrt{3}\right)$.

Combine Eqs. (53) and (54) we could find the relation between duration and area of spikes:

$$S_{av} = \frac{9}{4}\sqrt{\frac{1}{2\pi}\frac{\sigma^2}{C}}\tau_{av}. \tag{58}$$

If we compare S_{av} with the required buffer memory volume, then the average memory volume $\bar{S}_б$ can be determined in the following way:

$$\bar{S}_b = k_p \cdot \frac{\sigma^2}{C} T_p, \tag{59}$$

where k_p is a scaling factor.

The restriction level of $C = C_n$ can be chosen in the way that area of spikes with a probability of P_0 would be less than predetermined value of S_n.

Let us define the lower limit value of shaded area in Fig. 4 as $(\lambda S)_n$, where the latter is peaks appearance probability and the area of such peaks shall exceed some predetermined value. Therefore,

$$P_0(S) = P(S \le S_n) = 1 - \int_{+S_n}^{\infty} f(S,C)ds = 1 - \exp\left(-\frac{1}{3}\lambda^{2/3}S_n^{2/3}\right), \tag{60}$$

where λ can be defined using Eq. (54).

Influence curves (60) where $\lambda = Const$ are represented in Fig. 5.

They allow to determine the threshold of C_n, which defines allowable transmission rate. Transmission rate is dependent on values of ρ_0 and S_n, which also have an impact on λ_n parameter, which, in turn, can be uniquely determined using the following expression:

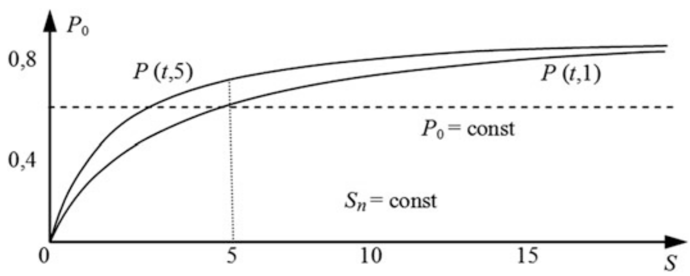


Fig. 5 Dependence of P_0 probability on spikes area

$$C_n = \sqrt{\frac{\lambda_n \sigma^3}{D\Delta f_e}}. \tag{61}$$

In addition, the threshold is measured from traffic average value, that means $C_n = C - m$ and corresponds to geometrical locus of (60) curve points, that passes through a point where of the straight lines of $P_0 = Const$ and $S_n = Const$ intersect.

Proposed method of aggregated traffic spikes characterization allows prediction the aggregated traffic containing big number of spikes inside Big Data center.

3.2 Analysis of Big Data Centers Energy Efficiency Increase Based on Constructed Models

During allocation of the network resource represented by virtual connections for Big Data center traffic we use methods, which allow short-term prediction of the traffic related to the virtual connection. In such a case, prediction error has an impact on quality of Big Data center traffic control. This, in turn, can reduce the effectiveness of network resource utilization.

Let $\bar{\bar{X}}_i^{(m)}(n+1)$ and $X_i^{(m)}(n+1)$ be predicted value, considering n last measurements, and real value of capacity for i-th data flow within m-node, respectively. Then prediction relative error could expressed as

$$\tau_i(n) = \frac{\bar{\bar{X}}_i^{(m)}(n+1) - X_i^{(m)}(n+1)}{X_i^{(m)}(n+1)}. \tag{62}$$

Let us define the maximal one from all prediction relative errors

$$\gamma = \max_i(|\tau_i(n)|).$$

Maximum value for the T_h capacity could be determined in the following way:

$$T_h = \lim_{M\to\infty} \frac{\sum_{n=1}^{M} W_i(n)}{M \times T_p} = \lim_{M\to\infty} \frac{M_c\overline{W_c} + M_d\overline{W_d} + M_u\overline{W_u}}{M \times T_p}, \tag{63}$$

where T_p is transmission time of a single packet, M is total amount of traffic samples, $M = M_c + M_d + M_u$; M_c—number of intervals, on which congestion takes place; M_d is a number of intervals that follow congested interval; $M_u = M - M_c - M_d$ is a number of the rest of intervals; $\overline{W_c}$, $\overline{W_d}$, $\overline{W_u}$ are average rates within the M_c, M_d, M_u intervals, respectively.

Dependences of T_h value on prediction relative error for elastic and aggregated traffic are given in Fig. 6.

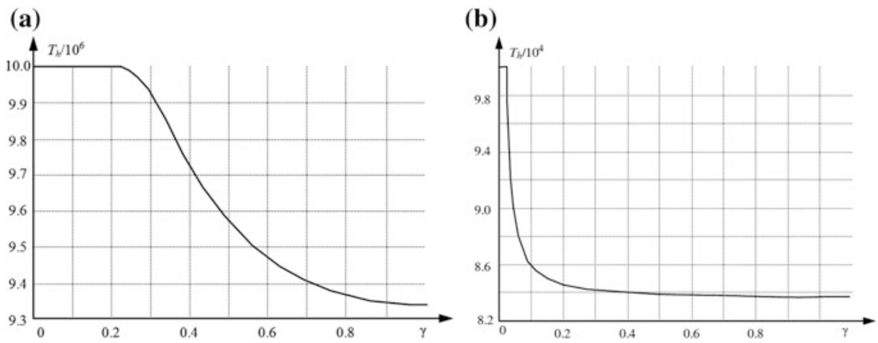


Fig. 6 Prediction relative error impact on T_h value for **a** elastic traffic, **b** aggregated traffic

With the prediction relative error value for elastic traffic below the value of 0.2 (Fig. 6a) the impact of such error was insignificant. With relative error increase from 0.2 to 0.5, the value of T_h varied between 10 and 9.6 Mb/s, that is close to 4%. At least, with relative error increase above value of 0.5, the value of T_h varied between 10 and 9.3 Mb/s, or close to 7%.

Analysis of prediction relative error for aggregated traffic (Fig. 6b) revealed that the impact on the value of T_h is slightly found with small values of γ, and with further increase in γ the value of T_h quickly falls off from 1 to 0.84 Mb/s, or close to 16%.

Fig. 7 Big Data center structure

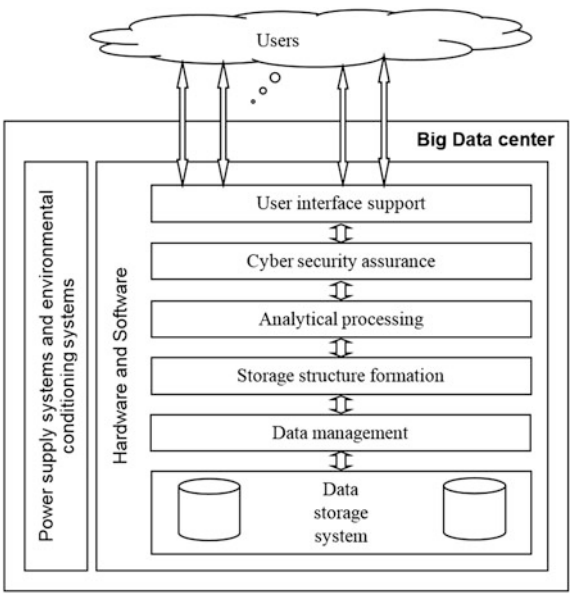


Table 1 Experimental results

Method	Traffic engineering				Complex of traffic models				Decrease in energy consumption (%)
Model	One flow		Several flows		One flow		Several flows		
	γ_{min}	γ_{max}	γ_{min}	γ_{max}	γ_{min}	γ_{max}	γ_{min}	γ_{max}	
M1	0.01	0.05	0.05	0.22	0.01	0.06	0.03	0.20	0.3
M2	0.05	0.20	0.15	0.35	0.02	0.18	0.05	0.25	0.9
M3	0.20	0.60	0.35	0.90	0.05	0.20	0.03	0.25	2.2

To analyze the possibilities of the proposed approach, it was studied a model of Big Data center operation, which corresponds to a structure presented in Fig. 7.

Further, using traffic models for a single source (M1), aggregated traffic (M2) and traffic containing big number of spikes (M3), it was conducted a calculation of the prediction relative error using standard features of Traffic Engineering and constructed models. Improvement of the prediction resulted in improved network balance, and, therefore, increased energy efficiency of a Big Data center (Table 1).

As Table 1 shows, the traffic containing big number of spikes is better predicted using models from the complex, proposed in this chapter. Application of such models, in contrast to existing ones (used in Traffic Engineering), do not result in degradation of aggregated flows rate inside virtual connections of Big Data centers. Respectively, such traffic resulted in increased energy efficiency of a Big Data center.

4 Conclusions

In this chapter, we considered solution to the problem of Big Data centers energy efficiency increasing by use of enhanced traffic control mechanism. Thus, the goal is achieved through traffic prediction accuracy increase via its modeling on the basis of appropriate samples. Such solution is oriented toward transmission and processing processes of Big Data. It is proposed a complex, which includes several models and a method; it uses traffic samples of several sessions as input data to perform short-term traffic prediction. At first, we define the character of real traffic. Further, control is transferred to one of the models contained inside the proposed complex. Finally, the complex allows predicting characteristics and volume of aggregated traffic's spikes, as well as amount of traffic, which exceeds pre-defined threshold, and maximal data transmission rate are being determined.

For the case of elastic traffic, the application of the developed complex will reduce the DC's resources usage, which is required to process data, down by 3%. As a result of the resources reallocation of the DC on the basis of the short-term prediction, the gain on the effectiveness of their use with such type of traffic will be 4–8%. Therefore, application of the proposed complex allows using the resources of the DC in more efficient way (reduces transmission time and, correspondingly, increases energy efficiency of the whole system).

References

1. Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.): Green IT engineering: concepts, models, complex systems architectures. Studies in Systems, Decision and Control, vol. 74. Springer (2017). https://doi.org/10.1007/978-3-319-44162-7
2. Chen, A.J.W., Boudrean, M.-C., Watson, R.T.: Information systems and ecological sustainability. J. Syst. Inf. Technol. **10**(3), 186–201 (2008)
3. The Software Improvement Group measures software energy use. The Green IT Review (2017). Last accessed 01 Mar 2017. Available at: http://www.seflab.com/news
4. Kuchuk G.A., Kovalenko A.A., Kharchenko V., Shamraev A.: Resource-oriented approaches to implementation of traffic control technologies in safety-critical I&C systems. In: Green IT Engineering: Components, Networks and Systems Implementation, Springer, pp. 313–338 (2017). https://doi.org/10.1007/978-3-319-55595-9_15
5. Michele, R., Raffaella, V., Michele, Z.: Accurate analysis of TCP on channels with memory and finite round-trip delay. IEEE Trans. Wirel. Commun. **3**(2), 627–640 (2004)
6. De Ghein, L.: MPLS fundamentals, p. 672. Cisco Press, USA p (2006)
7. D'Apice, C., Manzo, R., Likhanov, N., Salerno, S.: Network traffic modelling and packet loss probability approximation: stability problems for stochastic models. J. Math. Sci. **132**(5), 590–601 (2006)
8. Mozhaev, O., Kuchuk, H., Kuchuk, N., Mozhaev, M., Lohvynenko, M.: Multiservice network security metric. In 2nd International Conference on Advanced Information and Communication Technologies (AICT): Proceedings of International Conference, Lviv, Ukraine, pp. 133–136 (2017), 4–7 July 2017. https://doi.org/10.1109/aiact.2017.8020083
9. Kurose, J., Ross, K.: Computer networking: a top–down approach, 7th ed., Harlow: Pearson, 864 p. (2017)
10. Kuchuk, G.A., Akimova, Y.A., Klimenko, L.A.: Method of optimal allocation of relational tables. Eng. Simul. **17**(5), 681–689 (2010)
11. Kuchuk G.A., Kovalenko A.A., Mozhaev A.A.: An approach to development of complex metric for multiservice network security assessment. In Proceedings of International Conference on Statistical Methods of Signal and Data Processing (SMSDP-2010): 13–14 Oct 2010, Kiev: NAU, RED, IEEE Ukraine section joint SP, pp. 158–160 (2010)
12. Alaa, Z., Doulat, A.S., Khamayseh, Y.M.: Khamayseh al-howaide: performance evaluation of different scheduling algorithms in WiMax. IJCSEA Int. J. Comput. Sci. Eng. Appl. **1**(5), 81–94 (2011)
13. Teixeira, V., Guardieiro, A.: A new and efficient adaptive scheduling packets for the uplink traffic in WiMAX networks. EURASIP J. Wirel. Commun. Netw. **2011**(1), 112–123 (2011)
14. Annadurai, C.: Review of packet scheduling algorithms in mobile ad hoc networks. Int. J. Comput. Appl. **15**(1), 7–10 (2011)
15. Jandaeng, C., Suntiamontut, W., Elz, N.: Review PSA: the packet scheduling algorithm for wireless sensor networks. GRAPH-HOC Int. J. Appl. Graph Theory Wirel. Ad Hoc Netw. Sensor Netw. **3**(3), 1–12 (2011)
16. Karasaridis, A., Hatzinakos, D.: Network Heavy Traffic Modeling Using a-stable Self-Similar Process. IEEE Trans. Commun. **49**(7), 1203–1214 (2001)

A Markov Model of IoT System Availability Considering DDoS Attacks, Patching and Energy Modes

Maryna Kolisnyk and Vyacheslav Kharchenko

Abstract An important task in the conditions of the rapid spread of the Internet of things (IoT) is the assessment of its availability under the impact of various attacks. The chapter analyzes the energy modes of network devices operating system (OS) in the smart business center (SBC), and specifies the organization of the energy-efficient IoT-SBC subsystem. The aim of the research is to develop and study Markov models of SBC availability in view of energy regimes, the impact of cyber-attacks, the reliability of SBC components and patches on the firewalls vulnerabilities. The main questions of the chapter's research are an analysis of possible methods for reducing the power consumption of SBC devices, estimating the availability factor of SBC in the conditions of successful attacks.

Keywords IoT · DDoS attack · Smart business center · Energy modes
A Markov model · Availability

1 Description of Motivation

Internet of things (IoT) extends the scope of the Internet by people working at the computer in the direction of the autonomous intelligent smart devices connected to the Internet for remote monitoring and diagnosis [1–5]. Standards and technologies for connecting devices in the IoT, proposed by IEEE, ISO, ANSI, IETF (6LoWPAN, ROLL and CoRE working groups), IEC, ITU-T. The number of devices connected to the IoT is growing every day [5]. These can be complex smart

M. Kolisnyk
Department of Automation and Control in Technical Systems, National Technical University "KhPI", Kharkiv, Ukraine
e-mail: kolisnyk.maryna.al@gmail.com

V. Kharchenko (✉)
Department of Computer Systems, Networks and Cyber Security, National Aerospace University "KhAI", Kharkiv, Ukraine
e-mail: v_s_kharchenko@csn.khai.edu; v.kharchenko@csn.khai.edu

© Springer Nature Switzerland AG 2019

V. Kharchenko et al. (eds.), *Green IT Engineering: Social, Business and Industrial Applications*, Studies in Systems, Decision and Control 171,
https://doi.org/10.1007/978-3-030-00253-4_9

industrial complexes, smart transport, smart city lighting systems, car parks, smart hospitals, smart buildings. IoT consists of a large number of heterogeneous devices connected to network devices and networks built on the basis of various technologies and faces a variety of threats and attacks that adversely affect its proper functioning. Malicious impact and attacks on the vulnerabilities of IoT device components, software and databases can be applied on IoT systems. The target of the attackers can be stored data, video and audio recordings, disabling of IoT hardware and software components, industrial espionage. Protecting the security of IoT components and ensuring their high dependability is a very important task.

IoT network (wired and wireless sensor systems, which transmit information from one device to another in IoT (M2M solutions, applications to process data from sensors, mobile electronic devices, and cloud infrastructure)), consists of components such as: data processing centers (Datacenters) or stand-alone servers, routers, switches, sensors, communication lines [1–5]. For each of them it is necessary to provide the required power supply. Energy consumption should be minimal [6, 7].

There are two problems that need to be resolved quickly:

- increase in the number of denial-of-service (DoS) and distributed denial-of-service attacks (DDoS attacks) on IoT servers and data center servers;
- increase in energy consumption in case of successful attacks.

Some of scientific papers consider the solution of these problems [2–7]: solutions based on less energy-intensive streaming data, based on the application of advanced directions in the development of information technologies, components, networks and complex systems. IoT деюбvices can be affected by special attacks on the power system, DDoS attacks, spyware-attacks. It is advisable to consider the impact of the power consumption modes of the router and the server as a result of DDoS attacks and installing patches on the detected software vulnerabilities of devices to the Smart Business Center (SBC) availability.

2 Goals of Researching

The goal of the research is to develop a Markov model that must take into account the different states of the system functioning and the transitions between these states with different rates, and also research the IoT systems availability taking into account different energy modes, successful DDoS attacks and patches on the vulnerability of the firewall and the server software.

3 Requirements to SBC Organization

IoT involves networks of data-gathering sensors, cloud computing and machine-to-machine communication. IoT networks can be organized using wired or wireless technologies. The purpose of the work is to analyze and develop on the basis of this analysis, architecturally important from the point of view of dependability, security and energy saving IoT subsystem—the SBC. Any architectural features must have both high performance and low power modes, with the power mode selection under software control. The problem is to minimize energy consumption while not significantly impacting the effective performance [8–10]. Building the architecture of standard SBC with diversification of the power supply system and its various operating modes, analyzing the possible types of software vulnerabilities for the objects of this system, analyzing the means of ensuring the security of SBC's network components, evaluating SBC dependability indicators are the main research tasks.

When organizing SBC, its components are: a router, a server, a switch, digital sensors, IP cameras. The server stores management and diagnostic programs, collects, processes and stores data from sensors. It works in a constant availability mode, so it's important to reduce the server power. The router provides communication between the switch to which the digital sensors are connected and the server on which the control programs for controlling SBC are installed. Therefore, it is necessary to reduce the power consumption of the router.

To IoT for office solutions are presented such basic requirements [1–5]:

(a) in order to save energy in SBC be able to perform the installation of automatic temperature control systems, connection to the mobile network of intelligent systems Smart Metering accounting (electricity, gas and water), which allows you to make decisions on the use of certain energy modes in the office, as well as to save staff time through the use of remote water consumption data collection, electricity, gas, etc.;

(b) the possibility of using the various sensors and control units. It is necessary not just to automate certain functions (control of lighting, HVAC, etc.), but to integrate virtually any SBC equipment into a single system, works on the algorithm which will set the installer and designer SBC;

(c) a complete feedback, which will allow to operate virtually all systems SBC, analyze the situation, make conclusions and to be able to control the SBC without external intervention (without pressing the control panel button), but only upon the occurrence of an event (for example should be provided, the emergence of the human in the corridor include of lighting, on-off ventilation and air conditioning system, power source switching to an alternative power supply, etc.);

(d) SBC system should give staff full control over their offices and to provide protection against emerging new threats and threats due to the fact that new computer technologies with connection to the Internet allow attackers to connect to the system;

(e) increase in the number of computers and servers resulting in significant power consumption, it is necessary to provide greater flexibility and adaptability of the infrastructure of power facilities.

To reduce the load on the IoT power supply, can to use various methods, including [6–23]:

1. the active implementation of alternative energy sources. Alternative energy helps to improve the economic situation in the country and contribute to environmental improvement. Appropriate use of renewable or locally generated energy in SBC: solar, wind, hydro, geothermal, fuel cell, heat pumps, incorporating liquid cooling in a data center environment will reduce the consumption of electrical energy [12–15];
2. to provide virtualization to reduce the number of computers and servers. Using this method can reduce the number of servers, which will decrease the load on the power supply and reduce the release of thermal energy [12];
3. the use of energy-efficient chips at designing SBC management systems. For example, a prototype energy-efficient Wi-Fi-chip Rockchip RKi6000, reducing the consumption of IoT-devices by 85% was introduced in early 2016, to BT4.0 LE level. One of the advantages of the chip—technology of adaptive dynamic power management. Ultra-low consumption can be achieved in the operating mode, and a standby mode [16–23];
4. the use, whenever possible, the low-speed, but reliable data transmission. The use of traditional cellular technology in this area is too expensive, it can use a network of Low-Power Wide-area Network (LPWAN) [18];
5. hibernation mode for control of those devices that specific time may not work [18–21];
6. using of standby mode [6–11]. This mode is implemented in servers, workstations, in some models of routers;
7. when designing the SBC, it is necessary to take into account both the safety of operation and the reliability of the hardware and software of the system;
8. the system can function as standalone, with the connection of devices to the local access network, and with connection to a distributed network;
9. due to the fact that the operation of the SBC completely depends on the power source, it is possible to reserve both the blocks and their batteries, alternative power supplies for the Uninterruptible Power Supply (UPS) to function smoothly. The main in the SBC system are the stability of the operation, the possibility of data backup, the possibility of software development, rapid recovery in case of fail or failure;
10. debugging of one subsystem should not affect the operation of the other subsystems.

4 Features of Energy-Efficient IoT Systems Organization

The ecosystem of the IoT includes sensors, video cameras, which often have a power supply from batteries, and must work for several years without maintenance and replacement. To work over a long period of time with only one battery charge, these devices must consume a minimum amount of energy [1–4, 6–11, 14]. Also, IoT components can be powered by the energy received from an external source. These can be solar collectors, wind generators, diesel generators, heat pumps, the use of which leads to a decrease in the consumption of electrical energy.

Examples of energy-efficient IoT are abundant: Things may be simple sensors, more complex sensors (electrical power meter device), actuators (e.g. heating, venting and cooling (HVAC) room controller, motor), or complex devices (industrial circuit-breaker, building of industrial automation). The IoT application may range from a simple monitoring application, such as change of the temperature in a building, to a complex application, such as providing complete energy automation of a campus.

IoT communications may be required offline, where information is exchanged every day or on demand, or online allowing for real-time control. Building control applications can provide efficient use of the energy in a building while insuring comfort to building occupants. Traditional solutions are using complex building management systems (BMS) interconnected with Programmable Logical Controller (PLCs), which send orders to actuators based on sensor data. The system take into account multiple parameters such as weather forecasts on real-time energy costs [14, 15].

4.1 Low-Power, Wide-Area Technologies

Low power and Lossy Networks (LLNs) are made up of many embedded devices with limited power, memory, and processing resources [16, 17]. They are interconnected by a variety of links, such as IEEE 802.15.4, Bluetooth, Low Power Wi-Fi, wired or other low power PLC (Powerline Communication) links. LLNs are transitioning to an end-to-end IP-based solution to avoid the problem of non-interoperable networks interconnected by protocol translation gateways and proxies.

Low-Power, Wide-Area technologies are used in mobile communication networks and short-range technologies, they can be used to create IoT. Allows to reduce costs and minimal power consumption characteristics [18]. Major applications for LPWA connectivity include applications for connecting (or storing) energy-intensive assets to a managed electrical grid in order to increase the use of renewable energy sources and applications for consumer devices that are connected for home automation and life-support purposes.

LPWAN—energy-efficient network of long-range—wireless small data volume transmission technology over long distances, providing environment data collection from sensors, meters and sensors [19]. The main technologies, that includes LPWAN, are: Sigfox, LTE-M (LTE-MTC), NB LTE-M, NB-IoT, LoRa, Wi-Fi HaLow.

4.2 Standby Mode in the Components of IoT Network

The ecosystem of the IoT includes sensors, video cameras, which often have a power supply from batteries, and must work for several years without maintenance and replacement. To work over a long period of time with only one battery charge, these devices must consume a minimum amount of energy. The operation of the system depends entirely on the power source. For the uninterrupted operation of an uninterruptible power supply, backup of the both UPS (uninterruptible power supply) blocks and their accumulators, alternative power supplies. These can be solar collectors, wind generators, diesel generators, heat pumps, the use of which leads to a decrease in the consumption of electrical energy [6–15, 19–21].

In standby mode the power consumption of each individual device IoT is minimal, and power consumption increases when attacks to the server and network equipment are successful. Power management software enables or disables Active State Power Mode (ASPM) in each port of a component by programming the ASPM Control field [22, 23].

4.3 Operating System Power State Definitions

Similarly, the OS uses information from applications and user settings to put the system as a whole into a low-power state. The OS uses Advanced Configuration and Power Interface (ACPI) to control power state transitions in hardware [22, 23].

ACPI defines two cooling modes, Active and Passive: (a) Passive cooling OS reduces the power consumption of devices at the cost of system performance to reduce the system's temperature; (b) Active cooling OS increases the power consumption of the system to reduce its temperature. Active cooling requires increased power to reduce the heat within the system while Passive cooling requires reduced power to decrease the temperature. Active cooling allows maximum system performance, but it may create undesirable fan noise, while Passive cooling reduces system performance, but do it very quiet [22].

To further save power in the Working state, the operating system (OS) puts the CPU into low-power states (C1, C2, and C3) when the OS is idle. In these low-power states, the CPU does not run any instructions, and wakes when an interrupt, such as the OS scheduler's timer interrupt, occurs. The OS determines how much time is being spent in its idle loop by reading the ACPI Power

Management Timer. This timer runs at a known, fixed frequency and allows the OS to precisely determine idle time. Depending on this idle time estimate, the OS will put the CPU into different quality low-power states (which vary in power and latency) when it enters its idle loop.

4.4 Device Power States

Device power states are named D0, D1, D2, and D3. D0 is the fully on state, and D1, D2, and D3 are low-power states. The state number is inversely related to power consumption: higher numbered states use less power. Starting with Windows 8, the D3 state is divided into two substates, D3hot and D3cold [23].

D0 (Fully on)—the device is completely active and responsive. The link may be in either L0/L0s. L1 state may be achieved either by hardware-based ASPM or by requesting the link to enter L1.

D1 and D2: (rarely used)—no universal definition for these intermediate D-states. D1 is expected to save less power but preserve more device context than D2. L1 state is the required link power state in both of D-states.

D3 (Off) divided into two modes:

D3hot—Primary power is not yet removed from the device. D3hot maps to L1 to support clock removal on mobile platforms. PCI Bus clock is running.

D3cold—Primary power may be fully removed from the device. D3cold maps to L2 if auxiliary power is supported or, L3 if no power is delivered to the device. PCI Bus clock is stopped.

4.5 Router Power Modes

According to the [20, 21], there are energy modes of network equipment (router), which are used when creating IoT:

Active—sending packages with high power consumption.

Normal Idle (N_IDLE)—no packets (less energy).

Low-Power Idle (LP_IDLE)—no packets, less energy-intensive. Power consumption is reduced by turning off unused circuitry during LP_IDLE (part of the Physical layer, Media access control (MAC), interconnects, memory, central processor unit (CPU)), and only the necessary circuits (for example, clock recovery, alarm) should be included.

4.6 Processor Power State Definitions

Processor power states (Cx states) are processor power consumption and thermal management states within the global working state, G0. The process of entering and exiting the processor from state to state is described below. According to [22, 23], the CPU have 4 energy modes:

C0 Processor Power State—if processor is in this state, it executes instructions.

C1 Processor Power State—has the lowest latency. The processor in a non-executing power state. Platform scales the CPU clock frequency.

C2 Processor Power State—offers improved power savings over the C1 state. The processor in a non-executing power state. Platform scales the CPU clock frequency and voltage.

C3 Processor Power State—offers improved power savings over the C1 and C2 states. Processor's caches maintain state but ignore any snoops.

The worstcase hardware latency for this state is provided via the ACPI system firmware and the operating software can use this information to determine when the C2 state should be used instead of the C3 state. While in the C3 state, the processor's caches maintain state but ignore any snoops. The operating software is responsible for ensuring that the caches maintain coherency.

When a system is powered on, the following steps occur as the system resumes from hibernation [14–16]:

System Power-On Self-Test (POST).
System memory is decompressed and restored from the hibernation file.
Device initialization.
Drivers are restored to the state they were in prior to hibernation.
Services are restored to the state they were in prior to hibernation.
System becomes available for login.

There are two requirements for waking the system from a sleep or low-power idle state, or a device from a low-power state. First, the interrupt line must be Wake-Capable. Wake-capable interrupts are designed to be able to be delivered to the processor from low-power states. This implies that it must also cause the processor and any required platform hardware to power-up [16]. OS directs all system and device power state transitions. Employing user preferences and knowledge of how devices are being used by applications, the OS puts devices in and out of low-power states. Devices that are not being used can be turned off.

4.7 Server State Definitions

Server machines often have the largest hardware configurations and get the largest absolute power saving. To check and change the mode of power consumption, it is envisaged to use the "powercfg/hibernate", "powercfg/list", "powercfg/energy", "powercfg/srumutil", "powercfg/sleepstudy", "powercfg/deviceenablewake", "powercfg/devicedisablewake", "powercfg/lastwake", "powercfg/waketimers", "powercfg/availablesleepstates", "powercfg/devicequery" commands at the command line of the server OS, for example, MS Windows Server [17, 18, 20].

Servers often use two modes:

- Day Mode—servers are power-managed and staying in the Working state all the time, but putting unused devices into low-power states whenever possible. OSPM allows careful tuning of when to do this, thus making it workable.
- Night Mode—servers sleep as deeply as they can and are still able to wake and answer service requests coming in over the network, phone links, and so on, within specified latencies.

Because servers can be very large and have many disk and fan spindles, power management can result in large savings [19–22].

The server has the following modes of reduced energy consumption [20–22]:

S1 (Power On Suspend, POS, Doze)—power saving mode, which turns off the monitor, hard drive, but the CPU and RAM (random access memory modules) power is applied, reduced the frequency of the system bus. CPU cache is cleared, the CPU does not perform the instructions from the generator CPU.
S2 (Standby Mode)—reduced power consumption mode. In this mode, the monitor and the hard drive disable. From the CPU turns off the power supply. They stop clocks (continue to operate only those devices that are necessary for memory). Power is supplied only to the system memory (it contains information about the system status).
S3 (Suspend to RAM, STR, Suspend)—Standby. With this power saving mode, power is supplied only to the RAM (it stores information about the system status). All other PC components are disabled.
S4 (Suspend to Disk, STD, Suspend to Hard Drive, S4-Hibernation)—a deep sleep. With this power saving mode, the current state of the system is written to the hard drive, the power to all components of the PC is turned off.
S5 (Soft Off State) is similar to the S4 state except that the operation system (OS) does not save any context. The system is in the "soft" off state and requires a complete boot when it wakes. Software uses a different state value to distinguish between the S5 state and the S4 state to allow for initial boot operations within the basic input/output system (BIOS) to distinguish whether or not the boot is going to wake from a saved memory image.

4.8 Global System State Definitions

The full power system have some modes [17–25]:

G States reflect the user's perception of the machine.

G0 Working (S0)—a computer state where the system fully operational. It is not safe to disassemble the machine in this state.
G1 Sleeping (S1–S4)—power consumption is small and the system "appears" to be off. Work can be resumed without rebooting the OS. Large elements of system context are saved by the hardware and the rest by system software.
G2 Soft Off (S5)—the computer consumes a minimal amount of power. This state requires a large latency in order to return the Working state. The system's context will not be saved. The system must be restarted.
G3 Mechanical Off—this state that is entered by a mechanical means (i.e. power switch). The OS must be restarted and no hardware context is retained. Except for the real-time clock, power consumption is zero.

There are several sleep states of microprocessor systems in general [10, 14, 18–20]:

S1 Sleeping State—is a low wake latency sleeping state. In this state, no system context is lost (CPU or chip set) and hardware maintains all system context.
S2 Sleeping State—is a low wake latency sleeping state. This state is similar to the S1 sleeping state except that the CPU and system cache context is lost (the OS is responsible for maintaining the caches and CPU context). Control starts from the processor's reset vector after the wake event.
S3 Sleeping State—is a low wake latency sleeping state where all system context is lost except system memory. CPU, cache, and chip set context are lost in this state. Hardware maintains memory context and restores some CPU and L2 configuration context. Control starts from the processor's reset vector after the wake event.
S4 Sleeping State—is the lowest power, longest wake latency sleeping state supported by ACPI. In order to reduce power to a minimum, it is assumed that the hardware platform has powered off all devices. Platform context is maintained.
S5 Soft Off State—similar to the S4 state except that the OS does not save any context. The system is in the "soft" off state and requires a complete boot when it wakes. Software uses a different state value to distinguish between the S5 state and the S4 state to allow for initial boot operations within the basic input/output system (BIOS) to distinguish whether or not the boot is going to wake from a saved memory image.

In Table 1 [18] are shown features of the transition of the system from one state of electrical power consumption to another.

IoT is, by definition, composed by fully distributed systems, with discontinuous connection availability and, not rarely, with energy-consumption and computational power constraints [19].

According to [24, 25], there are some methods to build the power efficient SBC:

- using of energy-efficient systems—cooling, UPS, power supplies, servers, and non-IT equipment;

Table 1 Features of the transition of the system from one state of electrical power consumption to another

State	G, S, D states	Action	Power
Active	G0, S0, D0	Normal functionality	Main = on Aux = on
Prepare for sleep	G0, S0, D0 → D3	OS loads wake up packet filter into LAN controller via driver, sets device power state D3	Main = on Aux = on
Transition to sleep	D3, S0 → S3 or S4, G0 → G1	OS saves system state (memory image) to RAM or HDD. Main power off	Main → off Aux = on
Wake event	D3 → D0, S3 or S4 → S0, G0 → G1	Wake packet received. Device asserts PME# or wake#	Main = on Aux = on

- asset procurement—equipment is selected considering full lifecycle environmental impact, from technology providers with appropriate environmental policies and controls;
- building Construction—the physical building, its fixtures and fittings are designed and selected to be energy-efficient and to make use of relatively low impact materials and construction techniques;
- energy-efficient operation—the systems are monitored and operated to maximize energy efficiency;
- energy source—appropriate use of renewable or locally generated energy; solar, wind, hydro, geothermal, fuel cell, etc.;
- support services—selected and managed to minimize the environmental footprint.

5 Mitigation and Protection Methods from Attacks on Components of IoT System

IoT architecture of any system consists of five levels [2–5, 24–27]: the intellectual level of the connection, the data information at the connection level, the cybernetic level, cognitive level, the configuration level. Malicious attacks on impact and vulnerability components of IoT devices, software, and a DB can be applied to each of these levels.

The aim of intruders can be stored data, video and audio recording, shutdown of hardware and software components IoT, industrial espionage.

List of malicious actions types performed by the attacker: illegal use of user accounts; physical theft of office equipment and data carriers; theft software; run the executable code for the damage to the systems, for the destruction or corruption of data; modification data; identity theft; execution of actions that do not allow users to access network services and resources; execution of actions that absorb network resources and bandwidth.

The basis of any SBC system is the server on which the control software is stored. On the server undertaken the largest number of DoS- and DDoS attacks. Some types of attacks aimed at disabling the router and the switch, resulting in a malfunction of computers, tablets, smart phones and a variety of IoT devices connected to the system, as well as the basic components SBC system-sensors [1–4, 8, 20–29].

In IoT systems applies cybersecurity methods [22]: identity of things; homomorphic encryption; searchable encryption; trust establishment; secure systems collaboration technologies; privacy through usage control; continuous security audits; Identity and Access Management (IAM) technologies for IoT; application isolation and security boundary technologies; end-to-end policy management; integrates all policies; optimized framework with respect to available physical resources/security robustness (plan-do-check-act [15]); resilience including cyber-physical attacks; fault tolerance including cyber-physical attacks; detection and response to system threats; observe-orient-decide-act; monitoring of devices; coordination and analysis of threats; identity management; federated identity management, ID correlation between systems, Securing ID of devices; authenticity management; accountability/non-repudiation of data; anomaly detection. There are now cyber-attack protection methods that can also be used to protect against DDoS attacks on IoT components [23]: Hybrid DDoS Protection; Behavioral-Based Detection; Real-Time Signature Creation; Cyber-Security Emergency Response Plan. Also on the supply system can be carried out special attacks.

The IoT vulnerabilities may be more critical because they can impact the safety (e.g., industrial accidents) or the provision of essential services to the community (e.g., electrical power to hospitals). In some cases, the deployment of new technologies and devices in the home of the citizen or in the proximity may lead to new security or privacy issues (e.g., smart-meters) [22–31].

6 Development of the Model

Massive use cases have different network requirements and highly commercial needs: while reliability and availability might be more or less important, extended coverage is often a must and long latency is not problematic; infrequent, mainly downlink, and small data packages; devices need to be inexpensive (simple modules); device battery must last many years [27–29].

The main features of the SBC system are stability, the ability to backup data, the ability to modify the software, a quick recovery in case of failure or failure. All actions should interact with SBC subsystems, and therefore, should be aimed at starting and adjusting the SBC system itself. In this case, debugging any one subsystem should not affect the operation of the other subsystems.

When SBC is constructing, it is necessary to take into account energy modes, security, dependability of software and hardware of the SBC components.

The Markov model of wired SBC subsystem operation proposed in the chapter takes into account DDoS attacks and various power modes of the server and router, as well as possible patches on software vulnerabilities. For the proposed model, graphical dependencies of system availability are built and analyzed when the attack rate of the server and router changes, the rate of transition to the lower power consumption mode of the SBC server and the router.

6.1 A Markov Model of SBC Subsystems Functioning

Using for IoT SBC Internet wire network devices are: router with Ethernet-ports and wireless access ability, softswitch the second layer, firewall, power block, server with control software, IP-camera, sensors, cables [32–34]. The system can operates a standalone or with Internet connection. Assumptions for the developed model are the following:

– the flow of hardware system failures obeys the Poisson distribution law;
– there is reserve of central processor unit of the server and the router;
– failures caused by software design faults of SBC subsystems obeys Poisson distribution, as on the results of monitoring and diagnostics, testing corrected secondary error (the result of the accumulation of the effects of primary faults, software backdoors) to fix a malfunction or failure of the software, remove of impacts on software vulnerabilities, DoS- and DDoS attacks, the number of primary defects in the software permanently;
– the process, which occurs in the system, it is a process without aftereffect, every time in the future behavior of the system depends only on the state of the system at this time and does not depend on how the system arrived at that state. Therefore, the process has the Markov property. The mode of the server when software system shutdown and startup cycles in this model S4 is absent, because in this mode it is impossible to manage the server of SBC remotely. Was accepted the assumption that the means of control and diagnosis are ideal, which detect correctly the presence of fails and failures.

A Markov model of SBC subsystems functioning represented on Fig. 1, considering DDoS attacks and server's and router's energy modes without patches on possible vulnerabilities, which has the following states: Good-working state (1); The server is fully used with high power consumption S0 (2); The server is fully used, the hardware, that are not used, can enter the low-power mode S1 (3); Sleep mode of the server with low power consumption, a computer can wake up from a keyboard input, a LAN network or USB device S2 (4); Server appears off, power

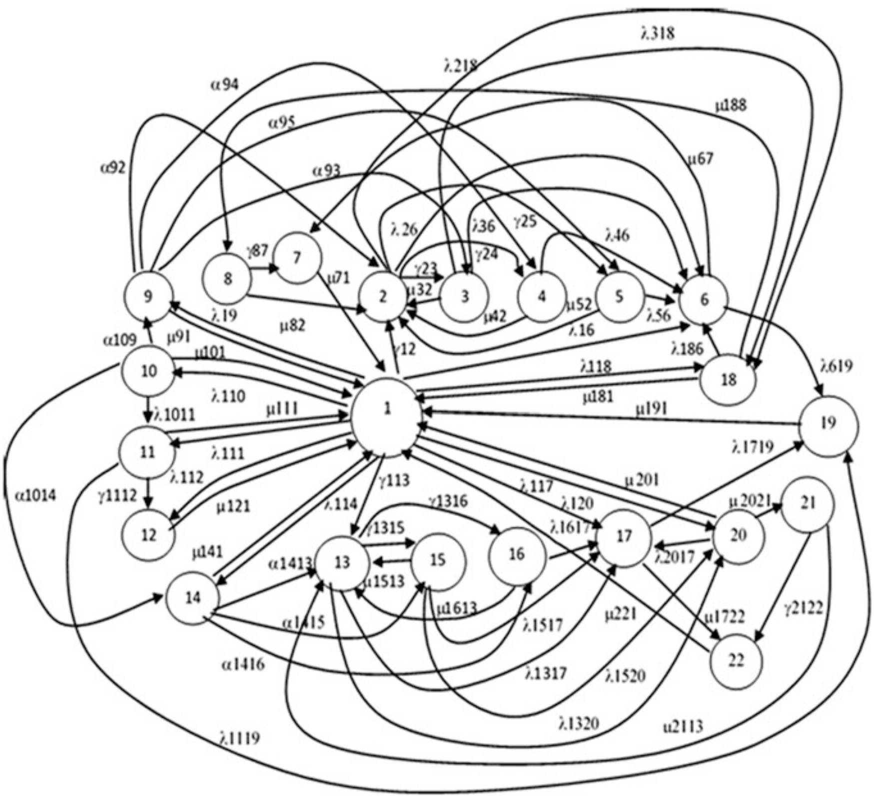


Fig. 1 A graph of a Markov model of SBC subsystems functioning

consumption is reduced to the lowest level S3 (5); Server failure (6); Switching to the backup server device after the server failure (7); Restarting of the server after the software fail (8); Successful DDoS attack on the server after the firewall failure (9); Firewall software or hardware failure (10); Attack on the power supply system after the firewall failure, that lead the failure of SBC general power system (11); Technical state of switch from the general power system after its failure on the alternative energy sources (solar, diesel generator, wind turbine) (12); Router Status Active—sending packages with high power consumption (13); Successful DDoS attack on the router (14); Good-working state of the router with transmitting packets —Normal Idle (15); Good-working state of the router without packet transmission Low-Power Idle (16); Router software or hardware failure (17); Server software or hardware fail (18); Router hardware or software fail (20); Switching to the backup router device after the router failure (21); Restarting the router software after the router software fail (22).

A system of linear differential equations of the Kolmogorov-Chapman composed and solved in the paper with the initial conditions:

$$P1(0) = 1, \quad \sum_{i=1}^{22} Pi(t) = 1. \tag{1}$$

An important indicator of dependability SBC under the influence of different kinds of DDoS attacks is the availability factor. As an index of reliability SBC we choose stationary availability factor AC, that is defined as the sum of the probabilities of staying the system in an up-states. Stationary AC is determined from equation:

$$\begin{aligned} AC = {} & P1(t) + P2(t) + P3(t) + P4(t) + P5(t) \\ & + P12(t) + P13(t) + P15(t) + P16(t) + P21(t) \end{aligned} \tag{2}$$

Pi(t)—probability of good-working state of SBC components.

Solving the system of Kolmogorov-Chapman equations, can to get the AC value of the SBC network's components after successful DDoS attacks, with considering energy modes of the server and the router, considering patches on vulnerabilities of SBC components. It follows that service availability, service continuity, cyber-security, data integrity, resilience and high dependability of software and hardware should be inherent in IoT networks.

The attack statistics analysis conducted by the NIST and CERT testifies to the fact that vulnerabilities can exist in IoT systems through which attackers perform DDoS attacks. Producers of network equipment and servers used in the construction of networks of IoT, after the discovery of the vulnerability, offer to install on the firewall an updated version of the software, which is patched the manifested vulnerabilities.

Let's consider how will change the proposed Markov model of SBC subsystems functioning when installing patches on the firewall the server (Fig. 2) and the firewall as a separate network device (Fig. 3).

For each of Markov model make up and solving the system of Kolmogorov-Chapman equations, and can get the AC value of SBC's components after successful DDoS attacks and with considering energy modes of the server and the router.

6.2 Simulation Results

On the basis of the statistical data analysis was assessed the main indicators of dependability and built graphics, shown in Figs. 4, 5, 6, 7, 8 and 9. As an example, was given graphical dependencies for different technical states of the server. We constructed the dependence of the system availability function (we denote it AC) from the transitions rates to different states (λij—failure rate, αij—attack rate, γij—rate of transition in different power modes of router and server, where i = 1 … 22, j = 1 … 22), which depend on events occurrence time.

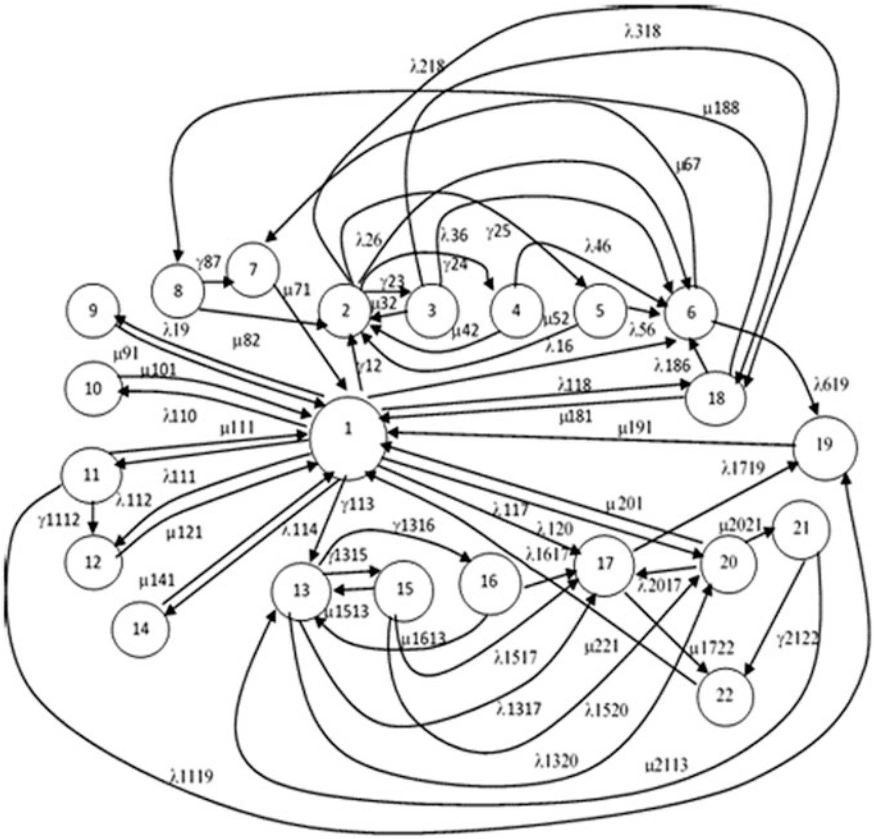


Fig. 2 A Markov model of SBC subsystems functioning when installing patches on the firewall of the server

Figures 4, 5, 6, 7, 8 and 9 shows the changing of availability function AC from changing the transition rates from one state to another in the Markov models with patches on vulnerabilities of: server firewall (AC 9) and firewall (AC 10) and model without patches (AC).

Analysis of the availability of the IoT—SBC system is performed taking into account component dependability, security (impact of DDoS attacks), recovery rates and various types of power modes of the server and router system, DDoS attacks on the router and server, and patches on vulnerabilities of firewall as devise and server's firewall. The research showed that the IoT system even with the required high value of AC is highly dependent on the correct, security and reliable operation of the firewalls. At the software maintenance stage, vulnerabilities can be detected in it, and the manufacturer suggests fixing the patches on it. If patches appear and install, the impact of attacks on the server and the router stops, so the proposed Markov model is rebuilt (Figs. 2 and 3).

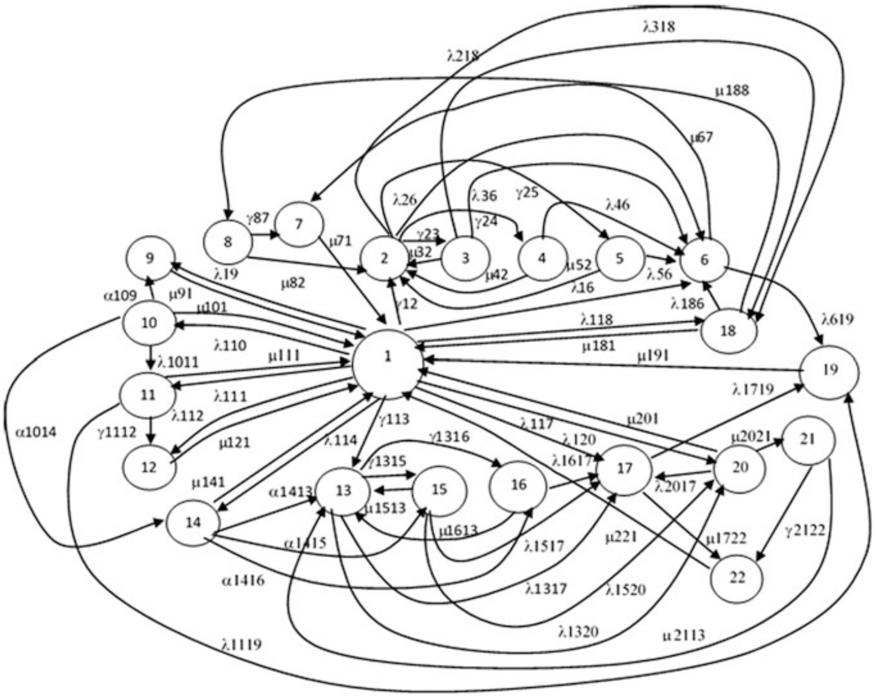


Fig. 3 A Markov model of SBC subsystems functioning when installing patches on the firewall and server firewall

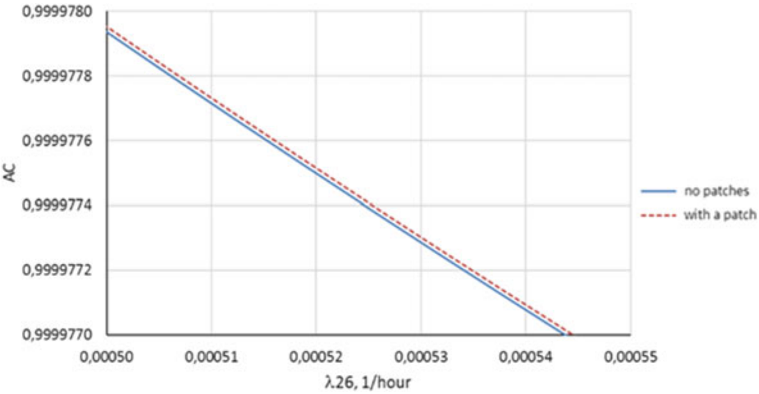


Fig. 4 Graph of dependence of SBC AC on the transition rate λ26 from active-power state of the server 2 to a state of the server failure 6 for model without patches on server firewall and model with patches if λ26 change values in range 5×10^{-4}–5.5×10^{-4} 1/h

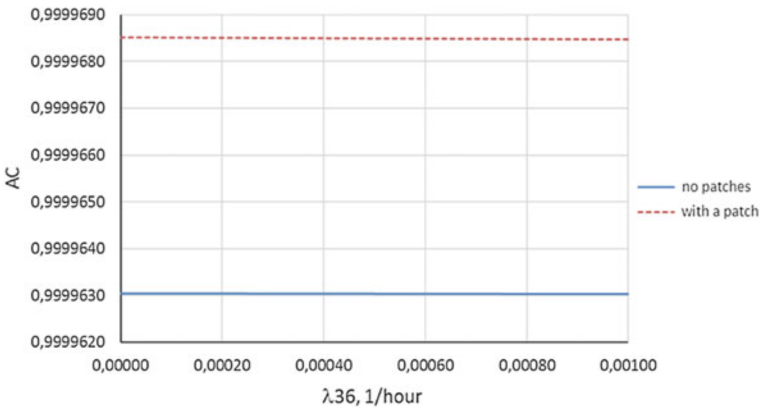


Fig. 5 Graph of dependence of SBC AC on the transition rate λ36 from low-power state of the server 3 to a state of the server failure 6 for model without patches and model with patches on server firewall if λ36 change values in range 0–10^{-3} 1/h

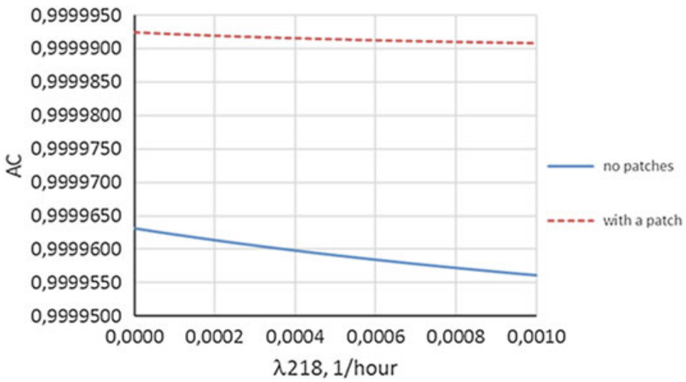


Fig. 6 Graph of dependence of SBC AC on the transition rate λ218 from active-power state of the server 2 to a state of the router fail 18 for model without patches and model with patches on server firewall if λ218 change values in range 0–10^{-3} 1/h

As initial data for the calculation of AC, whose values are based on the analysis of statistical data, was chosen: λ1317 = 0.00057 1/h; λ1517 = 0.00001 1/h; λ1617 = 0.000001 1/h; λl218 = 0.00001 1/h; λ318 = 0.00001 1/h; λ1320 = 0.000001 1/h; λ1520 = 0.000001 1/h; λ2017 = 0.00114 1/h; λ120 = 0.000001 1/h; μ67 = 60 1/h; μ141 = 0.125 1/h; μ111 = 0.5 1/h; μ32 = 40 1/h; μ42 = 30 1/h; μ52 = 30 1/h; μ1513 = 50 1/h; μ1613 = 60 1/h; μ71 = 0.02 1/h; μ87 = 2 1/h; μ81 = 30 1/h; μ101 = 1 1/h; μ121 = 5 1/h; μ181 = 1 1/h; μ191 = 0.02 1/h; μ91 = 1 1/h; μ171 = 1 1/h; μ188 = 60 1/h; μ61 = 0.02 1/h; μ2021 = 60 1/h; μ221 = 20 1/h; μ211 = 30 1/h; μ1722 = 60 1/h; μ201 = 40 1/h; μ2113 = 20 1/h.

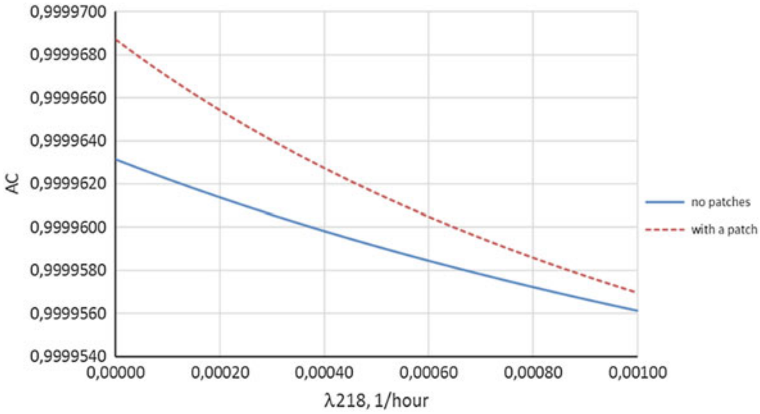


Fig. 7 Graph of dependence of SBC AC on the transition rate λ218 from active-power state of the server 2 to a state of the router fail 18 for model without patches and model with patches on firewall if λ218 change values in range 0–10^{-3} 1/h

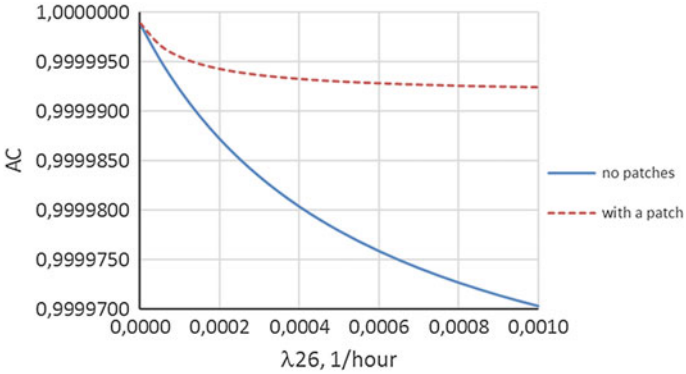


Fig. 8 Graph of dependence of SBC AC on the transition rate λ26 from active-power state of the server 2 to a state of the server failure 6 for model without patches on firewall and with patch on firewall if λ26 change values in range 0–10^{-3} 1/h

Reducing the AC value with an increase in transition rate from a good-working state with high power consumption of the server to the mode of software failure occurs due to the impact of external influences (DoS and DDoS attacks), and internal causes associated with defects in software and/or hardware of the server. If to change the transition rate λ26 value from 0.005 to 0.0055 1/h in the model without the patch, the value of the AC decreases, much like in the model with the patch installed on the server firewall from 0.999978 to 0.999977 (Fig. 4).

The AC value is 0.9999635, for the model, if the patch is not set when changing λ36 within 0–0.001 1/h. When installing a patch on the server firewall with the same initial data, the AC has a value of 0.9999925 (Fig. 5).

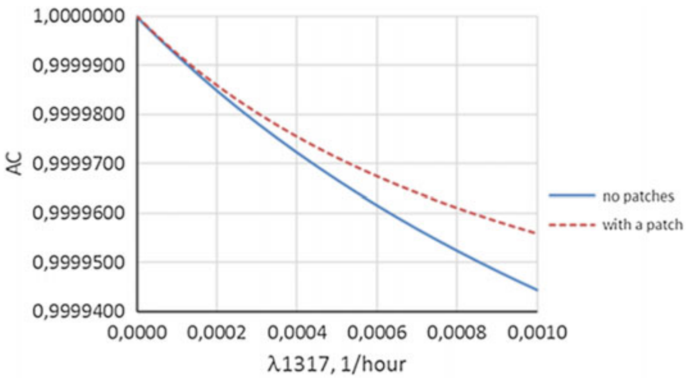


Fig. 9 Graph of dependence of SBC AC on the transition rate λ1317 from active-power state of the router 13 to a state of the router failure 17 for model without patches and with patches on firewall if λ1317 change values in range 0–10^{-3} 1/h

The availability function value is reduced from AC = 0.999963 to 0.99996, for the model, if the patch to the server firewall is not set, and from AC = 0.9999925 to 0.999991, if the patch is set, if the server's transition from the active power state of server to the fail state λ218 is changed within 0–0.001 1/h (Fig. 6).

The availability function value is reduced from AC = 0.999963 to 0.999956, for the model, if the patch to the server firewall is not set, and from AC = 0.999969 to 0.999957, if the patch is set, if the server's transition from the active power state of server to the fail state λ218 is changed within 0–0.001 1/h (Fig. 7).

If the transition rate value λ26 is changed from 0 to 0.001 1/h, the AC value decreases from 1 to 0.99997 for the model without patching and to 0.9999925 for the model with the patch installed on the firewall (Fig. 8).

Decrease in the AC value occurs 0.999945 for a model without a patch. If the values of λ1317 change within the range 0–10^{-3} 1/h, value of AC for model with patch on firewall decrease from value 1 to 0.999957 (Fig. 9).

Establishing a patch on the firewall (Fig. 8) allows to obtain the same AC values (1–0.999553) at λ1317 = 0–10^{-3} 1/h, but this value is significantly higher than in the model without patch: AC = 1–0.9999553 (Fig. 8).

If patches are not installed on the firewalls (Fig. 2), then the AC decrease from value 1 to 0.9999553 at λ26 = 10^{-3} 1/h. Patch installation on server firewall not significantly changes the AC of SBC value. If to install the patch on the firewall, the value of the AC increases compared to the model without patches, with the same initial data, from 0.9999553 to 0.9999925 (Figs. 4, 5, 6, 7, 8 and 9).

Under the influence of DDoS attacks, the server, which is in one of the energy-saving modes, will switch to the mode of increased power consumption.

Was researched and analyzed the function availability of IoT system—SBC, taking into account the reliability of components, rate of restoration, and different kinds of energy modes of server and router OS, DDoS attacks on the router and the server, and setting patches on firewalls vulnerabilities. Therefore, it is necessary to choose such values of SBC parameters at which the availability function of the proposed system for any changes in parameters taking into account the power consumption modes and under states of DoS and DDoS attacks will not change significantly.

7 Conclusion and the Future Work

This chapter provides an analysis of methods for reducing the power consumption of IoT devices. A new approach to the creation of SBC systems is proposed, taking into account the functioning of the power supply system with different modes of power consumption of server's and router's operating system.

To assess the availability of SBC, a Markov availability model of SBC was developed and investigated that takes into account successful DDoS attacks, fails and failures of hardware and software of various SBC components; the energy modes of the SBC system, their transition rate values are based on the analysis of the statistical data of the SBC components functioning; failures of IoT device management systems after the impact of attacks on the router software, power system, server, etc.

Attacks of DDoS on the network infrastructure of SBC or using IoT systems can last several hours and even a few days. Investigating the graphical dependencies obtained for the developed models, taking into account their rearrangement in case of appearance and installation of a patch on the vulnerability of the firewall, has shown that the value of AC SBC is most sensitive to fixing patches on the network firewall. When setting the patch, the value of AC remains high (0.9999925), even with the intensity of transition to the server failure state $\lambda 26 = 0.001$ 1/h.

Practical significance of the results allows to estimate the value of availability factor of IoT—SBC systems and develop recommendations for ensuring their high dependability in conditions of DoS and DDoS attacks impact, as well as reducing SBC's power consumption.

References

1. Al-Fuqaha, A., Guizani, M., Mohammadi, M., Aledhari, M., Ayyash, M.: Internet of things: a survey on enabling technologies, protocols, and applications. IEEE Commun. Surv. Tutorials **17**(4), 2347–2376 (Fourth Quarter 2015). Available at http://www.comsoc.org/files/Publications/Tech%20Focus/2016/iot/3.pdf. Accepted at 3 Aug 2017
2. Vermesan, O., Friess, P., Guillemin, P., et.al.: Internet of things—from research and innovation to market deployment. River Publishers Series in Communication, 141 p (2014). Available at http://www.internet-of-things-research.eu/pdf/IERC_Cluster_Book_2014_Ch.3_SRIA_WEB.pdf. Access date 15 Nov 2017

3. Delivering on the IoT Customer Experience. Business White Paper. Hewlett Packard Enterprise, 8 p. Available at http://h20195.www2.hpe.com/v2/GetDocument.aspx?docname = 4AA6-5128ENW. Accepted at 5 Aug 2016
4. Internet of Things and Its Future. Available at http://www.huawei.com/ilink/en/about-huawei/newsroom/press-release/HW_080993?dInID=23407&relatedID=19881&relatedName = HW_076569&dInDocName = HW_076557. Access date 20 Nov 2017
5. A Forrester Consulting Thought Leadership Paper Commissioned by Cisco, March 2015. Security: The Vital Element of the Internet of Things, 11 p (2015). http://www.cisco.com/c/dam/en_us/solutions/trends/iot/vital-element.pdf
6. Kharchenko, V., Kondratenko, Y., Kacprzyk, J.: Green IT engineering: concepts, models, complex systems architectures. In: Kharchenko, V., Illiashenko, O. (eds.) Concepts of Green IT Engineering: Taxonomy, Principles and Implementation, pp. 3–19 (2017). https://doi.org/10.1007/978-3-319-44162-7. https://www.springer.com/gp/book/9783319441610
7. Doukas, N.: Technologies for greener internet of things systems. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Components, Networks and Systems Implementation, pp. 23–41 (2017). http://www.springer.com/gp/book/9783319555942#otherversion = 9783319555959
8. Kondratenko, Y., et al.: Green-IT approach to design and optimization of thermoacoustic waste heat utilization plant based on soft computing. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Components, Networks and Systems Implementation, pp. 287–311 (2017). http://www.springer.com/gp/book/9783319555942#otherversion=9783319555959
9. Bardis, N.: Secure, green implementation of modular arithmetic operations for IoT and cloud applications. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Components, Networks and Systems Implementation, pp. 43–64 (2017). http://www.springer.com/gp/book/9783319555942#otherversion=9783319555959
10. Chemeris, A., et al.: Influence of software optimization on energy consumption of embedded systems. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Components, Networks and Systems Implementation, pp. 111–133 (2017). http://www.springer.com/gp/book/9783319555942#otherversion=9783319555959
11. Fesenko, H., et al.: Markov models of smart grid digital substations availability: multi-level degradation and recovery of power resources issues. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Components, Networks and Systems Implementation, pp. 339–355 (2017). http://www.springer.com/gp/book/9783319555942#otherversion=9783319555959
12. Internet Architecture Board (IAB). RFC 7452 Architectural Considerations in Smart Object Networking. https://www.rfc-editor.org/pdfrfc/rfc7452.txt.pdf
13. IEC. IoT 2020: Smart and Secure IoT Platform. White Paper, 181 p. http://www.iec.ch/whitepaper/pdf/iecWP-loT2020-LR.pdf
14. Rockchip RKI6000 Smart Home Products Unveiled at MWC 2016. http://chinagadgetsreviews.com/rockchip-rki6000-smart-home-products-unveiled-at-mwc-2016.html. Access date 03 Nov 2017
15. International Telecommunication Union (ITU). ITU–T Recommendation Y.2060 (2012). https://www.itu.int/md/T13-SG20–160725-TD-GEN-0467
16. Margaret Rouse. LPWAN (Low-Power Wide Area Network). http://internetofthingsagenda.techtarget.com/definition/LPWAN-low-power-wide-area-network
17. Lorch, J., Smith, A.J.: Software Strategies for Portable Computer Energy Management, 1 June 1988. https://www.microsoft.com/en-us/research/publication/software-strategies-for-portable-computer-energy-management/
18. IEEE Std 802.3az-2010.: IEEE Standard for Information Technology—Local and Metropolitan Area Networks-Specific Requirements—Part 3: CSMA/CD Access Method and Physical Layer Specifications Amendment 5: Media Access Control Parameters, Physical Layers, and Management Parameters for Energy-Efficient Ethernet. https://standards.ieee.org/findstds/standard/802.3az-2010.html

19. ASPM Optionality. PCIe Base Specification Revision 2.1. PCI-SIG Engineering Change Notice for Hewlett-Packard, Intel, Microsoft, 12 p, 19 June 2009. https://pcisig.com/sites/default/files/specification_documents/ECN_ASPM_Optionality_2009-08-20.pdf
20. Device Power Management. Microsoft Hardware Dev Center. https://docs.microsoft.com/en-us/windows-hardware/drivers/bringup/device-power-management
21. Chalupsky, D., Qi, E., Ganga, I.: A Brief Tutorial on Power Management in Computer Systems. Intel Corporation, 28 p, 13 Mar 2007. http://www.ieee802.org/3/eee_study/public/mar07/chalupsky_01_0307.pdf
22. Powercfg Command-Line Options. Microsoft Hardware Dev Center. https://docs.microsoft.com/en-us/windows-hardware/design/device-experiences/powercfg-command-line-options
23. Windows IT Pro Center. Powercfg Command-Line Options. https://technet.microsoft.com/en-us/library/cc748940(v=ws.10).aspx
24. Device Power States. Microsoft. https://docs.microsoft.com/en-us/windows-hardware/drivers/kernel/device-power-states
25. Kwa, S., Cohen, D.T.: INTEL Corporation. PCI Express Architecture Power Management. White Paper, 15 p, 8 Nov 2002. https://www.intel.com.br/content/dam/doc/white-paper/pci-express-architecture-power-management-rev-1-1-paper.pdf
26. Advanced Configuration and Power Interface. Specification Revision 5.1. Hewlett-Packard Corporation, Intel Corporation, Microsoft Corporation, Phoenix Technologies Ltd., Toshiba Corporation, 958 p, July 2014. http://www.acpi.info/DOWNLOADS/ACPIspec50.pdf
27. Real-time Protection Against Targeted Attacks Adaptive Behavior-Based Malware Protection. Whitepaper: Enterprise Security. Symantec. https://www.symantec.com/content/en/us/enterprise/white_papers/b-adaptive-behavior-based-malware-protection-WP-21241693.pdf
28. Internet of Things. IoT Governance, Privacy and Security Issues. European Research Cluster on the Internet of Things. Ovidiu Vermesan, Peter Friess, Coordinators of IERC Cluster, 128 p, Jan 2015
29. Cisco IoT System Brochure. Cisco IoT System Deploy. Accelerate. Innovate, 52 p (2015). http://www.cisco.com/c/dam/en/us/products/collateral/se/internet-of-things/brochure-c02-734481.pdf. Access date 17 Nov 2017
30. ISO/IEC 27000 Family—Information Security Management Systems. https://www.iso.org/isoiec-27001-information-security.html
31. Internet Security Threat Report, vol. 21, 81 p. Symantec, Apr 2016. Available at https://www.symantec.com/content/dam/symantec/docs/reports/istr-21-2016-en.pdf
32. Kolisnyk, M., Kharchenko, V., Piskachova, I., Bardis, N.: A Markov model of IoT system availability considering DDoS attacks and energy modes of server and router. In: ICTERI, 14 p (2017) http://ceur-ws.org/Vol-1844/10000699.pdf
33. Kharchenko, V., Kolisnyk, M., Piskachova, I.: Reliability and security issues for IoT-based smart business center: architecture and Markov model. In: MCSI 2016, Greece, Chania. IEEE, Computer of Science, Paper ID: 4564699 (2016)
34. Kharchenko, V., Kolisnyk, M., Piskachova, I., Bardis, N.: Markov model of the smart business center wired network considering attacks on software and hardware components. Int. J. Comput. Commun. **10**, 113–119. (2016). ISSN: 2074-1294. http://www.naun.org/cms.action?id=11961

Assessing the Impact of EEE Standard on Energy Consumed by Commercial Grade Network Switches

Joseph El Khoury, Eric Rondeau, Jean-Philippe Georges and Ah-Lian Kor

Abstract This book chapter is adapted from El Khoury in Assessing the benefit of deploying EEE on commercial grade network switches, Unpublished PERCCOM Masters Dissertation, University of Lorraine, Nancy, France, 2017, [1] and it is closely linked to work published in Kharchenko et al. (eds.) in Green IT engineering: concepts, models, complex systems architectures. Studies in systems, decision and control, vol. 74. Springer, Cham, 2017, [2], Kharchenko et al. (eds.) in Green IT engineering: components, networks and systems implementation. Studies in systems, decision and control, vol. 105. Springer, Cham, 2017, [3]. Reducing power consumption of network equipment has been both driven by a need to reduce the ecological footprint of the cloud as well as the immense power costs of data centers. As data centers, core networks and consequently, the cloud, constantly increase in size, their power consumption should be mitigated. Ethernet, the most widely used access network still remains the biggest communication technology used in core networks and cloud infrastructures. The Energy-Efficient Ethernet or EEE standard introduced by IEEE in 2010, aims to reduce the power consumption of EEE ports by transitioning Ethernet ports into a low power mode when traffic is not present. As statistics show that the average utilization rate of ethernet links is 5% on desktops and 30% in data centers, the power saving potential of EEE could be immense. This research aims to assess the benefits of deploying EEE and create a power consumption model for network switches with and without EEE. Our measurements show that an EEE port runs at 12–15% of its total power when in low power mode. Therefore, the power savings can exceed 80% when there is no traffic. However, our measurements equally show that the power consumption of a single

J. El Khoury · E. Rondeau · J.-P. Georges
CRAN, Université de Lorraine, Nancy, France
e-mail: eric.rondeau@univ-lorraine.fr

J.-P. Georges
e-mail: jean-philippe.georges@univ-lorraine.fr

A.-L. Kor (✉)
School of Computing, Creative Technologies, and Engineering,
Leeds Beckett University, Leeds, UK
e-mail: A.Kor@leedsbeckett.ac.uk

© Springer Nature Switzerland AG 2019

V. Kharchenko et al. (eds.), *Green IT Engineering: Social, Business and Industrial Applications*, Studies in Systems, Decision and Control 171,
https://doi.org/10.1007/978-3-030-00253-4_10

port represents less than 1% of the total power consumption of the switch. The base power consumed by the switch without any port is still significantly high and is not affected by EEE. Experiment results also show that the base power consumption of switches does not significantly increase with the size of the switches. Doubling the size of the switch between 24 and 48 ports increases power consumption by 35.39%. EEE has a greater effect on bigger switches, with a power (or energy) gain on the EEE-enabled 48-port switch compared to 2× EEE-enabled 24-port switch. On the other hand, it seems to be more energy-efficient to use 2 separate 24-port switches (NO EEE) than 2 separate 24-port switches (With EEE).

Keywords Power efficiency · EEE · Switch · Port · Power consumption Sleep · Hibernate · Network traffic · Burst traffic · Power consumption model

1 Introduction

Ethernet, a technology first deployed in 1980 is the most widely used access network in the world [4]. Although in recent years, the deployment of Ethernet has decreased in homes due to wifi, it still constitutes the main technology used in core networks and cloud infrastructures, particularly in data centers. In 2010, in the US, data center power consumption is estimated at 2.3% of the total country's power [5] while data centers worldwide are expected to grow approximately by 9% annually till 2015 [6]. Almost a quarter of the power consumption in these centers is attributed to network equipment [7] and with link speed expected to reach 200 Gbits/s [8], power consumption will undeniably, increase with time. According to our conducted experiments, the power consumption of a port increases with the increase of bandwidth. Our measurements reveal that a 1 Gbits/s port on a switch consumes 3 times more power than a 100 Mbits/s link and similarly, a fiber optic port consumes 5 times more power than a 1 Gbits/s link. This evidence the need to optimize the power consumption of network equipment which could further be enhanced by an optimal utilization rate of Ethernet links. On desktops, the average utilization of Ethernet links is at maximum 5% while that number is around 30% in data centers [9]. Consequently, for the past decade, researchers have been developing algorithms to achieve power consumption of Ethernet links that matches their utilization rates. The culmination of such research results in the standardization of EEE 802.03az in 2010 [10]. Energy-Efficient Ethernet (EEE) aims to reduce the power consumption of network equipment by running the ports in a low power mode with no imminent transmission. The ports could run in two different states: normal state where a link consumes the usual amount of power while transmitting information; and another low power mode aptly named Low Power Idle (LPI) when traffic is absent. The standard also details two transition phases namely Sleep and Wake-Up, that specifies the time it takes for the port to transition to each of the previously mentioned states. Note that the state of the ports on a network equipment is only affected by EEE. The base power consumption of a switch will not be affected by EEE, but the power consumed by a connected port is. The EEE standard also does not specify when state transitions should occur but have only provided details of which mechanism to employ bring about power saving. It is the manufacturer's onus to decide when to trigger a state change. Two mechanisms namely frame transmission (FTR) and burst transmission (BTR) are predominantly employed [11] in EEE. With FTR, the Ethernet interface is activated on the arrival of a new packet and the packet is then immediately handled. Contrastingly with BTR, the newly arrived packet is added to a queue and the interface switches to an active state once the queue is full. BTR allows packets coalescence within a certain timeout in order to avoid waking the interface excessively.

The research discussed in this paper aims to assess the benefits of deploying EEE in the core network by profiling the power consumption of an off-the-shelf Cisco 2960-X with and without power efficient protocol under different link speeds and traffic throughput.

1.1 Background

EEE affords implementation for different link speeds [12, 13]. For example, 100 Mbits/s and 10 Gbits/s links can be powered off even if the opposite side of the link does not support EEE. It is sufficient for one side of a link to detect a lack of traffic and invoke a sleep instruction without checking what occurs on the other end of the link. However, with 1 Gbits/s links, both ends of the link must agree before invoking the sleep instruction. Furthermore, state transitions cannot be interrupted in 10 Gbits/s links. Hence, 100 Mbits/s and 10 Gbits/s links can potentially yield higher power saving at the expense of a potential delay in delivery if one end transitions to sleep mode while the other end sends traffic.

In EEE, a link can have 4 different states [10] namely Active, Low Power Idle, Sleep and Wake: **Active** (A)—a link in this state has normal operation and power consumption. This is the switch's default operational mode in the event EEE is turned off; **Sleep** (S)—is a transition state invoked when a link in state A does not detect any incoming traffic. This state lasts a finite duration denoted by T_s and its power consumption during this transition equals that of A state; **Low Power Idle** (LPI)—a link operates in low power mode which is approximately 10% of its normal power. Power savings from EEE are achieved when the link operates in LPI; **Wake** (W)—is a transition state that takes a link from LPI to A. This transition is never interrupted. This state lasts a finite duration denoted T_w and its power consumption during this transition equals that of the A state. EEE related research has predominantly focused on maximizing the LPI duration in order to maximize power savings. As previously mentioned, mechanisms used to trigger change of states in EEE are implemented by the manufacturers. The two mechanisms, frame transmission (FTR) and burst transmission (BTR) are typically employed [11] in EEE. However, the difference between them is depicted in Fig. 1.

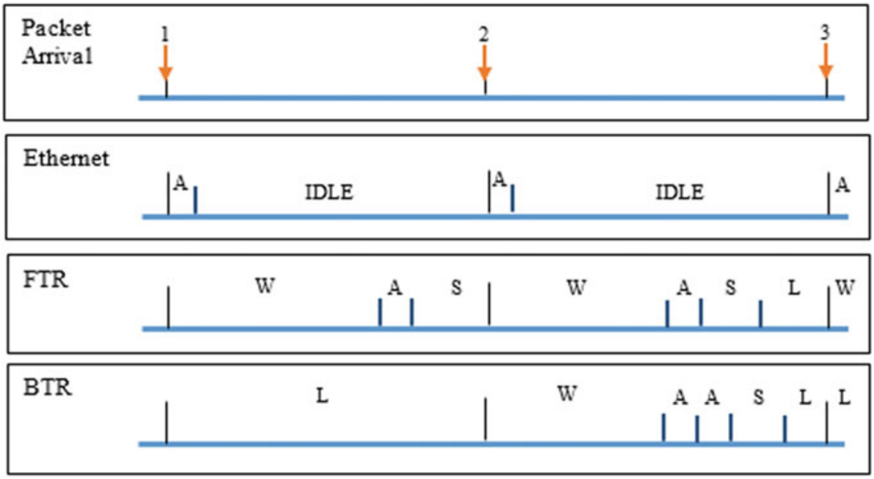


Fig. 1 An event line for ethernet, frame transmission and burst transmission of EEE. Note: *FTR* Frame transmission; *BTR* Burst Transmission; *A* Active; *L* Low power idle; *S* Sleep; *W* Wake

In normal legacy Ethernet, the idle phases consume an equal amount of power compared to the active phases. However, EEE aims to minimize the power consumption of idle times with FTR and BTR. With FTR, the Ethernet interface is activated the moment a new packet arrives and the packet is then immediately handled. On the contrary, for BTR, the newly arrived packet is added to a queue and the interface switches to an active state once the queue is full. BTR affords packets coalescence within a certain timeout in order to avoid excessively waking the interface. The nature of the BTR mechanism offers a great platform for further research. Two variables play a role in BTR [14] and they are: **buffer size**—this dictates how many packets to hold in the switch before coalescing them and releasing them as a burst; **timeout**—maximum time to keep the interface in LPI mode after the first packet is added to the buffer. If the buffers fill out before timeout, all packets in the buffer are released. Similarly, if the buffer is not full and the timeout expires, the packets are also released. Tweaking these two values yields many research opportunities with researchers attempting to maximize LPI time. Increasing the buffer size without setting a timeout can lead to significant packet delay since the equipment can wait infinitely for the buffer to fill up. On the other hand, setting a timeout that is too short can lead to a decrease in potential power savings.

1.2 Motivation

Ethernet still constitutes the main technology deployed in core networks and cloud infrastructures (for example, in data centers). In 2010, in the US, data center power consumption is estimated to be approximately 2.3% of the total country's power [5] with data center construction worldwide estimated to grow by 9% each year for the next 5 upcoming years [6]. Almost a quarter of the power of these data centers is consumed by the network equipment [7] and with link speeds expected to reach 200 Gbits/s soon [8], the power consumption will undeniably increase with time. EEE aims to bring about massive power savings to core networks and data centers, thus, it is imperative to investigate the effects of this technology. Orange Telecom, the largest telecommunications service provider in France, and a collaborator in this paper is interested in assessing the economic and environmental benefits of EEE deployment on their core networks. Based on this research finding, network administrators could choose to either upgrade their core networks with the new technology or continue their business as usual.

1.3 Problem Definition

Standardization of EEE occurred since 2010, however, to date there is limited investigation on its benefits. Consequently, this research aims to assess the benefits of deploying EEE in the core network by modelling the power consumption of an off-the-shelf Cisco 2960-X with and without power efficient protocol under different link speeds and traffic throughput. We will first measure the power consumption of the switch without EEE under different link speeds with 1 Gbits/s being the highest available bandwidth. Subsequently, we will enable EEE and test it under different traffic throughput in order to assess the improvement it attains over legacy Ethernet. We will also compare EEE to other power saving mechanisms to better assess its benefits.

1.4 Delimitations

As 10 Gbits/s links are still uncommon in commercial grade network equipment, no measurement could be made for this particular bandwidth. Our work revolves around profiling and modelling real equipment's power consumption but not for 10 Gbits/s (and beyond) links.

This book chapter is structured as follows: Sect. 2—Related Work; Sect. 3—Underlying Theories; Sect. 4—Implementation; Sect. 5—Results and Discussion; Sect. 6—Conclusion and Future Work.

2 Related Work

Since 2010 (i.e. after the standardization of EEE), researchers have explored ways to maximize the energy savings. This section will review existing research on EEE energy savings profiling. Reviriego et al. [15] conduct a research similar to our research. They create a power consumption model for small EEE enabled switches. Their results reveal that EEE switches consume similar amount of energy as legacy switches when no port is connected. They attribute this fact to hardware manufacturers not adapting their device components to EEE. Additionally, energy consumption of future switches would be very close to link utilization which is similar to work done by Paillassa et al. [16] and Rodríguez-Pérez [17]. However, both research work utilizes simulations in ns-3 software instead of real EEE switches. Saranavan et al. [18] assesses the benefit of EEE in High-Performance Computing scenarios. Their findings reveal that EEE reduces energy consumption of links by 70%. However, the on/off transitions of EEE greatly hamper the overall energy savings of the system. Consequently, the overall reported energy increase is merely 15% when using the default mechanisms of EEE. Saranavan and colleagues [ibid.] subsequently suggest a "Power-Down Threshold" feature that keeps the link on until a certain threshold is reached. With this technique, they report savings of around 7.5% on the overall energy consumption of the system as well as a reduction in the on/off transition overhead from 25 to 2%. In 2010, Lee et al. [19] propose a change to the MAC of optical network units in order to incorporate the interface called Slotted Delivery Mode. The Slotted Delivery method means that when traffic is low, the interface only wakes up periodically to handle the incoming and outgoing packets before going to sleep again. According to their simulations, they manage to have a power saving of 96%. Several researches explore the deployment of LPI mode. Kubo et al. [20] propose a hybrid mechanism that combines LPI and adaptive link rate which is aimed solely at 10 Gbits/s links. This is to sleep the switch when traffic is absent and lower the link rate when link utilization is low. This hybrid mechanism manages a savings up to 84% when traffic is absent and 63% when traffic is low. These findings concur with that of Reviriego et al. [21] and Jin et al. [22].

Mostowfi et al. [23] in 2015, suggest having two low-power modes in EEE in order to avoid latencies from on/off transitions. When traffic is low, EEE triggers the first stage of sleep. This stage can be woken up much faster than the default LPI. When traffic is absent after a certain threshold, a deeper sleep mode is activated. The energy consumption of this mode matches that of LPI. Therefore, this model does not save as much energy as the default model but instead improves the latencies and delays inherited from EEE implementation. Herreria-Alonso [24] extends this research by testing the two low-power modes on simulations based on real traffic scenarios. They evidence that the first stage of sleep is futile in most of their scenarios and that Deep Sleep state is preferred on the links instead of the two stage solution.

Other energy savings techniques are previously mentioned Frame Transmission (FTR) and Burst Transmission. Meng et al. [25] in 2013, propose a new EEE policy to exploit the strength of both frame and burst transmission. The policy dictates the use of frame transmission when traffic is low, in order to avoid delay issues from burst transmission. According to Reviriego et al. [14], using burst transmission on a 10 Gbits/s EEE link can yield up to 80% energy reduction in certain situations. Due to queueing delay from packets coalescence, these situations typically are not time sensitive (e.g. file download or file indexing on servers). However, it is worthy to note that the energy reduction is approximately 5% for video streaming and download where the packet delay can have significant impact on user quality of experience. In high traffic scenarios, Meng and colleagues' EEE policy [25] propose the use of a buffer and timeout based system. In cases where traffic is medial, the policy recommends using a buffer only system without timeout in order to receive optimum results. It is worth noting that this work has been accepted for publication in IEEE in March 2017. Mostowfi et al. [26] reveal that setting the buffer size to 1000 packets and setting the timeout for 100 ms result in an 80% decrease in energy consumption. Their work also takes into account delay resulting from packet coalescence and estimates an average packet delay of 50 ms under normal traffic load. Their findings show that sending traffic in bursts does not add delay to packet delivery but offer lower energy savings than their advertised 80%. It is worth noting that the scenario used for their experiments is a file download scenario. Similar work is conducted by Aksić et al. [27], in which they propose different buffer and time out values for different bandwidths. In 2012, Herreria-Alonso et al. [13] develop an EEE model to explore the optimum values for buffer and timeout given an energy consumption threshold denoted by 'n'. This threshold represents the maximum allowed energy consumption percentage over the ideal minimum. The optimum configuration of variables with 'n' = 5%, results in a timeout of 120 μs and a queue of 25.73 frames. According to their simulations, burst transmission with these parameter values performs better than frame transmission in all traffic scenarios with no mention of the queueing delay created as a result. Other related work includes Chakadkit and colleagues' work [28] which involves changes in routers' control plane in routers. They present EAGER and CARE, two OSPF metrics customized for augmenting the energy saving of IEEE 802.3az line cards. Both mechanisms take into consideration traffic and congestion and adaptive link rates. They claim an energy saving that is double to that of EEE with default mechanism.

In summary, most of the related research reviewed, aims to improve benefits of EEE using burst transmission. Research focuses on developing models using network simulations and testing the models against real world traffic from internet sources. Generally, research reviewed claims savings between 50 and 90% over traditional Ethernet. However, these models are have not used real equipment in order to validate the results. Several researches explore changes to existing EEE protocol in order to maximize savings. Some integrate EEE and adaptive link rate into a single mechanism. Changing link speed is the optimum solution for low traffic times since it only slows the flow of traffic without stopping. Meanwhile, EEE can be used in downtimes where there is no imminent traffic.

Finally, our work extends the work conducted by Hossain et al. [29] where they profile the power consumption of ethernet switches using experiments. They utilize switches similar to the ones we use for our experiments. The work concludes with a power consumption model of the switch with a multitude of variables. Our work aims to add EEE to that power consumption model in order to reflect the current status of industrial switches. Additionally, Reviriego et al. [15] have developed a linear model for energy consumption in switches. Similarly, we shall use a linear regression to model for our measured data.

3 Underlying Theories

A survey of relevant research reveals that a majority of them model energy consumption of EEE followed by validating them using network simulations. However, network simulators are not as reliable and accurate as real measurement obtained from physical network equipment. Therefore, for this research, we have obtained commercial grade network switches and evaluate their power consumption against all a set of parameters. These parameters include different bandwidths, 1 Gbps, 100 Mbps and 10 Mbps, as well as different power savings algorithms such as EEE and sleep/hibernation.

3.1 Unidirectional Versus Bidirectional EEE

According to the EEE standard [10], each EEE bandwidth has different implementation. For example, 10 Gbps link are unidirectional while 1 Gbps links are bidirectional in nature. A unidirectional link only needs to sense a lack of traffic on its own end before deciding to trigger an LPI mode. On the other hand, bidirectional links need to agree on both ends of the link on the lack of traffic before deciding to sleep. Imagine a scenario where two EEE capable PCs are directly connected to each other by a cable. If the link is 10 Gbps, each PC decides on its own when to sleep their respective ports. On the other hand, if the link that connects the two is a 1 Gbps link, both PCs have to communicate before deciding on when to trigger sleep mode. Transitions between sleep and wake cannot be interrupted in unidirectional links. Therefore, if a packet arrives to the equipment while it is transitioning to sleep, the packet has to wait for sleep to finish and wake to be consequently triggered. Depending on the network configuration and sleep/wake transition times, this particular packet could be dropped or delayed. Bidirectionality entails longer delays in 1 Gbps links when attempting to trigger sleep mode as both sides have to agree on when to trigger it. The switches used for our research have a maximum bandwidth of 1 Gbps, thus, the EEE we assess is bidirectional. This means that for our experiments, every equipment plugged into the switch is EEE capable else the interface reverts to legacy ethernet.

3.2 Modelling Power (and Energy) Consumption

Two types of power consumption models (with and without EEE) are abstracted from the measurements. Although different bandwidths (i.e. 10 Mbps, 100 Mbps, 1 Gbps, or Fiber optic links) are available to test without EEE, their behavior is similar—power consumption of these ports is fixed and independent of the incoming traffic. With EEE, the power (and energy) consumption of the port is directly linked to incoming traffic. The port operates in two modes: a low power LPI mode and full higher consuming power mode. As EEE merely changes the power consumption of the equipment's ports, it is crucial for our model to accurately represent this feature. Thus, our focus is on the measurements of the power consumption of ports. The EEE model has to both reflect the impact of the technology on the ports and the total consumption.

3.3 Linear Regression Analysis

Linear regression is the simplest and most commonly used prediction model. This type of regression analysis is used to explain the behavior of one fixed variable against one or more independent variables. The linearity of the predicted model can help shed light on the accuracy of the model. The basic shape of a linear equation is one with only one fixed variable and another unknown, independent variable. This basic linear model can be shown by this formula: $y = c + b * x$, where y is the predicted value, c, a constant, b represents a regression coefficient, and x is the independent variable. Besides depicting correlation between different variables, regression models can also be used to forecast trends. It can provide insight into the relationship between EEE and power consumption of our switches. The power consumption of single switch ports (note: the number of switch ports is an independent variable) is the dependent variable which could be forecasted using our model. However, a primary concern with regression analysis is overfitting a model into the data by taking into account multiple independent variables. Diverse range of variables could lead to inefficient models. Simplicity of the developed model is key to its efficiency. Another concern would be under fitting a model. This problem occurs when models attempt to reveal a dependence that does not exist. This problem could arise when independent variables do not affect the independent variable. Therefore, a bias towards certain aspects of the data could falsely lead a prediction of a non-existent linearity.

4 Methodology

In this section, we describe the tools and equipment used to conduct the experiments. They include measurement tools, graphing tools and network equipment.

4.1 Measurement Tools

PowerSpy2[1] by Alciom (see Fig. 2), is employed for the measurement of the switches' power consumption. Plugging the switch into PowerSpy2 yields real time data on the power consumption of the switch. PowerSpy2 is a wall plug that measures a myriad of information about the plugged devices and the device uses bluetooth to communicate the data to a special software that only runs on windows. The biggest advantage of using PowerSpy2 is its ability to detect subtle changes in power consumption. In order to measure the ports' power consumption, a measuring device that can detect variations in mW is required. PowerSpy2 has such a capability and it could measure as low as 10 mW with an accuracy of 1%. The device's baseline power consumption of 1 W is automatically deducted from its measurements in order to present the most accurate data. The Alciom software displays information on Voltage, Current and Power in real time. A multitude of logging features are available as well as providing a download facility for all measured data in a csv format. Three main functions of the Powerlog software are: automatic import of PowerSpy2 log files on real time power consumption (see Fig. 3 and the focus is on the Prms (W0 display of the dashboard), interactive power plot (see Fig. 4) and statistics (see Fig. 5). In Fig. 4, the focus is on the power graph because it depicts the power consumption of the device. It provides a visualization of power consumption trends and could help provide insight into anomalies or problems with the equipment. The software has control facility for resetting or restarting all measurements. Figure 5 shows the following information about power consumption of the equipment: average, standard deviation, maximum and minimum values. It provides a reset facility of the statistics. The crucial part of this research is to be able to generate real traffic in order to evaluate how the network equipment reacts under real pressure. Hence, we use JPerf to generate all traffic for our experiments. JPerf is a graphic wrapper for the traffic generating utility IPerf. Details are found here.[2]

Choice of Network Equipment

As our research goal is to test a piece of commercial grade equipment running EEE, we have opted for Cisco 2960-X switch. This particular switch is advertised as an energy saving solution for small offices and businesses. Besides EEE, the switch is also equipped with sleep/hibernation techniques that allow the network administrator to configure specific times where the switches would go into hibernation mode. This mode would not allow any traffic to pass through the switches and thus significantly reduces the power consumption. The inclusion of these features in the switch allows us to pit EEE against other power saving techniques in order to better understand the benefits it yields. Experiments conducted involve the use of two

---

[1] http://www.alciom.com/en/products/powerspy2-en-gb-2.html.

[2] http://wirelesslanprofessionals.com/wp-content/uploads/2011/02/How-to-Guide-on-JPerf-and-IPerf.pdf.

Fig. 2 PowerSpy2

24-port and two 48-port switches. In order to capture variations that could occur in the manufacture of switches, we have opted for two different switches of the same model. We have also opted for 24-port and 48-port variations in order to investigate how power consumption of the switch scales (if any) with the size of the switch. Two different sizes of switches could demonstrate if EEE could yield better results with bigger equipment (i.e. with more number of ports).

4.2 Power Profiling of Switch Ports

As previously mentioned, power profiling a switch's ports is essential for the development of the power consumption model. Hence, devised experiment procedures encompass placing a switch in an idle mode without any traffic followed by connecting the switch's ports incrementally (i.e. one by one) to a network equipment. Connecting the network equipment to the switch activates that port in the switch and total power consumption for utilizing that port is monitored by PowerSpy2 and displayed by its Powerlog software. For every port that is plugged, the statistics component is reset before the experiment commences (see Fig. 5) and the total power consumption of this port is logged. Each experiment is

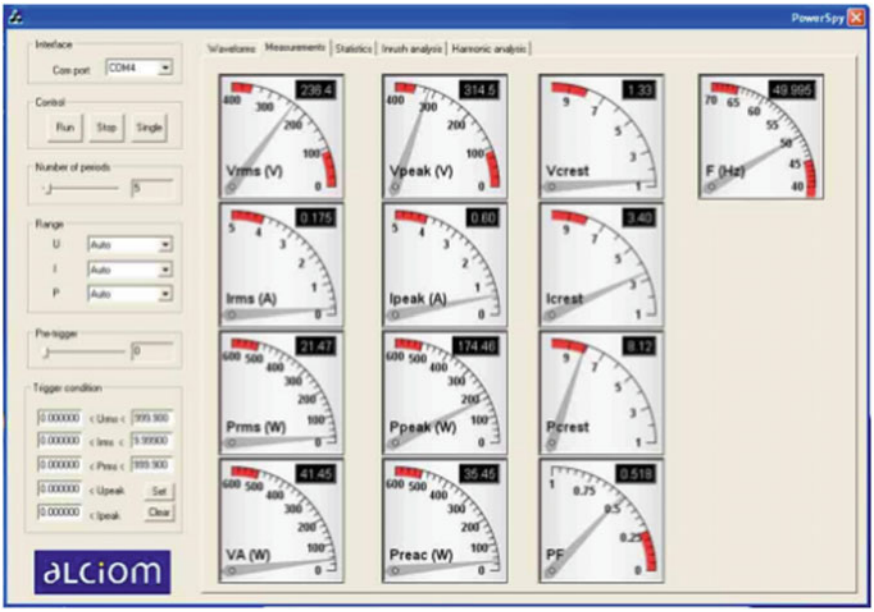


Fig. 3 Real time power consumption

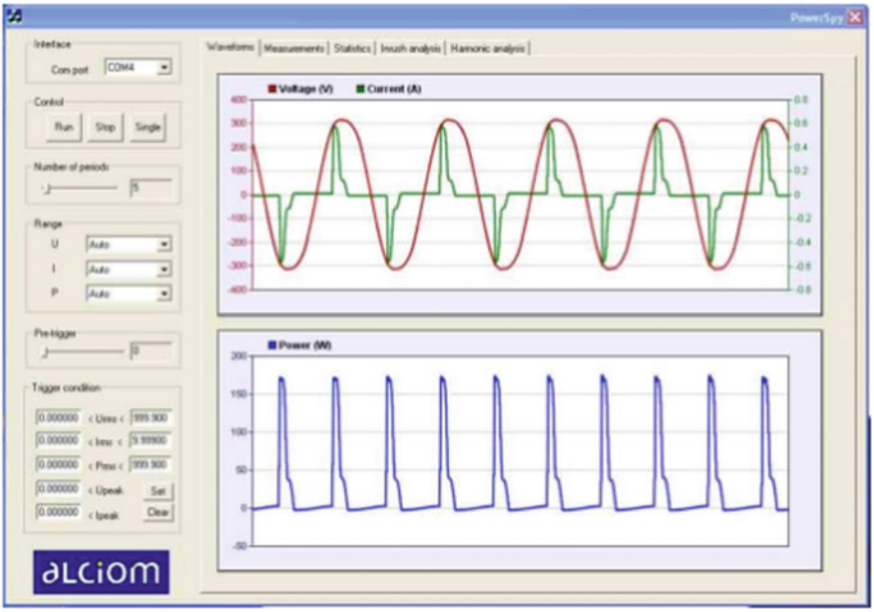


Fig. 4 Power consumption graphing tool

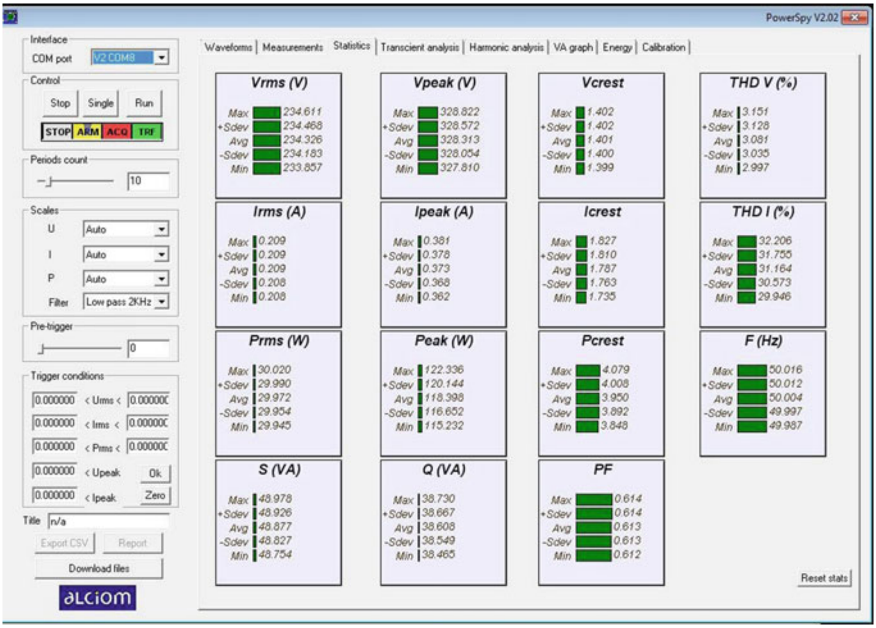


Fig. 5 Statistical calculation of powerlog software

systematically carried out by connecting a port of the switch at a time until all ports of the switch have been connected. Each set of experiments involves 10 repeated runs. The architecture of our experimental setup is shown in Fig. 6. As shown in Fig. 6, two switches are employed for the experiments. One switch is measured by PowerSpy2 while the other serves as a hub to connect the ports. Only one switch is measured at a time in order to be able to assess the power consumption of a single port. The setup is the same for both the 24-port and 48-port variants. However, the 48-port variant is matched by another 48-port variant in order to be able to connect all the ports. Our main aim of profiling the ports is to assess the impact of EEE. Hence, the set of experiments are coded into two categories namely: (with) EEE and No EEE. When EEE is not enabled, the bandwidths are switched between 10 Mbits/s, 100 Mbits/s, 1 Gbits/s and Fiber optic port. However, when EEE is enabled, the only focus is on the highest available bandwidth which is 1 Gbits/s. As EEE is supposed to shut the ports when traffic is not present, coupling EEE with lower bandwidths is not a commercially viable option and consequently, only 1 Gbits/s EEE links are utilized. Experiments conducted are tabulated in Table 1. Once data is collected from these experiments, an appropriate power model is developed and utilized for the rest of the experiments.

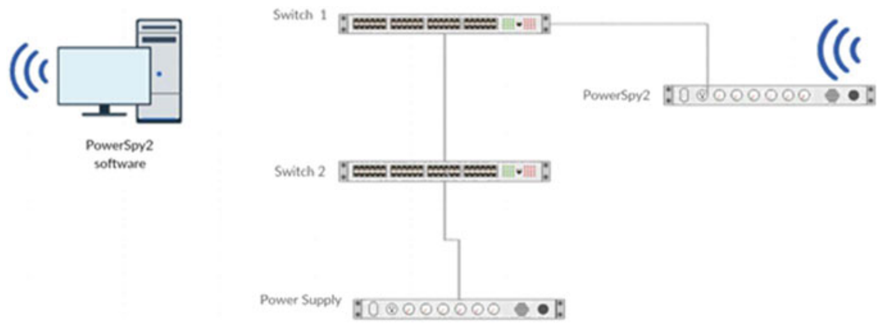


Fig. 6 Architecture of the experimental setup for power profiling switch ports

Table 1 Experiments (with) EEE and No EEE

(with) EEE	No EEE
1. 1 Gbits/s link speed 2. 100 Mbits/s link speed 3. 10 Mbits/s link speed 4. Fiber optic	1. 1 Gbits/s link speed

4.3 *Power Measurement of Switch with Different Traffic Patterns*

The next set of experiments we devise are to measure the power consumption of the switch under different traffic patterns in order to better estimate the benefits of deploying EEE in core networks. As EEE is heavily influenced by traffic, we devise a set of tests to put it through its paces. We have consequently designed 4 different traffic scenarios intended to test EEE in all its different aspects. The traffic patterns are as follows.

In Table 2, ST and LST are examples of burst traffic chosen to test the viability of sending traffic in bursts with EEE. Such traffic patterns are preferred patterns to balance performance and savings since EEE induce power savings in LPI states. Additionally, LST is chosen based on the delay researchers claim when EEE

Table 2 Different traffic patterns

Traffic patterns	Description
Full traffic (FT)	Consists of a 100% load on the link for the entire duration of the experiment
Sinusoidal traffic (ST)	Consists of a burst traffic for 20 s per each minute of the experiment
Short low sinusoidal traffic (LST)	Consists of a burst traffic for 1 s per each minute of the experiment
Minimum traffic (MT)	Consists of a constant load of 4 kbit/s over the link for the entire duration of the experiment

switches between sleep and wake states. Sending a 1 s traffic forces the switch to quickly change state and hence, demonstrating the delay inherent to it. As for the MT pattern, according to the EEE standard [10], there is a threshold for incoming packet sizes that prevents EEE links from waking up. This threshold is set to prevent links from waking up under unnecessary traffic from STP packets and similar protocols. Therefore, if we send packets which are sufficiently small, we can theoretically send traffic without waking up the links. The MT traffic pattern is set to test this aspect of EEE while FT pattern sets the upper limit of power consumption in EEE. The architecture of the experiment setup is depicted in Fig. 7. Based on Fig. 7, two EEE capable PCs are connected to EEE enabled switch with PC 1 sending various traffic patterns to PC 2. The experiments run for a duration of 10 min. When EEE is absent, the power consumption of a port is not affected by traffic as it always operates with maximum power. Hence, no traffic pattern is used for experiments with No EEE. Experiments conducted are tabulated in Table 3.

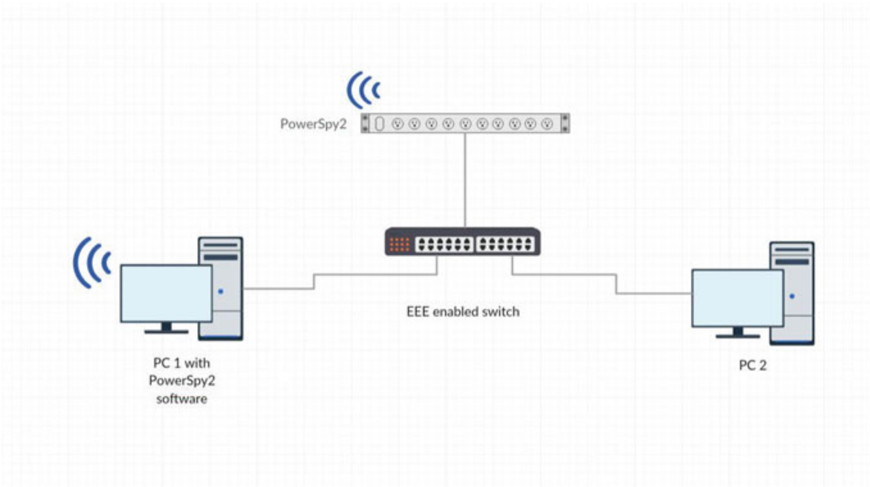


Fig. 7 Architecture of the experimental setup for power measurement of switch based on traffic throughput

Table 3 Experiments (with) EEE and No EEE (24-port and 48-port)

(with) EEE	No EEE
1. FT 2. ST 3. LST 4. MT	1. 1 Gbits/s bandwidth 2. 100 Mbits/s bandwidth

4.4 Compare Power Savings

The final part of the experiments is to utilize our developed model for assessing EEE against alternative solutions in a real scenario. Hence, we set up a scenario with a lab of 48 PCs. This scenario could be scaled up into a multitude of equivalent scenarios in a core network or data center environment. The options we have are either getting 2× 24-port switches or a single 48-port switch. Each switch can either be EEE enabled or a simple legacy ethernet switch. With legacy switches, ports could be shut down manually or switches completely turned off. The first scenario would be to test the power disparity between regular legacy switches of different sizes. This scenario represents a situation where traffic is unpredictable. The network administrator does not have access to traffic information and has no prediction of Up and Down times in the network. Therefore, the PCs in the lab are left continuously running for 24 h a day. On a legacy switch, this means that the ports will never sleep but just keep running. As the traffic is considered to be a constant 24-h traffic, there is no way we can measure the expected energy consumption using EEE. In order to calculate the energy consumption with EEE, the number of idle hours has to be known. Additionally, two NO EEE switches are used for this experiment. Thus, we can only use legacy switches for this scenario and this provides the opportunity to compare energy consumption of switches with differing sizes. The setup of the experiments is as follows: *two NO EEE 24-port switches versus one NO EEE 48-port switch*. Additionally, it is interesting to investigate how much to sleep the greedier option in order to come up with a better alternative. For example, how long should the two 24-port switches be shut down for their energy savings in order to match that of a single 48-port switch. This finding can inform network administrator of a better option for energy savings. On the other hand, if the network administrator knows that half of the PCs are not used at night, a second scenario could be devised in a manner whereby 24 PCs will continuously work for 24 h a day while the remaining 24 PCs could run for 16 h a day. Given similar switches and parameters of the first experiment, the following scenario could be arranged: *two NO EEE 24-port switches versus one EEE 48-port switch*; *two NO EEE 24-port switches versus two EEE 24-port switch.* Since a 48-port switch cannot be shut down at night as 24 ports are still needed, having a legacy 48-port switch in our scenario is not feasible. Therefore, choosing an EEE enabled 48-port switch for this scenario is a viable choice. On the other hand, as we have 24 PCs that can hibernate at night, we could then attach all these machines to the same switch and shut down the switch during the night. The second scenario arrangement is a direct comparison between EEE and legacy technologies while the first arrangement is a test of energy consumption scaling in switches. In this scenario, we can then test legacy methods of energy savings versus EEE in an attempt to see if EEE could bring about the needed benefits over its old fashioned manual counterpart.

5 Results and Discussion

In this section, experimental results are presented and discussed. To reiterate experiments conducted are: with and without EEE for 24-port as well as 48-port switches; different bandwidths; and finally, different traffic patterns.

5.1 *Power Profiling of Switch with 24 Ports (NO EEE and (with) EEE)*

For each variable set (i.e. bandwidth of 1 Gbits/s, 100 Mbits/s, and 10 Mbits/s) in our methodology, all 24 ports in the switch are plugged in one by one with their corresponding power consumption measured. The base power consumption of the switch with no ports connected is measured at 30.00 W. Each spike in the graphs shown in Figs. 8, 9, 10 and 11, represents a new link being plugged in for an EEE enabled (and without) switch. It is noted that the increase power consumption when a link is connected is consistent. A summary of the results is tabulated in Table 4. Link being plugged in for an EEE enabled (and without) switch. It is noted that the increase power consumption when a link is connected is consistent. A summary of the results is tabulated in Table 4.

Table 4 shows that the average power consumption of a 1 Gbits/s link (NO EEE) is approximately 3 times more than 100 and 10 Mbits/s (NO EEE) while the difference between the latter two is less pronounced. On the other hand, the average power consumption of a 1 Gbits/s link without EEE is approximately 7 times that of a 1 Gbits/s link with EEE. A graphical representation of the power consumption of the number of connected ports in a 24-port switch (NO EEE) are depicted in Fig. 12. Table 4 shows that the average increase in power consumption for each

Power Consumption of Switch with 24 Ports

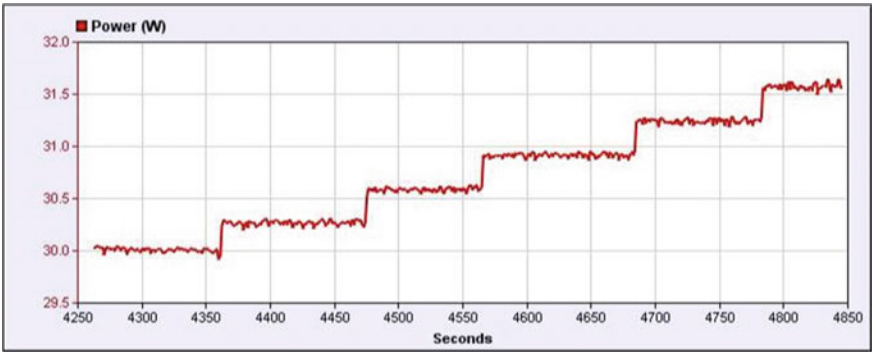


Fig. 8 1 Gbits/s link speed (no EEE)

Power Consumption of Switch with 24 Ports

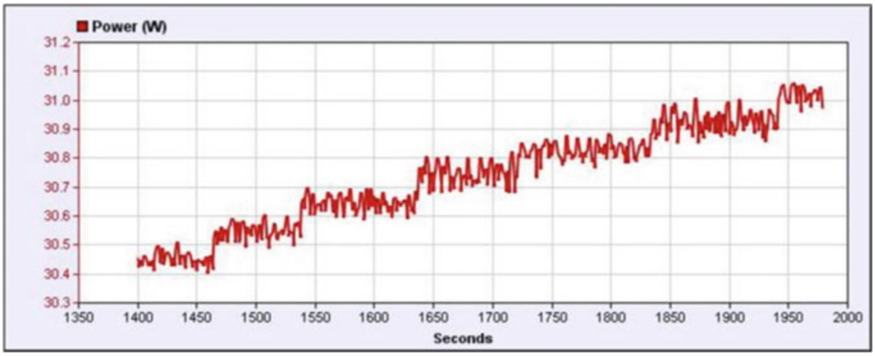


Fig. 9 100 Mbits/s link speed (no EEE)

Power Consumption of Switch with 24 Ports

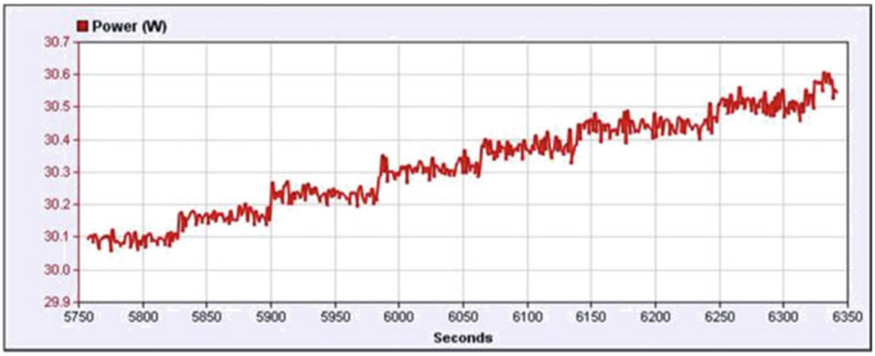


Fig. 10 10 Mbits/s link speed (no EEE)

Power Consumption of Switch with 24 Ports

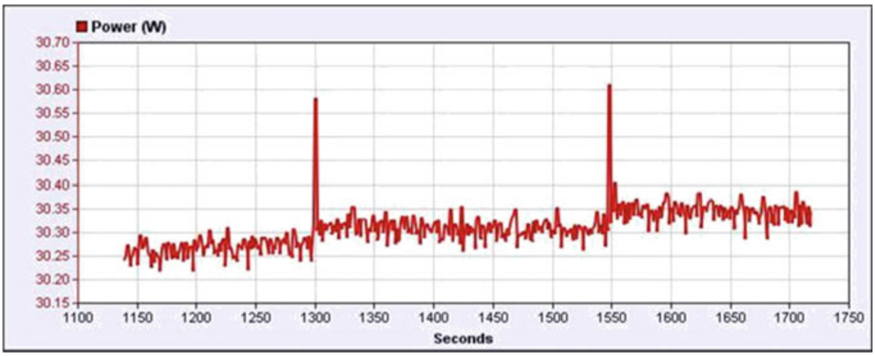


Fig. 11 1 Gbits/s link speed (with EEE)

Table 4 A summary of (with) EEE and NO EEE with 24 ports

Column A Bandwidth	Column B Average increase in power consumption for each new link (W)	Column C Normalized value for Column B	Column D Maximum consumption for switch 24 connected ports (W)	Column E Normalized value for Column D
1 Gbits/s (NO EEE)	0.33	1.00	37.88	1.00
100 Mbits/s (NO EEE)	0.10	0.16	32.28	0.16
10 Mbits/s (NO EEE)	0.07	0.07	31.66	0.07
1 Gbits/s [(with) EEE in LPI state]	0.05	0.00	31.22	0.00

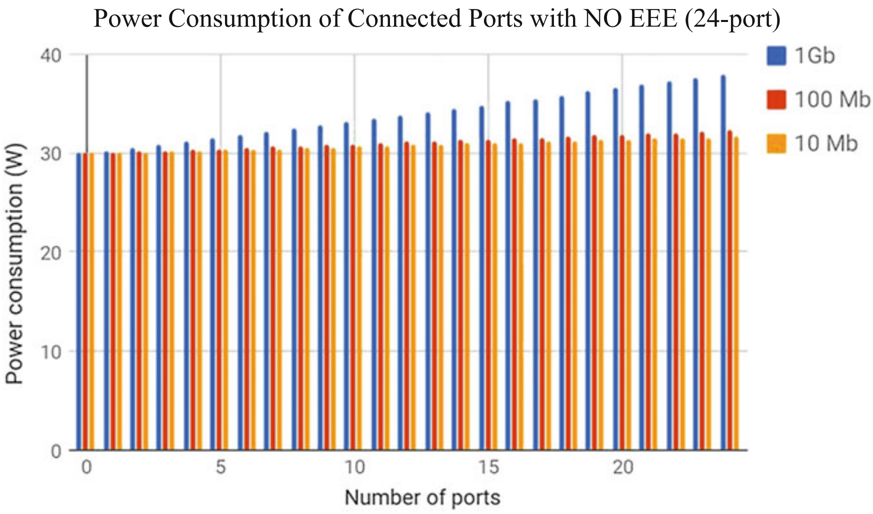


Fig. 12 Power consumption and the number of connected ports with NO EEE in a 24-port switch

connected link is approximately 0.05 W. As there is no traffic passing through the switch, this increase in power corresponds to the link in LPI mode. With the maximum power consumption of the link measured at 0.33 W for 1 Gbits/s link, LPI mode is about 15.1% of the maximum power consumption of the link.

Table 5 shows that the average power consumption of a 1 Gbits/s link (NO EEE) is approximately 3 times more than 100 and 10 Mbits/s (NO EEE) while the difference between the latter two is less pronounced. On the other hand, the average power consumption of a 1 Gbits/s link without EEE is approximately 9 times that of a 1 Gbits/s link with EEE. A graphical representation of the power consumption

Table 5 A summary of (with) EEE and NO EEE with 48 ports

Column A Bandwidth	Column B Average increase in power consumption for each new link (W)	Column C Normalized value for Column B	Column D Maximum consumption for switch 24 connected ports (W)	Column E Normalized value for Column D
1 Gbits/s (NO EEE)	0.35	1.00	51.10	1.00
100 Mbits/s (NO EEE)	0.10	0.20	39.34	0.20
10 Mbits/s (NO EEE)	0.07	0.10	37.90	0.10
1 Gbits/s [(with) EEE in LPI state]	0.04	0.00	36.46	0.00

of the number of connected ports in a 48-port switch (NO EEE) are depicted in Fig. 13. Table 5 shows that the average increase in power consumption for each connected link is approximately 0.04 W. As there is no traffic passing through the switch, this increase in power corresponds to the link in LPI mode. With the maximum power consumption of the link measured at 0.35 W for 1 Gbits/s link, LPI mode is about 11.4% of the maximum power consumption of the link.

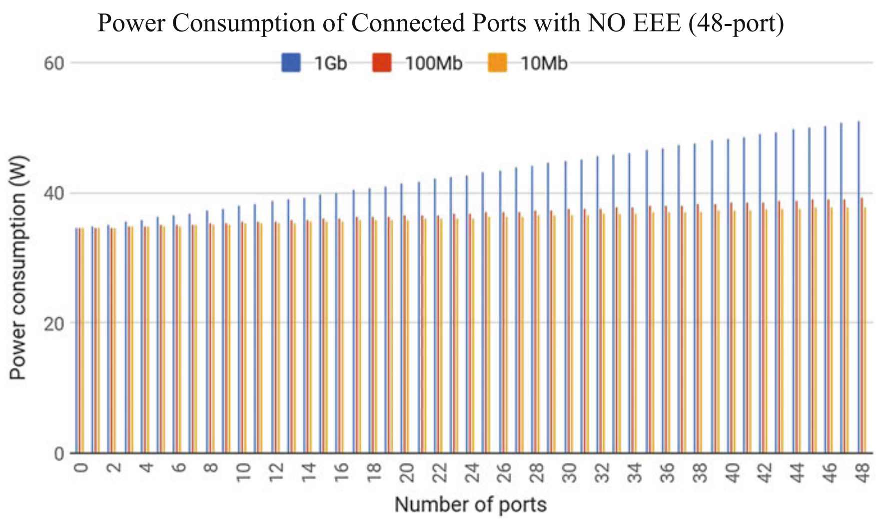


Fig. 13 Power consumption and the number of connected ports with NO EEE in a 48-port switch

5.2 Power Measurement of Switch According to Traffic Throughput

For this set of experiments, two EEE enabled PCs are connected to a switch, sending traffic to each other. The parameters set for these experiments (see Table 3) are: four traffic patterns (i.e. FT, ST, LST, MT); two different link bandwidths (i.e. 1 Gbits/s and 100 Mbits/s); two different technologies (i.e. EEE and No EEE). Each experiment runs for a duration of 10 min. Results of the experiments have been tabulated in Tables 6 and 7. Based on the results, it is evident that for a bandwidth of 1 Gbits/s, an EEE enabled switch will be more power efficient compared with one that is without EEE. Among all the different traffic patterns (for both the 24 and 48 ports), FT is the one that is not power efficient while MT is the most power efficient.

Table 6 Total power consumption of a 24-port switch

Parameters (24-port)			Total power consumption (W)	Power gain (%)
Bandwidth	EEE	Traffic pattern		
1 Gbits/s	NO EEE		5.1050	0.00
100 Mbits/s	NO EEE		5.0250	1.57
1 Gbits/s	With EEE		5.0573	0.93
1 Gbits/s	With EEE	FT	5.1479	−0.84
1 Gbits/s	With EEE	ST	5.0846	0.40
1 Gbits/s	With EEE	LST	5.0613	0.86
1 Gbits/s	With EEE	MT	5.0591	0.90

Note 1 Gbits/s (NO EEE) is used as the benchmark

Table 7 Total power consumption of a 48-port switch

Parameters (48-port)			Total power consumption (W)	Power gain (%)
Bandwidth	EEE	Traffic pattern		
1 Gbits/s	NO EEE		5.8730	0.00
100 Mbits/s	NO EEE		5.7916	1.39
1 Gbits/s	With EEE		5.7656	1.83
1 Gbits/s	With EEE	FT	5.8988	−0.44
1 Gbits/s	With EEE	ST	5.8196	0.91
1 Gbits/s	With EEE	LST	5.7755	1.66
1 Gbits/s	With EEE	MT	5.7693	1.77

Note 1 Gbits/s (NO EEE) is used as the benchmark

5.3 Overview of Power Consumption of EEE-Enabled Switches

The base power consumption of a 24-port switch with EEE (i.e. 30.10 W) is slightly higher than the switch without EEE (30.00 W see Sect. 5.1). Once our PCs are connected, this power consumption rises to 30.19 W (see Fig. 18). Our previous experiments show that a port in LPI consumes around 0.05 W power. The increase in power consumption approximately corresponds to the sum of two LPI consumption (hence, this data confirms previous LPI measurements). Once traffic reaches the switch, it seems that the power consumption corresponds accordingly and rises to 30.72 W (see Fig. 18). This rise can be attributed to the ports moving to an LPI to an active state. This increase in power consumption is measured at around 0.25 W per port. The sum of this figure and LPI consumption is 0.30 W, which approximately corresponds to the power consumption of 1 Gbits/s link with NO EEE (see Table 4). Therefore, with no traffic, the EEE link consumes around 2.02% of its possible maximum power value and the link consumes the maximum power value once a traffic arrives. It seems there is no visible real time delay between the sleep and wake transitions. The base power consumption of a 48-port switch with EEE is at 34.54 W (see Sect. 5.2).

Once our PCs are connected, this power consumption rises to 34.62 W. Our previous experiments show that a port in LPI consumes around 0.04 W (see Table 5). The increase in power consumption approximately corresponds to the sum of two LPI consumption (once again, this data confirms our previous measurements of LPI).

When traffic reaches the switch, it seems that power consumption corresponds accordingly and rises to 35.21 W (see Fig. 19). This rise could be attributed to the ports moving from an LPI to an active state. This increase in power consumption is measured at around 0.30 W per port. If we add the LPI consumption to this figure, the result is 0.34 W, which corresponds to the power consumption of 1 Gbits/s link with No EEE (see Table 5). Therefore, with no traffic, the EEE link consumes approximately 1.90% of its maximum power value and the link consumes the maximum power value on the arrival of the traffic. To reiterate, there is no visible real time delay between the sleep and wake transitions (Figs. 14, 15, 16, 17, 18 and 19).

5.4 Comparison of Power Consumption of 24-Port and 48-Port Switches (with and Without EEE)

To reiterate, the base power consumption of the 24-port and 48-port switches without any ports connected is 30.00 and 34.54 W respectively. The power consumption of the switches after connecting the ports, are shown in Fig. 20. Without EEE, the difference in the power consumption is quite significant with the 48-port consuming 51.1 W compared to 37.88 W for the 24-port. The difference is less stark with

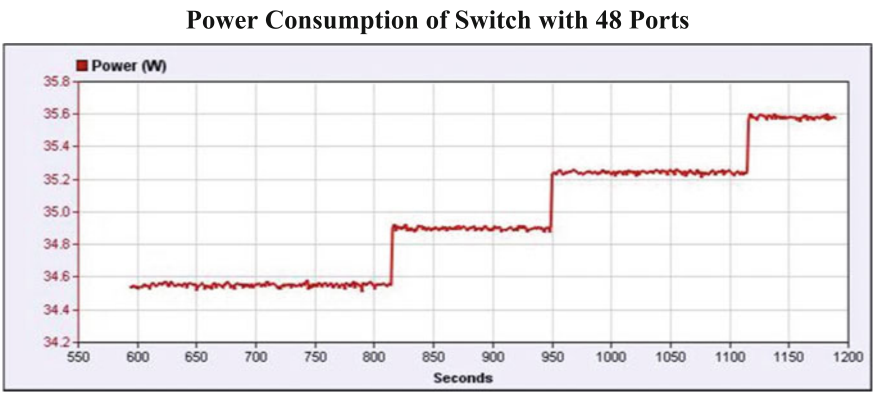


Fig. 14 1 Gbits/s link speed (no EEE)

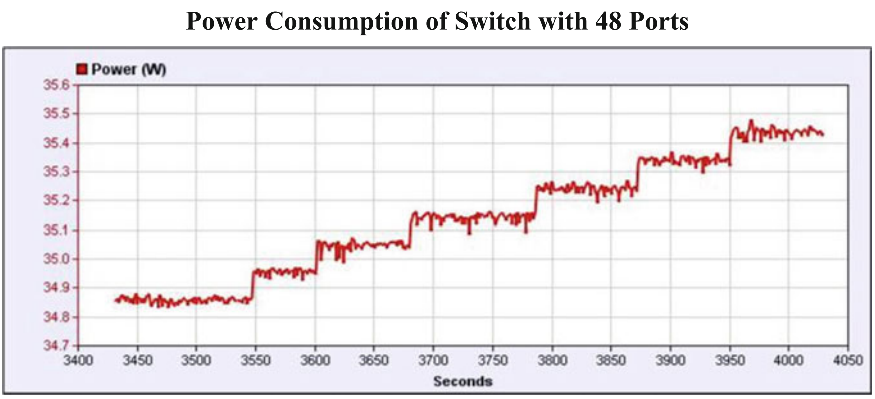


Fig. 15 100 Mbits/s link speed (no EEE)

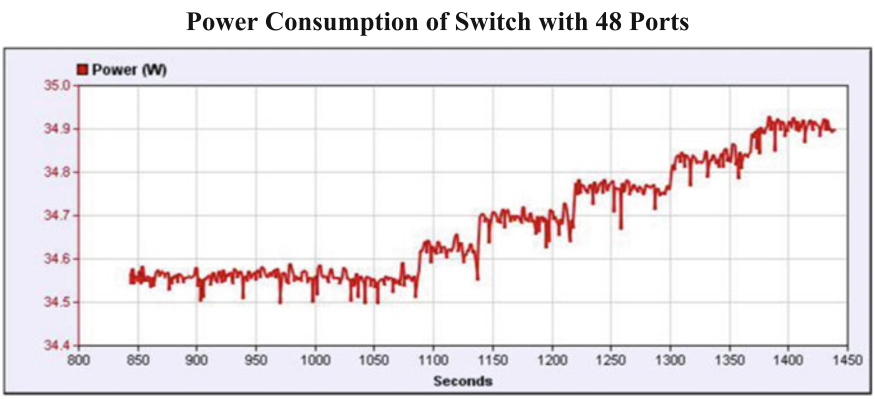


Fig. 16 10 Mbits/s link speed (no EEE)

Power Consumption of Switch with 48 Ports

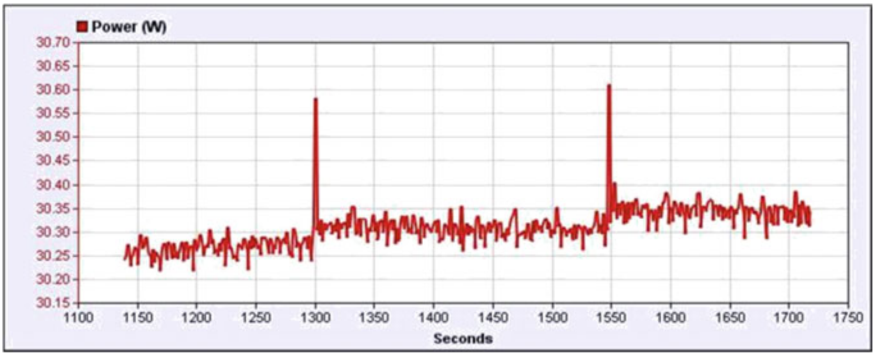


Fig. 17 1 Gbits/s link speed (with EEE)

Power Consumption of EEE enabled Switch with ST Traffic Pattern

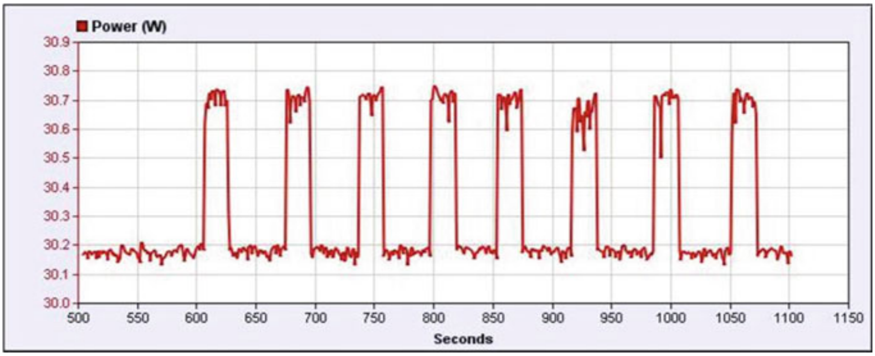


Fig. 18 24-port

Power Consumption of EEE enabled Switch with ST Traffic Pattern

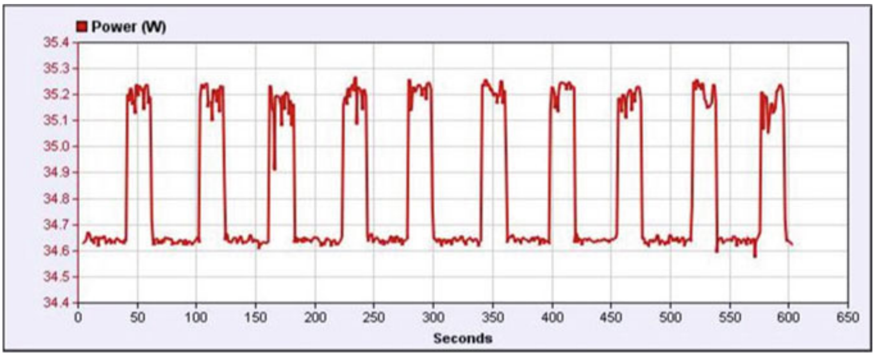


Fig. 19 48-port

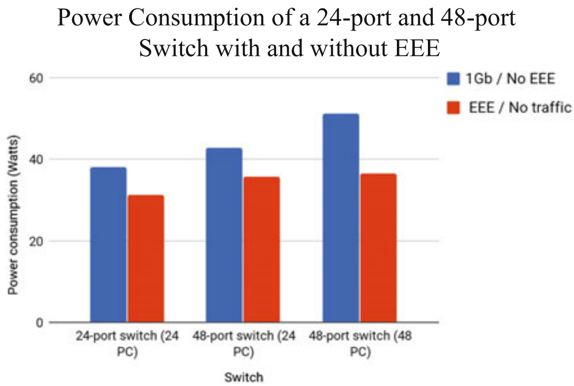


Fig. 20 24-port and 48-port switches

EEE-enabled where the consumption of the switch is 31.30 and 35.61 W for the 24-port and 48-port variants respectively. It is also noted that the power consumption of a single switch port is slightly higher on the 48-port variant measuring at 0.35 W compared to its 24-port counterpart measuring at 0.33 W. Another point to note is that when connecting 24 PCs to the 48-port switch, the difference between the power consumption of the switch and that of a 24-port one is merely the difference between their base power. As the links for both switches consume almost the same amount of power, the only difference between the two is the increased number of ports on the 48-port switch and the slight increase in base power consumption when no port is connected. These results are tabulated in Table 8 and further comparative analysis has been tabulated in Table 9. As the 48-port switch with all the ports connected has the highest total power consumption, it is used as the benchmark for the other configurations. The power gain indicates the power efficiency of using EEE-enabled switches. Based on the analysis shown in Table 8, a 48-port with 48 connected pcs seem to indicate the highest power gain. Table 9 depicts calculated power ratios to facilitate easy comparison. Based on the tabulated results, the following conclusions could be drawn. Firstly, a 24-port EEE-enabled switch with 24 connected pcs is more power efficient than a 48-port EEE-enabled switch with 24 connected pcs. On the other hand, a 48-port EEE-enabled switch with 48 connected pcs is more power efficient that two 24-port EEE-enabled switch with 48 connected pcs.

5.5 Estimated Power (and Energy) Consumption Models

The development of the power and energy models (see Table 10) for the switches in this research are derived from measurements discussed in preceding sections in this book chapter. As previously mentioned, a traffic of 4 kbps (see Table 2) or less

Table 8 Power consumption for different configurations (with and without EEE)

Parameters				Power consumption (W)		Power gain (%)
Switch	No. of connected PCs	Bandwidth (Gbits/s)	Traffic	NO EEE	WITH EEE	
24-port	24	1	No traffic	37.88	31.30	17.37
48-port	24	1	No traffic	42.82	35.52	17.05
48-port	48	1	No traffic	51.10	36.51	28.55

Note Power gain is calculated efficiency gained for switches WITH EEE compared to NO EEE

Table 9 Power ratio for different configurations (with and without EEE)

Parameters				Power consumption			
Switch	No. of connected PCs	Bandwidth (Gbits/s)	Traffic	NO EEE (W)	Power ratio (NO EEE)	WITH EEE (W)	Power ratio (WITH EEE)
24-port	24	1	No traffic	37.88	1.21	31.30*	1.00
48-port	24	1	No traffic	42.82	1.37	35.52	1.13
48-port	48	1	No traffic	51.10	1.63	36.51	1.17
2× 24-port	48	1	No traffic	75.76	2.42	62.60	2.00

Note The reference value is the lowest power consumption value with *. Additionally, the values for 2× 24-port are derived from one 24-port for comparison purposes

does not trigger a port to wake up. Hence, the power (and energy) consumption equation has to consider this condition. Consequently, when traffic is 4 kbps or less, all ports remain in LPI mode.

5.6 Deployment of the Power (and Energy) Models to Estimate Power (and Energy Savings)

In this section of the chapter, we shall use two scenarios to demonstrate how the Power (and Energy) Consumption Models tabulated in Table 10 are deployed (see Tables 11, 12 and 13).

Table 10 Power (and energy) consumption models for the switches

Switch	Model	Key	
NO EEE	$P_S = (N_P \times P_P) + P_B$ (1)	P_S	Power consumption of switch (W)
		P_P	Power consumption of a single port (W)
		P_B	Base power consumption of switch with no connected port (W)
With EEE	$P_S = (N_{SP} \times P_{LPI}) + (N_{AP} \times P_{LPI}) + P_B$ (2)	P_{LPI}	Power consumption of a single port in LPI state (W)
		N_P	Number of connected ports
		N_{SP}	Number of ports in sleep state (ports with no traffic running through them)
		N_{AP}	Number of active ports (ports with traffic running through them)
With EEE	For B > 4 kbps $E_S = P_B \times T_D + N_{AP} \times (P_P \times T_A) + N_{SP} \times (P_{LPI} \times T_S)$ (3)	E_S	Energy consumption of switch (Wh)
		B	Bandwidth (kbps)
	For B <= 4 kbps $E_S = (P_B + N_P \times P_{LPI}) \times T_D$ (4)	T_A	Total duration with traffic (h)
		T_S	Total duration without traffic (h)
		T_D	Total duration of experiment (h) $T_D = T_A + T_S$

In summary, results from power profiling the switch's ports firstly showcase that without EEE, the power consumption of 1 Gbits/s link on a 24-port switch is on average 0.33 W for each new link added. Results from power profiling the ports on the 48-port switch on the other hand, firstly showcase, that without EEE, the consumption of 1 Gbits/s link is on average 0.35 W for each new link added. A 1 Gbits/s link then consumes 3 times more than a 100 Mbits/s with the latter consuming slightly more than 10 Mbits/s. The total energy consumption for a NO EEE 48-port is less that two separate NO EEE 24-port switches. If both run continuously for 24 h per day, the energy gain is 32.30%. Several scenarios have been created for the deployment of EEE-enabled 24-port and 48-port switches. One of the results to be highlighted for a scenario in Table 13 (2× NO EEE 24-port switch—one run for 24 h while the other 16 h per day; and 2× EEE 24-port switch—same running conditions) is that the latter incurs an energy loss of 16.46% when compared with the former.

Table 11 Scenario 1—deployment of power (and energy) consumption models for the switches (NO EEE)

Scenario (1 Gbits/s)	Formula substitution	Measured values
Scenario 1.1 One NO EEE 24-port switch	Use Eq. 1 (Table 11) $P_{24} = (24 \times P_P) + P_B$ $= 24 \times 0.33 + 30.00$ W $= 37.92$ W	**Per port** P_P = 0.33 W (for 24-port, from Table 4) P_P = 0.35 W (for 48-port, from Table 5)
Scenario 1.2 Two NO EEE 24-port switches	$P_{24\times2} = 2 \times P_{24} = 2 \times 37.92$ W $= 75.84$ W $E_{24x2} = 75.84$ W $\times$ 24 h $= 1820.26$ Wh (per day)	**Per switch** P_B = 30.00 W (for 24-port from Sect. 5.1) P_B = 34.54 W (for 48-port from Sect. 5.2)
Scenario 1.3 One NO EEE 48-port switch	Use Eq. **1** (Table 11) $P_{48} = (48 \times P_P) + P_B$ $= 48 \times 0.35 + 34.54$ W $= 51.34$ W $E_{48} = 51.34$ W $\times$ 24(h) (per day) $= 1232.16$ Wh	N_P = 24 (for 24-port switch) N_P = 48 (for 48-port switch) T_D = 24 h (per day)
Conclusion 1 drawn for Scenarios 1.1–1.3	$P_{48} > P_{24}$, and the increase in power consumption is 35.39% $E_{48} < E_{24\times2,}$ therefore, it is more energy-efficient to use one 48-port switch then 2 separate 24-port switches and the energy gain is 32.30%	
Scenario 1.4 Analysis for NO EEE 24-port and 48-port switches	**Assumption** There are two 24-port and one 48-port switches. Based on conclusion 1, the energy consumption (per day) for a 48-port is less than two 48-ports. The goal of the following analysis is to investigate how many hours to sleep the 2× 48-port switches (per day) in order to bring about energy savings compared to 1× 48-port switch $2 \times P_{24} \times T_A <= E_{48}$ (from Scenario 1.3) $2 \times P_{24} \times (24 - T_S) <= E_{48}$ $T_S => 24 - [E_{48}/(2 \times P_{24})$ (substitute for E_{48} and $P_{24)}$ with values from Scenarios 1.1 and 1.3) T_S => 7.753 h per day	T_A = duration for an active port T_S = duration for a sleeping port
Conclusion 2 drawn for Scenario 1.4	The 2× 24-port switches will have to sleep for at least 7.753 h per day in order to be more energy-efficient than 1× 48-port switches (with NO EEE)	

Table 12 Scenario 2—deployment of power (and energy) consumption models for the switches (EEE and NO EEE)

Scenario (1 Gbits/s)	Formula substitution	Measured values
Scenario 2.1 Two NO EEE 24-port switches	**Scenario 2.1.1 assumption** One 24-port runs for 24 h while another runs for 16 h per day From Scenario 1.1, the power consumption of a NO EEE 24-port switch is: $P_{24} = 37.92$ W $T_{D1} = 24$ h and $T_{D2} = 16$ h $E_{24\times2} = (24 \times P_{24}) + (16 \times P_{24})$ $= (24 \times 37.92) + (16 \times 37.92)$Wh $= 1516.80$ Wh	Per port $P_P = 0.33$ W (for 24-port, from Table 4) $P_P = 0.35$ W (for 48-port, from Table 5) $P_{LPI} = 0.05$ W (for 24-port, from Sect. 5.1) $P_{LPI} = 0.04$ W (for 48-port, from Sect. 5.2)
	Scenario 2.1.2 assumption Both 24-port switches run for 16 h per day From Scenario 1.1, the power consumption of a NO EEE 24-port switch is: $P_{24} = 37.92$ W $T_{D2} = 16$ h $E_{24\times2} = 2 \times (16 \times P_{24})$ $= 1213.44$ Wh	Per switch $P_B = 30.00$ W (for 24-port from Sect. 5.1) $P_B = 34.54$ W (for 48-port from Sect. 5.2)
Scenario 2.2 One EEE 48-port switch	**Assumption** All 48 ports run continuously for 16 h (per day) Only 24 ports (out of the 48 ports) run continuously for the remaining 8 h while the other 24 ports are in sleeping mode	$N_P = 24$ (for 24-port switch) $N_P = 48$ (for 48-port switch) $N_{AP1} = 48$ $N_{AP2} = 24$ $N_{SP1} = 0$ $N_{SP2} = 24$
	Part 1 use Eq. 3 (see Table 11) and $T_{D2} = 16$ h $E_{48(D2)} = P_B \times T_{D2} + N_{AP1} \times (P_P \times T_{A1}) + N_{SP1} \times (P_{LPI} \times T_{S1})$ $E_{48(D2)} = 34.54 \times 16 + 48 \times (0.35 \times 16) + 0 \times (0.04 \times 0)$ $E_{48(D2)} = 821.44$ Wh **Part 2** use Eq. 3 (see Table 11) and $T_{D3} = 8$ h $E_{48(D3)} = P_B \times T_{D3} + N_{AP2} \times (P_P \times T_{A2}) + N_{SP2} \times (P_{LPI} \times T_{S2})$ $E_{48(D3)} = 34.54 \times 8 + 24 \times (0.35 \times 8) + 24 \times (0.04 \times 8)$ $E_{48(D3)} = 351.20$ Wh **Part 3** The total energy consumption per day	$T_{D1} = 24$ h (per day) $T_{D2} = 16$ h (per day) $T_{D3} = 8$ h (per day) $T_{A1} = 16$ h (per day) $T_{A2} = 8$ h (per day) $T_{S1} = 0$ h (per day) $T_{S2} = 8$ h (per day) $B > 4$ kbps

(continued)

Table 12 (continued)

Scenario (1 Gbits/s)	Formula substitution	Measured values
	$E_{48(Total)} = E_{48(D2)} + E_{48(D3)}$ $E_{48(Total)} = 821.44\text{ W} + 351.20\text{ W}$ $= 1172.64\text{ W}$	
Conclusion 3 drawn for Scenarios 2.1.1–2.2	Based on Scenario 2.1.1, $E_{48} < E_{24\times2,}$ therefore, it is more energy-efficient to use one 48-port switch than 2 separate 24-port switches so that some of the ports can go to sleep when there is reduced traffic. The energy gain for Scenario 2 is 22.68% Based on Scenario 2.1.2, $E_{48} < E_{24\times2,}$ therefore, it is more energy-efficient to use one 48-port switch than 2 separate 24-port switches for the same running condition of 16 h per day. The energy gain for Scenario 2 is 2.69%	

Table 13 Scenario 3—deployment of power (and energy) consumption models for the switches (EEE and NO EEE)

Scenario (1 Gbits/s)	Formula substitution	Measured values
Scenario 3.1 Two NO EEE 24-port switches	**Assumption** One 24-port switch runs for 24 h (per day) while the other runs for 16 h (per day). T_{D1} = 24 h and T_{D2} = 16 h Use the outcome of Scenario 1.1 $E_{24\times2}$ = 1516.80 Wh	Per port P_P = 0.33 W (for 24-port, from Table 4) P_P = 0.35 W (for 48-port, from Table 5) P_{LPI} = 0.05 W (for 24-port, from Sect. 5.1) P_{LPI} = 0.04 W (for 48-port, from Sect. 5.2) Per switch P_B = 30.00 W (for 24-port from Sect. 5.1) P_B = 34.54 W (for 48-port from Sect. 5.2)
Scenario 3.2 Two EEE 24-port switches	One 24-port switch runs for 24 h (per day) while the other 24-port switch runs for 16 h only (per day) with 8 h in sleep mode. T_{D1} = 24 h and T_{D2} = 16 h **Part 1** use Eq. 3 (see Table 11) and $T_{D1} = 24$ h (all 24 ports are active) $E_{24(D1)} = P_B \times T_{D1} + N_{AP1} \times (P_P \times T_{A1})$ $+ N_{SP1} \times (P_{LPI} \times T_{S1})$	N_P = 24 (for 24-port switch) N_P = 48 (for 48-port switch) N_{AP1} = 24 N_{AP2} = 24 N_{SP1} = 0 N_{SP2} = 0 N_{SP2} = 24

(continued)

Table 13 (continued)

Scenario (1 Gbits/s)	Formula substitution	Measured values
	$E_{24(D1)} = 30.00 \times 24 + 24 \times (0.33 \times 24) + 0 \times (0.05 \times 0)$ $E_{24(D1)} = 910.08$ Wh **Part 2** use Eq. 3 (see Table 11) and T_{D2} = 16 h (all 24 ports are active) $E_{24(D2)} = P_B \times T_{D2} + N_{AP2} \times (P_P \times T_{A2}) + N_{SP2} \times (P_{LPI} \times T_{S2})$ $E_{24(D2)} = 30.00 \times 16 + 24 \times (0.33 \times 16) + 0 \times (0.05 \times 0)$ $E_{24(D2)} = 606.72$ Wh **Part 3** use Eq. 4 (see Table 11) and T_{D3} = 8 h (all 24 ports are sleeping) $E_{24(D3)} = (P_B + N_P \times P_{LPI}) \times T_{D3}$ $= (30.0 + 24 \times 0.05) \times 8$ Wh $= 249.60$ Wh The total energy consumption per day $E_{24x2(Total)} = E_{24(D1)} + E_{24(D2)} + E_{24(D3)}$ $E_{24(Total)} = 910.08 \text{ Wh} + 606.72 \text{ Wh} + 249.60 \text{ Wh}$ $= 1776.40$ W	T_{D1} = 24 h (per day) T_{D2} = 16 h (per day) T_{D3} = 8 h (per day) T_{A1} = 24 h (per day) T_{A2} = 16 h (per day) T_{A3} = 0 h (per day) T_{S1} = 0 h (per day) T_{S2} = 0 h (per day) T_{S3} = 8 h (per day) B > 4 kbps (for Part 2) B <= 4 kbps (for Part 3)
Conclusion 4 drawn for Scenarios 3.1–3.2	$E_{24\times2}$ (With EEE) < $E_{24\times2}$ (NO EEE), therefore, it is more energy-efficient to use two 24-port switches (NO EEE) than 2 separate 24-port switches (With EEE) though one 24-port switch go to sleep for 8 h. The energy loss for Scenario 3 is 16.46%	

6 Conclusion

During our experimentation, we have explored the power (and energy) consumption of links in relation to their bandwidths and conclude that the power (and energy) consumption of ports will exponentially increase with the increase of speed. With the emerging 10 Gbits/s links, the expected power (and energy) efficiency of those links will surpass our measured values for 1 Gbits/s links. Our experimental measurements reveal that an EEE port runs at 12–15% of its total power when in low power mode. Therefore, the power savings (when no traffic is present) can constantly exceed 80%. However, our measurements equally show that the power (energy) consumption of a single port represents less than 1% of the total power (energy) consumption of the switch. The base power consumed by the switch without any port is still significantly high and is not affected by EEE. Experiment results also show that the base power consumption of switches does not significantly increase with the size of the switches. When doubling the size of the switch between 24 and 48 ports, we measure an increase of 35.39% in power (or energy) consumption. EEE also holds greater effect on bigger switches, with the power (or energy) gain on the EEE-enabled 48-port switch compared to 2× EEE-enabled 24-port switch. However, it seems to be more energy-efficient to use 2 separate 24-port switches (NO EEE) than 2 separate 24-port switches (With EEE).

The best use case for EEE is hence one where the equipment is very large and the traffic has a lot of downtime. According to statistics from Technavio [6], big data centers with link utilization rates of 30% and where the equipment is very large, presents a perfect customer for a mass adoption of EEE. However, office spaces and computer labs will not sufficiently receive energy savings to justify a mass adoption of the technology. The implementation of EEE on the other hand, has delivered protocol related performance. There is no noticeable delay in transitioning from Sleep to Wake and the protocol delivers small but noticeable energy savings over regular 1 Gbps. Consequently, significant interest in re-implementing these tests on link speeds of 10 Gbps and higher in order to assess potential benefits from EEE for such bandwidths. For future work, similar tests should be conducted for larger switches in order to test the benefits of deploying EEE in large data centers. Similarly, conducting similar tests on switches with different models and manufacturers will help generalize the results presented in this work.

Acknowledgements This work was supported by the Erasmus Mundus PERCCOM programme [30].

References

1. El Khoury, J.: Assessing the benefit of deploying EEE on commercial grade network switches. Unpublished PERCCOM Masters Dissertation, University of Lorraine, Nancy, France (2017)
2. Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.).: Green IT Engineering: Concepts, Models, Complex Systems Architectures. Studies in Systems, Decision and Control, vol. 74. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-44162-7
3. Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.). Green IT Engineering: Components, Networks and Systems Implementation. Studies in Systems, Decision and Control, vol. 105. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-55595-9
4. Meng, J., Ren, F., Jiang, W., et al.: Modeling and understanding burst transmission algorithms for energy efficient ethernet. In: IEEE/ACM 21st International Symposium on IEEE on Quality of Service (IWQoS), Montreal, QC, Canada, 3–4 June 2013. https://doi.org/10.1109/iwqos.2013.6550278
5. Brill, K.G. (2007). The Invisible Crisis in the Data Center: The Economic Meltdown of Moore's Law, White Paper, Uptime Institute. URL: http://www.mm4m.net/library/The_Invisible_Crisis_in_the_Data_Center.pdf. Accessed date 27 Oct 2017
6. Technavio. (2017). Global Data Center Construction Market 2017–2021. URL: https://www.technavio.com/report/global-data-center-global-data-center-construction-market-2017-2021. Accessed date 27 Oct 2017
7. Armbsrust, M., et al.: Above the Clouds: A Berkeley View of Cloud Computing, Technical Report UCB/EECS-2009-28, EECS Department, U.C. Berkeley, Feb 2009. URL: http://cacs.usc.edu/education/cs653/Armbrust-CloudComp-Berkeley09.pdf. Accessed date 27 Oct 2017

8. Seoane, I., Hernandez, J.A., Reviriego, P., Larrabeiti, D.: Energy-aware flow allocation algorithm for energy efficient ethernet networks. In: Proceedings of 19th International Conference on Software, Telecommunications and Computer Networks, SoftCOM, Split, Yugoslavia, 15–17 Sept 2011
9. Gunaratne, C., Christensen, K., Nordman, B., Suen, S.: Reducing the energy consumption of ethernet with adaptive link rate (ALR). IEEE Trans. Comput. **57**(4), 448–461 (2008). https://doi.org/10.1109/tc.2007.70836
10. IEEE. (2010). IEEE P802.3az Energy Efficient Ethernet Task Force. URL: http://www.ieee802.org/3/az/. Accessed date 27 Oct 2017
11. Christensen, K., et al.: IEEE 802.3az: the road to energy efficient ethernet. IEEE Commun. Mag. **48**(11), 50–56 (2010). https://doi.org/10.1109/mcom.2010.5621967
12. Marsan, M.A., et al.: A simple analytical model for energy efficient ethernet. IEEE Commun. Lett. **15**(7), 773–775 (2011). https://doi.org/10.1109/lcomm.2011.060111.110973
13. Herreria-Alonso, S., Rodrıguez-Perez, M., Fernandez-Veiga, M., Lopez-Garcıa, C.: Optimal configuration of energy-efficient ethernet. Elsevier Comput. Netw. **56**(10), 2456–2467 (2012). https://doi.org/10.1016/j.comnet.2012.03.006
14. Reviriego, P., et al.: Burst transmission for energy-efficient ethernet. IEEE Internet Comput. **14**(4), 50–57 (2010). https://doi.org/10.1109/mic.2010.52
15. Reviriego, P., et al.: An energy consumption model for energy efficient ethernet switches. In: Proceedings of IEEE International Conference on High Performance Computing and Simulation (HPCS), Madrid, Spain, 2–6 July 2012. 0.1109/HPCSim.2012.6266897
16. Paillassa, B., et al.: Performance evaluation of energy efficient policies for ethernet switches. In: Proceedings of IEEE International Wireless Communications and Mobile Computing Conference (IWCMC), Sardinia, Italy, 1–5 July 2013. https://doi.org/10.1109/iwcmc.2013.6583659
17. Popescu, I., et al.: Application-centric energy-efficient ethernet with quality of service support. Electron. Lett. **51**(15), 1165–1167 (2015). https://doi.org/10.1049/el.2015.0999
18. Saravanan, K.P., Carpenter, P.M., Ramirez, A.: Power/performance evaluation of energy efficient ethernet (EEE) for high performance computing. In: IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS), Austin, Texas, USA, 21–23 Apr 2013
19. Lee, S., Chen, A.: Design and analysis of a novel energy efficient ethernet passive optical network. In: Proceedings of the 9th International Conference on Networks (ICN), Menuires, France, 11–16 Apr 2010. https://doi.org/10.1109/icn.2010.9
20. Kubo, R., et al.: Study and demonstration of sleep and adaptive link rate control mechanisms for energy efficient 10G-EPON. J. Opt. Commun. Netw. **2**(9), 716–729 (2010). https://doi.org/10.1364/JOCN.2.000716
21. Reviriego, P., et al.: Study of the potential energy savings in ethernet by combining energy efficient ethernet and adaptive link rate. Trans. Emerg. Telecommun. Technol. **23**(3), 227–233 (2012). https://doi.org/10.1002/ett.1526
22. Jin, S., Fan, R., Yue, W.: A hybrid energy saving strategy with LPI and ALR for energy-efficient ethernet. In: Proceedings of IEEE 2nd International Conference on Computer Science and Network Technology, (ICCSNT), Changchun, China, 29–31 Dec 2012. https://doi.org/10.1109/iccsnt.2012.6525945
23. Rodríguez-Pérez, M., et al.: Optimum traffic allocation in bundled energy-efficient ethernet links. IEEE Syst. J. (99) (2015). https://doi.org/10.1109/jsyst.2015.2466086
24. Herreria-Alonso, S., et al.: Optimizing dual-mode EEE interfaces: deep-sleep is healthy. IEEE Trans. Commun. **65**(8). https://doi.org/10.1109/tcomm.2017.2700310
25. Larrabeiti, D., et al.: Towards an energy efficient 10 Gb/s optical ethernet: performance analysis and viability. Opt. Switching Netw. **8**(3), 131–138 (2011). https://doi.org/10.1016/j.osn.2011.03.009
26. Mostowfi, M., Christensen, K.: Saving energy in LAN switches: new methods of packet coalescing for energy efficient ethernet. In: Proceedings of IEEE IGCC Orlando, USA, 25–28 July 2011. https://doi.org/10.1109/igcc.2011.6008547

27. Aksić, M., Bjelica, M.: Packet coalescing strategies for energy-efficient ethernet. Electron. Lett. **50**(7), 521–523 (2014). https://doi.org/10.1049/el.2014.0386
28. Chakadkit, T., Jakllari, G., Paillassa, B.: Augmenting the energy-saving impact of IEEE 802.3 az via the control plane. In: Proceedings of IEEE International Conference on Communication Workshop, (ICCW), London, UK, 8–12 June 2015. https://doi.org/10.1109/iccw.2015.7247610
29. Hossain, M., Rondeau, E., Georges, J.-P., Bastogne, T.: Modeling the power consumption of ethernet switch. In: International SEEDS Conference 2015: Sustainable Ecological Engineering Design for Society, Leeds, UK, 17–18 Sept 2015
30. Klimova, A., Rondeau, E., Andersson, K., Porras, J., Rybin, A.V., Zaslavsky, A.: An international master's program in green ICT as a contribution to sustainable development. J. Cleaner Prod. **135**, 223–239 (2016). https://doi.org/10.1016/j.jclepro.2016.06.032

Mobile Phones and Energy Consumption

Cristea Vlad Vasile, Colin Pattinson and Ah-Lian Kor

Abstract This paper is an extension of (Cristea et al. in International SEEDS conference, 2015) [1] and abstracted from (Cristea in Energy consumption of mobile phones, 2015) [2]. This research contributes to the potential of greening software application as discussed in (Kharchenko et al. in Green IT engineering: concepts, models, complex systems architectures, vol 74, 2017a) [3] and (Kharchenko et al. in Green IT engineering: components, networks and systems implementation, vol 105, 2017b) [4]. Additionally, green design principles abstracted from this research will be relevant for designing the systems in (Kondratenko et al. in Green IT Engineering: Components, Networks and Systems Implementation, 2017) [5] and (Kuchuk et al. in Green IT Engineering: Components, Networks and Systems Implementation, 2017) [6]. Battery consumption in mobile applications development is a very important aspect and has to be considered by all the developers in their applications. This study will present an analysis of different relevant concepts and parameters that may have impact on energy consumption of Windows Phone applications. This operating system was chosen because there is limited research even though there are related studies for Android and iOS operating systems. Furthermore, another reason is the increasing number of Windows Phone users. The objective of this research is to categorize the energy consumption parameters (e.g. use of one thread or several thread for the same output). The result for each group of experiments will be analyzed and a rule will be derived. The set of derived rules will serve as a guide for developers who intend to develop energy-efficient Windows Phone applications. For each experiment, one application is created for each concept and the results are presented in two ways: a table and a chart. The table presents the duration of the experiment, the battery consumed in the experiment, the expected

C. V. Vasile
Garmin, Cluj-Napoca, Romania

C. Pattinson · A.-L. Kor (✉)
School of Computing, Creative Technologies, and Engineering, Leeds Beckett University, Leeds, UK
e-mail: A.Kor@leedsbeckett.ac.uk

C. Pattinson
e-mail: C.Pattinson@leedsbeckett.ac.uk

© Springer Nature Switzerland AG 2019
V. Kharchenko et al. (eds.), *Green IT Engineering: Social, Business and Industrial Applications*, Studies in Systems, Decision and Control 171,
https://doi.org/10.1007/978-3-030-00253-4_11

battery lifetime and the energy consumption, while the charts display the energy distribution based on the main threads: UI thread, application thread and network thread.

Keywords Energy efficiency · Mobile phone · Smartphone energy consumption Rules

1 Introduction

In recent years, the smartphones market has a significant boost. According to eMarketer, the number of smartphone users has grown from 1.13 billion in 2012 to 2.03 billion in 2015 [7]. This ascending trend yields a prediction of around 2.5 billion smartphone users by 2017. This means that around 30% from the world's population will own such a device. There are two dominant operating systems that run on these smartphones: iOS and Android. According to the same source, in the last quarter of 2014 the percentage of smartphones which support Android is 76.6%, while the smartphones which support iOS is represented by only 19.7%. The rest of 3.7% is split between Windows Phone operating system with 2.8%, BlackBerry operating system with 0.4% and others operating systems.

Although the difference between the first two operating systems and the rest is large, in the future these statistics will change. Statistica portal predicts that operating system market in 2017 will look like this: the Android market will decrease to a value around 68.3%, the iOS market will decrease to a value around 17.9% and the Windows Phone market will increase up to 10.2%. These data suggest the fact that Windows Phone operating system is in continual development and in the future it could be a competitor for Android and iOS operating systems.

According to Statistica portal, in October 2014 [8] there is a number of 1.3 million applications in App Store, 1.3 million applications in Google Play and only around 300,000 applications in Windows Store. TheNextWeb.com presents an article [9] in which a spokesperson from Microsoft confirms that the number of application from Windows Store reached 300,000 in June 2014 and the fact that "in the past year alone the Windows and Windows Phone app catalog has grown 94%, while the number of active developers has grown by 50%.". According to the newest statistics from Microsoft [10], in March 2015, there is a number of 585,000 applications in Windows Store. It is noted that the increasing rate of application development is very high, thus promoting Windows Store to become a competitor for App Store and Google Play. This is the reason for conducting experiments for Windows Phone in this study and to conduct detailed analysis of the concepts and controls used by Windows Phone developers.

According to Smart2020 report [11] information technology and communication (ICT) consumes around 2% of the world's energy, however, the ICT sector's emissions 'footprint' is expected to decrease to 1.97% of the global emissions by 2030 [12]. This number can be compared to the total energy consumed by airline industry. In 2020 the mobile phones will represent 1% from the ICT carbon footprint and the mobile network will represent 13%. It is very difficult to calculate very precisely the energy consumed by a smartphone, because this is not only an object used for communication. When a user charges his phone every day or maybe two times per day the total amount of energy consumed by a smartphone will become considerable. Another important factor that should be considered when the energy consumption is calculated, is the whole internet infrastructure. Nowadays the data generated by smartphones and transferred across the internet is significant and it grows continually, because the number of users that access the internet through a smartphone is in an upward trend.

The aim of this study is to compare concepts and controls that are used for developing Windows Phone applications, and to establish a set of rules that can be used by any developer who intends to build an energy efficiency application. There will be a predefined number of rules that will be tested and which will encompass the following components: UI, processing, and network.

This study makes the following contributions: it investigates the energy consumption of Nokia smartphones running on Windows Phone 8.1 operating system; it investigates the energy consumption of specific Windows Phone controls; it investigates the energy consumption of specific programming concepts; it provides a set of rules, which will optimize a mobile application from energy point of view.

2 Related Work

Smartphones' energy efficiency research is gathering momentum and it is growing in parallel with the development of the smartphones. Nowadays there are many components like processor or screen that can be optimized, but the battery is not one of them yet. This is why it is very important to have control over the battery and to know exactly which part of the application consumes more energy and why.

Related studies to this paper address the following issues: tools that measure energy consumption [13–15] and provide breakdown of energy consumption the mobile device's main hardware components [16]; overall consumption [17–20]; cloud services [21]; and network measurement [21–23] and power consumed whilst using LTE as well as WiFi network connections [24]. An analysis of the characteristics of the power consumption for context-aware mobile applications has been conducted to form the basis for energy-efficient context-aware services in mobile environments [25] while [26] develop a battery behavior model to explore the effects (i.e. energy consumption and batter drain) of different usage patterns in sensory operations for context-aware smartphones. An investigation on the energy consumption of wireless communication interfaces (e.g. Bluetooth, WiFi, and 3G) during various scenarios (e.g. standby, scanning, and transferring) [27]. Moreover, there are some studies that attempt to improve the battery life. One of these studies is investigated by Parkkila and Porras [28]. The mobile phones field is not the only one where researchers are trying to find some "green" optimizations. Networking is another area of research where a lot of optimizations are made. An example in this category the research by Drouant et al. [29] and Pattinson and Robinson [30].

Notably, most of the studies focus on the hardware components (e.g. optimization of power consumption in multi-core smart phones through power-aware scheduling algorithms, [31]) or on the network (e.g. power consumption analysis of transmission over 3G networks, [32]). Elliott et al. [33] has conducted an investigation of battery and energy consumption of media-related applications in Android smartphones. To date, there is limited detailed analysis of the energy consumption of software components. All of the studies are platform independent, so they can be made for Android, iOS or Windows Phone. For example, one study presents the energy consumption of a display in general but not the factors that influence this consumption. The research in this paper address this identified gap. It attempts to go a step deeper and to analyze different factors that can influence the energy consumption of a mobile application. From [17] work, it is known the fact that the display component is one of the components that consumes the most energy in an application. What is not known is the underlying cause of this phenomenon and how to improve the energy consumption. The purpose of this paper is to identify a part of the element that consumes most of the energy.

3 Methodology

As it was already mentioned in the Introduction Section, the purpose of this research is to provide a set of rules that can be used by developers in order to obtain mobile applications that are more energy-efficient. Nowadays, there are a lot of operating systems for smartphones, such as: Android, iOS, Windows Phone or Jolla. Each of these operating systems has many particularities, so it is very difficult to obtain a set of rules that can be applied to all operating systems. This study will focus only on one specific operating system, Windows Phone 8.1, a product of Microsoft Company released in April 2014 [34].

3.1 Tools

Applications for Windows Phone 8.1 can be developed using Microsoft Visual Studio 2013 [35]. This software is an IDE (integrated development environment) from Microsoft. It can be used for developing desktop applications, websites, web services, Windows applications and mobile applications. Microsoft Visual Studio 2013 includes the following programming languages: C, C++, VB.NET (Visual Basic), C# and F#. Besides Visual Studio, another tool is required in the development process: Windows Phone 8.1 SDK. This tool installs everything that is necessary for developing and testing Windows Phone applications. For the User Interface part, each application can be opened in Microsoft Blend, which is a specialized tool in UI design. Figure 1 presents a basic Windows Phone application in Visual Studio 2012.

The main components [36] that can be found in Visual Studio for Windows Phone are:

- ***Toolbox***—contains a list with all the controls that can be found in the basic installation. Extra components can be added to the project if they are referenced from the solution and from the current page;
- ***Design View***—shows the design of the application. The controls from Toolbox can be dragged directly to the design view and the XAML code will be automatically updated;
- ***XAML View***—shows the code that is generated for the interface. After each modification, the Design view part will be refreshed;
- ***Properties Windows***—offers the possibility to see and to modify the properties of different controls or files;
- ***Solution Explored***—shows all the projects and files that are included in the current solution, in a hierarchical structure;
- ***Target Device***—offers the possibility to choose the device on which the application will run. This device can be a virtual emulator or a real device. The

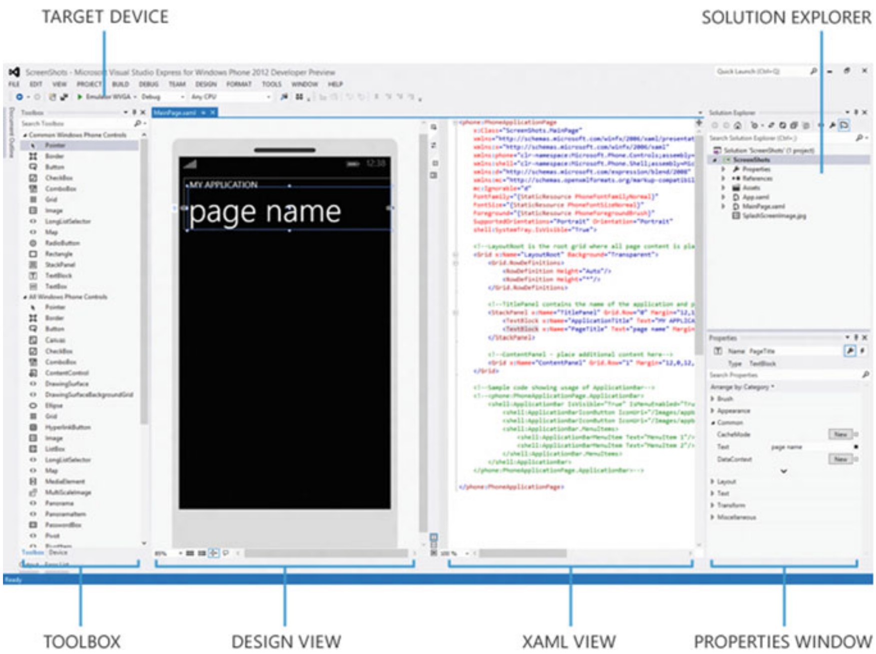


Fig. 1 Visual Studio 2012 for Windows Phone [34]

virtual emulator is a desktop application that offers the possibility to simulate a real environment for an application. The emulator is configurable and can simulate any real device, in terms of hardware and software components.

3.2 Windows Phone Application Analysis Tool

Another tool that is really useful is the Windows Phone Application Analysis tool and the interface is depicted in Fig. 2. This tool is used for monitoring and profiling an application:

- ***Profiling***—evaluate either execution-related or memory-usage aspects of a mobile application;
- ***Monitoring***—evaluate the behavior of the application.

The output generated by this tool can be general (see Fig. 3) or in detail (see Fig. 4). The general output is a summary of all parameters that are measured while the detailed output contains graphs that present the application during the execution time.

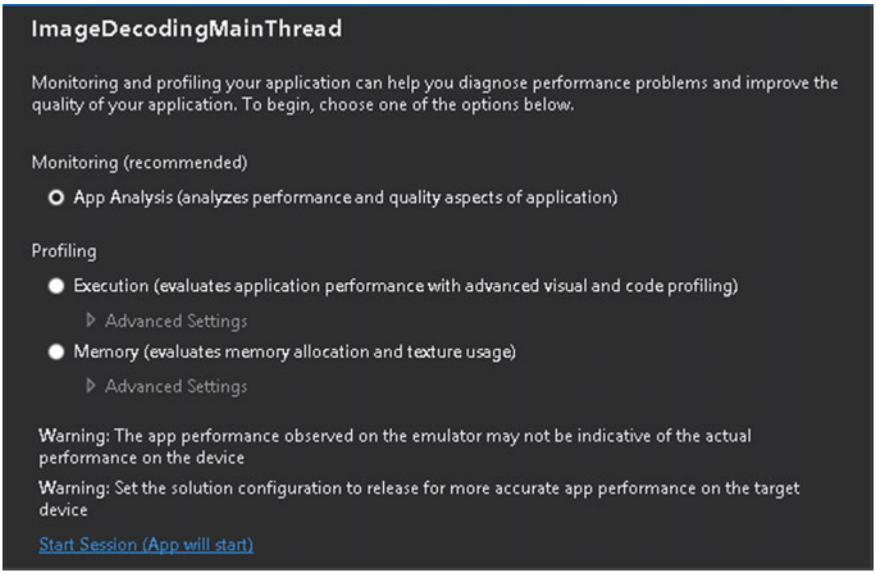


Fig. 2 Windows Phone Application Analysis tool interface

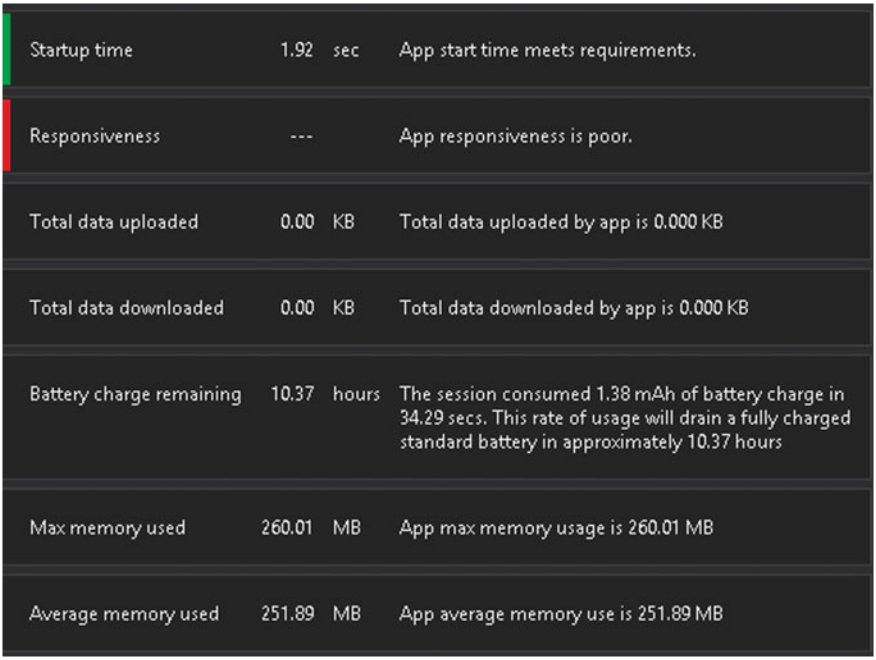


Fig. 3 Windows Phone Application Analysis tool—general output

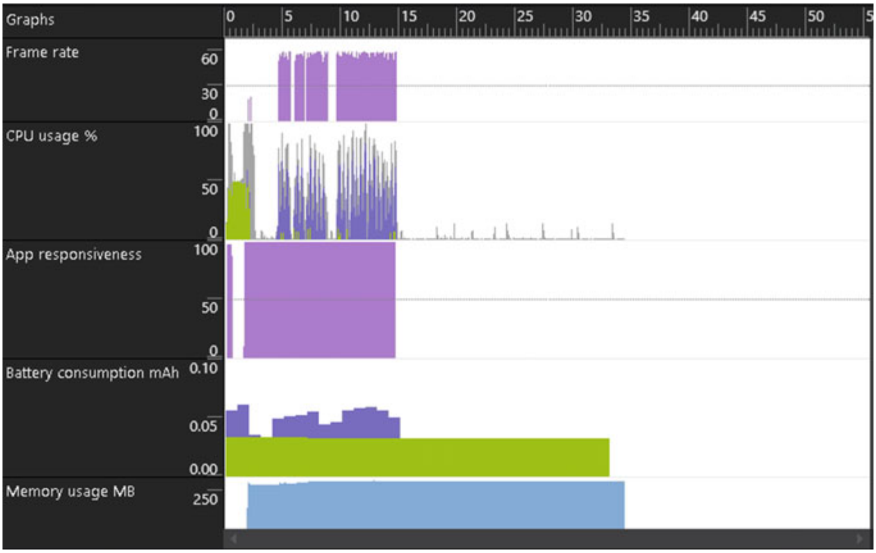


Fig. 4 Windows Phone Application Analysis tool—detailed output

3.3 Microsoft Expression Design 4

The last tool used for this thesis is Microsoft Expression Design 4, which is specialized in graphic design. It is used for complex objects that can be exported in different formats, like: XAML format or PNG format.

3.4 Experiments

The set of rules obtained is based on some common concepts that are used in programming or on the improvements that Microsoft brought into Windows Phone SDK. Oren Nachman, developer for Microsoft, said in one of his talks entitled "Windows Phone 8: Performance and Optimization for Developers" [37] that the performance of an application can be measured in "feelings". This means that a user who uses an application feels that the application is fast, that every action is processed immediately, that scrolling through pictures will not block the application and that navigating through pages is really smooth. This is the reason developers are focusing a lot on these aspects and try to optimize them. Also, the tools that are used by developers offer new controls that should be faster, more responsive and consume less memory. One aspect that is not always taken into consideration when a mobile application or a new control is developed is the battery consumption. The method chosen for this research is an experimental method. According to Oxford dictionary, an experiment is "a scientific procedure undertaken to make a discovery, test a hypothesis or demonstrates a known fact". This method is the most suitable for our research because currently only assumptions are made about whether the new controls are more efficient than the old ones, or whether one concept is more efficient than another one.

Experiments: On Application-Related Components

The main criterion that is applied in the selection of the elements that constitute the experiments is the diversity in terms of application's components. It is very important to have at least one element from each component of a mobile application tested.

The basic structure of a mobile application contains three components:

- **Frontend component** or the User Interface—it refers to the controls that are displayed to the user;
- **Backend component**—it refers to all the processing made by an application: data processing, command handlers and services connections;
- **Web services component**—it refers to all the services that are stored on servers, and which expose the Create/Read/Update/Delete functionality.

Accordingly, we can group the elements listed above in the following three groups (Table 1).

Table 1 Experimental elements

Frontend components	VirtualizedStackPanel [38], StackPanel [39], ListBox [40], LongListSelector [41], ProgressBar [42], Opacity [43], Visibility [44], Storyboard [45], Image background creation, background property [46]
Backend components	Assembly, recursive function, iterative function, page constructor, onNavigatedTo event for [47], Thread, multithread [48], while [49], base64 string format [50], Image build action [51], synchronous loading, asynchronous loading, image decoding [52], image format: PNG [53], JPG [54], XAML [55]
Web services components	Clouds [56]

Hypotheses

After the decision has been made on experiments in this research, the next step is to identify the hypothesis. Due to the fact that the controls and concepts that are to be tested, are used in different contexts, it is impossible to have only one hypothesis. For this reason, the components are grouped based on their functionalities and followed by the formulation of a hypothesis for each group. Based on these groups a number of 25 hypothesis have been derived, tested and discussed in this paper. The hypotheses are presented in Table 2.

For each of these experiments one or two applications are created and executed. These applications are executed several times and an average value is shown as the final result. For collecting the results the Windows Phone Application Analysis software is used. The data collected are: battery charge remaining, the execution time and the battery consumption. After obtaining the battery consumption, the energy consumption is calculated using the following formula:

$$E = QV$$

where E is energy (Wh), Q is charge (Ah), and V is Voltage (V).

The value for voltage depends on the phone that we are using. Consequently, the voltage for a specific phone: Nokia Lumia 1320 is 3.7 V.

Experiment Configurations

The experiments for this study are device dependent. This means that the collected results are specific for a device. The configurations that are used for the experiments can be found in the following Table 3.

As it can be noted in the above table the only phone dependent values are: the battery and screen resolution. This means that we should obtain different numbers for different emulators, but the rules obtained are universal (i.e. can be applied to any device). Three threads are being measured: UI thread, application thread and network thread. The UI thread phone-dependent because it is dependent on the resolution screen. The battery properties are important for the transformation of battery consumption to energy consumption. Since the battery is the same type for a specific device it does not influence the final result. All the experiments are tested on an emulator. The interface of the emulator looks depicted in Fig. 5.

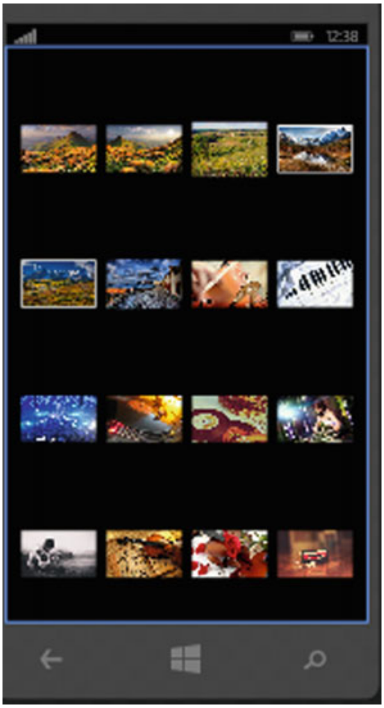
Table 2 Hypotheses for the experiments

No.	Hypotheses
1.	The darker colors used as background for a mobile application consume less energy than the brighter ones
2.	A JPG file format consumes less energy than a PNG file format in a mobile application
3.	Storing a visual object as image consumes less energy than storing the same object as XAML
4.	Using background threads consumes less energy than using the UI thread
5.	A static object consumes less energy than an animated object
6.	Using image decoder to size consumes less energy than using the default decoder
7.	Using asynchronous methods consumes less energy than using synchronous methods
8.	Using "Visibility" property consumes less energy than using "Opacity" property
9.	Using a determinate progress bar consumes less energy than using an indeterminate progress bar
10.	Using a "LongListSelector" control consumes less energy than using a "ListBox" control
11.	Setting "Build type" property to "Resource" for an image, consumes less energy than setting the same property to "Content"
12.	Storing a set of images in JPG format consumes less energy than storing the same images as base64 format
13.	A "for" loop consumes less energy than a "while" loop
14.	Using several threads to complete an operation consume less energy than using one thread to complete the same operation
15.	Executing a heavy processing operation in constructor consumes less energy than executing the same operation in "OnNavigateTo" event
16.	Using an iterative function consumes less energy than using a recursive function
17.	Using a "StackPanel" control consumes less energy than using a "VirtualizingStackPanel" control
18.	Using one assembly, for storing the resources, consumes less energy than using several assemblies
19.	An animated object that is created in the XAML file consumes less energy than an animated object that is created in procedural code
20.	An image stored locally consumes less energy than an image stored in the clouds
21.	A video file stored locally consumes less energy than an image stored in the clouds
22.	An audio file stored locally consumes less energy than an image stored in the clouds
23.	A JPG file format stored in clouds consumes less energy than a PNG file format stored in clouds
24.	Downloading an image and access it locally consumes less energy than accessing the picture multiple times in clouds
25.	Processing an operation locally consumes less energy than processing the same operation in clouds

Table 3 Device configuration

Property	Value
Battery voltage	3.8 V
Nominal voltage	3.7 V
Battery type	BV-4BW
Emulator type	720p
Emulator resolution	1280 × 720
Brightness	100%

Fig. 5 Application snapshot



4 Results and Discussion

For each experiment there are two types of output: first output is a table which presents the duration of the experiment, the battery consumption, the energy consumption and an estimated value of the remaining battery life. The second output is a graph, which presents the distribution of battery consumption based on the main threads: UI thread, application thread and network thread. In order to obtain a result, several executions of the same experiment are made. This paper presents details of several experiments and a set of rules obtained.

4.1 Visual Object Storing

Aim To investigate the impact of storing a visual object as Extensible Application Markup Language (XAML) and as image on energy consumption (Table 4).
Based on the results of these experiments it can be concluded that it is more efficient to work with images than with XAML objects. The difference is not very big in terms of energy consumption, but if millions of applications that display images are considered, this can be a considerable improvement. Also from the user's experience point of view, it is a big improvement considering the fact that the battery will last longer. This difference occurs because when using XAML, the application will create an object for each tag and this can load the processor more, while in the case of image files the processor has to render an image that is stored locally and this will happen faster. For more complex objects the difference will grow. In Figs. 6 and 7 it is noted that the energy consumed by the UI thread (green color) is the same in both cases. The only noticeable difference is the energy consumed by the application thread. In this case, it can be concluded that more energy is required for creating the XAML object than to decode a picture.

4.2 Control Hiding

Aim To investigate the impact of "visibility" property and "opacity" property on energy consumption (Table 5).
Both the applications executed are the same, it is observed that the energy consumption is different. The difference is 0.07 mAh, which happens because the Opacity property will keep the rectangles in memory, in order to improve the speed of the application. Even though the application with enabled Opacity is faster, it costs more in terms of energy consumption. In the first graph (Fig. 8) it is noted that the energy consumption of the UI thread is lower because the objects are deleted. In the second case (Fig. 9), even if the objects cannot be seen on the screen, they are stored in memory so more energy is consumed. From Figs. 8 and 9 some interesting facts have been observed. The energy consumed by the application thread (purple color) is similar in both cases. There are small differences, but not significant ones. The energy difference that appears in this experiment is related to the UI thread

Table 4 Visual object storing—energy consumption

	Time (s)	Battery consumption (mAh)	Battery charge remaining (h)	Energy consumption (Wh)
XAML format	10.50	0.28	15.90	0.001036
PNG format	10.34	0.25	16.41	0.000925

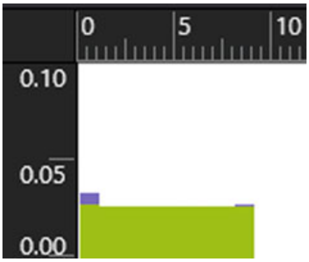


Fig. 6 XAML format

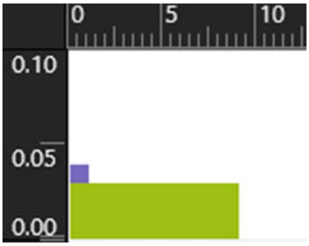


Fig. 7 PNG format

Table 5 Control hiding—energy consumption

	Time (s)	Battery consumption (mAh)	Battery charge remaining (h)	Energy consumption (Wh)
Visibility	20.71	1.26	6.83	0.004662
Opacity	20.63	1.33	6.44	0.004921

(green color). It can be seen in Fig. 8 that the UI thread consumes less energy while the objects are hidden. If the Opacity property is set, the energy consumed by the UI thread does not drop like in the previous case.

4.3 *Progress Bar Consumption*

Aim To investigate the energy efficiency of a determinate progress bar and an indeterminate progress bar (Table 6).

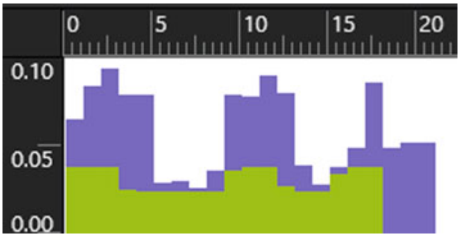


Fig. 8 Visibility property

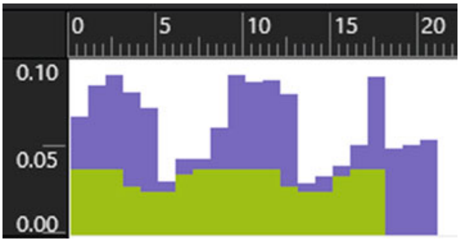


Fig. 9 Opacity property

Table 6 Progress bar—energy consumption

	Time (s)	Battery consumption (mAh)	Battery charge remaining (h)	Energy consumption (Wh)
Determinate	15.68	0.37	17.57	0.001369
Indeterminate	15.46	0.42	15.24	0.001554

As it can be seen from the charts above the determinate progress bar is more energy-efficient than the indeterminate one. This happens because the indeterminate bar is an animation which is shown all the time and which requires some processing. The determinate progress bar is based on a value so it does not require any repetitive pattern. This fact can be noticed in Figs. 10 and 11. The application thread (purple color) consumes more energy for an indeterminate progress bar because it supports the animation during the execution. In Fig. 10 it can be seen that it is required energy only when the application is launched. The UI thread (green color) consumes the same amount of energy in both cases. These controls can be used in different cases, but the determinate one ought to the preferred option.

4.4 Image Format

Aim To investigate the impact of displaying a set of images that is in a JPG (Joint Photographic Experts Group) format or in a base64 string format (Table 7).

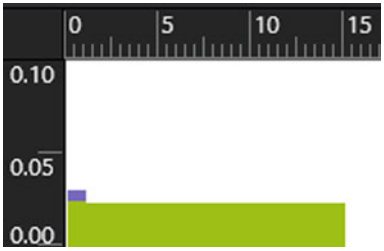


Fig. 10 Determinate progress bar

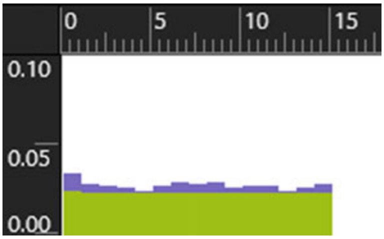


Fig. 11 Indeterminate progress bar

Table 7 Image format—energy consumption

	Time (s)	Battery consumption (mAh)	Battery charge remaining (h)	Energy consumption (Wh)
JPG	11.68	0.30	15.99	0.00111
Base64	11.30	0.30	15.90	0.00111

The battery consumption is equal in the both cases considered above, so it is not relevant if images are kept as JGP or as strings. Although the battery consumption is equal, it can noticed the fact that the distribution of application thread is different. In Fig. 12, it can be seen that it requires a lot of energy for computation (purple color) at the beginning, but after that it drops significantly. In the second case, it can be seen that the time for all the computation is longer. The energy consumed by UI thread (green color) is similar in both cases (Fig. 13).

4.5 Loop Instructions

Aim To investigate the energy efficiency of two loops instructions: for and while. Based on the results in Table 8, there is no difference between these two commands. This happens because, as previously mentioned, the only difference between the two instructions is the syntax. From Figs. 14 and 15 it can be seen that the energy consumption distribution of both UI thread (green color) and application thread (purple color) are the same in both cases.

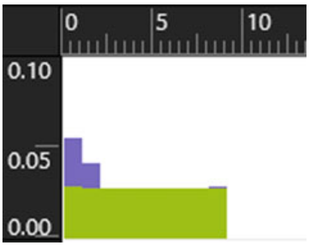


Fig. 12 Base64 format

Fig. 13 JPG format

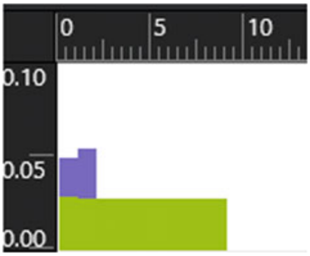


Table 8 Loop instruction—energy consumption

	Time (s)	Battery consumption (mAh)	Battery charge remaining (h)	Energy consumption (Wh)
For	21.67	0.56	16.10	0.002072
While	21.73	0.56	16.12	0.002072

Fig. 14 For loop

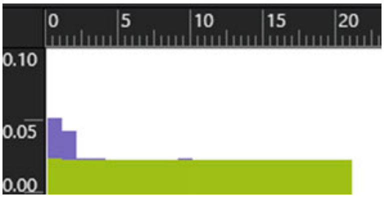
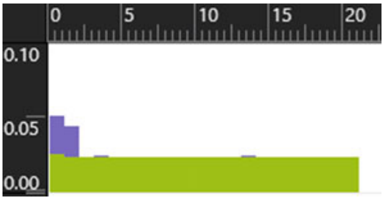


Fig. 15 While loop



4.6 *Threads*

Aim To investigate the energy efficiency of an application that uses one thread and of an application that uses more threads (Table 9).

Table 9 Threads—energy consumption

	Time (s)	Battery consumption (mAh)	Battery charge remaining (h)	Energy consumption (Wh)
Single thread	53.33	1.98	11.23	0.007326
Multithread	52.32	1.26	16.58	0.004662

As it can be seen from the charts above, the difference between the two approaches is significant. From Figs. 16 and 17 it can be noticed that the energy used by the UI thread (green color) is the same in both cases. There is a big difference in the application thread (purple color). For the single thread, it can be observed that it has required a lot of time to calculate all the numbers, which means a lot of energy wasted because the CPU is working. In the second case, the energy consumed by the application is very small because all the computations are done during the same time, in different threads. In the first case, 25 s are needed for processing while in the second case the results are shown immediately (Figs. 18 and 19).

4.7 *Function Type*

Aim To investigate the energy efficiency of an application that uses an iterative function compared to an application that uses a recursive function (Table 10).

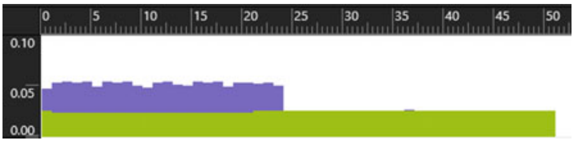
The application that uses an iterative function is more efficient according to the graphs above. It is noticed that the recursive function requires more time to compute and it also consumes more energy (purple color). Moreover the user has to wait until all the results are loaded before the application can be used. In the case of the iterative function the amount of energy that is required is very low. Furthermore, it is noticed that in this case, the application is faster due to the fact that the thread is busy for a shorter span. The energy consumed by the UI thread (green color) is similar in both cases.

4.8 *Storing Images*

Aim To investigate the impact of displaying a set of images that is stored locally in comparison with a set of images that are stored in a web page (Table 11).

Loading images from different sources has a big impact on the total energy consumed by a mobile application. The application that stores the images locally consumes less energy than an application that requests the images from a web page. From Figs. 20 and 21 it can be noticed the fact that the UI thread (green) and the CPU thread (purple) consume the same amount of energy in both applications. The difference between the applications is made by the network (gray): the experiment

Fig. 16 Single thread



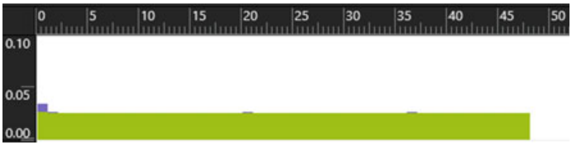


Fig. 17 Multithreading

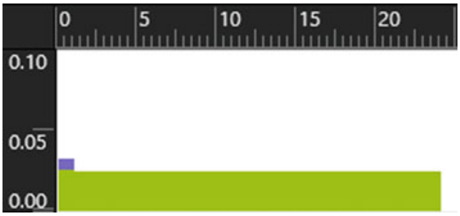


Fig. 18 Iterative function

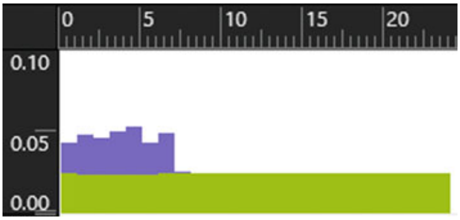


Fig. 19 Recursive function

Table 10 Function type—energy consumption

	Time (s)	Battery consumption (mAh)	Battery charge remaining (h)	Energy consumption (Wh)
Iterative	25.28	0.61	17.29	0.002257
Recursive	26.73	0.77	14.55	0.002849

Table 11 Storing images—energy consumption

	Time (s)	Battery consumption (mAh)	Battery charge remaining (h)	Energy consumption (Wh)
From internet	21.96	0.92	10.00	0.003404
Stored locally	21.43	0.69	13.00	0.002553

presented in Fig. 20 shows there is no energy consumed by the network while the one in Fig. 21 shows a significant amount of energy that is consumed by the network.

Fig. 20 Image stored locally

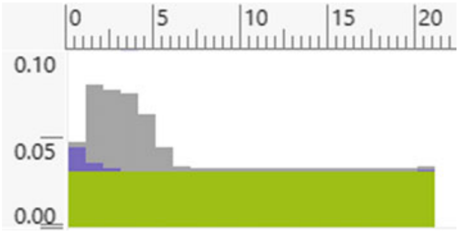


Fig. 21 Image downloaded

4.9 Image Format (JPG and PNG) in Clouds

Aim To investigate the impact of displaying a PNG (Portable Network Graphics) file format and a Joint Photographic Experts Group (JPG) file format, that is stored on a web page, on energy consumption.

This experiment reveals the fact that working with JPG format is "greener" than working with PNG format, if the images are stored on a website. From Table 12, it can noticed that the difference between these two formats is significant. In Figs. 22 and 23, it can be observed that the difference in the consumed energy is made by the network thread (gray). The UI thread (green) and the application thread (purple) have similar values. The distribution of the energy consumed by these two threads is also similar. The energy consumed by the network thread differs because of the images' file sizes. After the transformation from JPG in PNG (using http://image.online-convert.com/convert-to-png website), the files stored as PNG have a bigger size than the JPG files, and that is why the application that displays the PNG files consumes more energy.

Table 12 Image format in clouds—energy consumption

	Time (s)	Battery consumption (mAh)	Battery charge remaining (h)	Energy consumption (Wh)
JPG	21.01	0.74	11.90	0.002738
PNG	25.36	1.09	9.67	0.004033

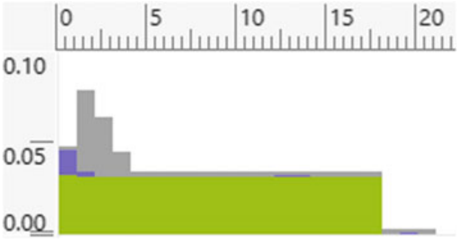


Fig. 22 JPG file format

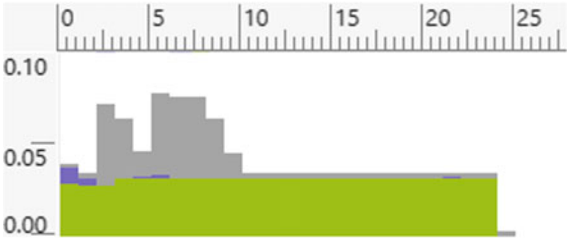


Fig. 23 PNG file format

4.10 Images—Multiple Access

Aim To investigate the impact on energy consumption of displaying multiple times the same picture from a web sites and the impact on energy consumption of downloading a picture and displaying it from a local source (Table 13).

From this experiment it can be noticed that the application which displays the images without saving them consumes less energy than the application which first downloads the picture. Figures 24 and 25 show that the energy consumed by the UI

Table 13 Multiple access—energy consumption

	Time (s)	Battery consumption (mAh)	Battery charge remaining (h)	Energy consumption (Wh)
From the same URL	31.28	0.96	13.54	0.003552
Download and display locally	31.19	1.02	12.74	0.003774

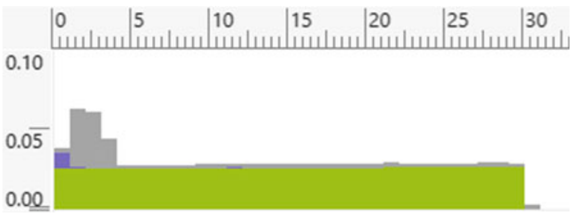


Fig. 24 Application that display the image from the same URL

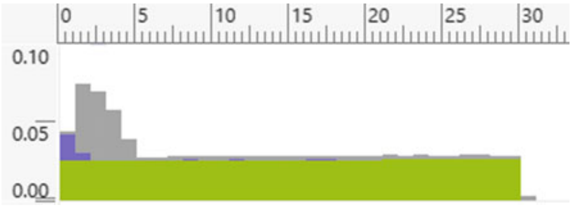


Fig. 25 Application that save and display the images

thread (green) is similar in both cases. The energy consumed by the application thread (purple) differs in these cases because it requires extra processing for saving the picture. The network thread (gray) consumes, also, less energy in the first case. Another fact that can be noticed is that each application makes a single request for the picture. In the first application this happens because of the cache mechanism that is implemented by default in Windows Phone 8. In the second case there is one request because we are downloading the image and using it after that from a local source.

4.11 Heavy Processing Operation

Aim To compare the impact on energy consumption of an operation that is run locally to an operation that is run in clouds.

The execution of some operations can significantly influence the energy consumption of an application. It can be seen in this experiment that executing some operations locally can save a lot of energy. From Table 14 it is noted that the difference between these two applications is significant. Figures 26 and 27 show that the UI thread (green) consumes the same amount of energy in both cases. In Fig. 27, the application thread (purple) request some energy only at the beginning while processing the data. For the other application the application thread consumes

Table 14 Heavy processing operation—energy consumption

	Time (s)	Battery consumption (mAh)	Battery charge remaining (h)	Energy consumption (Wh)
Cloud	42.34	1.73	10.23	0.006401
Locally	40.08	1.02	16.41	0.003774

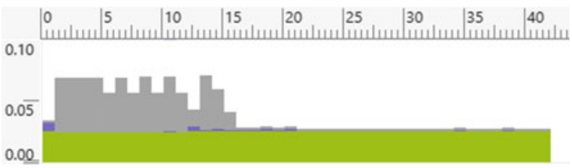


Fig. 26 Cloud processing

Fig. 27 Local processing

energy during the execution of the application because the data received from server has to be processed. The network thread (gray) shows a difference between these two applications, because in the first case there is a significant amount of energy consumed by this thread, while in the second case, the energy consumed by the network thread is 0.

4.12 Decoding Threads

Aim To investigate the impact of displaying images using backgrounds threads and using the UI thread on energy consumption (Figs. 28 and 29).

This experiment shows that the energy consumed by these two applications is different. From Table 15, it is noted that decoding an image in a separate thread is more efficient than using only one thread. Regarding the energy distribution, it can be seen that UI thread (green color) generates the same amount of energy in both cases while the application thread (purple color) generates less energy when using background threads. Another fact that can be noticed in the charts is the processing time. In the first case the application thread is working for 15 s while in the second case the application thread is working for 7 s. This happens because when using more than one thread, the tasks are executed in a parallel way. When all the processing is made by one thread it takes more time to decode all the pictures.

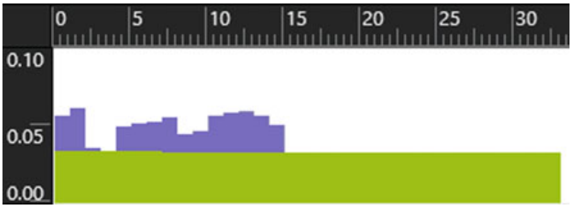


Fig. 28 CreateOption attribute set to BackgroundCreation

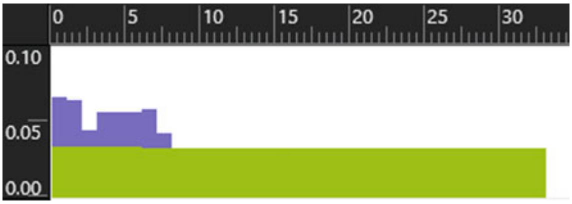


Fig. 29 Without CreateOption attribute set

Table 15 Decoding threads—energy consumption

	Time (s)	Battery consumption (mAh)	Battery charge remaining (h)	Energy consumption (Wh)
With CreateOption attribute	33.49	1.27	10.96	0.004699
Without CreateOption attribute	34.19	1.38	10.37	0.005106

4.13 Animations

Aim To investigate the energy efficiency of an application that displays an animation created in XAML file compared to an application that displays an animation created in procedural code (Table 16).

The charts above show that the energy consumption of the two applications is the same. This happens because the animation is the same in the both cases. Consequently the energy consumed is equal. It is noted that running the animation in the composition thread or in the UI thread gives the same effect. It might be possible to find some differences if the UI thread is overloaded. From Figs. 30 and 31 it is observed that both the UI thread (green color) and application thread (purple color) have a similar distribution of the consumed energy and of the amount of energy consumed.

From the total number of 25 experiments [1, 2], the assumed hypothesis was true in 14 cases. The hypothesis is not relevant in 5 experiments and it is false in 4 cases.

Table 16 Animations—energy consumption

	Time (s)	Battery consumption (mAh)	Battery charge remaining (h)	Energy consumption (Wh)
XAML	10.80	0.29	15.51	0.001073
Procedural code	10.51	0.29	15.30	0.001073

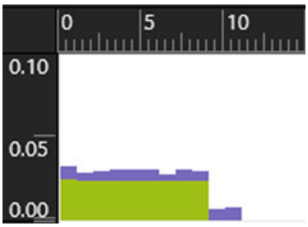


Fig. 30 Code behind

Fig. 31 XAML format

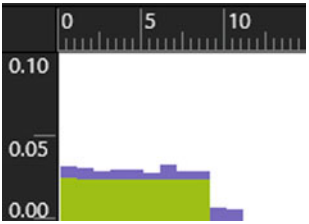


Table 17 Rules obtained after running the experiments

Hypotheses	Status	Rule
Hypothesis no. 1	Confirmed	Use darker colors in Windows Phone applications
Hypothesis no. 2	Not relevant	The PNG or JPG file format does not influence the energy consumption of a mobile application
Hypothesis no. 3	Confirmed	Use PNG format instead of XAML format for displaying images
Hypothesis no. 4	Confirmed	Use “CreateOption” attribute for all the pictures
Hypothesis no. 5	Confirmed	Use static objects instead of animated ones as much as possible
Hypothesis no. 6	Confirmed	Use decoder to size when the dimension of the image control is known
Hypothesis no. 7	Confirmed	Use asynchronous loading for pictures
Hypothesis no. 8	Confirmed	Use Visibility property for hiding an object instead of Opacity property
Hypothesis no. 9	Confirmed	Choose a determinate progress bar if the context allows this
Hypothesis no. 10	Rejected	For the basic use of a list use a “ListBox” control
Hypothesis no. 11	Confirmed	Use “Resource” value when developing mobile applications
Hypothesis no. 12	Not relevant	Either JPG format or Base64 format can be used for displaying pictures
Hypothesis no. 13	Not relevant	Either “for” or “while” loop can be used in developing a “green” application
Hypothesis no. 14	Confirmed	Use multi-threads in a mobile application
Hypothesis no. 15	Rejected	Use “OnNavigateTo” method for data initialization
Hypothesis no. 16	Confirmed	Use iterative functions instead of recursive ones
Hypothesis no. 17	Rejected	Use “VirtualizingStackPanel” inside “ItemsControls” elements
Hypothesis no. 18	Not relevant	Either storing the resources in a different assembly or in the same assembly, the energy consumption is the same

(continued)

Table 17 (continued)

Hypotheses	Status	Rule
Hypothesis no. 19	Not relevant	An animated object can be created either in XAML file or in procedural code
Hypothesis no. 20	Confirmed	User images stored locally
Hypothesis no. 21	Inconclusive	–
Hypothesis no. 22	Inconclusive	–
Hypothesis no. 23	Confirmed	Use JPG format if the picture are stored in clouds
Hypothesis no. 24	Rejected	Access the images directly from web service rather than downloading them
Hypothesis no. 25	Confirmed	Process data locally

Two hypotheses are inconclusive. Table 17 presents a summary of the results obtained from these experiments (note the third column is the energy efficiency rule).

5 Conclusions

Developing a mobile application has to be based on the user experience. Nowadays a user expects an application that is fast and responds to any input. The battery consumption is another aspect which is really important for a user, but which is associated most of the times with the phone and not with an application. It is true that the energy consumption of an application is not the same for two different mobile phones, but most of the energy consumption is application dependent. This study reveals the fact that there are some concepts, such as single threading, which consumes more energy than similar concepts (multithreading), which give the same output. For a developer it is very important to choose the right approach in order to offer the user the best experience when using an application and a longer battery life.

The second reason for this study is sustainability. Each experiment shows the energy consumed by each tested concept or control. The value obtained can be used for calculating the total impact that an application can have on the environment. This is an important aspect because nowadays ICT produces 2% from the total energy consumed in the world. This percentage will grow, because the ICT domain is in a continuous development, so it is very important to reduce the energy in all the aspects. There are studies conducted in this domain, but most of them focus on Android phones or on iOS phones. Windows Phone is not very popular at the moment, but, according to the sources presented in the Introduction section, there will be an increase in the next years. One aspect that could be very interesting to study is the energy consumption of each operating system and to compare the differences between them. Another future work will be exploring the relationship between energy consumption and different hardware components on the same platform. For example, it would be interesting to know the relationship between the energy consumption and the size of the screen, or the screen type. This study could help the producers to choose the right components for the future models of phones.

Acknowledgements This work was supported by the Erasmus Mundus PERCCOM programme.

References

1. Cristea, V.-V., Pattinson, C., Kor, A.L.: Energy Consumption of Mobile Phones. In: International SEEDS Conference, 17–18 Sept 2015, Leeds
2. Cristea, V.: Energy consumption of mobile phones. Unpublished Masters Degree Dissertation, Leeds Beckett University, UK (2015)
3. Kharchenko, V., Kondratenko, Y., Kacprzyk J. (eds.): Green IT Engineering: Concepts, Models, Complex Systems Architectures. Studies in Systems, Decision and Control, vol. 74. Springer, Cham (2017a). https://doi.org/10.1007/978-3-319-44162-7
4. Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.): Green IT Engineering: Components, Networks and Systems Implementation. Studies in Systems, Decision and Control, vol. 105. Springer, Cham (2017b). https://doi.org/10.1007/978-3-319-55595-9
5. Kondratenko, Y., Korobko, V., Korobko, O., Kondratenko, G., Kozlov, O.: Green-IT approach to design and optimization of thermoacoustic waste heat utilization plant based on soft computing. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Components, Networks and Systems Implementation. Studies in Systems, Decision and Control, vol. 105, pp. 287–311. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-55595-9_14
6. Kuchuk, G., Kovalenko, A., Kharchenko, V., Shamraev, A.: Resource-oriented approaches to implementation of traffic control technologies in safety-critical I&C systems. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Components, Networks and Systems Implementation. Studies in Systems, Decision and Control, vol. 105, pp. 313–337. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-55595-9_15
7. Emarketer.com: Smartphone Users Worldwide Will Total 1.75 Billion in 2014—eMarketer. [online] Available at: http://www.emarketer.com/Article/Smartphone-Users-Worldwide-Will-Total-175-Billion-2014/1010536. Accessed 15 Apr 2015
8. www.statista.com: Topic: App Stores. [online] Available at: http://www.statista.com/topics/1729/app-stores/. Accessed 15 Apr 2015 (2014)
9. Protalinski, E.: Microsoft Confirms Windows Phone Store Passed 300,000 Apps. The Next Web. [online] Available at: http://thenextweb.com/microsoft/2014/08/08/microsoft-confirms-windows-phone-store-300000-apps/. Accessed 15 Apr 2015 (2014)
10. news.microsoft.com: Microsoft by the numbers. [online] Available at: http://news.microsoft.com/bythenumbers/ms_numbers.pdf. Accessed 15 Apr 2015
11. Webb, M.: SMART 2020: Enabling the Low Carbon Economy in the Information Age (2008)
12. GeSI: Smarter2030: ICT Solutions for 21st Century Challenges. URL: http://gesi.org/report/detail/smarter-2030. Accessed 20 Mar 2018 (2015)
13. Pathak, A., Hu, Y., Zhang, M.: Where is the energy spent inside my app? In: Proceedings of the 7th ACM European Conference on Computer Systems—EuroSys '12 (2012)

14. Hao, S., Li, D., Halfond, W., Govindan, R.: Estimating mobile application energy consumption using program analysis. In: 2013 35th International Conference on Software Engineering (ICSE) (2013)
15. Jung, W., Kang, C., Yoon, C., Kim, D., Cha, H.: DevScope. In: Proceedings of the Eighth IEEE/ACM/IFIP International Conference on Hardware/Software Codesign and System Synthesis—CODES + ISSS '12 (2012)
16. Carroll, A., Heiser, G.: An Analysis of power consumption in a smartphone. In: Proceedings of the 2010 USENIX Conference on USENIX Annual Technical Conference, pp. 21–21, Boston, MA, 23–25 June 2010
17. Corral, L., Georgiev, A., Sillitti, A., Succi, G.: A method for characterizing energy consumption in Android smartphones. In: 2013 2nd International Workshop on Green and Sustainable Software (GREENS) (2013)
18. Xia, F., Hsu, C., Liu, X., Liu, H., Ding, F., Zhang, W.: The power of smartphones. Multimedia Syst. **21**(1), 87–101 (2013)
19. Carroll, A., Heister, G.: An Analysis of Power Consumption in a Smartphone (2010)
20. Hahnel, M., Dobel, B., Volp, M., Hartig, H.: Measuring energy consumption for short code paths using RAPL. ACM SIGMETRICS Perform. Eval. Rev. **40**(3), 13 (2012)
21. Namboodiri, V., Ghose, T.: To cloud or not to cloud: a mobile device perspective on energy consumption of applications. In: 2012 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM) (2012)
22. Wilke, C., Piechnick, C., Richly, S., Poschel, G., Gotz, S., Abmann, U.: Comparing mobile applications' energy consumption. In: Proceedings of the 28th Annual ACM Symposium on Applied Computing—SAC '13 (2013)
23. Andreucetti, R., Chen, S., Yuan, Z., Muntean, G.: Smartphone energy consumption of multimedia services in heterogeneous wireless networks. In: 2014 International Wireless Communications and Mobile Computing Conference (IWCMC) (2014)
24. Malik, M.Y.: Power Consumption Analysis of a Modern Smartphone. URL: https://arxiv.org/abs/1212.1896. Accessed 20 Mar 2018 (2012)
25. Lee, M., Kim, D., Lee, J.: Analysis of characteristics of power consumption for context-aware mobile applications. Information **5**(4), 612–621 (2014). https://doi.org/10.3390/info5040612
26. Yurur, O., Liu, C., Moreno, W.: Modeling battery behavior on sensory operations for context-aware smartphone sensing. Sensors (14248220) **15**(6), 12323–12341 (2015). https://doi.org/10.3390/s150612323
27. Abreu, D.P., Villapol, M.E.: Measuring the energy consumption of communication interfaces on smartphones using a moderately-invasive technique. In: Global Information Infrastructure and Networking Symposium (GIIS) 2012, 1–6 Dec 2012, Choroni, Venezuela. https://doi.org/10.1109/giis.2012.6466662
28. Parkkila, J., Porras, J.: Improving battery life and performance of mobile devices with cyber foraging. In: 2011 IEEE 22nd International Symposium on Personal, Indoor and Mobile Radio Communications (2011)
29. Drouant, N., Rondeau, E., Georges, J., Lepage, F.: Designing green network architectures using the ten commandments for a mature ecosystem. Comput. Commun. **42**, 38–46 (2014)
30. Pattinson, P., Robinson, L.: A study of the effectiveness of "Wake up on LAN" as a means of power management. In: ICE-B 2008, pp. 73–76 (2008)
31. Li, S., Mishra, S.: Optimizing power consumption in multicore smartphones. J. Parallel Distrib. Comput. **95**, 124–137 (2016). https://doi.org/10.1016/j.jpdc.2016.02.004
32. Ukhanova, A., et al.: Power consumption analysis of constant bit rate video transmission over 3G networks. Comput. Commun. **35**(14), 1695–1706 (2012). https://doi.org/10.1016/j.comcom.2012.05.010
33. Elliott, J., Kor, A.L., Omotosho, A.: Energy consumption in smartphones: an investigation of battery and energy consumption of media related applications on android smartphones. In: International SEEDS Conference, 2017, Leeds, UK
34. Microsoft: Windows Phone 8.1 Mobile Device Management Overview. Available at: https://www.microsoft.com/en-gb/download/details.aspx?id=42508. Accessed date: 15 Apr 2015 (n.d.)

35. Msdn.microsoft.com: Welcome to Visual Studio 2013 [online]. Available: https://msdn.microsoft.com/en-us/library/dd831853(v=vs.120).aspx. Accessed date: 15 Apr 2015
36. Msdn.microsoft.com: Visual Studio Express 2012 for Windows Phone 8. [online] Available at: https://msdn.microsoft.com/en-us/library/windows/apps/ff630878(v=vs.105).aspx. Accessed 15 Apr 2015
37. Channel 9: Windows Phone 8: Performance & Optimization for Developers (Channel 9). [online] Available at: http://channel9.msdn.com/Events/Build/2012/3-048. Accessed 15 Apr 2015 (2012)
38. Msdn.microsoft.com: VirtualizingStackPanel Class—Windows App Development. [online] Available at: https://msdn.microsoft.com/en-us/library/windows/apps/windows.ui.xaml.controls.virtualizingstackpanel. Accessed 15 Apr 2015
39. Msdn.microsoft.com: StackPanel Class (System.Windows.Controls). [online] Available at: https://msdn.microsoft.com/en-us/library/system.windows.controls.stackpanel(v=vs.110).aspx. Accessed 15 Apr 2015
40. Msdn.microsoft.com: ListBox Class (System.Windows.Controls). [online] Available at: https://msdn.microsoft.com/en-us/library/windows/apps/system.windows.controls.listbox(v=vs.105).aspx. Accessed 15 Apr 2015
41. Msdn.microsoft.com: LongListSelector Class (Microsoft.Phone.Controls). [online] Available at: https://msdn.microsoft.com/en-us/library/windows/apps/microsoft.phone.controls.longlistselector(v=vs.105).aspx. Accessed 15 Apr 2015
42. Msdn.microsoft.com: ProgressBar Class (System.Windows.Controls). [online] Available at: https://msdn.microsoft.com/en-us/library/windows/apps/system.windows.controls.progressbar(v=vs.105).aspx. Accessed 15 Apr 2015
43. Msdn.microsoft.com: UIElement.Opacity Property (System.Windows). [online] Available at: https://msdn.microsoft.com/en-us/library/system.windows.uielement.opacity(v=vs.110).aspx. Accessed 15 Apr 2015
44. Msdn.microsoft.com: UIElement.Visibility Property (System.Windows). [online] Available at: https://msdn.microsoft.com/en-us/library/system.windows.uielement.visibility(v=vs.110).aspx. Accessed 15 Apr 2015
45. Msdn.microsoft.com: Storyboard Class (System.Windows.Media.Animation). [online] Available at: https://msdn.microsoft.com/en-us/library/system.windows.media.animation.storyboard(v=vs.110).aspx. Accessed 15 Apr 2015
46. Msdn.microsoft.com: Control.Background Property (System.Windows.Controls). [online] Available at: https://msdn.microsoft.com/en-us/library/system.windows.controls.control.background(v=vs.110).aspx. Accessed 15 Apr 2015
47. Msdn.microsoft.com: Page.OnNavigatedTo Method (System.Windows.Controls). [online] Available at: https://msdn.microsoft.com/library/system.windows.controls.page.onnavigatedto(VS.95).aspx. Accessed 15 Apr 2015
48. Msdn.microsoft.com: Multithread. [online] Available at: https://msdn.microsoft.com/en-us/library/3c8c4cxa.aspx. Accessed 15 Apr 2015
49. Msdn.microsoft.com: while (C# Reference). [online] Available at: https://msdn.microsoft.com/en-us/library/2aeyhxcd.aspx. Accessed 15 Apr 2015
50. Tools.ietf.org: RFC 4648—The Base16, Base32, and Base64 Data Encodings. [online] Available at: https://tools.ietf.org/html/rfc4648#section-4. Accessed 15 Apr 2015
51. Developers.de: Windows Phone 7: Content vs. Resource Build Action - Damir Dobric Posts - developers.de. [online] Available at: http://developers.de/blogs/damir_dobric/archive/2010/09/18/windows-phone-7-content-vs-resource-build-action.aspx. Accessed 15 Apr 2015
52. Msdn.microsoft.com: PictureDecoder.DecodeJpeg Method (Stream, Int32, Int32) (Microsoft.Phone). [online] Available at: https://msdn.microsoft.com/en-us/library/windows/apps/ff708027(v=vs.105).aspx. Accessed 15 Apr 2015
53. W3.org: Portable Network Graphics (PNG) Specification (Second Edition). [online] Available at: http://www.w3.org/TR/PNG/#1Scope. Accessed 15 Apr 2015

54. Whatis.techtarget.com: What is JPG? What Opens a JPG? File Format List from WhatIs.com. [online] Available at: http://whatis.techtarget.com/fileformat/JPG-JPEG-bitmap. Accessed 15 Apr 2015
55. Msdn.microsoft.com: What is XAML? [online] Available at: https://msdn.microsoft.com/en-us/library/cc295302.aspx. Accessed 15 Apr 2015
56. SearchCloudComputing: What is cloud computing?—Definition from WhatIs.com. [online]. Available at: http://searchcloudcomputing.techtarget.com/definition/cloud-computing. Accessed 15 Apr 2015

Energy-Efficient Multi-fragment Markov Model Guided Online Model-Based Testing for MPSoC

Jüri Vain, Leonidas Tsiopoulos, Vyacheslav Kharchenko, Apneet Kaur, Maksim Jenihhin, Jaan Raik and Sven Nõmm

Abstract Term greenware capitalizes the energy efficiency, performance of processes and durability of hardware. Online monitoring/testing is a measure to support these features in autonomous mission critical systems (AMCS). The amount of time needed for sampling AMCS service quality modes has direct impact to energy consumption. We propose an efficient online testing method of AMCS service modes where the Multi-Fragment Markov Models (MFMM) are used for specifying the system reliability and quality related behavior on high level of abstraction, and the more concrete state and timing constraints are specified explicitly using Uppaal Probabilistic Timed Automata (UPTA). To interrelate these models we demonstrate how the MFMM is mapped to UPTA. Second contribution is the test case selection mechanism for online identification of AMCS service modes by model-based conformance testing. The efficiency of active mode sampling is achieved by serializing the test cases for each sampling session using hypotheses provided by MFMM. The hypotheses are tested using UPTA for online conformance. The approach is illustrated with the Bonfire Multi-Processor System-on-Chip (MPSoC).

J. Vain (✉) · L. Tsiopoulos · A. Kaur · M. Jenihhin · J. Raik · S. Nõmm
Tallinn University of Technology, Tallinn, Estonia
e-mail: juri.vain@ttu.ee

L. Tsiopoulos
e-mail: leonidas.tsiopoulos@ttu.ee

A. Kaur
e-mail: apneet.kaur@ttu.ee

M. Jenihhin
e-mail: maksim.jenihhin@ttu.ee

J. Raik
e-mail: jaan.raik@ttu.ee

S. Nõmm
e-mail: sven.nomm@ttu.ee

V. Kharchenko
National Aerospace University KhAI, Kharkiv, Ukraine
e-mail: vyacheslav.kharchenko@csn.khai.edu

© Springer Nature Switzerland AG 2019

V. Kharchenko et al. (eds.), *Green IT Engineering: Social, Business and Industrial Applications*, Studies in Systems, Decision and Control 171,
https://doi.org/10.1007/978-3-030-00253-4_12

Keywords Formal methods · Uppaal timed automata · Verification
Markov models · Model-based testing · Multi-processor system-on-chip

1 Introduction

Online monitoring and model-based validation techniques are commonly accepted quality assurance measures for mission critical systems. Online testing is considered to be the most appropriate technique for Model-Based Testing (MBT) of embedded systems where the Implementation Under Test (IUT) is modeled using nondeterministic models [1]. The nondeterminism stems from the physical nature of the IUT, particularly, its internally interacting parallel processes, loose timing conditions, and hardware related jitter. Often, the term *on-the-fly* is used instead of online testing to describe the test generation and execution at runtime.

In recent years, in addition to fault detection, power constraints have increasingly gained attention during system development and operation. This has led to the new concepts of design and term greenware. Greenware as the whole is a combination of green hardware and green software, and services [2]. In autonomous systems, for instance, too high energy consumption due to frequent online testing during mission would drain the battery and cause mission failure. Moreover, in at-speed testing, excessive power in testing causes overtesting for a circuit produced with nanoscale technologies. Integrated circuits usually consume much more power in test mode than in normal mode, introducing problems like IR drop and circuit destruction, etc. [3].

If the critical system behavior is complex it is hard or practically impossible to predict all potential causes why the system can deviate from its specified behavior, e.g. would it be due to the bugs, hardware faults or cyber-attacks, unexpected load profiles or other reasons that have impact on system's Quality of Service. Therefore, in model-based approaches the causal relations are often approximated with probabilistic ones. On the other hand, for error detection and correction the online verification of design correctness should explore the more concrete state space for explicit conditions under which the faults expose. Following from the results presented in [4], in this paper we focus on energy-efficient online testing implemented by combining predictive Markov modeling with online MBT.

In our previous work, we proposed an integrated approach with Multi-Fragment Markov Models (MFMM) [5] and Uppaal Probabilistic Timed Automata (UPTA) [6]. MFMM are used for specifying the system reliability related probabilistic behavior on high-level of abstraction. This is relevant for the assessment of the availability of system functions depending on scenarios for online execution and verification. The more concrete state and timing constraints related with MFMM are specified explicitly with UPTA. To interrelate these two model classes we demonstrated how the MFMM is mapped to UPTA which is more relevant for model checking and model-based test generation. The second contribution concerned the test case selection mechanism for online testing. This prioritizes the test cases by probabilities of modes specified in MFMM. To confirm which mode has actually occured the conformance between the mode models and observable behavior of the IUT is sampled by online testing.

We extend our previous approach targeting energy minimization during online testing. Running mode identification tests frequently or having extensive test cases in energy and/or time critical missions might be costly computationally as well as regarding energy consumption. Such an overhead created by executing tests may compromise the main functionality of the system or reduce the performance, e.g. cause premature depletion of batteries. On the other hand, running the mode identification tests too seldom, may cause delayed compensation of failures or improper reaction to some emergent situation. Therefore, the *minimization of the mode identification decision time* is crucial for energy critical autonomous systems life span and their mission success.

The new contributions of this paper are:

- We generalize the method of Mission Critical Systems (MCS) operational mode identification proposed in [4]. Instead of using only probabilities of system mode switching during execution for prioritizing mode identification tests, we also incorporate the probability of successful identification of a mode by a given test and the test case duration. The restrictive assumption about the uniform duration of tests is relaxed so that tests of any duration and modes with any probability constraints are used for computing priorities to minimize the average mode identification time.
- The second contribution of this paper is that, instead of applying the full model of the mode for its identification, we apply only the minimal necessary part of it called Minimal Distinguishing Test Fragment (MDTF). Applying MDTFs reduces the number and length of test sequences needed for mode identification.

This is illustrated with the Bonfire Multi-Processor System-on-Chip (MPSoC) example maintained as an open-source project, available at [7].

The rest of the chapter is organized as follows. In Sect. 2 we discuss related work and in Sect. 3 we present the preliminaries of our approach. In Sect. 4 we introduce the availability MFMM of Bonfire and its mapping to UPTA mode switching automaton for guiding the online conformance testing of modes. In Sect. 5 a set of UPTA models that refines the behavior of MFMM modes is developed. The method of creating the Minimal Distinguishing Test Fragment of UPTA models is presented in Sect. 6. In Sect. 7 we introduce criteria and a procedure for ordering mode identification tests to minimize the average mode identification time. In Sect. 8, we evaluate our method with the Bonfire case study and in Sect. 9 conclude the chapter.

2 Related Work

Markov chains is a well established model-based analytical tool for understanding stochastic systems behavior [8]. Stochastic approaches based on Markov models have been applied on multi-processor systems, e.g. to analyse fault-tolerance [9] and performance [10]. Also Markov model-based approaches have been recently applied to analyse the degradation and recovery of resources in Smart Grid applications [11].

Mode switching has been proposed as one of the methods for addressing the correct execution of mixed-critical applications on embedded multi-core platforms due to failures at runtime. Socci et al. [12] presented the Real-Time BIP (Behavior Interactions Priorities) framework for providing alternative executions in case of errors. The priority assignment is based on the applications' models which support different modes, corresponding to different levels of QoS. Observers monitor the state of the applications at runtime and report violations to the runtime manager which in turn updates the mode of the applications such that complete failures are prevented.

In similar context, Negrean et al. [13] provided an approach for handling shared resources across mode changes of applications on multi-core platforms conforming to the AUTOSAR standard. On such multi-core platforms the applications are scheduled using a partitioned fixed-priority preemptive scheduling, they share resources and they are subject to runtime mode management.

Several works which are related to the integration of Markov models and system MBT approaches, aim at improving the testing efficiency. For example [14] being one of the first works presented a method for statistical testing based on a Markov chain model of software usage. A more recent related work by Kashyap et al. [15] proposes to create system models from available system usage data using Markov chains. The goal is to create test plans that employ a "likelihood-based" test prioritization scheme that considerably improves the coverage of generated test suites.

Corresponding testing methodology for the automotive domain is presented by Siegl et al. [16]. A Markov Chain Usage Model is studied to describe all possible usage scenarios of the IUT and to provide the bases for systematic development of test cases. These test cases can be further processed inside the Extended Automation Method employed by AUDI AG. The approach makes it also possible to get both indicators for the test-planning in advance and to obtain dependability measures after the test cases have been processed.

During the last fifteen years several approaches targeting efficient online testing of MPSoCs have been published. Some of these works focus on respecting the power budget without affecting much the computation performance. For example, Haghbayan et al. [17] presented a power-aware approach for online testing by scheduling software-based self-test routines on the cores during their idle periods. Moreover, a test criticality metric based on a device aging model is used to guide/prioritize the selection of the cores to be tested.

All the aforementioned works that take into account Markov models are concerned with offline MBT while our approach is targeting efficient online MBT with minimum energy overhead. Like in the work by Haghbayan et al., we propose to test the system modes in idle times to reduce testing overhead. On the other hand, we propose a formal model-based online testing approach for MPSoCs. In addition to exploiting Markov model probabilities for reducing testing time we provide a method to generate minimal distinguishing test cases to facilitate testing energy efficiency even further.

To the best of authors' knowledge this is the first work to combine Markov models and UPTA for online MBT of MPSoCs.

3 Preliminaries

3.1 Online Model-Based Testing

The state space explosion problem experienced by many model-based offline test generation methods is avoided by the online techniques because only a limited part of the state-space needs to be kept track of when a test is running. However, exhaustive planning would be difficult on-the-fly because of the limitations of computational resources and energy constraints at the time of test execution. Thus, developing a planning strategy for online testing should address the trade-off between reaction time and online planning depth to reach the practically feasible test cases in an energy-efficient manner.

The simplest approach to on-the-fly selection of test stimuli in online MBT is to apply so called random-walk strategy where no computation sequence of IUT has an advantage over the others. The test is performed usually to discover violations of input/output conformance relation IOCO [18] or timed input/output conformance relation TIOCO [19] between the IUT and its model. Random exploration of the state space may lead to test cases that are unreasonably long and nevertheless may leave the test purpose unachieved.

In order to overcome the deficiencies of long lasting testing usually additional heuristics, e.g. "anti-ant" [20, 21] and "reactive planning" [22], are applied for guiding the exploration of the IUT state space. The extreme of guiding the selection of test stimuli is exhaustive planning by solving at each test execution step a full constraint system set by the test purpose and test planning strategy. For instance, the witness trace generated by model checking provides the test stimuli that yield optimal tests. The critical issue in the case of explicit state model checking algorithms is the size of the model and the complexity of the checked formula that leads to the explosion of the state space. Therefore, model checking based approaches are used mostly in offline test generation.

Another option for reducing the online test generation and planning overhead is to partition the test models into smaller ones where the test cases do not need the full blown model coverage. Following this idea in this paper we define the test models by system availability modes so that each of the test models refines only one state (availability mode) of the abstract Markov model. Next reduction step extracts from each mode model a minimal model fragment which differs from those of all others and, thus, can be used for generating mode distinguishing tests. By reducing the size of test models in this way we can apply simple random-walk strategy instead of strategies that require computationally expensive state space exploration.

3.2 Multi-fragment Markov Models

Kharchenko et al. [23] developed a set of potential scenarios of reaction to verification of detected faults during execution of a system and defined a basic set of availability models. In order to assess the availability depending on such scenarios a set of corresponding single- [24] and multi-fragment [25] Markov models were developed. MFMM are used for specifying the system reliability related probabilistic behavior which is relevant for the assessment of the availability of system functions depending on scenarios for online execution and verification.

Some scenarios that can be specified by MFMM are as follows: SC1—all possible types of offline verification are performed. Online verification and correction of faults are impossible at runtime; SC2—all possible types of offline verification are performed. Online verification is impossible at runtime. Some correction of faults is possible and made in case of detecting of fault; SC3—all possible types of verification are performed (offline and online). Correction of all detected faults is performed; SC4—all possible types of verification are performed (offline and online). Correction of detected faults is performed. Blocking of parts of functions and system degradation is possible.

For the work in this chapter, we relate to scenarios SC3 and SC4. The availability model of Bonfire MPSoC corresponding to these scenarios will be presented and discussed in Sect. 4.

3.3 Uppaal Probabilistic Timed Automata

Uppaal Timed Automata (UTA) [26] address behavioral and timing aspects of systems providing efficient data structures and algorithms for their representation and analysis through model checking.

An Uppaal Timed Automaton (UTA) is given as the tuple $(L, E, V, CL, Init, Inv, TL)$, where L is a finite set of locations, E is the set of edges defined by $E \subseteq L \times G(CL, V) \times Sync \times Act \times L$, where $G(CL, V)$ is the set of constraints in guards, $Sync$ is a set of synchronization actions over channels and Act

is a set of sequences of assignment actions with integer and boolean expressions as well as with clock resets. V denotes the set of integer and boolean variables. CL denotes the set of real-valued clocks $(CL \cap V = \emptyset)$. $Init \subseteq Act$ is a set of assignments that assigns the initial values to variables and clocks. $Inv : L \rightarrow I(CL, V)$ is a function that assigns an invariant to each location and $I(CL, V)$ is the set of invariants over clocks CL and variables V. $TL : L \rightarrow \{ordinary, urgent, committed\}$ is the function that assigns the type to each location of the automaton.

Uppaal Probabilistic Timed Automata (UPTA) [6] is a stochastic and statistical modeling extension of Uppaal Timed Automata (UTA). UPTA preserves the standard UTA constructs such as integer variables, data structures and user-defined C-like functions. Additionally, UPTA support branching edges where weights can be added to define a probability distribution on discrete transitions. It is important to note that the weights may be general expressions that depend on the states and not just simple constants. For the work in this chapter we use the branching edges with probability weights.

4 Availability MFMM of Bonfire MPSoC and Mapping to UPTA

Let us first introduce the Bonfire MPSoC before we re-visit the modeling approach presented in [4]. This will be the base for the extensions of this approach we propose in Sects. 6, 7 and 8.

4.1 Bonfire MPSoC

The Bonfire MPSoC employs a Network-on-Chip (NOC) with a 2D mesh topology. Each router uses wormhole switching equipped with fault tolerant mechanisms. A processor is connected to each router via a network interface. The router comprises of an input buffer, a routing computation unit, a switch allocator granting requests to the same output based on Round-Robin policy and a crossbar switch to forward the packets. An abstract view of Bonfire is presented in Fig. 1a. For this chapter we are considering an instantiation with four processors and four routers. Since the routing mesh of Bonfire is the IUT the availability of which is studied in this chapter, the test interface (denoted by dashed line in Fig. 1a) is defined by the interfaces between the network interfaces and processors, i.e., the processors constitute an environment for the IUT.

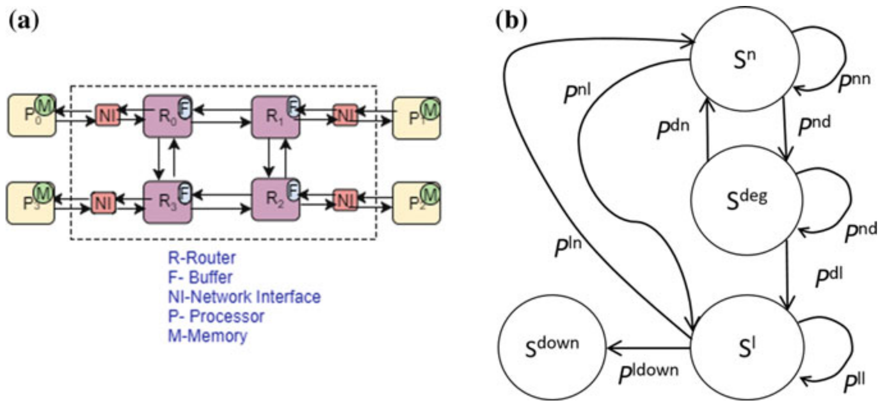


Fig. 1 **a** Bonfire MPSoC, **b** availability Markov model for Bonfire MPSoC

4.2 *Availability MFMM of Bonfire MPSoC*

The two dominant concerns of hardware system's reliability are Soft-Error Reliability (SER) and Life Time Reliability (LTR). The first one addresses transient faults at any time of the system's lifespan and can rapidly compromise the service while the second one addresses intermittent and permanent faults due to system's degradation over time. Both aspects became severe with miniaturization of implementation technologies in nanoelectronics. In contrast to our previous work [4] where we considered also software faults and software updates, we are considering soft errors, i.e., hardware transient faults, as the main faults in this chapter, because these are the most interesting for analysing and testing the execution mode changes of a system.

In order to develop an availability model for Bonfire, we consider a scenario of a space probe travelling through a possibly hostile environment of cosmic radiation. The on-board computer of the probe is implemented as a quad-core processor version of the Bonfire MPSoC. The faults that can occur are assumed to be transient faults caused by cosmic particle hits. With moderate radiation levels such faults can be handled by resubmitting the failed packets. If the fault rate becomes higher than a certain threshold then the on-board computer will continue operation in low mode until reconfigured.

During the travelling of the space probe, with the probability of $P_1 = 0.09$ the environment worsens with a fault rate F_1 per sampling period. On the other hand, with a probability of $P_2 = 0.001$, we enter a physical environment of extremely high radiation, with an error rate $F_2 \gg F_1$ per sampling period. The probability of the probe continuouing travelling in an environment without cosmic radiation is, thus, $P_0 = 1 - (P_1 + P_2) = 0.909$.

Figure 1b presents the availability Markov model of the scenario described above. The model includes three operational states, which are normal (represented by state S^n), degraded (state S^{deg}) and low (state S^l) mode, respectively. State S^{down} models the *System Down* mode where the system is not operable anymore. Due to possible recovery the system operation can be restored from other states. For the rest of this chapter state S^{down} is not considered further.

Note that for simplicity we present a snapshot of a Multi-Fragment Markov availability model where the probability distributions are stationary. Naturally the full model is non-stationary and the multiple fragments expand to the right of the first fragment after software corrections. It is assumed that the right-most fragment would not include any more software faults. Furthermore, since we consider only HW transient faults with steady failure and recovery rates, such a fragment of the availability model is sufficient for demonstrating our method.

The probability of a fault occurring during the sample period in the degraded mode is $\lambda^1 = F_1 \times A$, where A is the average silicon area involved in transferring packets through the 4-router mesh NoC. Similarly, $\lambda^2 = F_2 \times A$.

Initially, the space probe is in the normal state S^n. With the probability of $P^{nd} = \mathrm{P}_1 \times \lambda^1$ a fault occurs and the transition to the degraded state S^{deg} happens. From the normal state S^n a transition is also possible to the low mode state S^l with the probability $P^{nl} = \mathrm{P}_2 \times \lambda^2$. With probability $P^{nn} = 1 - \left(P^{nd} + P^{nl}\right)$ the system continues operating in normal mode. At the degraded state S^{deg}, a recovery via packet retransmission is attempted, which will succeed with a probability of $P^{dn} = \mu^{dn} = 1 - P^{nd}$ bringing the system back to the normal state. However, the recovery will fail with a probability of P^{nd} keeping the system in the degraded state. From the degraded state the system can be taken down to the low state S^l with the probability P^{nl} similarly as from the normal state. There, the chip is operating in low mode with probability $P^{ll} = P^{nl}$ until reconfigured from an onboard memory which will then cause the system to continue from the normal state with the probability $P^{ln} = 1 - P^{nl}$.

The Markov availability model of Fig. 1b will be next mapped to UPTA for modeling of the mode switching.

4.3 Mapping Bonfire Availability MFMM to UPTA

The UPTA model corresponding to the Markov Bonfire model is presented in Fig. 2a. The transitions of the Markov model in Fig. 1b are mapped to probabilistic transitions (denoted by dotted lines and labeled with probabilities) of UPTA shown in Fig. 2a. The probabilities of UPTA are the same as the probabilities in Fig. 1b. They specify the probabilities of events that cause transition from one mode to another. The probabilities of events that trigger transitions are normalized with respect to the model time unit denoted by *tu* which corresponds to one clock cycle in Bonfire MPSoC.

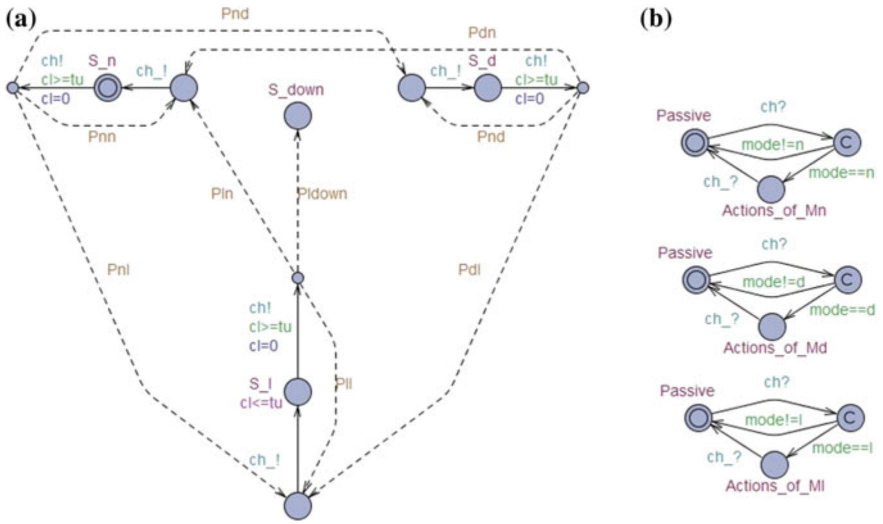


Fig. 2 **a** Mode UPTA for Bonfire MPSoC based on the Markov availability model, **b** mode adapter automata for the different operational behaviors

The UPTA automata that interface the abstract availability model in Fig. 2a with detailed mode behavior models are shown in Fig. 2b. Initially all automata are in location *Passive* waiting for synchronisation from the Mode switching automaton. Upon synchronisation they will either return back to the *Passive* location or they activate the automata which specify the relevant internal behavior of a mode. Locations *Actions_of_Mn*, *Actions_of_Md* and *Actions_of_Ml* represent system staying in respective mode which are refined in the next section.

5 UPTA Models of the Bonfire MPSoC Execution Modes

We have created the models of the main components of the Bonfire MPSoC platform, namely, the processor and the router considering an instantiation of the platform with four processors and four routers. These models can be used for online MBT of each different operational mode of the system. Since only router's faults are considered the processor component model is assumed to be the same in all modes. The router component models comprise the functionality in normal, degraded and low mode that represent respectively functioning under ideal conditions, functioning under hardware transient faults causing packet corruption along a link and functioning under transient faults of links resulting in excluding them from the routing paths until possible recovery.

5.1 Normal Mode

The UPTA templates representing the normal functionality mode of the processors and routers are shown in the upper and lower part of Fig. 3, respectively. The environment to the routing system consists of four processors which are instantiated with four processes sending and four processes receiving data packets respectively to and from the routing system. The routing system is modeled as UPTA processes parameterized with source and destination router id-s. The routing template specifies (i) the case where there is no packet in its input queue (the edge from *Wait* location to *Idle_tact* location) and (ii) the case where there is at least one packet in the input queue to be forwarded to the next router (the edge from *Wait* location to right and its subsequent edges and locations reachable before return to location *Wait*). The lower-left and lower-right template in Fig. 3 correspond to the input link of a router from its adjacent processor and link between two routers, respectively.

The composed model of the normal functionality mode represents design assumptions A1–A10 and guarantees G1–G3 of routing as follows.

- A1—Processors are indexed clock-wise from 0 to 3 in the mesh and denoted respectively with P_i, where $i : [0, 3]$. The same holds for local routers R_i;
- A2—Processors communicate by sending and receiving data packets;
- A3—Any processor P_i can send at most one data packet per *tu* to its local router R_i;
- A4—Any pair of neighbor routers has direct communication link in two directions, a *link*(i, j) denotes a link from R_i to R_j and *link*(j, i) a link from R_j to $R_{i,}$ where each link *link*(i, j) has a buffer *buf*(i, j) to store the data packet postponed due to congestion;

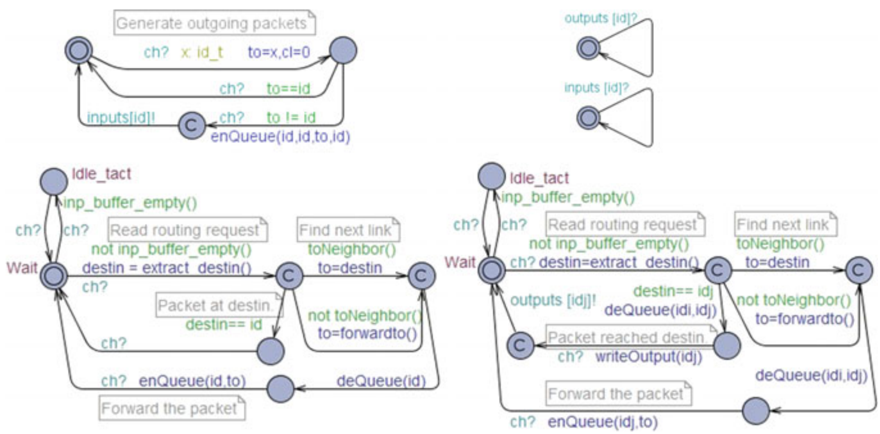


Fig. 3 UPTA for normal mode functionality of processor (upper part) and router links (lower part)

- A5—An oldest element in the buffer *buf*(i, j) is forwarded along the link *link*(i, j);
- A6—Buffer emptiness is checked and content sent once in a *tu*;
- A7—A data packet is written by router R_i to buffer *buf*(i, j) of its outgoing link *link*(i, j) during the same *tu* when it is read from some of its incoming link *link*(k, i);
- A8—Congestion rises when the data packets of different incoming links *link*(i, j) and *link*(k, j) of router R_j are arriving in the same *tu* and need to be forwarded along the same outgoing link *link*(j, l);
- A9—The congestion can rise at most between two incoming links in one *tu*;
- A10—The data packets are identified by their source and destination processor id-s.

The explicit design *guarantees* for the processors and routers are as follows.

- G1—Any data packet sent from a processor P_i does not have congestion on any link with some other data packet sent from the same processor P_i;
- G2—The data packets once routed along *link*(i, j) are never routed backwards along *link*(j, i) in normal mode; backtracking is allowed only in case of link faults;
- G3—In the routing system satisfying assumptions A1–A10 any data packet can be routed to its destination with at most three hops in normal mode, i.e. one along local link and two along links between routers while the congestion can occur in inter router link only.

We additionally assume that processors never send the data packets to themselves and reading packets by processors from the routing system never blocks.

Each router in normal mode implements the design assumptions A4–A7 and provides the guarantees G2, G3 and ensures that no data packet loss occurs.

5.2 Degraded Mode

The UPTA template for the degraded mode of the system is presented in the upper part of Fig. 4. While the processor template remains the same, the routing template for the degraded mode extends the normal mode assumptions with the assumptions for the correct delivery of the data packet over the link between routers succeeding with probability p and packet transmission failure occuring with probability $1 - p$. The refinement of the edge with label “Forward the packet” with two alternative outcomes—one for successful and the other for failure of transmission between a pair of routers introduces probabilities p and $1 - p$ respectively of these options.

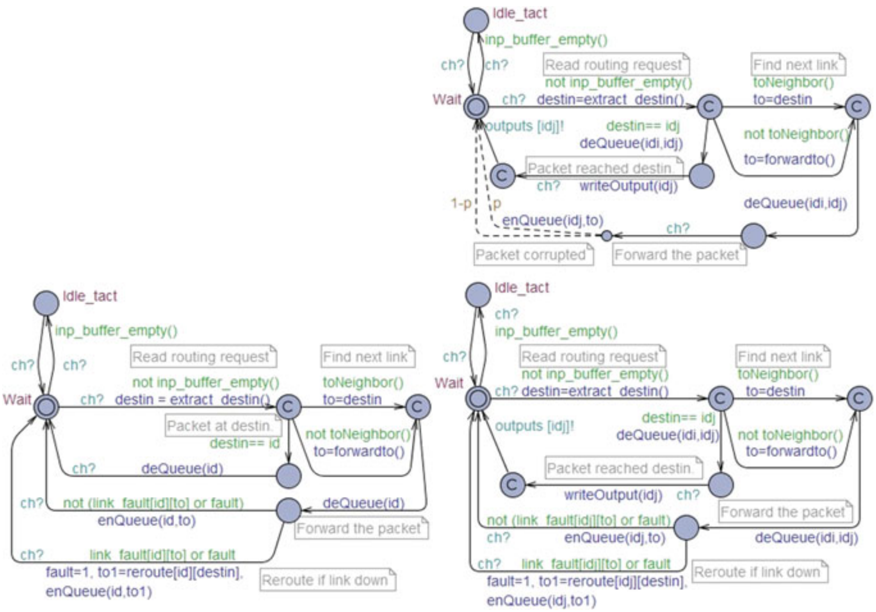


Fig. 4 UPTA for router link degraded mode functionality with probability for data packet loss (upper) and low mode functionality with modeling of link failure (lower)

5.3 *Low Mode*

The UPTA model template for the low mode of the system is presented in the lower part of Fig. 4. The assumptions for the low mode maintain the assumptions for the normal mode regarding writing and reading data packets to and from processors. In low mode lower left part of Fig. 4 links between routers (lower right part of Fig. 4) are expected to be faulty after the system has switched to this mode with some probability which means that the packets are retransmitted and rerouted, respectively. This is modeled with edge "Reroute if link down" in the templates.

6 Minimal Distinguishing Test Fragment

In this section we present the theoretical background for reducing the number and length of test sequences needed for identification of modes. Assume current mode m_i at ith sampling point has possible successor modes $m_{i1}, \ldots, m_{ik}$ that have to be identified at ith + 1 sampling point. This will provide $k+1$ test hypothesis and corresponding conformance test cases $Ŧ_i, Ŧ_{i1}, \ldots, Ŧ_{ik}$ which are defined by mode models $M_i, M_{i1}, \ldots, M_{ik}$.

Let $\mathbf{TR}(M_i)$ denote the set of traces that are used as test sequences generated by model M_i and $\cup_j \mathbf{TR}(M_j)$ is the union of such traces generated by mode models $M_{i1}, \ldots, M_{ik}$. Then sufficient condition for a trace tr^i to distinguish mode M_i from other modes is that $tr^i \in \mathbf{TR}(M_i) \setminus \cup_j \boldsymbol{TR}(M_j)$, where $M_j \in \{M_{i1}, \ldots, M_{ik}\}$, provided tr^i is controllable.

In special case MDTF can be extracted only by syntactic comparison of $M_i, M_{i1}, \ldots, M_{ik}$. It is when M_i has some structural element st_x (location or transition) unique to M_i and it is reachable by some shortest test sequence. Then the MDTF of M_i is the minimal of such sequence tr^x, which includes st_x. The distinguishing fragment can be found by search procedure of complexity $\prod_{i=1}^{k} |ST_i|$, where ST_i denotes the set of structural elements of M_i from which the MDTF is searched and i ranges over indices of models $M_i, M_{i1}, \ldots, M_{ik}$.

From the test controllability point of view there are two options:

- Deterministic MDTF. The MDTF of M_i is deterministically controllable if there exists at least one distinguishing test sequence per mode. To find the MDTF of M_i we have to solve $\frac{n!}{2(n-2)!}$ bisimulation equivalence checking tasks taking models pair-wise from the set $\{M_i, M_{i1}, \ldots, M_{ik}\}$. The bisimulation is defined relative to test I/O alphabet. If two models are not bisimilar the bisimilarity checks provide counter-example traces which are candidate for MDTF. For a counter-example trace tr^i s.t. $tr^i \in \mathbf{TR}(M_i)$ and $tr^i \in \mathbf{TR}(M_j)$, where $j \neq i$, to be a MDTF its uniqueness should be checked with all $tr^i \in \mathbf{TR}(M_j)$. If any of these checks provides counter-example different from tr^i it means the search for MDTF has to backtrack and new counter-example trace should be generated. Since the mode models $M_i, M_{i1}, \ldots, M_{ik}$ are distinguishable by definition and the mode hypothesis provide $k+1$ options each backtracking means in the worst case repeating k checks.
- Non-deterministic MDTF. In non-deterministic models, i.e., while test inputs label two or more transitions departing from the same location the test is controllable non-deterministically. It means that if unintended transition is non-deterministically triggered by given test input the test sequence has to return to the same branching point and the input has to be tried again until intended transition will be executed. Another prerequisite of non-deterministic controllability of test is that IUT satisfies fairness assumption. A system is fair if after infinite number of trials each of the enabled alternative edges is executed infinitely many times. In practice, the fairness can be granted or checked only in its bounded form. Bounded fairness means that after finite number of trying a test input to trigger non-deterministic choices the difference in the number of selected edge execution does not exceed some finite bound k. Assuming that the MDTF of a model M_i includes locations l with non-deterministic outgoing edges where the branching factor at each location l_j is b_j then the MDTF of M_i should provide at most $\sum_{j=1}^{l} b_j.k$ traces to cover all non-deterministic transitions.

7 The Ordering of Mode Tests

The goal of this section is to introduce criteria and a procedure of ordering mode identification tests to minimize the average mode identification time.

We assume an application where the system is periodically sampled to identify its QoS mode. Let us assume that the ith sampling has identified some mode m^i. We have to arrange the tests for identifying the ith + 1 sample so that the tests identify mode m^{i+1} in average with shortest time. Let there be $n > 0$ options of switching from m^i to m^{i+1}, where $m^{i+1} \in \{m_1, \ldots, m_n\}$ including the case which corresponds to staying in the same mode as m^i. To identify in which of the modes $m_1, \ldots, m_n$ the system will possibly switch to (or stay) we apply mode tests $Ŧ_1, \ldots, Ŧ_n$ respectively, each being unique and conforming with its corresponding mode only.

From MFMM analysis the probabilities $p_1, \ldots, p_n$ of transitions between modes during a sampling period are known. Also the execution durations $t_1, \ldots, t_n$ of tests $Ŧ_1, \ldots, Ŧ_n$ are known. Mode identification tests are executed sequentially, therefore identification of ith mode with test $Ŧ_i$ cannot be done before the tests of higher priority are performed. Assume serializing single mode tests by priorities provided the priority list $Ŧ_1, \ldots, Ŧ_i$. It means that the ith test in the sequence if terminating in some conclusive state (pass or fail) takes time T_i that is calculated according to formula (1).

$$T_i = \sum_{k=1}^{i} t_k \tag{1}$$

Let us also introduce the probability of detecting the mode that is identifiable by ith test which is calculated by formula (2).

$$P_i = \frac{p_i \prod_{k \in [1,n] \setminus \{i\}} (1 - p_k)}{\sum_{i=1}^{n} \prod_{k \in [1,n] \setminus \{i\}} (1 - p_k)} \tag{2}$$

The average test suite execution time needed for mode identification is computed then as

$$\overline{T} = \sum_{i=1}^{n} P_i T_i \tag{3}$$

In special cases where either (i) $p_i = P_i$ for all $i \in [1, n]$ or (ii) $t_1 = \cdots = t_n$, the tests have to be ordered by their durations in ascending order (case (i)), or the tests have to be ordered by their corresponding modes probabilities in descending order (case (ii)). Unfortunately both test ordering strategies do not provide optimum in the general case, i.e., when the mode probabilities as well as their test durations have arbitrary values.

For the general case where no constraint is imposed on mode test durations and mode probabilities, we propose to solve the tests prioritizing problem by numerical analysis. This method presumes that, at first, all possible test orderings (permutations) are generated and for each permutation corresponding sequences of P_i and T_i are computed. Thereafter, the priorities of tests are chosen based on this order in the permutation that provides minimum average testing time (3). Regardless the fact that the solution cannot be analytically derived and its computation complexity is exponential in the number of modes n, for the small values of n, such computations can be performed with reasonable time bound. Since computation is done offline exact computation time of optimal serializing of mode tests is not critical.

The approach proposed can be substantiated with the following numerical example. Suppose the number n of candidate modes from which the actual one has to be identified is 6 and their probabilities respectively $p_1 = 0.6, p_2 = 0.2, p_3 = 0.13, p_4 = 0.05, p_5 = 0.015, p_6 = 0.005$. The durations of mode tests are known as well, e.g. $t_1 = 0.4, t_2 = 0.67, t_3 = 0.45, t_4 = 0.23, t_5 = 2, t_6 = 0.001$.

Ordering testing sequence just by test's durations in ascending order will lead to the following initial probabilities ordering $P_1 = 0.005, P_2 = 0.6, P_3 = 0.05, P_4 = 0.2, P_5 = 0.13, P_6 = 0.015$ which clearly is not neither monotonously increasing nor decreasing. Corresponding average testing time $\overline{T} = 1.22$.

At the same time minimum testing time $\overline{T_{min}} = 0.992$ is achieved by the ordering in which test durations are respectively $t_6 = 0.001, t_4 = 0.23, t_3 = 0.45, t_2 = 0.67, t_1 = 0.4, t_5 = 2$ and the mode probabilities by tests respectively $p_6 = 0.005, p_4 = 0.05, p_3 = 0.13, p_2 = 0.2, p_1 = 0.6, p_5 = 0.015$.

Figure 5 visualizes the expected mode identification times with all permutations of 6 tests as described in the example above. The minimum testing time shown with blue horizontal line in Fig. 5 corresponds to the permutation marked with bold asterisk. It differs from the average value 2.2 of expected identification times over all permutations (thick red horizontal line in Fig. 5). The average identifications

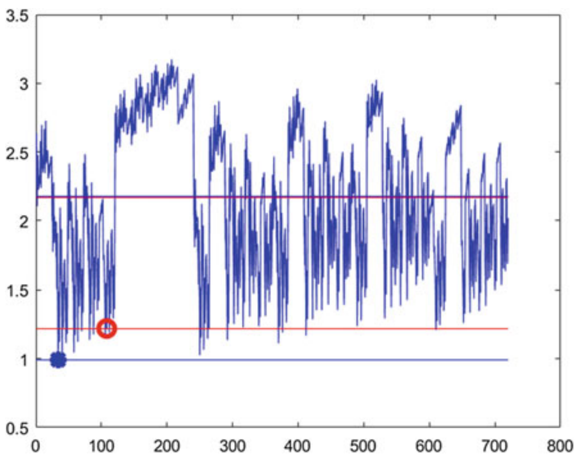


Fig. 5 Expected mode identification times with all permutations of 6 tests. Minimum testing time shown with bold asterisk and the average identifications time when ordering tests by their duration increasingly is indicated by line marked with circle

time when ordering tests by their duration increasingly is indicated by another horizontal line intersecting with circle. It can be seen that both are strictly larger than the minimum which can be achieved by the ordering method proposed.

8 Evaluation of the Method with Bonfire Case Study

To evaluate the method introduced in Sects. 6 and 7 we explore the Bonfire case study (further called *System*) presented in Sects. 4 and 5. The mode of *System* is sampled periodically via test interface by running tests that identify its current mode. Suppose the *i*th sampling identifies that *System* is in normal mode m_N. To identify the mode at $i\text{th} + 1$ sample three options according to the MFMM model have to be considered: staying in the normal mode m_N, transition to the degraded mode m_D, or transition to the low mode m_L.

The UPTA models M_N, M_D, M_L (Figs. 3 and 4) of modes m_N, m_D, m_L respectively specify the expected behavior of the *System* under various operational conditions of modes and they are used for checking the conformance between the system behavior and that of the models. Still, the complete models of modes may have considerable overhead compared to that needed for mode identification. Like it has been described in Sects. 6 and 7 we can minimize the mode identification time by the following procedure. At first, we extract from the mode models M_N, M_D, M_L the minimal distinguishing test fragments MDTFs, denoted $\widehat{Ŧ}_N, \widehat{Ŧ}_D, \widehat{Ŧ}_L$ respectively, to shorten the execution times of individual mode tests $Ŧ_N$, $Ŧ_D$, $Ŧ_L$.. Second, we use the mode identification probabilities P_N, P_D, P_L and durations of $\widehat{Ŧ}_N, \widehat{Ŧ}_D, \widehat{Ŧ}_L$ to serialize them when identifying the mode at sampling $i + 1$.

8.1 *Definition of Minimal Distinguishing Test Fragments*

In the following we define the MDTFs for Bonfire case study based on the following arguments:

8.1.1 Normal Mode

According to model M_N, in normal mode the packets routed in the system are assumed to be transmitted without corruption along shortest routes. By definition of the MFMM the modes to be identified are mutually exclusive. To distinguish the normal mode m_N from degraded and low modes we have to take the negation over the disjunction of occurence conditions of m_D and m_L.

In Bonfire routing system, depending whether the source and destination processors in the 2D mesh are neighbors or not, the packet delivery may need either

1-hop or 2-hop routes. By assumption, each hop takes one clock cycle. To simplify the test construction, assume that packets are sent by processors to addresses which need 1- and 2-hop routes equally.

Given the sampling period is N clock cycles and the probability of transition to degraded mode during that period is p_d we can calculate the packet corruption probability p_{D1} during transmission.

Let the packet corruption probability during transmission be p_{D1}. Since both 1- and 2-hop transmissions should be equally taken into account in the calculation of packet corruption probability, we sum them getting $p_{D1} = p + (1 - p)p = 2p - p^2$, where p is the probability of one packet corruption in one hop. Solving this equation for $p_{D1} = 0.1571$ under constraint that solution is within interval $[0, 1]$, we get $p = 0.0819$. It means that the packet corruption occurs in average once during $\frac{1}{p} = 12.2$ 1-hop transmissions. Since processors can emit at most one packet in a clock cycle and there are 4 processors that can emit packets simultaneously the average number 12.2 of 1-hop transmissions can be overapproximated with 16 without increasing the test duration. Thus, the test $\widehat{\text{Ŧ}}_N$ needs to perform in average 4.4 = 16, 1-hop transmissions for packet corruption to occur. Note also that the overapproximation increases the confidence of mode identification decision made based on that test.

To exclude possible side effects due to packet buffering if packets collide in routing we choose the test case such that packet collisions are excluded. For that the packets are addressed by all source processors alternatively to their left and then to their right neighbors. This forces the transmissions test to cover each link with 1-hop packet transmissions 2 times. In addition to 4 clock cycles during which 16 transmissions over links are made, 2 more cycles are needed, one to send the initial packets from processors to their home routers, and the other cycle for passing the last packets from destination routers to their local processors. Thus, in total the $\widehat{\text{Ŧ}}_N$—test with 16 simultaneous transmissions needs execution of $t_N = 4 + 2 = 6$ clock cycles. The model that implements $\widehat{\text{Ŧ}}_N$ is shown in Fig. 6. Note that only the processor model needed to be updated to implement the above test case.

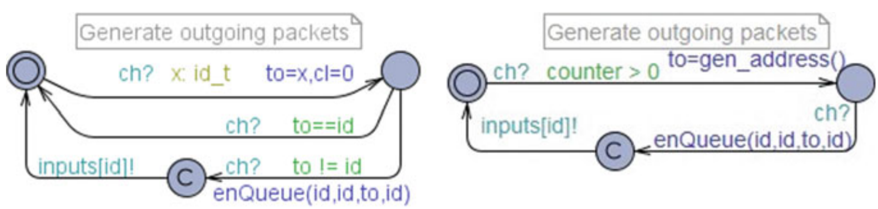


Fig. 6 Processor template modified for MDTF

8.1.2 Degraded Mode

Since the identification condition of the degraded mode m_D is symmetric to that of the normal mode m_N, the number of clock cycles needed in average for detecting the occurence of one packet corruption in transmission is the same as for m_N, i.e., $t_D = 4 + 2 = 6$ clock cycles. Just while the test $\widehat{Ŧ}_N$ returns fail, in case of packet corruption the $\widehat{Ŧ}_D$ test returns pass. Thus, the corresponding $\widehat{Ŧ}_D$ test model consists of the same update in the processor template as shown in Fig. 6, but extends the model of $\widehat{Ŧ}_N$ with an edge of packet corruption as depicted in the upper right part of Fig. 4 in the template for the router link.

8.1.3 Low Mode

In low mode one of the links between routers is expected to be faulty which means that the packets routed over this link have to be rerouted. For instance, if 1-hop packet transmission from processor to processor over operational link takes only 3 clock cycles, the broken link between the routers causes rerouting that takes all together 5 cycles. By applying same packet addressing mechanism as in $\widehat{Ŧ}_N$ and $\widehat{Ŧ}_D$ tests, one 1-hop packet has to be sent by each processor to its left and one to its right neighbor processor. If routing along the broken link will be needed latest in the third clock cycle, and rerouting takes 3 cycles, plus passing the packet from destination router to its processor takes 1 cycle, we get that in case of link fault the normal packet transmission time $1 + 1 + 1$ cycles will extend to $1 + 1 + 3 + 1$ clock cycles. If normally starting from 3rd cycle all processors receive one packet in a cycle then the link fault is detected if any of the processors does not receive a packet during 4 cycles. It confirms that it suffices sending two 1-hop packets from each test port, one to left and one to right neighbor, and detecting that either at 3rd or 4th cycle one of the output ports does not receive a packet. This is implemented in a modified model of $\widehat{Ŧ}_L$ with restriction on the upper bound of the counter variable used in the C-like function *gen_address* of edge "Generate outgoing packets" in the right template of Fig. 6.

8.2 Ordering of Test Cases by Their Priorities and Durations

Let us consider the case where Bonfire is in normal mode and transitions to degraded and low mode are possible. The probabilities of these modes are respectively $P^{nd} = 0.1571$ and $P^{nl} = 0.0127$. Thus, the probability of staying in normal mode is $P^{nn} = 1 - P^{nd} - P^{nl} = 0.8302$. These probabilities are calculated based on the description in Sect. 4.2.

Table 1 Permutations of tests, probabilities of modes and test durations indexed by permutations

No. of test suite	Order of tests in the suite	P^{nd}	P^{n1}	P^{nn}	t_1	t_2	t_3
1	$Ŧ_D, Ŧ_L, Ŧ_N$	0.1571	0.0127	0.8302	6	4	6
2	$Ŧ_D, Ŧ_N, Ŧ_L$	0.1571	0.8302	0.0127	6	6	4
3	$Ŧ_L, Ŧ_D, Ŧ_N$	0.0127	0.1571	0.8302	4	6	6
4	$Ŧ_L, Ŧ_N, Ŧ_D$	0.0127	0.8302	0.1571	4	6	6
5	$Ŧ_N, Ŧ_D, Ŧ_L$	0.8302	0.1571	0.0127	6	6	4
6	$Ŧ_N, Ŧ_L, Ŧ_D$	0.8302	0.0127	0.1571	6	4	6

Table 2 Permutations of tests, P_i, T_i—probabilities and durations of mode identification (indexed by permutations), $\overline{T}$—average mode identification time by suite

No. of test suite	Order of tests in the suite	P_1	P_2	P_3	T_1	T_2	T_3	$\overline{T}$
1	$Ŧ_D, Ŧ_L, Ŧ_N$	0.1467	0.1252	0.7281	6	10	16	13.7820
2	$Ŧ_D, Ŧ_N, Ŧ_L$	0.4686	0.4626	0.0688	6	12	16	9.4638
3	$Ŧ_L, Ŧ_D, Ŧ_N$	0.1985	0.6342	0.1673	4	10	16	9.8129
4	$Ŧ_L, Ŧ_N, Ŧ_D$	0.2239	0.7151	0.0610	4	10	16	9.0231
5	$Ŧ_N, Ŧ_D, Ŧ_L$	0.1869	0.5971	0.2160	6	12	16	11.7424
6	$Ŧ_N, Ŧ_L, Ŧ_D$	0.1730	0.5527	0.2743	6	10	16	10.9536

The experimental data collected in Table 1 and used for calculating formulas (1)–(3) provide average mode identification times for tests permutations as shown in Table 2. It can be seen that executing MDTF tests in the order $\widehat{Ŧ}_D, \widehat{Ŧ}_N, \widehat{Ŧ}_L$(test suite 4) provides minimum average mode identification time 9.0231 clock cycles that differs clearly from the arithmetic mean 10.7963 clock cycles calculated over all the permutations in Table 2. Second observation from the experiments is that by increasing the variance of individual test durations and mode probabilities will increase also the average of the absolute deviation and that makes the gain of applying the proposed method more feasible for modes that have higher variance of probability and individual tests duration.

9 Bonfire MPSoC Online Testing Harness

In the case of online test generation the model is executed in lock step with the IUT. The latter is represented by a hardware RTL implementation of the NoC communication part of the Bonfire MPSoC wrapped by a testbench and simulated in the Mentor Graphics ModelSim environment [27]. The communication between the model and the IUT involves controllable inputs of the IUT and observable outputs of the IUT. For UPTA there exists the online MBT tool Uppaal Tron [28].

The UPTA models developed in Sect. 5 for conformance testing of Bonfire MPSoC execution modes consist of IUT which is the routing system and the environment of IUT which consists of processors. Interactions between the environment and IUT constitute the symbolic representation of the test run against which the conformance of real I/O sequences are checked. Since we have a different IUT model for each mode the test cases concern only those models that correspond to the mode. As mentioned before the models of environment, i.e., the processors are the same in all test cases.

A visualization of the test setup for our case study is provided in Fig. 7. First, the analysis on availability Markov Model for the case-study implementation is performed and probabilities for switching to the next execution mode are calculated resulting in the execution mode hypotheses. Uppaal Tron is used as the test generation and execution engine and dTron [29] as the adapter generation framework. In Uppaal Tron environment, the test model represents symbolically the interaction between IUT and its environment components. The test adapter provided by dTron is responsible for translating symbolic I/O of the test model to communication packets passed to the IUT in the ModelSim environment and mapping the packets received at the destination processor back to model symbolic output. Next, Uppaal Tron is responsible for evaluation of these symbolic outputs and deciding if the execution mode test has passed or failed, thus either confirming the current execution mode hypothesis or initiating another iteration with next mode hypothesis and its test model.

The test configuration used in the current work consists of test execution environment dTron and one test adapter per processor-router link that transforms abstract input/output symbols of the model to input/output data of the IUT ModelSim interface. Uppaal Tron is used as a primary test execution engine which simulates symbolic interactions between the IUT and its environment. The interactions are monitored during model execution. When the environment (one of the processors in the model) initiates an input action i Tron triggers input data generation in the adapter and the actual test data is written to the IUT ModelSim interface. In response to that, the receive event at destination processor port

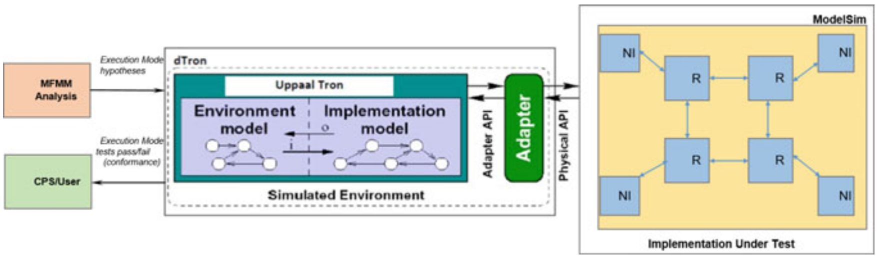


Fig. 7 The test setup involving the Uppaal Tron test engine and the distributed adapter library dTron

produces output data that is transformed back to symbolic output o. Thereafter, the equivalence between the real output returned and the output o specified in the model is checked. The run continues if there is no conformance violation, i.e. there exists an enabled transition in the model with parameters equivalent to those passed by the IUT. In addition to input/output conformance, the real-time input/output conformance is checked by Uppaal Tron.

Regarding measurement of power consumption during online testing we need to take into account the total power consumption of a CMOS chip which can be split into two components, i.e. the static power and dynamic power. The first one is due to the leakage current dependent on the implementation technology and the environmental parameters such as the supply voltage and can be simplified to a constant over time for the presented approach. Dynamic power is highly correlated to the (transistor) switching activity of the chip [30]. Its estimation can be done by taking into account the operating frequency of the chip and the number of signal toggles. In our approach we use Bonfire design RTL simulation in ModelSim with Toggle Coverage analysis enabled. The latter reports the total number of all signal toggles occurred during the application time of a particular test or a set of tests.

Our experiments have reported roughly similar switching activity per time unit for all individual tests. Therefore the ratio of power consumption between the tests can be approximated to the ratio of the test lengths.

10 Conclusions

In this chapter we have addressed the online monitoring and model-based mode identification techniques for mission critical systems. The goal of this approach is to reduce the active testing time and related to that energy consumption both being critical factors in autonomous missions. We propose an integrated approach where the MFMM are used for specifying the system reliability and quality related behavior on high-level of abstraction relevant for Markov analysis, and the more concrete state and timing constraints related with MFMM are specified explicitly using UPTA for stochastic modeling and model-based test generation. To interrelate these two model classes we have shown the mapping from MFMM to UPTA. The second contribution is the test case selection mechanism for online testing where the test cases are prioritized by the probabilities of modes switching during system execution, probabilities that modes will be identified during conformance testing and durations of mode tests. The hypotheses on which mode the system switches next are provided by MFMM and the hypotheses are tested for conformance between UPTA models that specify the mode behavior and that of IUT.

The approach is illustrated with the Bonfire MPSoC routing system where three modes and their tests are characterized with mode transition probabilities and test durations respectively. The results confirm that serializing mode tests by these characteristics allows reducing average mode identification time during active mode sampling. Second improvement along this line is reducing individual tests themselves by extracting from them minimal fragments needed for confirmation of mode hypothesis. The approach has been illustrated by a set of UPTA models that refines the behavior of MFMM modes and applied thereafter as a test suite for model-based conformance testing.

As future work, more practical experiments are planned to quantitatively support the presented analytical reasoning and to estimate advantages of the proposed MMFM guided online testing under operational conditions and scenarios for realistic MCSs.

Acknowledgements This research was supported by the Estonian Ministry of Education and Research institutional research grant no. IUT33-13, in part by project H2020 RIA IMMORTAL, and by European Union through the European Structural and Regional Development Funds.

References

1. Veanes, M., et al.: Formal methods and testing. Chap. Model-based Testing of Object-oriented Reactive Systems with Spec Explorer, pp. 39–76. Springer, Berlin, Heidelberg (2008)
2. Kharchenko, V., Illiashenko, O.: Concepts of Green IT Engineering: Taxonomy, Principles and Implementation, pp. 3–19. Springer International Publishing, Cham (2017). https://doi.org/10.1007/978-3-319-44162-7_1
3. Miyase, K., Sauer, M., Becker, B., Wen, X., Kajihara, S.: Search space reduction for low-power test generation. In: Proceedings of the 22nd Asian Test Symposium (ATS), 2013. IEEE (2013)
4. Vain, J., Tsiopoulos, L., Kharchenko, V., Kaur, A., Jenihhin, M., Raik, J.: Multi-fragment Markov model guided online test generation for MPSoC. In: Ermolayev, V., Bassiliades, N., Fill, H.G., Yakovyna, V., Mayr, H.C., Kharchenko, V., Peschanenko, V., Shyshkina, M., Nikitchenko, M., Spivakovsky, A. (eds.) ICT in Education, Research and Industrial Applications: Integration, Harmonization and Knowledge Transfer: Proceedings of the 13th International Conference, ICTERI 2017, pp. 594–607. CEUR Workshop Proceedings RWTH Aachen University (2017)
5. Trivedi, K.S., Kim, D.S., Roy, A., Medhi, D.: Dependability and security models. In: Proceedings of 7th International Workshop on Design of Reliable Communication Networks, 2009. DRCN 2009. IEEE (2009)
6. Bulychev, P.E., et al.: UPPAAL-SMC: statistical model checking for priced timed automata. In: Wiklicky, H., Massink, M. (eds.) Proceedings 10th Workshop on Quantitative Aspects of Programming Languages and Systems (2012)
7. Bonfire project wiki: URL https://github.com/Project-Bonfire/Bonfire/wiki (2017)
8. Kemeny, J.G., Snell, J.L.: Finite Markov chains. Springer, New York (1976)
9. Aupperle, B., Meyer, J.: State space generation for degradable multiprocessor systems. In: Fault-Tolerant Computing, 1991. FTCS-21. Digest of Papers, Twenty-First International Symposium. IEEE (1991)
10. Chiola, G., Dutheillet, C., Franceschinis, G., Haddad, S.: Stochastic well-formed colored nets and symmetric modeling applications. IEEE Trans. Comput. **42**(11), 1343–1360 (1993). https://doi.org/10.1109/12.247838

11. Fesenko, H., Kharchenko, V., Brezhnev, E., Zaitseva, E., Levashenko, V.: Markov Models of Smart Grid Digital Substations Availability: Multi-Level Degradation and Recovery of Power Resources Issues, pp. 339–355. Springer International Publishing, Cham (2017). https://doi.org/10.1007/978-3-319-55595-9_16
12. Socci, D., Poplavko, P., Bensalem, S., Bozga, M.: Modeling Mixed-Critical Systems in Real-time BIP. In: 1st Workshop on Real-Time Mixed Criticality Systems. Taipei, Taiwan (2013). URL: https://hal.archives-ouvertes.fr/hal-00867465
13. Negrean, M., Klawitter, S., Ernst, R.: Timing analysis of multi-mode applications on AUTOSAR conform multi-core systems. In: Proceedings of the Conference on Design, Automation and Test in Europe, DATE'13, pp. 302–307. EDA Consortium, San Jose, CA, USA (2013)
14. Whittaker, J.A., Thomason, M.G.: A markov chain model for statistical software testing. IEEE Trans. Softw. Eng. **20**(10), 812–824 (1994). https://doi.org/10.1109/32.328991
15. Kashyap, A., Holzer, T., Sarkani, S., Eveleigh, T.: Model based testing for software systems: an application of markov modulated markov process. Int. J. Comput. Appl. **46**(14), 13–20 (2012)
16. Siegl, S., Dulz, W., German, R., Kiffe, G.: Model-driven testing based on markov chain usage models in the automotive domain. In: Waeselynck, H. (ed.) 12th European Workshop on Dependable Computing, EWDC 2009, Toulouse, France (2009)
17. Haghbayan, M.H., Rahmani, A.M., Miele, A., Fattah, M., Plosila, J., Liljeberg, P., Tenhunen, H.: A power-aware approach for online test scheduling in many-core architectures. IEEE Trans. Comput. **65**(3), 730–743 (2016)
18. Tretmans, J.: Testing concurrent systems: a formal approach. In: Baeten, J.C.M., Mauw, S. (eds.) CONCUR'99 Concurrency Theory: 10th International Conference Eindhoven, The Netherlands, 1999 Proceedings, pp. 46–65. Springer, Berlin, Heidelberg (1999)
19. Brinksma, E., Tretmans, J.: Testing transition systems: an annotated bibliography. In: Cassez, F., Jard, C., Rozoy, B., Ryan, M.D. (eds.) Modeling and Verification of Parallel Processes: 4th Summer School, MOVEP 2000, France, 2000 Revised Tutorial Lectures, pp. 187–195. Springer, Berlin, Heidelberg (2001)
20. Li, H., Lam, C.P.: Using Anti-Ant-like Agents to Generate Test Threads from the UML Diagrams, pp. 69–80. Springer, Berlin, Heidelberg (2005)
21. Veanes, M., Roy, P., Campbell, C.: Online testing with reinforcement learning. In: Havelund, K., Núñez, M., Rou, G., Wolff, B. (eds.) Formal Approaches to Software Testing and Runtime Verification: First Combined International Workshops, FATES 2006 and RV 2006, Seattle, WA, USA, August 15–16, 2006. Revised Selected Papers, pp. 240–253. Springer, Berlin, Heidelberg (2006)
22. Vain, J., Kaaramees, M., Markvardt, M.: Online testing of nondeterministic systems with the reactive planning tester. In: Petre, L., Sere, K., Troubitsyna, E. (eds.) Dependability and Computer Engineering: Concepts for Software-Intensive Systems, pp. 113–150. IGI Global, Hershey, PA (2012). https://doi.org/10.4018/978-1-60960-747-0.ch007
23. Kharchenko, V., Ponochovnyi, Y., Boyarchuk, A.: Availability Assessment of Information and Control Systems with Online Software Update and Verification. Springer International Publishing (2014)
24. Guimaraes, A.P., Oliveira, H.M.N., Barros, R., Maciel, P.R.: Availability analysis of redundant computer networks: a strategy based on reliability importance. In: IEEE 3rd International Conference on Communication Software and Networks (ICCSN), 2011. IEEE (2011)
25. Kharchenko, V., Odarushchenko, O., Odarushchenko, V., Popov, P.: Selecting mathematical software for dependability assessment of computer systems described by stiff markov chains. In: Proceedings of 9th International Conference on ICT in Education, Research and Industrial Applications: Integration, Harmonization and Knowledge Transfer (ICTERI 2013), vol. 1000, pp. 146–162. CEUR-WS (2013)

26. Behrmann, G., David, A., Larsen, K.G.: A tutorial on Uppaal. In: Bernardo, M., Corradini, F. (eds.) SFM-RT 2004. LNCS, vol. 3185, pp. 200–237. Springer, Berlin (2004)
27. Modelsim ASIC and FPGA Design—Mentor Graphics: URL https://www.mentor.com/products/fv/modelsim/ (2017)
28. Uppaal Tron: URL http://people.cs.aau.dk/marius/tron/
29. dTron: URL https://cs.ttu.ee/dtron/
30. Patrick Girard Nicola Nicolici, X.W. (ed.): Power-Aware Testing and Test Strategies for Low Power Devices. Springer, US (2010)

Decrease of Energy Consumption of Transport Telecommunication Networks Using Stage-by-Stage Controlling Procedure

Gennady Linets and Sergey Melnikov

Abstract A new approach to the development of advanced automatic monitoring system and adaptive control of transport telecommunication network, allowing to reduce energy consumption of switching centers during the analysis and identification of faults and failures in equipment operation is offered in this paper. Energy reduction is achieved due to the use of step-by-step procedure of finding out abnormal situations, errors optimization of identification system states, the use of multi-agent and permission of agents dependent on hierarchy level of transport telecommunication network.

Keywords Transport telecommunication network · Energy consumption reduction Type I and Type II errors · Control · Step-by-step procedure

1 Introduction

Today the energy efficiency reduction and introduction of energy-saving technologies into various spheres of human activity is one of the key problems for modern civilization. Conducted studies in Green-technologies sphere allow to make conclusion, that actually a great attention is paid to the given studies [1–3]. The scholars such as Bianzino, Chaudet, Rossi, and Rougier studied Green technologies problems for wired communication networks upon following directions: information rate adaptation, proxy and energy interfaces, and energy compatible applications [1].

Such authors like Bolla, Bruschi, Davoli, and Cucchetti studied the possibility of energy consumption monitoring in the Internet [4]. Bianzino et al. [5] suggested to improve the energy efficiency of the Internet resources by optimizing energy

G. Linets (✉) · S. Melnikov
North-Caucasus Federal University, Stavropol, Russian Federation
e-mail: kbytw@mail.ru

© Springer Nature Switzerland AG 2019

V. Kharchenko et al. (eds.), *Green IT Engineering: Social, Business and Industrial Applications*, Studies in Systems, Decision and Control 171,
https://doi.org/10.1007/978-3-030-00253-4_13

consumption of routers, dependent on the real load of served data traffic. Zhang et al. [6] carried out a review of energy-saving technologies condition as well as problems of energy efficiency standardization of optical networks.

Lange et al. [7] made a prediction of power consumption future distribution for broadband universal operator of cellular network.

Chabarek et al. [8] have proposed to take into account power consumption at the stage of transport telecommunication networks design and planning of access networks.

Panarello et al. [9] introduced the concept "green router" (Grouter), under which they meant a router with accumulation control functions along with physical level of source/power, and a measuring device was added to [10] in the router that estimates the minimum accessible volume of served traffic and consumed energy at that.

However, these papers do not consider the problem of energy consumption reduction by transport telecommunication (TN) networks equipment owing to the more effective detection of abnormal situations arising in event of faults and failures in the equipment that it is of priority importance for telecommunications practice. A huge information traffic is generated by abnormal situations, which is to be processed by monitoring and control system at a stated time in order to identify the emerged problem that requires maximum energy costs in its traditional formulation.

Firstly, among the known methods of abnormal situations identification we should point out the methods based on estimation of attribute values distribution density. Subjects of research are considered as implementations of multidimensional random variables, distributed in attribute space by a definite law. They are based on Bayesian scheme of taking decision with the highest accuracy of clarification, and are reduced to the definition of plausibility ratio in the field of multi-dimensional space when separating boundaries are constructed between them.

The main difficulties of in the using of these methods are [11]:

- need of storing all training set to calculate estimations of local densities probability distribution;
- high response to non-representativity of training set.

It means that in the decision-making process it is required that all devices accessible for measuring should be constantly switched on. The information, which requires the switching on and functioning under voltage of all network available measuring tools, is used for the abnormal situations identification.

The given work offers a new approach to the development of advanced automatic monitoring system and adaptive control of transport telecommunication networks, which allows reducing its energy consumption during the analysis and identification of emerging faults and failures in operation. Equipment energy reduction by the system of automatic monitoring and adaptive control is achieved:

– by means of abnormal situations gradual detection use, on assumption that all of measuring devices are switched on as far as they are identified;
– by means of multi-agent systems use and authorizing the agents dependent on the hierarchy level of transport telecommunication network.

2 State of the Art

The importance of TN automatic monitoring system and adaptive control development is defined by the complexity of current tasks by assessment of their current conditions and energy costs associated with it. The main difficulty emerging in the design process is to ensure stability properties and required monitoring quality in all operation modes. This is due to that at changing of the structure or dynamic properties of subsystems, the TN parameters and energy consumption are changed too. Characteristics of control and disturbance actions can also be changed over a wide range.

TN is characterized by a great uncertainty of working conditions. The details about their real setting for some of its parameters are inaccurate, while their potential change patterns can be often known at the quality level only. In some cases, the external effects and some TN parameters are to be described statistically. At a large degree of uncertainty, the procedure of TN management system synthesis should stipulate the simultaneous solution of problems of an effective actuating unit and rational (in some sense, optimal) control algorithm using both a priori and a posteriori the information obtained in the course of its operation.

Transport network management system consists of two basic parts: monitoring object and management system. The functions of management system are the following: identification of TN status; generation of control actions on the basis of control goals taking into account TN status and the environment.

TN status identification is closely linked to the concept of pattern recognition. Therefore, the possibility of applying a pattern recognition system at the stage of its state identification seems to be quite obvious and natural. Deterministic mathematical models that uniquely and simply define control actions if there are certain external disturbances are used in the systems of controlled object state identification and control actions generation in previously developed and modern automation systems. Moreover, the problem of how disturbing factors parameters are linked with one or another states of a controlled object is not analyzed and solved as well [11]. In general, the relationship between the observed parameters of TN and their states is of dynamic and probabilistic nature. Traditional network resource management systems are in substance parameters control systems, i.e. systems that control not the TN states but its observable parameters. It was proved that for the effective management of complex systems, the management of states is required but not of parameters.

To increase functioning quality it is required that automatic monitoring system of states should be adaptive. In this case, the term "adaptive system" means the approach, when the lack of a priori information is compensated due to (in comparison with non-adaptive system) fuller using of current information. In other words, to reduce the uncertainty degree and to achieve preset parameters of information exchange quality, management processes, a purposeful forced change of parameters and operating device structure is carried out based on the current (a posteriori) information. Thus, adaptation is the process of purposeful change of parameters, properties or transport network structure based on the information obtained in the course of its main functions execution in order to achieve its optimality state in a changing environment at the initial uncertainty.

Some methods of state estimation, identification and technical systems management are known today. For instance, it is known a stochastic control method which is based on the optimal control in the context of random interferences at which, according to the specified characteristic of a certain system and to familiar probabilistic characteristics of interferences, a control algorithm is defined ensuring the extremum of selected quality criterion. This method is based on the principle of distribution, which allows showing a stochastic control problem represented by a combination of two tasks: an optimal estimation and identification of the system status and an optimal deterministic control. The measuring system, which implements this method, submits to the resolver all available for measuring information concerning the status of the system, which is influenced by some random perturbations and it defines the maximum power consumption. The control algorithm is formed on the basis of this information.

In the task of optimal stochastic control, the controlled object is set by state equation. As a rule, the optimal stochastic controller is like a Kalman filter for status assessment and deterministic optimal controller with stochastic input. For putting into effect this method, it is necessary to have information about signal correlation functions of error that makes practically impossible its using for monitoring a distributed system, because in this case such information is not available. In addition, it is impossible to obtain and estimate de facto the extent of information amount reduction required for distributed systems management. The disadvantage of this method is the fact that for its implementation it is necessary to be constantly armed with all available for measuring information. This means that all measuring devices of identification attributes should be permanently switched on and be in working condition.

However, there is one disadvantage in this control method. All available for measuring information is constantly used during the process of controlling action generation, which in distributed system it not only saturates communicative channels with transit data flow but also leads to the situation that during the network functioning all measuring tools are under voltage. The system does not have a control loop associated with the influence on the source of disturbances, which can result in the network blockage (blocked alignment).

It is known the agent-oriented approach that allows coming to a higher level of management. The agent is a stand-alone program, which interacts not only with other agents but also with the environment, in order to obtain a single solution of the problem. The task can be created both by the user and by the agent itself, which is able to plan their actions based on the analysis of the information received. This approach can be used on different levels of the automatic control systems hierarchy. However, its usage for TN assumes a large amount of management information generated by the information exchange by agents of various levels. Therefore, the agent-oriented approach should be used only after taking measures to eliminate abnormal TN situations.

3 An Approach

We used the approach, which joins the collection of information for resource management and TN status identification based on simultaneous application of step-by-step detection of abnormal situations and multi-agent system. The reduction of management information volume as well as energy consumption is taken place due to authorized agents dependent on the level of TN hierarchy. Agents are allotted a task of monitoring not only the state of TN separate objects but also inter-layer information exchange. It is expected to use the agents of two types: agents with simple behavior and learning agents. Agents with simple behavior act only on the basis of current knowledge. Their function is based on the IF scheme (condition) THEN (the action) Fig. 1.

Learning agents Fig. 2 are independent and able to learn and adapt to changing conditions. These agents perform the following functions: training and development in the process of environment interaction; adaptation to environment changes in real

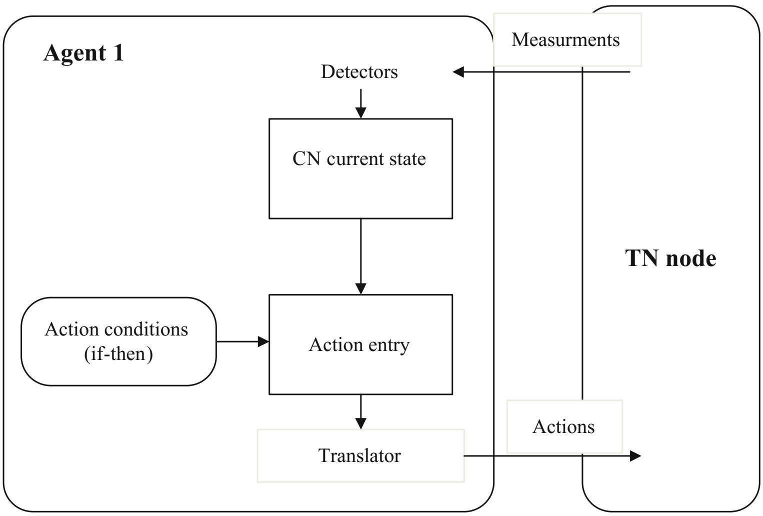


Fig. 1 Structure of the agent with the simple behavior

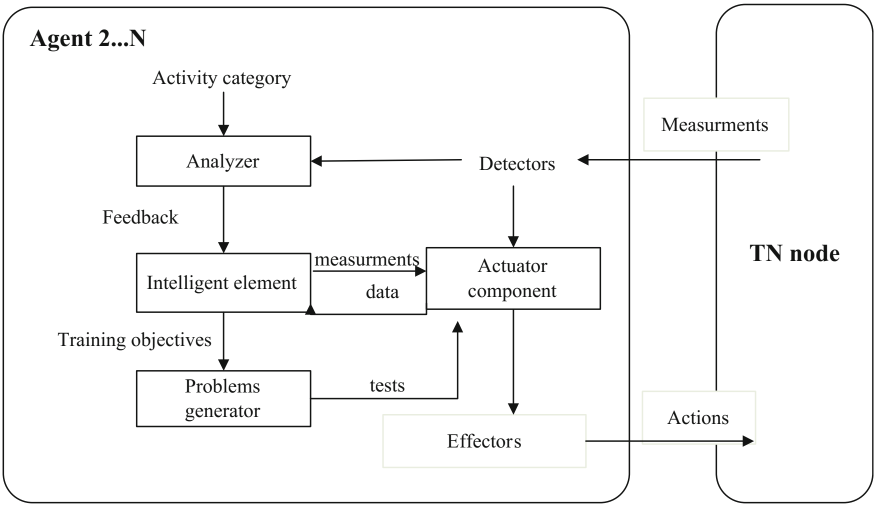


Fig. 2 Learning agents structure

time; adaptation to new ways of evolving problem solution; maintaining a database with the possibility of its acquisition; analysis of their behavior in terms of error and success.

Multi-agent systems (MAS) are used to solve problems, which can't be solved with a single agent. MAS possess the following features: autonomy, independence of agents; limitation of vision, agents have no vision about the whole system; decentralization, there are no agents to control the entire system. Since MAS have shown them- selves to be useful in telecommunications sphere, they can be used also to provide high capacity, scalability and recoverability of TN. In the case of MAS, the using the energy consumption is determined by the type of using agents, the appliance of which is defined by the complexity of solving tasks.

Figure 3 offers a functional diagram of automatic monitoring system and adaptive control of distributed systems status. It consists of: a transport network presented as an hierarchical structure comprised of *N* levels; an smart converter; a measuring and clustering unit; an identification unit; a registration and management unit; a source of disturbances; a bus of switching nodes interworking; a bus of issuing a priori information as well as a communication bus with measuring center of network.

The transport network includes switching nodes of different levels (1 … *N*) (Fig. 3).

A smart transmitter forms and sends out intellectual agents of proper level, which penetrate into network structure and coordinate the activity according to vested powers. At first stages of TN status identification the agents with a simple behavior (Fig. 1) are used, while at the following stages—learning agents (Fig. 2), which are carried out the above mentioned functions.

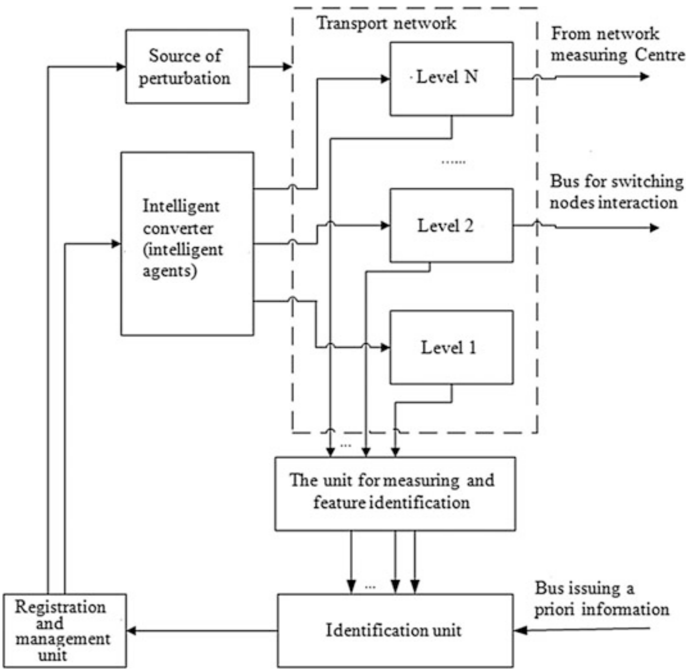


Fig. 3 Functional scheme of system for TN automatic control and adaptive management (agent-oriented programming)

The measuring and feature selection unit is linked with its inputs to the distributed system output (Fig. 3). Its outputs (1 … *N*) are linked to the identification unit inputs.

The identification unit determines TN status transmitting the data to the registration and management unit. If this unit detects TN abnormal condition, the registration and management unit affects on smart mode changer that sends intellectual agents to clear up the reasons of abnormal conditions and get TN into normal condition. Simultaneously a controlling impact is made on the source of random perturbations in order to eliminate the cause of abnormal situation.

The registration and management unit is designated to collect, store and analyze statistical data regarding the states at all levels as well as TN management and transfer of such data to the network measuring center (via TN and smart transmitter).

The source of disturbances is the reason of abnormal situations in TN.

The possibility of TN units interaction as well as intellectual agents interaction is provided by using the bus of switching nodes. The registration and management unit makes decision concerning TN state taking into account the a priori data transmitted through the respective bus (Fig. 3).

Intellectual agents are sent out addressed through the network only after the detection of abnormal situation. The purpose of their distribution is to identify the reasons of interaction and localize abnormal situation.

The decision on the TN status is made at the first stage based on a local information analysis available for measurement in the first level nodes. At the second and next stages, the decision about TN network is taken on the basis of the additional information analysis obtained from the bus of switching nodes interaction (information exchange with adjacent nodes). At the last stage, the identification system receives extra information from the network measurement center (NMC), which allows using all the information available for measurement in order to make a grounded decision.

4 Problem Statement

In general, the task aimed to identify objects could be reduced to the verification of numerous hypotheses $H_1, H_2, \ldots, H_i, \ldots, H_n$, where H_i is a hypothesis implying the object's belonging to Class A_i. Let's assume that the a priori distributions of these hypotheses probabilities are set. It is known what's the likelihood the object $P(H_i)$ can belong to class A_i (or how often the object of this class is appeared). Moreover, $\sum_{i=1}^{n} P(H_i) = 1$, as the object is to be pertained to a certain class. The conditional density of distribution is $p_i(x) = p(x_i/H_i)$.

Two hypotheses $H_1 = N$ and $H_2 = \overline{N}$ are used in the identification system under design at corresponding to them a priori probabilities of situation, emerging in the network as a normal $p_1 = p(H_1) = p(N)$ one and an abnormal $p_2 = p(H_2) = p(\overline{N})$. And also $p_1 + p_2 = 1$.

It is required to find a decision rule ensuring the top accuracy in the identification system. Using the Neyman-Pearson criterion we fix the probability of «false alarm» $P_{f.a.}$ at stable level C and claim the minimum of pass error $P_{pass}^{\min}$ of TN operating trouble. Then

$$P_{pass}^{\min} = p_2 \left[1 - \prod_{i=1}^{n} \bar{\beta}(x_{oi}) \right] \tag{1}$$

at restriction of $P_{f.a.} = p_1 \prod_{i=1}^{n} \alpha(x_{oi}) = C$, const, where $\alpha(x_0)$ are Type I errors; and $\beta(x_0)$ are Type II errors.

In order to improve the energy efficiency of transport networks it is necessary to determine the possible degree of decline in the control information ρ_c, which is to be transmitted between nodes to specify the type of abnormality controlled conditions and using a MAS.

5 Problem Solution

5.1 Defining of Decision Rule for Identification System

During operation of TNs the status of which could be described by a large number of parameters, performance monitoring of their working efficiency is to be carried out using stage-by-stage principle of classification.

At the first stage, the TN status is checked by the summarized index and if the abnormality is detected, a stricter control is carried out at the second and next stages on which its real status is defined.

At each of the stages the control system makes Type I $\alpha(x_{oi})$ and Type II $\beta(x_{oi})$ errors (Fig. 4). Type I and II errors («false alarm» and «abnormal situation» errors), are defined as follows:

$$\alpha(x_{oi}) = \int_{x_{oi}}^{\infty} f(x_i/N)dx_i \tag{2}$$

$$\beta(x_{oi}) = \int_{-\infty}^{x_{oi}} f(x_i/\bar{N})\,dx_i \tag{3}$$

The activity of the monitoring error evaluation system can be described by a probability graph (Fig. 5).

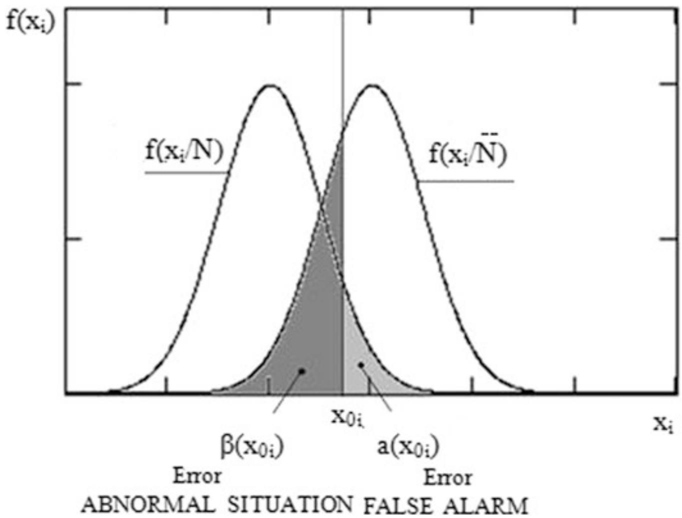


Fig. 4 Density graph for feature distribution of x_i stage

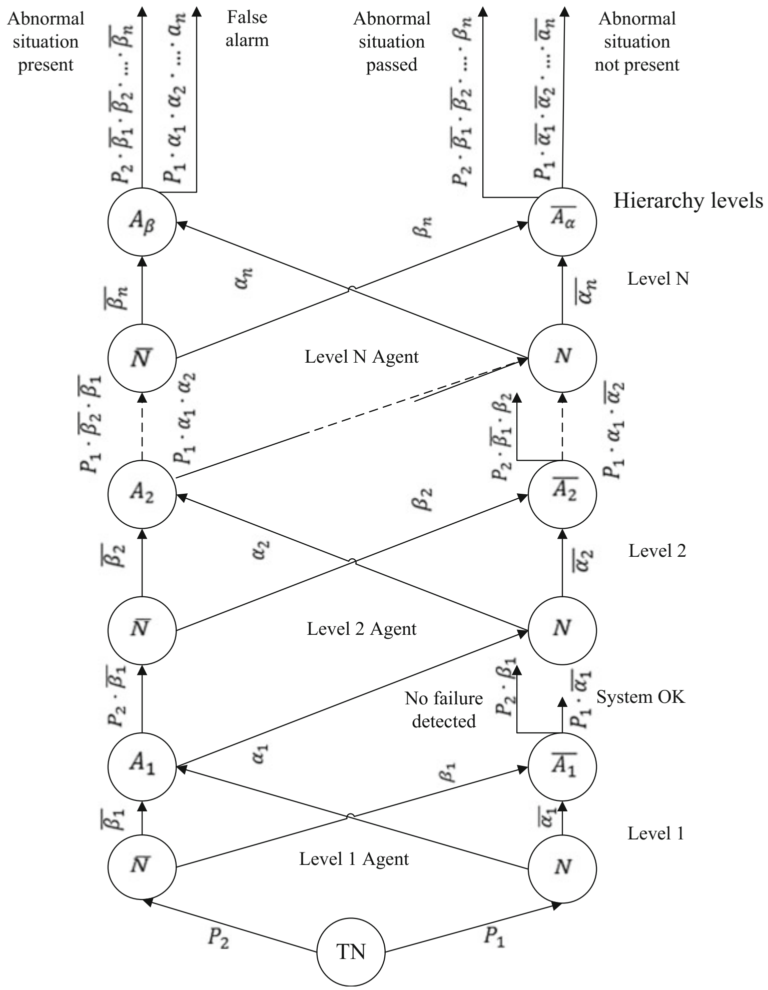


Fig. 5 Hierarchy-based principle for building agent-oriented probability graph of TN status

The TN is characterized with the certain states of N and $\bar{N}$, under which we will mean normal and abnormal operation in the process of its functioning which (in the monitoring system) are respectively displayed into the normal A and abnormal $\bar{A}$ state (Fig. 5).

A priori probabilities of the normal and abnormal states in TN are detected correspondingly by a priori probabilities of the p_1 and p_2 states. For all the stages, with abnormal situation omission, were obtained resultant errors

$$P_{pass} = p_2\beta_1 + p_2\bar{\beta}_1\beta_2 + \cdots + p_2\bar{\beta}_1\bar{\beta}_2\ldots\beta_n = p_2\left[1 - \prod_{i=1}^{n}\bar{\beta}(x_{oi})\right] \quad (4)$$

and «false alarm» errors:

$$P_{f.a.} = p_1\alpha_1\alpha_2\ldots\alpha_n = p_1 \cdot \prod_{i=1}^{n}\alpha(x_{oi}) \quad (5)$$

where: $\overline{\beta_i} = \bar{\beta}(x_{0i})$, $\alpha_i = \alpha(x_{0i})$, $p_1 = 1 - p_2$—a priori probability of a normal situation occurrence; p_2—a priori probability of its absence.

The problem of thresholding x_{oi} for each stage is currently central. Let's make tradeoffs of thresholds. Following the Neyman-Pearson criterion we will set the probability of «false alarm» at certain given C level. Then, for the entire network, we get

$$P_{f.a.} = p_1\prod_{i=1}^{n}\alpha(x_{oi}) = C \quad (6)$$

Having minimized the probability of abnormal situation omission we get

$$P_{pass}^{\min} = \min_{x_{oi}^*} p_2\left[1 - \prod_{i=1}^{n}\bar{\beta}(x_{oi})\right] \quad (7)$$

As a result, the decision rule ensuring the highest level of accuracy for the identification system is the Neyman-Pearson criterion which is detected for the TN by expressions (6) and (7).

5.2 Method of Errors Optimization of Identification System Conditions Monitoring and Determination of Optimal Thresholds Classification

5.2.1 Solution of Optimization Task for n Stages

The minimization problem of the function (7), where variables x_{oi} are linked by functional dependence (6) is the constrained optimization problem. Let's form a functional of optimization

$$F = p_2\left[1 - \prod_{i=1}^{n}\bar{\beta}(x_{oi})\right] + \lambda p_1\prod_{i=1}^{n}\alpha(x_{oi}) \quad (8)$$

where λ is an undetermined Lagrange multiplier.

Calculating the derivatives $\frac{\partial F}{\partial x_{oi}} = 0$, we get a system of n equations:

$$\begin{aligned} & p_2 \frac{\partial \bar{\beta}(x_{o1})}{\partial x_{o1}} \bar{\beta}(x_{02})\ldots\bar{\beta}(x_{0n}) - \lambda p_1 \frac{\partial \alpha(x_{01})}{\partial x_{01}} \alpha(x_{02})\ldots\alpha(x_{on}) = 0; \\ & \ldots \\ & p_2 \frac{\partial \bar{\beta}(x_{on})}{\partial x_{on}} \bar{\beta}(x_{o1})\ldots\bar{\beta}(x_{0(n-1)}) + \lambda p_1 \frac{\partial \alpha(x_{on})}{\partial x_{on}} \alpha(x_{01})\ldots\alpha(x_{0(n-1)}) = 0, \end{aligned} \tag{9}$$

which, along with Eq. (6) allow to find the undetermined Lagrange multiplier λ and n of x_{oi} variables.

Since the system (9) contains n of x_{oi} variables and undetermined Lagrange multiplier λ, which must be defined, while the equation system (9) has n equations and $n + 1$ of variables, then for finding its unambiguous solution one more equation should be added, in the capacity of which we take the tolerance probability equation for «false alarm» (6).

Equations (9) and (6) including (2) and (3) after differentiation on the lower and higher threshold take the form:

$$\begin{aligned} & \frac{\int_{-\infty}^{x_{o2}} f(x_2/\bar{N})dx_2 \ldots \int_{-\infty}^{x_{on}} f(x_n/\bar{N})dx_n}{\int_{x_{o2}}^{\infty} f(x_2/N)dx_2 \ldots \int_{x_{on}}^{\infty} f(x_n/N)dx_n} = \lambda \frac{p_1}{p_2} \frac{f(x_{o1}/N)}{f(x_{o1}/\overrightarrow{N})}; \\ & \ldots \\ & \frac{\int_{-\infty}^{x_{o1}} f(x_1/\bar{N})dx_1 \ldots \int_{-\infty}^{x_{o(n-1)}} f(x_{n-1}/\bar{N})dx_{n-1}}{\int_{x_{o1}}^{\infty} f(x_1/N)dx_1 \ldots \int_{x_{o(n-1)}}^{\infty} f(x_{n-1}/N)dx_{n-1}} = \lambda \frac{p_1}{p_2} \frac{f(x_{on}/N)}{f(x_{on}/\overrightarrow{N})}; \\ & p_1 \prod_{i=1}^{n} \int_{x_{oi}}^{\infty} f(x_i/N)dx_i = C. \end{aligned} \tag{10}$$

Under certain distribution laws and $f(x_i/\overline{N})$ this problem has a single value solution. In Eqs. (10) the unknowns are optimal thresholds for classification $(x_{o1}^*, x_{o2}^*, \ldots, x_{oi}^*, \ldots, x_{on}^*)$ at each of the stages, offering a minimum of the function (7), i.e. the minimal probability of «abnormal situation» omission. The solution clearly allows defining the probability of TN operating trouble absence (probability of normal operation):

$$\overline{P}_{pass} = 1 - P_{pass}^{\min} \tag{11}$$

At each of the following stages, the information about correct solution of $\overline{\beta}_i = 1 - \beta_i$ (information about normal TN functioning) is analyzed.

When the Type I or Type II errors are detected, their specification is carried out by means of the following stage procedure. If the situation is confirmed at the next stage, we can affirm with a high probability its real presence. In such case, a smart transmitter sends an intellectual agent of respective level to troubled unit, which operates on according to its powers, and prepares TN for service.

5.2.2 Solution of Condition Monitoring Identification Problem on the Example of a Two-Stage Procedure

Let us define the procedure of error detection by the feature x as the first stage of two-stage procedure, while for the second stage we will detect the error by the feature y. Generally, the identification problem solution is determined by Type I and Type II errors. Now let's write down the functions for distribution density of the feature x at TN problem-free functioning $f(x/I) = f_1(x)$ and at trouble functioning —$f(x/\bar{I}) = f_2(x)$. Then the errors of Type I and Type II of the detector (stage 1) are:

$$\alpha_o = \int_{x_0}^{\infty} f_1(x)dx \quad \beta_o = \int_{-\infty}^{x_0} f_2(x)dx \tag{12}$$

Similarly, Type I and Type II errors for the recognizer (Stage 2) are defined:

$$\alpha_p = \int_{y_0}^{\infty} f_1(y)dy \quad \beta_p = \int_{-\infty}^{y_0} f_2(y)dy \tag{13}$$

Transport network (system) has the following conditions: "1"—the system is out of order, the failure was not detected; "2"—the system is operational, it was found as workable; "3"—failure was detected and recognized (abnormal situation presence); "4"—the system is workable, false detection and recognition (false alarm); "5"—system is out of order, failure was detected but not recognized (abnormal situation omission); "6"—the system is in order, a false detection and correct recognition.

In view of formulae (12) and (13), formulae (6) and (7) take the form:

$$P_{f.a.} = p_1 \int_{x_0}^{\infty} f_1(x)\,dx \int_{y_0}^{\infty} f_1(y)dy = C = const \tag{14}$$

$$P_{pass}^{\min} = \min p_2 (1 - \int_{x_0}^{\infty} f_2(x)dx \int_{y_0}^{\infty} f_2(y)dy) \tag{15}$$

Since in the expression (14) the thresholds x_0 and y_0 are linked with one functional dependence $x_0 = \phi(y_0)$, having differentiated (15) by y_0 and set it to zero, we get:

$$\frac{dx_0}{dy_0} \cdot f_2(x_0) \cdot \int_{y_0}^{\infty} f_2(y)dy + f_2(y_0) \cdot \int_{x_0}^{\infty} f_2(x)dx = 0 \tag{16}$$

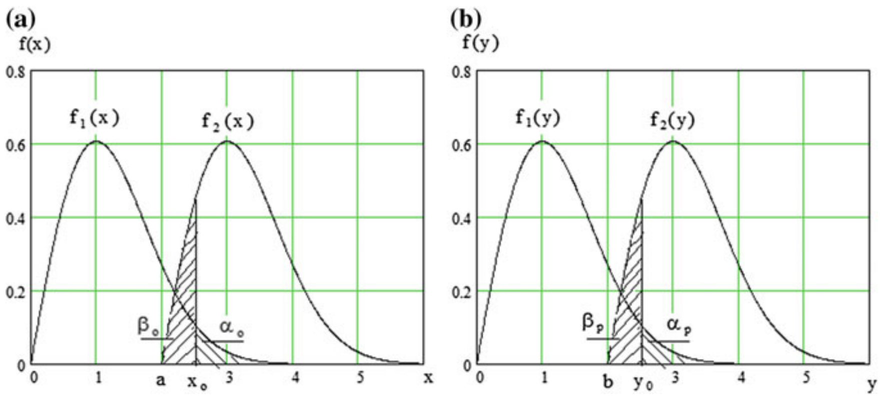


Fig. 6 Distribution density features *x* and *y*

At free-hand laws of *x* and *y* features distribution, particularly, at normal law, there is no possibility to obtain an exact solution of the classification thresholds optimization problem. However, in certain cases, at Rayleigh distribution laws, in particular, the solution could be obtained in its final form.

Let's set densities of features *x* and *y* distribution probability in the form of Rayleigh distribution laws, Fig. 6a, b. The Type I and Type II errors are in the form of:

$$\alpha_o = \int_{x_0}^{\infty} f_1(x)\,dx;\ \beta_o = \int_{a}^{x_0} f_2(x)dx;\ \alpha_p = \int_{y_0}^{\infty} f_1(y)\,dy;\ \beta_p = \int_{b}^{y_0} f_2(y)dy.$$

$$f_1(x) = xe^{-\frac{x^2}{2}};\ f_2(x) = (x-a)e^{-\frac{(x-a)^2}{2}};\ f_1(y) = ye^{-\frac{y^2}{2}};\ f_2(y) = (y-b)e^{-\frac{(y-b)^2}{2}}.$$

Equations (14) and (16) are transformed into expressions:

$$x_0^2 + y_0^2 = 2\ \ln\frac{p_1}{C} \tag{17}$$

$$\frac{dx_0}{dy_0}\cdot(x_0 - a) + (y_0 - b) = 0 \tag{18}$$

Having differentiated (17) by y_0 by substituting the result in (18), we obtain optimal classification thresholds:

$$x_0^* = \sqrt{\frac{2\ln p_1/C}{1+\left(\frac{b}{a}\right)^2}}; \tag{19}$$

$$y_0^* = \sqrt{\frac{2\ \ln\ p_1/C}{1+\left(\frac{a}{b}\right)^2}}; \tag{20}$$

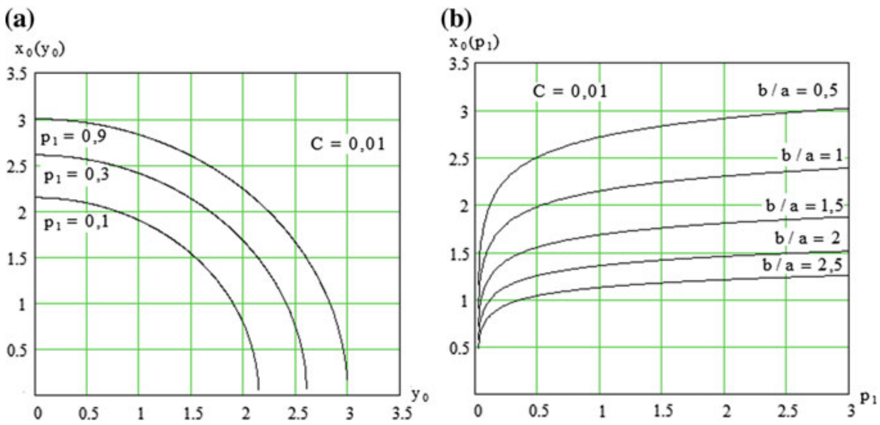


Fig. 7 Dependence curves: **a** optimal threshold of detector from optimal threshold of recognizer; **b** optimal threshold of detector from a priori probability of network's normal status

Dependences in Fig. 7a, built according to the formula (17) are circumferences of the $\sqrt{2\ \ln p_1/C}$ radius. With the increase of a priori probability of network normal status p_1, the radius is increased subject to the logarithmic law.

The dependence curves in Fig. 7b have clearly non-linear nature and show a tendency to get reduced along with a growth in the degree of the classes crossing on both features. Apparently, a similar trend is observed in other signs distribution laws.

Using (19), (20), and (15) we get the minimum probability value of «abnormal situation» omission:

$$P_{pass}^{\min} = p_2\left\{1 - C/p_1\ \exp\left[\sqrt{2(a^2+b^2)\ \ln p_1/C} - 0,5(a^2+b^2)\right]\right\} \tag{21}$$

Now let's compare the two- and one-stage procedures for detecting abnormal condition in TN by means of the respective power ratio of the solution $k_c = \overline{\beta_2}/\overline{\beta_1}$ at the same «false failure» probability.

Then for Rayleigh's law and the optimal threshold for the detector we get

$$x_0^* = \sqrt{2\ \ln p_1/C},$$

where $b = 0$ (one-stage procedure) from Eq. (21) we find out:

$$\overline{\beta}_1 = C/p_1\ \exp\left[a\sqrt{2\ \ln p_1/C} - a^2/2\right]. \tag{22}$$

Similarly, we define $\overline{\beta}_2$ for the two-stage control procedure:

$$\overline{\beta}_2 = C/p_1 \, \exp\left[\sqrt{2(a^2+b^2)\,\ln p_1/C} - 0,5(a^2+b^2)\right] \tag{23}$$

Having set the obtained result in combination with (23) in the expression for k_c, we:

$$k_c = \exp\left\{\sqrt{\ln p_1/C}\left[\sqrt{2(a^2+b^2)} - a\sqrt{2}\right] - b^2/2\right\} \tag{24}$$

The dependence curves $k_c = f(C)$ (at $a = 1$ and different values b) illustrating the value of control errors reduction at two-stage procedure and preset probability of normal operation p_1, can be seen in Figs. 8 and 10.

The advantage of k_c is defined to a large extent by a priori probability value of a system condition normal condition p_1. The higher this probability is the higher is the advantage of k_c Figs. 8 and 10. The physical meaning of parameter b change expresses the information content of feature recognition (Stage 2). The advantage of k_c depends on information content of the second-stage features and is increased along with the parameter b. The more k_c the higher the information content of used features at the situation recognition in the network (see Figs. 8, 9 and 10). k_c is growing with the decrease of «false alarm» C probability, therewith k_c is changed from one to tens of times.

The advantage of k_c depends also on features information content of the first-stage, which is grown along with the parameter α increase. Figures 11, 12 and 13.

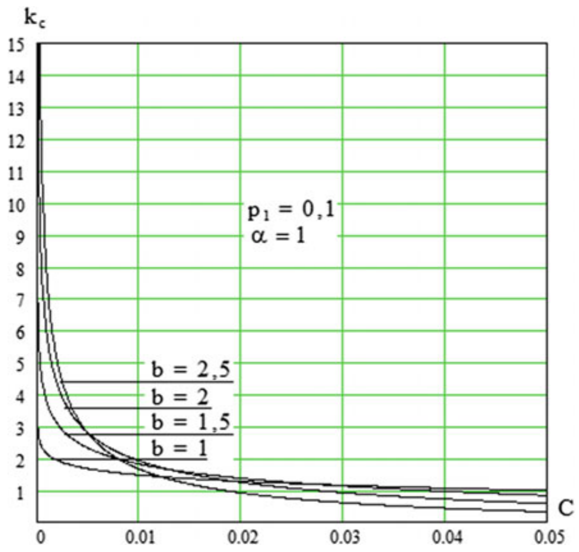


Fig. 8 Dependence of k_c on C at $p_1 = 0.1$

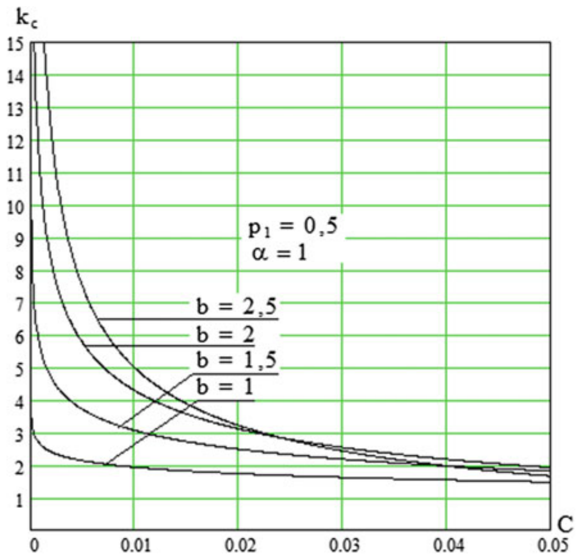


Fig. 9 Dependence of k_c on C at $p_1 = 0.5$

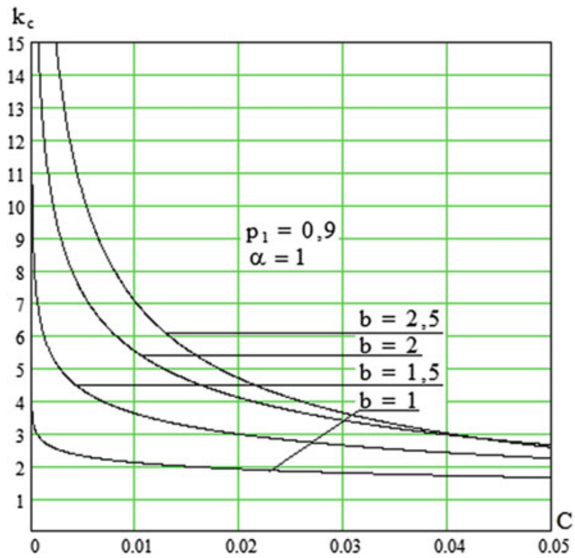


Fig. 10 Dependence of k_c on C at $p_1 = 0.9$

The value of the errors, therefore, depends on features information content, at both the first and the second stages (see formulae (19) and (20) and Figs. 11, 12 and 13).

A stage-by-stage control procedure ensures the maximum accuracy in detecting abnormal situations in the network because it uses independent recognition features at each of the stages and, therefore, is no worse than the Bayesian procedure. At

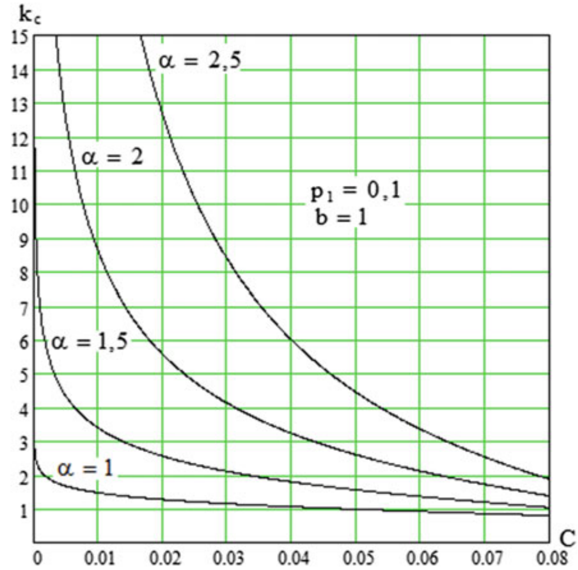


Fig. 11 Dependence of k_c on C at $p_1 = 0.1$ and different parameters of a

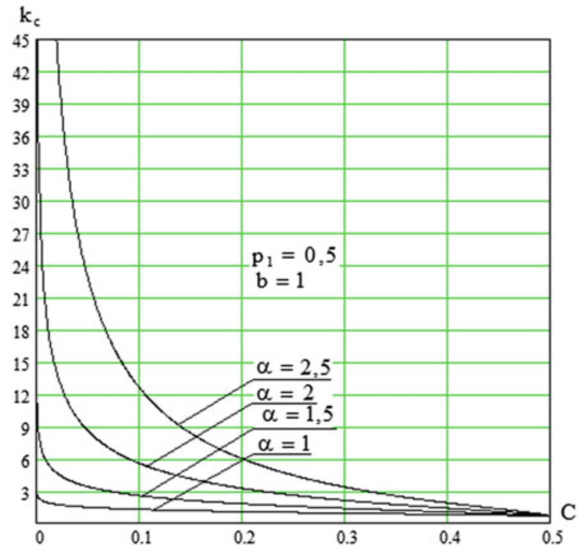


Fig. 12 Dependence of k_c on C at $p_1 = 0.5$ and different parameters of a

stage-by-stage control, the making decision on TN status is carried out with attracting extra features so far as necessary. A number of stages switching is reduced in the wake of stage number increase. The control is over when decision on TN on proper functioning is made.

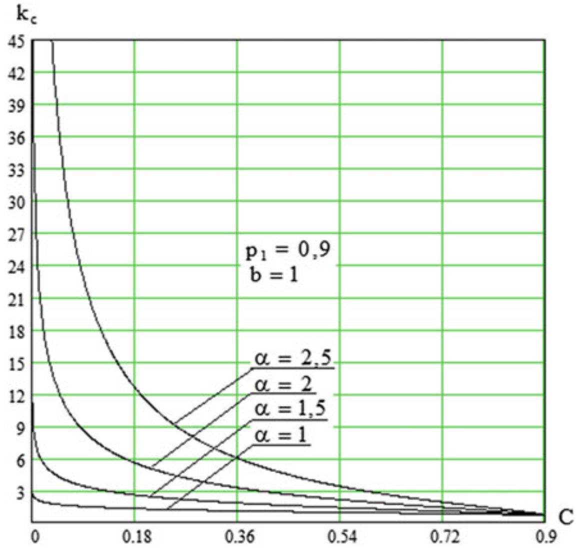


Fig. 13 Dependence of k_c on C at p_1 = 0.9 and different parameters of a

5.2.3 Determination of Intellectual Agents Sending Numbers

Proceeding from real conditions of the TN states identification system functioning, the number of intellectual agents sendings defines energy consumption and only depends on states "1" and "3" of the probability graph Fig. 6.

In state "5" (system is invalid), the failure is detected, but is not recognized—identification system does not inform network users about the abnormal situation missing and continues to work as a serviceable one. Therefore, this state is not taken into account at sending number of intellectual agents.

Therefore, the number of intellectual agents sendings is determined by states "1" and "3" of probability graph and in practice is equal to the ratio of missing probability $P_{pass} = P_2(1 - \overline{\beta_1 \beta_2})$ to the total number of identification system N switching. Keeping in mind that at the minimization $\min P_{pass} = P_{pass}^{\min}$ we get the sending number formula of intellectual agents:

$$\varepsilon = N \cdot P_{pass}^{\min} = N \cdot p_2 \left\{ 1 - C/p_1 \exp\left[\sqrt{2(a^2 + b^2) \ln p_1/C} - 0,5(a^2 + b^2) \right] \right\}.$$

Intellectual agents of the first level are agents with simple behavior and their power consumption is far less than power consumption of learning agents sent at later stages. Let us suppose that power consumption of simple agents is E_1 and learning agents power consumption is E_2. So the total power consumption by agents is $E = E_1(1 - \overline{\beta_1}) + E_2(1 - \overline{\beta_2})$.

6 Reduction of Control Information Volume Due to the Usage of Stage-by-Stage Classification

Since the solution on TN normal functioning at the first stage may be taken based on the local state information about the node condition (for example, used volume of buffer memory, the state of the outgoing channels) there is no need to the information exchange with other network nodes. This leads to a reduction of the management information circulating through the network.

For as much as the information exchange is usually performed with the neighboring network nodes and followed by the transfer of management information, so this also leads to the reduction of its volume as the information is not brought to each remote switching node.

In addition to (11) at the second and next stages that part of information which is stipulated by occurrence of “false alarm” probabilities $p_1\alpha_1, p_1\alpha_1\alpha_2, \ldots, p_1\alpha_1\alpha_2\ldots\alpha_n$.

The value $P_{f.a.} = p_1\alpha_1(1 + \alpha_2 + \alpha_2\alpha_3 + \cdots + \alpha_2\alpha_3\ldots\alpha_n)$ determines that part of the total information received per unit, which is to be analyzed on the second and subsequent stages. It determines the degree of reduction of information volume to be transmitted between the transport network nodes to specify the type of violation:

$$\rho_c = \frac{1}{p_1\alpha_1(1 + \alpha_2 + \cdots + \alpha_2\alpha_3\ldots\alpha_n)}. \quad (25)$$

Thus, the degree of volume reduction of control information circulating through the network depends on the TI error occurring at each stage. The advantage regarding the reduction of management information volume depends on the information content of features recognition at the second and next stages since the information content of feature α_1 at the first stage is a fixed value. It is defined by free buffer space capacity, the value of which can be strictly checked by local information of each certain node. However, the increase of features information content at later stages is linked with the measurements on the network, the capacity of which determines the quality of decision-making in a gradual monitoring and leads to power consumption increase. These measurements are required to improve the information content of features. They are linked with the need to attract additional measuring resources and natural increase of analysis time.

At two-stage control procedure that part of information which is in the second stage is analyzed just the part of information which is the probability of detector “false alarm” is analyzed on the second stage:

$$P_{f.a.} = p_1 \int_{x_0}^{\infty} f_1(x)dx. \quad (26)$$

The values (26) define that part of the total information flow that is to be analyzed at the second stage. It defines directly the degree of information volume reduction which is to be transmitted from node to node for violation type specification.

At Rayleigh laws of features distribution the formula (25) takes the form:

$$\rho_c = 1/p_1 \exp(-\frac{\ln(p_1/C)}{1+(b/a)^2}).$$

7 An Example

Let us suppose that power consumption, spent per volume unit of the processed management information, is a constant fixed value. Then, by the value of management information volume reduction in switching nodes ρ_c we can consider power consumption reduction expended on processing of management information. The degree of reduction of management information volume in relation to the "false alarm" C preset probability is shown in Fig. 14.

Thus, the reduction degree level is equal to 3 Fig. 14, this value is reached at C from 0.001 to 0.006 when the relative degree of classes intersection b/a from 1.1 to 0.8 and a priori probability of the normal functioning p_1 = 0.5.

When C = 0.0022 Fig. 15 (*d* point) the advantage of error k_c reduction is not less than two times when the management information volume reduction degree is equal to 3 Fig. 15 (*d* point).

Therefore, at gradual control procedure we can observe not only the advantage of power consumption but also the classification error reduction of the object states under control.

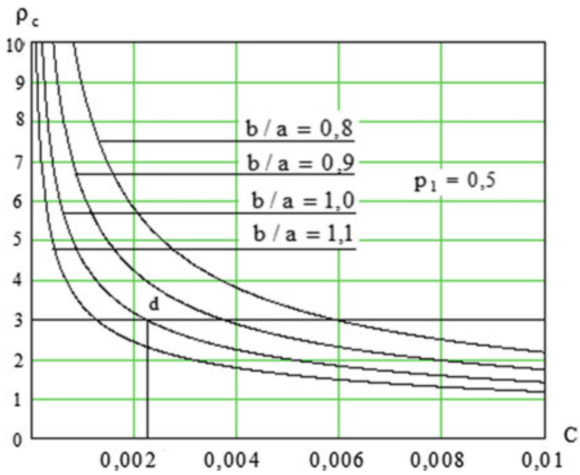


Fig. 14 Reduction degree of volume control information depending on a given probability of "false alarms"

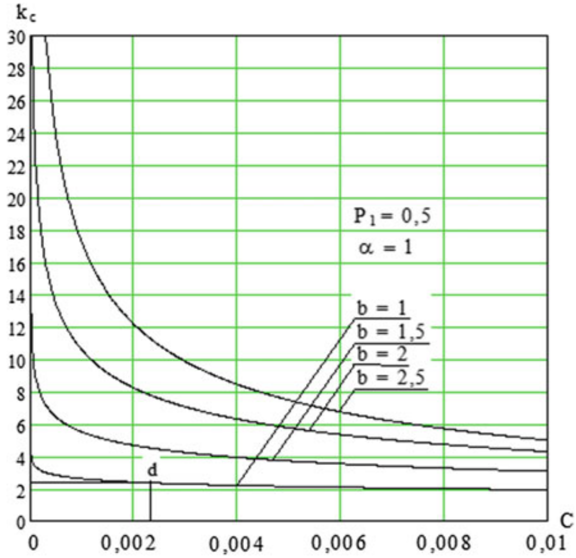


Fig. 15 Odds of reducing errors due to two-stage procedure for a given a priori probability of the normal functioning

8 Conclusion

The carried out studies allow to draw the following conclusions:

1. The reduction of power consumption by transport telecommunications network at the analysis and identification of faults, malfunctions can be achieved owing to the use of gradual detection of abnormal situations. For the analysis, only that part of features measuring tools information which belong to a given stage. At the first stage according to local information which is contained in given switch node, the operating trouble presence is defined, while on the second and next stages the degree and type of failure are clarified.
2. The problem of monitoring errors optimization and detection of classification optimal threshold for general cases was solved and the example its application to two-stage procedure of abnormal situation detection at Rayleigh features distribution laws.
3. The power consumption of automatic monitoring and adaptive control system depends on the type and form of features x and y distribution density Fig. 7. Power consumption is the more than the less information content features a (detector) and b (analyzer).
4. The degree of management information volume reduction owing to the application of a stage-by-stage principle of abnormal situations detection in the network was estimated. It is defined by a given probability of a “false alarm” and its value can be from 1.1 to 3 times. It was shown that the management information volume reduction (hence also power consumption), when using abnormal situation gradual detection, depends on the given probability of a

"false alarm" and correlations of features a and b information content. It is increased with the decrease of "false alarm" given probability Fig. 14. The advantage of k_c is determined by a priori probability value of p_1 system normal functioning (see Fig. 8, 9 and 10). The higher the probability p_1, the greater the advantage of k_c.

5. The advantage k_c depends on the features information content both the first and second stages, it grows with the increase of parameters a and b. It was shown that the advantage is equal to values of two or more times due to the application of two-stage procedure estimating it by coefficient k_c.
6. The use of multi-agent systems and vesting the agents with authorities, depending on the level of the transport telecommunication network hierarchy, allows carrying out address screening of information features and attracting additional features for analysis of origin reasons and abnormal situations location. Only required part of complex system netmetric is used. Such approach allows to reduce power consumption of automatic monitoring and adaptive control system due to the reduction of management information volume that is to be to be transmitted between nodes to specify the deviation from norm of the network controlled parameter.
7. In order to reduce power consumption of modern TN it is advisable to carry out studies that are more fundamental: (a) by the use of agent-oriented approach; (b) the application of x and y features distribution spontaneous laws in the control system and adaptive control.

References

1. Bianzino, A., Chaudet, C., Rossi, D., Rougier, J.: A survey of green networking research. Commun. Surv. Tutor. IEEE. (2010)
2. Gupta, M., Singh, S.: Greening of the internet. In: Proceedings of the 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications—SIGCOMM'03, pp. 19–26 (2003)
3. Kharchenko, V., Kondratenko, Y., Kacprzyk, J., (eds.): Green IT Engeneering: Components, Networks and Systems Implementation, 355 p. Springer International Publishing (2017)
4. Bolla, R., Bruschi, R., Davoli, F., Cucchietti, F.: Energy efficiency in the future internet: A survey of existing approaches and trends in energy-aware fixed network infrastructures. IEEE Commun. Surv. Tutor. **13**, 22 (2011)
5. Bianzino, A., Raju, A., Rossi, D.: Apples-to-apples: a framework analysis for energy-efficiency in networks. ACM SIGMETRICS Perform. Eval. Rev. **38**, 81–85 (2011)
6. Zhang, Y., Chowdhury, P., Tornatore, M., Mukherjee, B.: Energy efficiency in telecom optical networks. IEEE Commun. Surv. Tutor. **12**, 441–458 (2010)
7. Lange, C., Kosiankowski, D., Weidmann, R., Gladisch, A.: Energy consumption of telecommunication networks and related improvement options. IEEE J. Select. Topics Quantum Electron. **17**, 285–295 (2011)
8. Chabarek, J., Sommers, J., Barford, P., Estan, C., Tsiang, D., Wright, S.: Power awareness in network design and routing. In: IEEE INFOCOM 2008—The 27th Conference on Computer Communications, pp. 457–465 (2008)

9. Panarello, C., Lombardo, A., Schembra, G., Chiaraviglio, L., Mellia, M.: Energy saving and network performance: a trade-off approach. In: Proceedings of the 1st International Conference on Energy-Efficient Computing and Networking—e-Energy'10, pp. 41–50 (2010)
10. Lombardo, A., Panarello, C., Schembra, G.: Achieving energy savings and QoS in internet access routers. ACM SIGMETRICS Perform. Eval. Rev. **38**, 76–80 (2011)
11. Simankov, V., Lutsenko, E.: Adaptive management of complex systems based on image recognition theory. Technical University of Kuban State Technological University, Krasnodar (1999)

Model and Methods of Human-Centered Personal Computers Adaptive Power Control

Igor Turkin and Oleksandr Vdovitchenko

Abstract The problem of economical use of modern PCs' electric energy is solved at all levels, starting with hardware. Power consumption of the computing platform can be structured by dividing it into two parts: static and dynamic, where last of them is defined by processor load. There are two main approaches in solving the problem of reducing the computers' power consumption at the hardware and at the operating system levels:

- the dynamic change of the processor voltage and frequency;
- the advanced configuration's interface and power management technology.

Only a few publications that take into account the user's model when power saving PC's control are known. Mostly, the terms «user's model», «user's behavior model» are used in works on ergonomics of the user's interface, in search and advisory systems, in systems of user's authentication, and also in tasks of identifying abnormal behavior of computing systems. Proposed solutions use models based on the measurement of the user's motor activity with typical and most used devices for entering information: a keyboard and a manual manipulator «mouse». Also the existing models ignore the semantic and target components of the user's behavior for which the elementary motor activity is carried out. Such components are manifested at the level of the operating system in the form of application's starts/stops, switching of focus, etc., and they are also an integral part of the general model of user's activity of the PC. The purpose of the article is to develop a model of user's activity and the method of its identification in the system of PC's power consumption. In addition the evaluation of the implementation of this model and methods in the system of the PC's power consumption is of interest.

Keywords Adaptivity · The dynamic change of the condition · Hidden markov model · BAUM-Welch method · Energy savings

I. Turkin · O. Vdovitchenko (✉)
National Aerospace University «KhAI», 17 Chkalova Street, Kharkiv, Ukraine
e-mail: o.vdovitchenko@khai.edu

© Springer Nature Switzerland AG 2019

V. Kharchenko et al. (eds.), *Green IT Engineering: Social, Business and Industrial Applications*, Studies in Systems, Decision and Control 171,
https://doi.org/10.1007/978-3-030-00253-4_14

1 Introduction

1.1 Motivation

In computing systems dynamic power control technologies which are implemented at the software-hardware level provide great opportunities for dynamic control of productivity and consumed energy, but the control of these indicators is limited only by the use of resources and the factors of user's behavior that are not reflected in using of system resources such as processor load, time of inactivity, etc. are not always taken into account.

Such researchers as A. Chandrakasan, R. Brodersen, J. M. Rabaey, M. Pedram, S. Udani, J. Smith, L. Benini, G. De Michaels, J. Lorch, A. Smith, D. Snowdon made significant contribution to the theory of dynamic power consumption control of personal computers (PCs). The research is conducted at the level of international corporations, with Intel, Microsoft, Toshiba, Samsung, AMD, Asus as leaders.

There are few publications that take into account the user's model for energy-saving personal computer (PC) management. Mostly, the terms "user model", "user activity model" are used in works on ergonomics of the user interface, search and recommendation systems, user authentication systems, as well as in the tasks of identifying the abnormal behavior of computing systems, etc. Among the important scientific works, which use the user's model for energy-efficient PC control, let's mention publications by L. Benini, J.-B. Durand, B. Kveton, Chich-Han Yu, S. Mannor. The solutions which are proposed in these articles the models based on measuring the motor activity of a user with the typical and most usable devices for entering information—a keyboard and manual manipulator "mouse" are used. In the existing models, the semantic and target components of the user's activity for which the elementary motor activity is carried out, are ignored. These components are detected at the operating system level in the form of application starts/stops, focus switching, etc., and are also an inseparable part of the general model of PC user's activity, so developing a user's model and adaptive power control technology for personal computers is an important scientific and applied problem.

1.2 Goal and Approach

The purpose of the research is to reduce the energy costs of modern PCs, maintaining the acceptable for user characteristics of PC using comfort. To achieve this goal, the following tasks should be solved:

1. To do the analysis of publications concerning the problem of energy-efficient use of PCs;

2. To develop a user focused model of the computer system PC's adaptive power control;
3. To develop a user-centered adaptive power consumption control system for PCs;
4. To perform an experimental evaluation of the feasibility and effectiveness of the user model and the method of adaptive power control of PCs.

2 Related Work

In the era of information technologies, the problem of energy consumption has come to a new level. New technologies and standards are being developed. Companies that have joined the Climate Savers Computing Initiative have been working to reduce the worldwide energy consumption of computers by half in 2018. CSCI has already joined Apple, Dell, Google, HP, Intel, Lenovo, Microsoft and many others [1].

The power consumption of the computing platform can be structured by dividing it into two parts:

– the static part includes the costs of maintaining the platform in working state (for example, energy consumed by the cooler, motherboard, etc.). It has a weak correlation with the nature of executable applications, so we are less interested in it;
– the dynamic part is associated with the execution of programs. As this part directly depends on the characteristics of the loading, the degree of resources' utilization and the policy of power consumption, it can vary greatly. Exactly here there are opportunities for optimization, as the work of the software makes a significant contribution to the overall power consumption of the system [2].

It is known that there is no need for the constant use of the processor with maximum performance. When optimizing power consumption, we can consider the energy spent to perform all tasks in strictly defined time intervals as the criterion of efficiency. The constant work of the computer during idle time, even with reduced processor consumption, leads to excessive consumption of electricity.

Based on this, two energy-saving approaches were developed:

– the dynamic change of frequency and voltage of the processor's power supply;
– idle-waiting mode control.

SpeedStep is Intel's energy-saving technology, based on a dynamic change in the frequency and voltage of the processor. The development under the working name Geyserville includes modifications: SpeedStep, SpeedStep II and SpeedStep III [3, 4].

The similar energy-saving technologies from AMD are known. PowerNow! technology which is developed for application in processors' versions K6-2+, K6-III+ i Athlon, which are installed in laptops. The CPU clock frequency and voltage automatically decrease when the computer is idle or load insufficiently, which reduces power consumption and heat dissipation. Cool'n'Quiet is a PowerNow! technology version designed for ordinary, non-mobile processors [5].

These energy saving technologies are based on the method of dynamic voltage and frequency scaling [6]. The supplied voltage determines the frequency of the processor, and, therefore, affects the power consumption. At the same time voltage switching is energy-intensive and rather long (expensive) transient processes. Accordingly, identifying the optimal frequency of the processor for performing the incoming tasks is a problem.

Advanced Configuration and Power Interface (ACPI) [7]—a common interface for hardware detection, power control and configuration of motherboard and devices. The task of ACPI is to provide interaction between the operating system, equipment and motherboard BIOS. ACPI was designed specifically for the implementation of OnNow's ideology. OnNow is an ideology of a computer device that is ready to recover and begin to work quickly at any time.

Models and methods of dynamic power control are based on policies. The purpose of the power control policy is turning off the required components when they are not in use and turning them on at the moment when they become necessary. The main problem with the power control policy is uncertainty at the time of the transition from the on state to the off state and vice versa. The problem of prognostication is as follows: it is necessary to determine whether the idle period will be long enough to compensate for the overhead of the transition between device states. The minimum length of the idle time for energy savings is called the break-even time [8]. Existing power control policies can be divided into three groups: time-out policies; predictive policies; stochastic policies.

Time-out policies are built upon measuring time of inactivity t. They put the device into sleep mode, if it is in standby mode, more than t. The main supposition is that if the device is not used during time t, then it will be put into idle state. The disadvantage of this policy is that devices may not be used while consuming energy until the end of a given time [9]. The solution to this problem is an adaptive time-out policy—a dynamic change t based on certain parameters. In the paper [10] regulation t based on the correlation between the performance delay and the waiting time for the previous idle period is proposed. If the coefficient is high, t will increase, otherwise—it decreases.

Prognostication policies determine the duration of idle time's future period. Thus, you can decide immediately about the transition to sleep, depending on the prediction. In the paper [11] the scheme of exponential averages is offered. The future period of idle time is predicted on the basis of measuring the exponential average provided and actual lengths of the previous period of idle time:

$$\mathrm{In} + 1 = \mathrm{a} \cdot \mathrm{Jn} + (1 - \mathrm{a}) \cdot \mathrm{In}, \tag{1}$$

where

In + 1 is a new predicted value of idle time;

In the last predicted value;
Jn the last period of idle time;
a the constant damping factor in the range from 0 to 1

Chung and others in the paper [12] propose a prediction policy that uses adaptive tree learning. An adaptively trained tree encodes a sequence of idle time periods in the nodes of the tree. Sheet nodes preserve the level of reliability of prognostication (RDP) associated with the corresponding sequence. The RDP is updated with the help of a finite state machine.

A policy that represents a mixture of adaptive time-outs and prognostication schemes is considered in the paper [13]. The policy is based on separating the cluster sequence into groups which are referred as sessions, depending on user queries. It provides the length of the current session, based on the prognosticated and actual lengths of previous sessions. In the absence of requests from the user, the time-out time will be reduced by the adjustment factor instead of disconnecting the device after a specified time. Conversely, if a request is received, time is increased by the same adjustment factor. The off command is given when the device is in standby mode for a long time compared to the predicted session duration.

Stochastic policies are based on models of stochastic processes. When an accidental event occurs, the device changes the power state. Minimizing energy consumption and productivity delays are criteria for stochastic optimization.

In the paper [14] modeling the process of queries, queue queries and service provider (device) in the form of Markov processes with discrete time was carried out. They determine the state of the system, such as the concatenation of queuing states, queue, and service provider. The policy is a matrix, on the basis of which a team corresponding to the specific state of the system is selected. Policy optimization for maximizing stored energy while limiting productivity delays is implemented.

In the paper [15] service provider and query queues are modeled as Markov processes with continuous time, and the time between requests arrives is modeled as an exponential distribution. This model gives the opportunity not to evaluate the periodically necessary power supply. Instead, the arrival and maintenance of queries are events that initiate the transition between states.

In the papers [22, 23] the model of queries and their maintenance using «memoryless distributions», which is not accurate in real situations is proposed [16]. In the paper [17] the query arrival is modeled as Pareto distribution, and the times of device transitions between power states are modeled as a uniform distribution.

Existing power consumption policies have a number of disadvantages, for example:

– time-out policies continue to consume energy when performance is no longer needed, but the time to an end of the PC work has not come yet;
– prognostication policies can solve this problem, but they work on the fundamental assumption that service requests have a high degree of correlation;

– stochastic policies are good at minimizing power consumption, taking into account loss of productivity, but they assume an accurate distribution for the queue of queries, if the distribution changes, then the model will become inadequate.

3 Background of the Work

The basis for the article was the thesis «One design point that fits all use cases won't be sufficient to scale down the energy consumption to the required levels. Therefore, design and run-time tools supporting dynamic adaptation and adoption of energy under quality of experience requirements will be needed. The full complexity of dynamic adaptation should not percolate up to the application programming level but be part of the system level. This will require comprehensive energy models for computing.» expressed in the work «Green IT Engineering: Concepts, Models, Complex Systems Architectures» [18].

The concept of adaptability is dominant at the modern theory of energy-saving control of personal computer systems. In this paper, we touch upon a little researched area, exactly the user. He is regarded as the main influencing person in the power control system of a personal computer.

Analysis of user models in the task of managing PC's power consumption is one more essential point. The term «User model» is considered from different positions in modern information technologies. The user can be defined as a subject (action maker or state carrier), a simplified imaginary schema (subjective model) created by a user that is powerful enough to describe the interaction with a device, but does not always reflect its real internal mechanics [19]. The user as an object is understood as the one to whom the subject's thought or action is directed.

By purpose, user models as objects are divided into:

– adaptive interfaces. They are changing the display of information. They are built on the basis of queries (documents, words, history), human characteristics (demographic, individual, psychological) [20];
– information behavior models. They are changing the system reaction. They are built on the basis of human characteristics, parameters of specific tasks, i.e., the main characteristics of actions (time to solve the problem, time to act). They are composed of three levels: syntactic, semantic, pragmatic;
– model of identification user's rights.

By the method of obtaining a user models are divided into:

– template model is a set of criteria selected by experts which characterize the behavior of the user;
– dynamic model is a «user track» rating based on system reviews.

Turning to the management of PC's power consumption, you can identify the main parameters on which the control depends. In this interpretation the user model in the task of managing the power consumption of personal computers is considered in terms:

– User's waiting time is a value that characterizes the time that passed from the current state of the system in the working state;
– performance is the value that characterizes the performance of a personal computer, depending on the mode of power consumption;
– power consumption is a value that shows how much power your PC consumes.

The problem of the user model in the system of power consumption management of personal computers hasn't been studied well, so from the existing works it is possible to distinguish two approaches.

The first is the development of a user model based on the Bayesian network [21]. The main hypothesis is that the latent state significantly influents on the distribution of the work's duration. The future idle time is defined as the duration between the current state and the time of the next active state. The user model is described as the state in time with the parameters: the previous state, the time in the previous state, the peak load hours, the current state.

The second one is the user model, which represents the queue of requests for the service system. In the paper [22] it is assumed that the user can be in two states: active and passive. In the active state of the user there should be a queue of requests, which is exponentially distributed. The simulation of user interaction and energy control system is described as a Markov process with discrete time.

The disadvantage of existing models is that these models ignore the semantic and target components of user behavior for which elementary motor activity is carried out. Such component is manifested at the level of the operating system in the form of start/stop applications, switching focus, etc. and it is also an integral part of a model of PC's user activity. To evaluate the efficiency of this model, we use the energy saving characteristic proposed in the paper [23].

4 Proposed Solution for Adaptive Power Control of Pc

Model of human-centered personal computer adaptive power control system is presenred below (Fig. 1). According to available classifications of adaptive systems, this system belongs to model identification adaptive controllers (MIACs), which perform system identification while the system is running.

Adaptability is realized by correcting the algorithm parameters of the decision-making block on controlling the power supply circuits (policies), as well as by using adapted (corrected) user activity models and PC power consumption processes. The interaction of a detailed stochastic model of a system with identification blocks and adaptive control is depicted in Fig. 2.

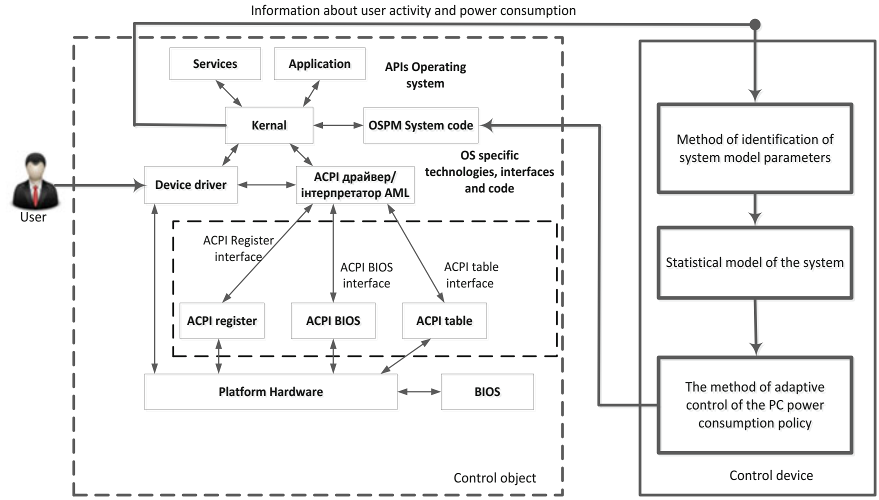


Fig. 1 Model of human-centered personal computer adaptive power control system

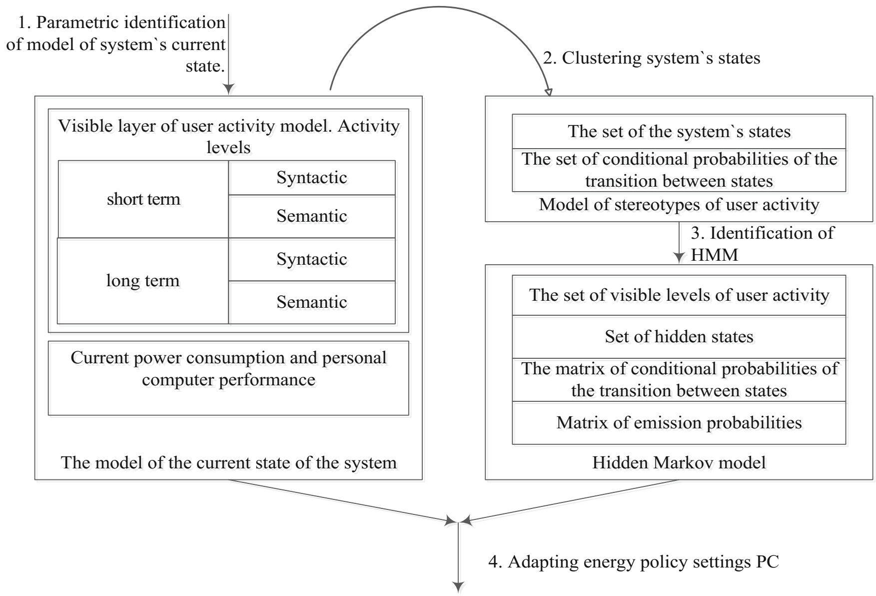


Fig. 2 The interaction of a detailed stochastic model

4.1 The Model of the PC's Current State

This model will be considered as a sequence of instant states $\Xi = \langle S_k, k = 1\ldots K\rangle$. Each S_k state is characterized by energy efficiency

$$En_k = \langle P_k, N_k\rangle, \tag{2}$$

where

P_k is a current performance;
N_k is a current power consumption.

A user event activity UA_k:

$$S_k = \langle UA_k, En_k\rangle. \tag{3}$$

User event activity is considered at two levels (syntactic and semantic) and in two intervals: short and long term:

$$UA_k = \left\langle \underline{Syn}_k, \overline{Syn}_k, \underline{Sem}_k, \overline{Sem}_k \right\rangle, \tag{4}$$

where

$\underline{Syn}_k$ is a syntactic level of the user's interaction with the keyboard, mouse, clipboard, etc. on the short-term interval;
$\overline{Syn}_k$ is a syntactic level of user's activity in the long-term interval;
$\underline{Sem}_k$ is a semantic level of user's activity that characterizes the results of user control of the start/stop of applications, switching the focus, etc. on the short-term interval;
$\overline{Sem}_k$ is a semantic level of user's activity in the long-term interval.

Metrics of user activity on syntactic and on semantic levels are the number of events per unit of time. Then the vertex of the stereotypes' model is a cluster, a subset of "close to each other" objects of the A vertex. The distance $\rho(s_i, s_j)$ between the objects S_i and S_j depends on the accepted metric in the space of characteristics.

4.2 The Stereotype Model

The stereotype model is a marked graph of state space that characterizes the multiplicity of trajectories in the state of the system's space:

$$MSt = \langle V, E\rangle, \tag{5}$$

where

$V = \{v_{i,l}\}$ a multiplicity of ordered by the vertices levels (the system's states).

$E = \{\langle v_{i-1,l}, v_{i,m}, Pr_{i-1,l,i,m}\rangle\}$ the multiplicity of edges that connect the clusters of the previous layer with the clusters of the next layer with the calculated transition probability.

Each vertex contains information about the kernel of the cluster in the model's components of the system's current state:

$$v_{i,l} = \left\langle \underline{Syn}_{i,l}, \overline{Syn}_{i,l}, \underline{Sem}_{i,l}, \overline{Sem}_{i,l}, P_{i,l}, N_{i,l} \right\rangle \quad (6)$$

The theory of clustering is used to automatically determine the ratio of objects to the model of the system's current state to the stereotype. The object of data for clustering is the current state (S_k). A vector of characteristics is identified with each object $S_k = \left\langle \underline{Syn}_k, \overline{Syn}_k, \underline{Sem}_k, \overline{Sem}_k, P_k, N_k \right\rangle$, S_k components are the separate object characteristics. The number of characteristics is the size of characteristics' space. The multiplicity consisting of all characteristics' vectors we mark as A: $A = (s_1, \ldots, s_m)$.

4.3 The User Activity Model

The user activity model is based on the hidden Markov model (Fig. 3):

$$MAU = \langle S, O, A, B \rangle, \quad (7)$$

where

S is a multiplicity of hidden states, $S = \{S_1, S_2, S_3, S_4\}$, which characterize the expected time before shutdown (shutdown of a personal computer);

O is the multiplicity of user activity's visible levels, $O = \{O_1, O_2, O_3, O_4, O_5\}$ (the visible layer of the hidden Markov model is given by variables, which are measured by the intensity of events for a certain time interval, taking into account the group of data used, also in the visible layer passed two states—the beginning and the end of work, when the user activity is zero);

$A = \{a_{ij}\} = P(S_t = j | S_{t-1} = i)$ matrix of conditional probabilities of the transition between states S at the time t;

B is the matrix of emission probabilities.

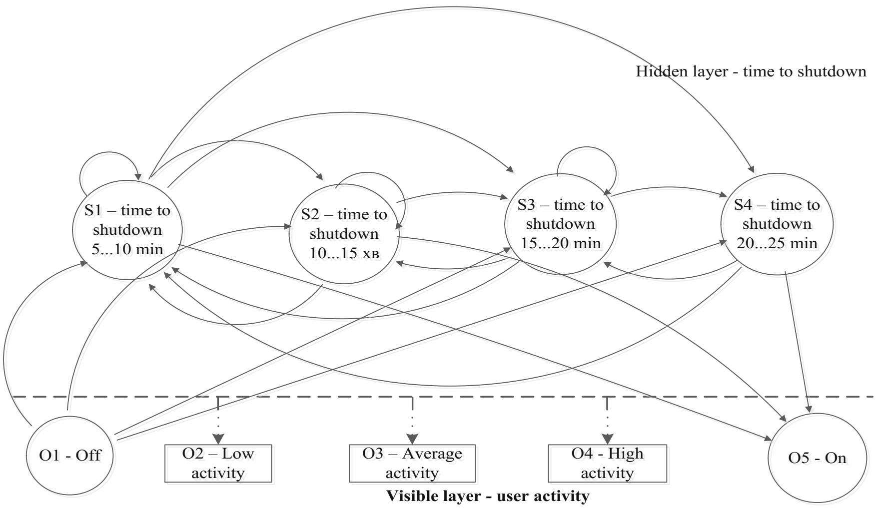


Fig. 3 The structure of the hidden Markov model

The event activity of MU user is considered on two levels: syntactic and semantic:

$$\mathrm{MU} = \langle \mathrm{Sy}, \mathrm{Se} \rangle. \tag{8}$$

The method of a personal computer's adaptive power control system is developed. The key processes in this case are parametric identification of the system's current state model, clustering of system states, identification of the parameters of the user activity model, adaptation of the PC power consumption policy parameters (Fig. 2).

4.4 Parametric Identification of Models of System's Current State

The model of the system's current state involves the identification the parameters of two components: the user's activity and the current consumption and productivity of the personal computer. Let's consider the results of observations of a random variable S_k, that are presented in the form of statistical series. Let's divide the whole range of available values S_k at the intervals and count the number of values per each i-th digit, divide the resulting number by the total number of observations n and define the frequency that corresponds to a different digit:

$$p_i = \frac{m_i}{n}. \tag{9}$$

To calculate the number of histogram's intervals empirical formula of Sturgess is used: $m \approx (1 + 3,322 \lg n)$.

Approximation of histograms, which are obtained experimentally, has an approximate nature and the solution of practical compliance of theoretical histogram's function is based on the criteria for approval of Kolmogorov-Smirnov, Anderson-Darling, χ^2.

When solving practical tasks on analysis of real user event activity, the mathematical law of its distribution is unknown.

In this case, an assumption of compliance with the theoretical law of experimental data are extended. This hypothesis requires the statistical test, as a result it either confirmed or refuted.

4.5 Clustering System`S States

Given that: Ξ—a set of vector characteristics, $\Xi = \langle \langle S_k \rangle, k = 1..K \rangle$; V—set of clusters labels, $V = \{v_{i,l}\}$. The vector of characteristics (object) $S = \langle UA, En \rangle$—the unit of data for clustering algorithm. It is the element of *6*-dimension space: $S = \left\langle \underline{Syn}, \overline{Syn}, \underline{Sem}, \overline{Sem}_k, P,\ N \right\rangle$. The characteristic (attribute) S_k is the component of the scalar vector S. The dimension 6 is the number of object's characteristics S. k-th object with S is defined as $S_k = (s_{k,1}, \ldots, s_{k,d})$. The required distance function between objects is $\rho(s_i, s_j)$.

We need to build an optimal partitioning of objects into groups i.e. split the sample into subsets that do not intersect which are called clusters, so that each cluster consists of objects that are close in the metric ρ, and objects from different clusters significantly differ. Thus each object $s_i \in S$ top model graph of stereotypes is attributed $v_i \in V$. The optimality of partitioning is defined as a requirement to minimize mean square error of breakdown (minimizing of the average distance inside the clusters):

$$F_1 = \sum_{v \in V} \frac{1}{|v|} \sum_{s \in v} \rho^2(s, \mu_v) \rightarrow \min, \tag{10}$$

where μ_v—is the center of the v (centroid).

It is necessary to maximize the distance between clusters:

$$F_2 = \sum_{v \in V} \rho^2(\mu_v, \mu) \to \max, \tag{11}$$

where μ—is the common center of mass of the whole sample.

To consider the approximation to the minimum and maximum let's point to a minimum the ratio:

$$\phi = \frac{F_1}{F_2} \to \min. \tag{12}$$

In this case we take into account limits on the number of clusters because in the optimal case, each object will be a separate cluster.

The solution of the problem. To determine an object belonging to the cluster the method of k-means++ is used.

The first stage: By T steps in the model of the current state of the system $S = \langle UA, En \rangle$ the sequence of objects is generated $S_k = \left\langle \underline{Syn}_k, \overline{Syn}_k, \underline{Sem}_k, \overline{Sem}_i, P_k, N_k \right\rangle$, which is the object with characteristics. The input data are a set top model of the current state of the system (S) and the anticipated number of clusters (the number of graph's stereotypes models) N. We will be guided in the normalization not by extreme values, but by the typical ones, statistical characteristics of data such as mean and dispersion:

$$\widetilde{s}_i = \frac{s_i - \overline{s}_i}{\sigma_i}; \quad \overline{s}_i = \frac{1}{n}\sum_{a=1}^{n} s_i^a; \quad \sigma_i^2 = \frac{1}{n-1}\sum_{a=1}^{n} (s_i^a - \overline{s}_i^a)^2, \quad i = 1, \ldots, k. \tag{13}$$

where

$\widetilde{s}_i$ a new item row;
s_i an element of output row;
$\overline{s}$ the average row value;
σ_i dispersion;
n the number of row items.

To limit the range of data the nonlinear transformation is used:

$$\widetilde{s}_i = f\left(\frac{s_i - \overline{s}_i}{\sigma_i}\right), \tag{14}$$

where $f(a) = \frac{1}{1 + e^{-a}}$, which guarantees placement of data in a range [0,1].

The second stage. Since the space is multidimensional, and variable characteristics are normalized, that is increasing impact of dispersions between certain characteristics are not necessary, the Euclidean distance is used:

$$\rho = \sqrt{\sum_{i=1}^{D} (s_i - \bar{s}_i)^2}, \tag{15}$$

where

s_i is i-th characteristic of the system s;
$\bar{s}$ is i-th characteristic of the system $\bar{s}(s \neq \bar{s})$, $i = 1 \ldots D$, where D = 6—number of characteristics.

The third stage. One of the k-means problem is the initial cluster centers choice. The solution of this problem is proposed in the modified method k-means++.

The sequence of the method k-means++.

Step 1. Select the first cluster center c_1 randomly (among all states of the system S).

Step 2. Choose a new cluster center c_i from $s \in S$, for which the square of the distance to the selected centroids will be the largest:

$$\frac{D(s)^2}{\sum_{s \in S} D(s)^2}. \tag{16}$$

Repeat step 2 until you select N centers of the clusters.

Step 3. For each $i \in \{1, \ldots, K\}$ select the cluster C_i as a set of points with S, which are closer to c_n, than to c_j for all $n \neq j$. The measure of proximity is Euclidean distance between objects (states of the system) and centroids (cluster centers):

$$C_i = \min_n \sqrt{\sum_{p=1}^{D} (s_{k,p} - c_{n,p})^2}. \tag{17}$$

Step 4. For each $i \in \{1, \ldots, N\}$, enumerate c_i, as the center of mass at all points in C_i:

$$c_i = \frac{1}{|C_i|} \sum_{s \in C_i} s. \tag{18}$$

Step 5. Repeat steps 3–4 until the clusters will change, that is objects will move between clusters or functional change will be minimal ϕ.

4.6 Method of Parameter Identification User's Activity Model

The observed value is known (user's activity) N(t) as a sequence $N_T = N_0, N_1, \ldots,$ N_{T-1} length T, generated by the sequence of time states to the end of user's experience $O = O_1, O_2 \ldots O_T$. Probability to watch $N(t)$ is $P(N_T) = \sum_X P(N_T|X_T)$ $P(X_T)$, where the sum is given by all possible hidden variables x (t) as a sequence of hidden nodes $X_T = X_0, X_1, \ldots, X_{T-1}$. According to available data N(t) it is necessary to determine hidden parameters of the most probable sequence of Markov chain states $S = S_1 \ldots S_T$.

Thus, if an obtained visible learning sequence of the output power O is such that $O = O_1, O_2, \ldots, O_T$ ($O_i \in N_{expr}$), then the parameters of the model MAU = (P, B, π) should maximize a posteriori probability p(O|MAU).

The solution of the problem. For definition of the model's parameters MAU = (P, B, π) in this paper iterative procedure which is based on the method of Baum-Welch is used.

First stage: By steps T in the model MAU = (P, B, π) the sequence of observations is generated $O = O_1, O_2, \ldots, O_T$ (O_i $N_{изм}$), moreover at time t for state s: $X_i = i$ probability $P(O_{1,t-1}|X_t = s)$ that during the conversion the sequence of observations was formed $O_{1,t-1}$, a $P(O_{t,T}|X_t = s)$—the probability that after this condition there is a sequence of observations $O_{t,T}$. The probability $P(X_t = s|O) = P$ $(X_t = s|O_{1,t-1} \cap O_{t,T})$ at time t the chain is in a state s is searched.

For any s state into a random t step probability $P(O_{1,t}|X_t = s_i)$ the fact that the way it made to sequence $O_{1,t}$ for the next t can be recurrently calculated:

$$\begin{aligned} P(O_{1,t}|X_t = s_i) &= \sum_{j \in S} P(O_{1,t}|X_t = s \cap X_{t-1} = j) \\ &= \sum_{j \in S} P(O_{1,t-1}|X_{t-1} = j) P(X_t = s|X_{t-1} = j) P(O_t = o_t|X_t = s). \end{aligned} \tag{19}$$

The probability to get into the state S at the t-th step, given that after the transition event occurs O_t, will be equal to the probability to be in a state j at t-th step, multiplied by the probability to go out of the state j at s, having realized an event O_t for all j∈S.

The probability that after a random state S the sequence will be made $O_{t+1,T}$ is defined recurrently:

$$P(O_{t,T}|X_t = s) = \sum_{j \in S} P(O_{t+1,T}|X_{t+1} = j) P(X_{t+1} = j|X_t = s) P(o_{t+1}|X_t = s). \tag{20}$$

To find the probability that the chain of events will be held, *P (O)*, the product should summarize both probabilities for all states at an arbitrary step *t*:

$$P(O) = \sum_{s \in S} P(O_{1,t}|X_t = s)P(O_{t,T}|X_t = s), \tag{21}$$

As the future of Markov chain does not depend on the past and the observation of probability events O_t does not depend on the past observations on the sequence $O_{1,t\text{-}1}$, then the probability that at time *t* the chain will be in a position *s*:

$$P(X_t = s|O) = P(X_t = s|O_{1,t-1} \cap O_{t,T}) = \frac{P(X_i = s|O_{1,t-1})P(X_{t=i} = s|O_{t,T})}{P(O)}. \tag{22}$$

The second stage: For a given HMM θ = (P, B, π) with space of states S = {s_1, s_2,…, s_K}, initial probabilities π_i of being in a state i and probabilities $P_{i,j}$ of transition from state i to state j, observed by y_1,…, y_T and a plurality of output information J_{in} the most plausible sequence of states of hidden nodes is searched S = S_1…S_T (Viterbi way) that most accurately describe this model. Then the most likely sequence of states x_1,…,x_T given recurrent relation:

$$V_{1,k} = P(y_1|k)\pi_k, V_{t,k} = P(y_t|k) \max_{x \in S}(P_{x,k}V_{t-1,x}), \tag{23}$$

where $V_{t,k}$—the probability of the most plausible sequence of states responsible for the appearance of the first t visible symbols, culminating in a state *k*.

Viterbi path is searched by states *x*, satisfying the following equation:

$$x_T = \arg\max_{x \in S}(V_{T,x}), \tag{24}$$

$$x_{t-1} = \begin{cases} \arg(V_{t,k}), \text{ если } t > 1, \\ k, \text{ если } t = 1. \end{cases} \tag{25}$$

The third stage: according to the starting sequence of observations O the unknown parameters HMM θ are determined, that maximize the probability of observing O:

$$MAU^* = \max_{MAU}(P(O|MAU)). \tag{26}$$

Number of points in time r, in which observations are carried out is pre-set and consists of the following steps:

(1) determining the sequence of model states S_i = {s_{i1},…, s_{ir}}, i = 1,…, r system at a given time;
(2) assessment of probabilities P(S_i) of each sequence appearance S_i, i = 1,…, r. Discovered in the previous step, by calculating the products of probabilities of

transitions between states for the model ranges, limited by control set time points, namely:

$$P(S_i) = \prod_{t=1}^{r-1} P_{s,i,t,t+1}, \tag{27}$$

where $P_{s,i,t,t+1}$—the probability of transition from state S_{it}, in which the system was at time t, to the state $S_{i,t+1}$, at the moment t + 1;

(3) the probability of appearance of the observed sequence X = $\{x_1,\ldots, x_r\}$ for sequences of states S_i, i = 1,…, r:

$$P_{ix} = P(S_i) \prod_{j=1}^{r} P_{x,i,j}, \tag{28}$$

where $P_{x,i,j}$—probability of observable characteristics x_j in stage s_{ij};

(4) selection of the most probable sequence of states $S_{max} \in \{S_i\}_{i=1,\ldots,r}$, which corresponds the largest probability:

$$\{P\}_{x,\max} = \max_i P_{ix}, \quad i = 1, \ldots, r. \tag{29}$$

4.7 The Method of PC's Power Consumption Adaptation

It is given: MSt is a model of stereotypes of user's activity, which includes a plurality of states of the system and a set of conditional probabilities of transition between states. The power consumption policy:

$$PM = \langle t, ST \rangle, \tag{30}$$

where

t is time to finish of PC work;
ST stage of ACPI, which the system should move to at the end of time t

It is necessary to determine whether the downtime (Tbe) is long enough to offset the overhead of switching between states considering fines for lost productivity. To adapt the parameter t of the power consumption policies to maximize P performance while minimizing the power consumption of N it is required to minimize the ratio:

$$\sum_{i=1}^{n} \int_{0}^{\Delta} \widehat{f}_i(\tau) \cdot p(S_t = i) \cdot u(\tau) d\tau \to \min, \tag{31}$$

where

n	the number of the system's states;
Δ	time of transition from active to passive state;
$u(\tau)$	exponential function of the fine for premature loss of productivity;
$p(S_t = i)$	the probability of finding a state St;
$\hat{f}_i(\tau)$	experimental time that the system spent in one of the states.

Step 1. Define a user session end with a PC via conditional probabilities of model of E stereotypes. For each time point t gets the probability of being in a state V for model of stereotypes of user activity. For the system probability of transition from state to state will be defined as:

$$c_t = \max\{P(E_i = t)\}. \tag{32}$$

Step 2. The general approach to reducing the time for the transition to sleep mode is that if the probability of the transition to the standby mode is greater than the given threshold, then the time to standby mode is reduced. Reducing the time of transition to standby mode $pol(t) = \text{Stand by}$ occurs when the functional (31) is less than ρ (admissibility adjustment factor). If the calculated value is greater than ρ, the time does not change.

Step 3. Check t = 0, If "yes" there is a transition to a state of sleep, conversely go to step 1.

5 Simulation

Physical experiment is carried out in one of the computer classes of the software engineering department of NAU "KhAI". This class has 16 similar types of desktop configuration. Computers' configuration:

CPU: Intel Core i5-4690;
GPU: Intel HD Graphics 4600;
HDD: 1 Tb;
OS: Microsoft Windows 7 Service Pack 1 64-bit;
RAM: 8 Gb.

Data collection technologies. Hooking is the ability to hook in the OS's messaging system, which allows you to track a particular message (including an event) until it reaches the target function.

As a mechanism for hooking, we use the functionality of the operating system of the Windows family (hereafter OS Windows) «WindowsHook».

Thus, the method of hooking, based on an event model, is implemented next way.

Step 1. To register the event handlers in the event manager. In response to the external influence, generate an event.
Step 2. In response to an external event, a Windows system event is generated.
Step 3. To send the event to hooker.
Step 4. The hooker takes an event and looks for a handler for it, written beforehand.
Step 5. The event handler processes the event, saves the event data.
Step 6. The event is further transmitted in the event manager.
Step 7. The event manager takes an event and looks for a handler for it.
Step 8. The event manager calls the handler.
Step 9. If the method work isn't terminated then to go to step 2, else to finish.

Methods for obtaining data on ACPI states, transition times between states and power consumption in the current state, the appropriate tool for obtaining the required parameters are Intel Power Gadget.

A software prototype to collect and store user activity information was developed on the basis of the Hookings mechanism.

The prototype of user activity data collection software is intended for:

– tracking and recording input/output events;
– tracking and registration the start and stop of applications;
– definition of running and stopped applications at a given time point;
– saving data in xml format in the form of a certain structure.

Since the experiment was conducted in the computer classroom, the following characteristic of sequences have been singled out based on theoretical assumptions that the user's PC activity in the classroom depends on the learning process, as well as on the basis of experimental data:

– the lesson in the middle of the semester (Class A). This sequence can be interpreted as: low activity at the beginning of the lesson due to the fact that the teacher conducts introductory acquaintance in order to work at the lesson, high activity during work at the lesson changes with average activity, as well as low activity during the break between the half-periods and at the end of the lesson. An example of a sequence: {O1-O2-O2-O3-O5-O5-O5-O2-O1-O1-O5-O5-O6-O1-O1-O4-O4-O2-O1};

- the lesson before the session (Class B). Example: {O4-O5-O5-O6-O5-O5-O3-O1-O4-O4-O5-O4-O4-O5-O6-O6-O5-O5-O1}, such a sequence of activity is specific for a lesson before the session, the defense of a diploma or projects. You can notice that the user activity is high throughout the lesson and there are only a few changes during short breaks and consultations with the teacher;
- the lesson at the beginning of the semester (Class C), an example of a PC user activity sequence: {O2-O1-O1-O1-O3-O4-O5-O6-O4-O1-O1-O5-O4-O4-O4-O4-O4-O2-O1}. Such class is typical when it takes more time to work with the theoretical material at the beginning of the lesson, and during the lesson there is a reduced activity of the user, connected with the state of students after the pause in the educational process;
- the consultation lesson (Class D) {O1-O1-O1-O1-O5-O5-O5-O1-O1-O1-O6-O6-O5-O6-O1-O1-O2-O1-O3}, during which the PC is used from time to time, that is, there are no clearly selected transitions between the lowest and the very high activity.

For the experimental evaluation of the trained user activity model based on the collected statistical data during the experiment, we will classify the user activity sequences on the test data sample. As indicators of the adequacy of the trained user activity model, the matching factor between the output data and the obtained results was selected, namely the Cohen's kappa coefficient, the Kendall rank correlation coefficient, Pearson correlation, χ^2 and the overall consistency of the results of the training and the test sample. The Cohen's kappa coefficient is considered more reliable than the simple calculation of the compliance's percentage, because it takes into account the agreement of random variables which occur [24]. Cohen's kappa coefficient measures the agreement between two evaluators, each of which classifies N elements in C mutually exclusive categories. Cohen's kappa coefficient k:

$$k = \frac{p_o - p_e}{1 - p_e} = 1 - \frac{1 - p_o}{1 - p_e}, \tag{33}$$

where

p_o is the relative observed consistency between evaluators;
p_e is the hypothetical probability of random coherence, using observable data to calculate the probability of each observer randomly for each category.

If the coefficient k = 1, then the data are completely consistent. Otherwise, we can talk about consistency only at a certain level of trust.

The Kendall rank correlation coefficient. It is used to identify the relationship between quantitative or qualitative indicators if they can be ranked. The values of the index X are displayed in ascending order and assigned ranks to them. Ranking the value of Y and count the Kendall correlation coefficient [25]:

$$\tau = \frac{2S}{n(n-1)}, \tag{34}$$

where S = P–Q;

P is the total number of observations followed after the current observations with a large rank values Y;
Q is the total number of observations followed by current observations with lower rank values Y;
n is the number of elements in a row.

The Pearson correlation coefficient characterizes the existence of a linear relationship between two values [26]. The Pearson correlation coefficient is calculated by the formula (35):

$$r_{xy} = \frac{\sum_{i=1}^{m}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{m}(x_i - \bar{x})^2 \sum_{i=1}^{m}(y_i - \bar{y})^2}} = \frac{\mathrm{cov}(x, y)}{\sqrt{s_x^2 s_y^2}}, \tag{35}$$

where

$\bar{x}, \bar{y}$ are the selective averages x^m and y^m;
s_x^2, s_y^2 are the selective dispersions.

Mathematical modeling of training algorithms and classification is performed using the open source software [27].

The calculation of model parameters is carried out using the method of parametric identification of the user's activity model. The purpose of the calculation is to determine the matrix of transition probabilities' distribution $\mathbf{P} = \{p_{ij}\}$, matrix **B** distribution of probabilities of user activity's observed levels.

Table 1 Output data for learning the user activity model

Sequence	Label	Hidden states
0-1-1-2-4-4-4-1-0-0-4-4-5-0-0-3-3-2-0	A	4
0-0-1-2-4-4-4-1-0-0-4-4-5-0-0-3-3-1-0	A	4
0-1-0-3-4-4-4-1-0-0-4-4-5-0-0-3-3-0-0	A	4
0-0-0-4-4-4-4-1-0-1-4-4-5-0-0-3-4-0-0	A	4
3-4-4-5-4-4-2-0-3-3-4-3-3-4-5-5-4-4-0	B	4
4-4-4-4-4-4-4-2-2-0-4-4-5-3-3-4-5-4-5	B	4
3-4-4-4-4-5-5-1-0-2-4-4-5-0-0-4-4-4-4	B	4
1-0-0-0-2-3-4-5-3-0-0-4-3-3-3-3-3-1-0	C	4
0-0-1-1-2-2-2-3-0-0-2-3-4-2-1-2-2-1-1	C	4
0-0-0-0-0-0-1-2-3-3-2-1-0-1-2-4-0-0-0	C	4
0-0-0-0-0-4-4-4-5-1-1-0-0-5-5-4-5-0-0	D	4
0-0-4-4-4-0-0-0-0-5-5-4-4-0-1-0-4-5-0	D	4
4-4-5-0-0-0-3-5-4-4-0-1-0-0-4-4-4-4-5	D	4

To study the user activity model, the most common sequences of user activity levels were used for the lesson in NAU "KhAI". They were observed for the different classes of activity which were given earlier. Output data are shown in Table 1.

6 Results and Analysis

The adequacy of the model is verified using parametric and non-parametric statistics. We can note that the concept of adequacy in this case was interpreted as a criterion for the accuracy of the activity class definition with the help of a trained model based on test sets of data. Test results are shown in Fig. 4.

Evaluating the adequacy of parametric statistics, we find that the difference between the initial data and their relation to the corresponding class does not exceed 1%.

In order to detect the differences in power consumption while PC working in the classroom, a prototype of the PC's adaptive power control was used. With it the modeling on the basis of experimental data collected from the PC was carried out. We also give a comparison of the average energy consumed on a similar machine of the same computer class during the working day.

Figure 5 shows the PCs consumption with standard policy and adaptive control policy.

The main advantage in power consumption is taken for 4 h of PC operation, The main gain for energy consumption is at 4 h of the PC work, it can be correlated with

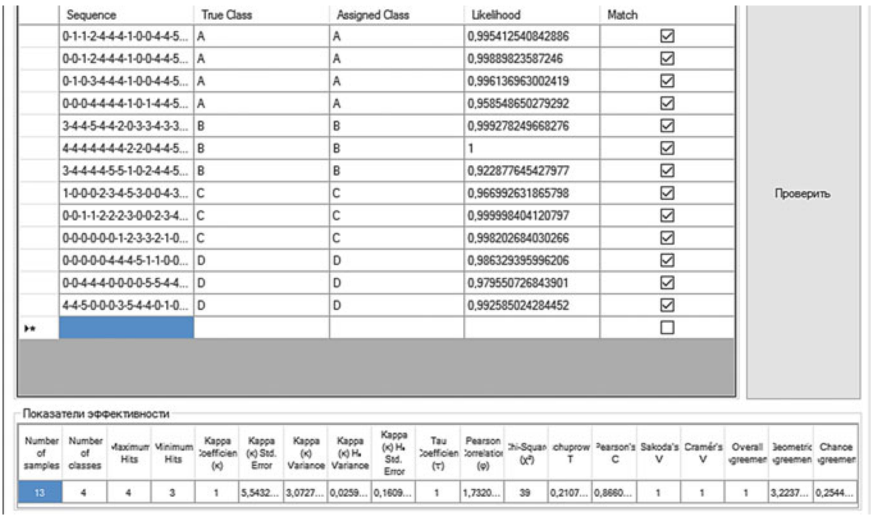


Fig. 4 The window of the trained model test results

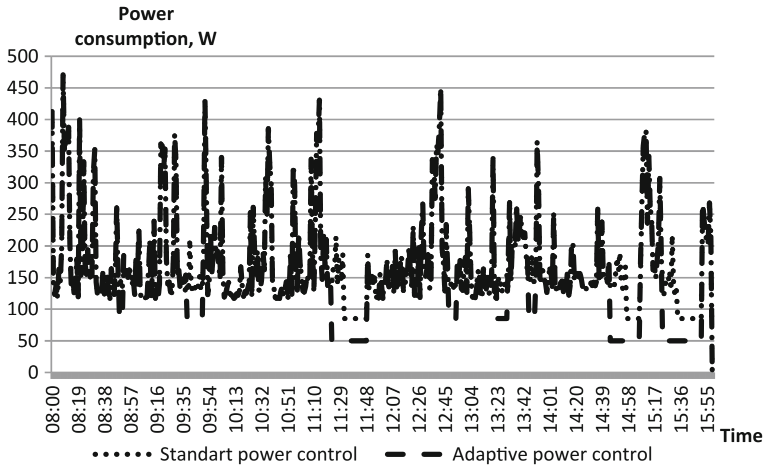


Fig. 5 Power consumption by a personal computer with different power control technologies

the fact that in the interval from 11.25 to 11.55 there is a break in the students schedule, under standard settings, the PC switched to the level S1 sleep stage, after 10 min of it's inactivity. With adaptive power consumption policies, the PC went into sleep after 3 min of inactivity, with at lower power consumption ACPI S2. At the 8th hour of working, when the students' lessons end at 15.25, and working day at 16.00, PC switched to S1 sleep state after 10 min of inactivity on basic settings of the time out policy. With adaptive policy it switched to S3 sleep state after 3 min, since user activity at the end of the 4th lesson was low, which reduced the penalty for loss of productivity and increased the probability of transition to a state of reduced power consumption for a longer period. It allowed to reduce power consumption for 1 PC by 0.7 kWh. Taking into account that in the computer class there are 18 computers, the savings can be about 12.6 kWh.

A comparison of hourly percent of idle time between two power control technologies is shown in the Fig. 6. The peculiarity of the user's work for the PC in the period of time in the middle of the study semester are frequent breaks, but they are less than time policy of time out. For adaptive politics this period is characterized by the average and low user's activity, which is characteristic for the transferring of the PC in less time intervals to sleep state, with the sleep state of the first S1 level. This state allows going into working condition for 2–5 s, while consuming 4 times less electricity. Sleep state time for adaptive power control was 1 h per day, and for basic technology—23 min.

In order to evaluate the differences in the comfort of working with a personal computer using the basic technology of power control and adaptive power control of the PC, a ranked anonymous interview was conducted. Students were in the role

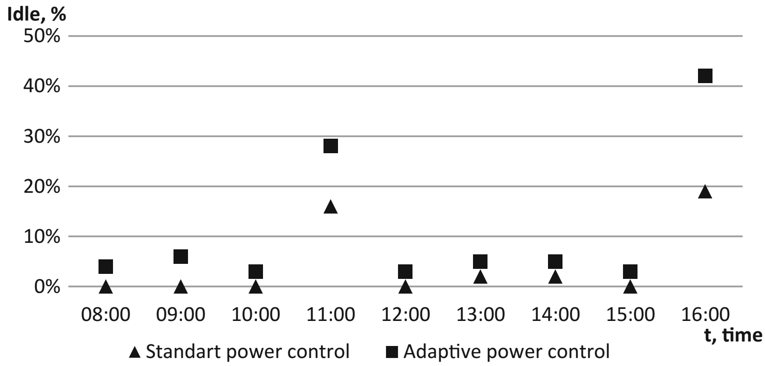
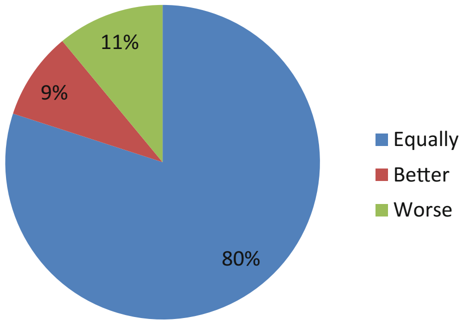


Fig. 6 The percentage of PC's idle time per hour with different power consumption policies in the middle of the semester

Fig. 7 Subjective valuation of the PC's comfortable work with adaptive power control, unlike basic technology



of experts who were asked to evaluate the comfort of working with the PC after the end of the working day. The questionnaire contained one question: "Evaluate the comfort of using the PC today compared with yesterday?". Possible answers: "Better (2), Equally (1), Worse (0)". To determine the coordinated degree of 200 experts' opinion, a special measure is used—Kendall concordation coefficient (W). As a result, W = 0,02, it follows that the differences between the opinions of experts are insignificant. The result of the interview is presented in Fig. 7.

Thus the electricity savings for the semester for one computer class with 16 (computer's amounted to 1025 kWh, which in monetary terms (in June 2016) was 25 UAH ($\sim$1 $) per every computer of the educational class every month, while the permissible indicators of the PC's using were saved.

In valuating the effectiveness of PC's adaptive power control, the metric of cash resources' savings by reducing the cost of replacing the components of the PC was not taken into account, as the amount of time spent on the worked hours for failure decreases. It is worth noting that, in addition to the material efficiency indicator, a reduction in the negative impact of the PC on the environment is achieved.

7 Conclusion

The main results include a method of adaptive power management of the personal computer, which allows you to achieve lower energy costs of the personal computer while maintaining acceptable performance comfort to the user. PC usage implemented a sequence of four steps.

1. Parametric identification of models in the exact state of the system is performed by methods of mathematical statistics and scores from the previous processing.
2. Clustering states of the system are based on the method k-means++ and allow to reduce the dimension and complexity of the problem.
3. Identification of hidden Markov model of user activity uses the Baum-Welch method and can be attributed to the current user behavior to one of the elements of the set of stereotypes user's activities.
4. The adaptation of the policy settings PC energy forms the management solutions considering of fines for loss of productivity and overheads.

References

1. Mingay, S.: Green IT: a new industry shock wave. Gartner, pp. 1–28. [online] Available at: http://www.ictliteracy.info/rf.pdf/Gartner_on_Green_IT.pdf (2007). Accessed 14 Oct 2017
2. Uchechukwu, A., Li, K., Shen, Y.: Improving cloud computing energy efficiency. In: 2012 IEEE Asia Pacific Cloud Computing Congress (APCloudCC) (2012)
3. Download.intel.com.: Enhanced Intel® SpeedStep® Technology for the Intel® Pentium® M Processor. [online] Available at: http://download.intel.com/design/network/papers/30117401.pdf (2017). Accessed 7 Oct 2017
4. Phoronix.com. Intel EIST SpeedStep—Phoronix. [online] Available at: https://www.phoronix.com/scan.php?page=article&item=397&num=1 (2017). Accessed 12 Oct 2017
5. KV.by. Cool'n'Quiet. [online] Available at: https://www.kv.by/archive/index2005030502.htm (2017). Accessed 13 Nov. 2017
6. Snowdon, D., Ruocco, S., Heiser, G.: Power management and dynamic voltage scaling: myths and facts. Nat ICT Aust. Sch. Comput. Sci. Eng. **12**, 1–7 (2005)
7. . Acpi.info.: ACPI—Advanced Configuration and Power Interface. [online] Available at: http://www.acpi.info (2017). Accessed 8 Nov 2017
8. Benini, L., Micheli, G.: Dynamic power management. Springer, US, Boston, MA (1998)
9. Karlin, A., Manasse, M., McGeoch, L., Owicki, S.: Competitive randomized algorithms for nonuniform problems. Algorithmica **11**(6), 542–571 (1994)
10. Helmbold, D., Long, D., Sconyers, T., Sherrod, B.: Adaptive disk spin-down for mobile computers. Mob. Netw. Appl. **5**(4), 285–297 (2000)
11. Hwang, C., Wu, A.: A predictive system shutdown method for energy saving of event-driven computation. ACM Trans. Des. Autom. Electron. Syst. **5**(2), 226–241 (2000)
12. Chung, E., Benini, L., De Micheli, G.: Dynamic power management using adaptive learning tree. In: IEEE/ACM International Conference on Computer-Aided Design, pp. 274–279. IEEE Press (1999)
13. Lu, Y., De Micheli, G.: Adaptive hard disk power management on personal computers. In: Ninth Great Lakes Symposium on, pp. 50–53. VLSI (1999)

14. Benini, L., Bogliolo, A., Paleologo, G., De Micheli, G.: Policy optimization for dynamic power management. IEEE Trans. Comput. Aided Des. Integr. Circuits Syst. **18**, 813–833 (1999). https://doi.org/10.1109/43.766730
15. Qiu, Q., Pedram, M.: Dynamic power management based on continuous-time Markov decision processes. In: Proceedings of the 36th ACM/IEEE Conference on Design Automation Conference—DAC'99 (1999). https://doi.org/10.1145/309847.309997
16. Lu, Y., Chung, E., Šimunić, T. et al.: Quantitative comparison of power management algorithms. In: Proceedings of the Conference on Design, Automation and Test in Europe—DATE'00 (2000). https://doi.org/10.1145/343647.343688
17. Simunic, T., de Micheli, G., Benini, L.: Event-driven power management of portable systems. 12th International Symposium on System Synthesis, pp. 18–23. IEEE Computer Society, Washington (1999)
18. Turkin, I., Vdovitchenko, A.: Energy-efficient scheduling for portable computers as bi-criteria optimization problem. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Concepts, Models, Complex Systems Architectures. Studies in Systems, Decision and Control, vol 74, pp. 87–100. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-44162-7_5
19. Saravjit, R.: Engagement experiences on the path to richer insight and truth. In: Russian UX Conference, p. 76 (2014)
20. Kurzantseva, L.: On the construction of intelligent adaptive interface based on agent-based technology for a general-purpose computer systems. Inf. Technol. Comput. Eng. **1**, 16–20 (2010)
21. Yu, C., Mannor, S., Theocharous, G., Pfeffer, A.: User model and utility based power management. AAAI'07: Matherials of the 22nd National Conference 1918–1919 (2007)
22. Graybill, R., Melhem, R.: Power Aware Computing, pp. 101–125 (2002)
23. Kharchenko, V., Illiashenko, O.: Concepts of green IT engineering: taxonomy, principles and implementation. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Concepts, Models, Complex Systems Architectures. Studies in Systems, Decision and Control, vol 74, pp. 3–19. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-44162-7_1
24. Bakeman, R., Gottman, J.: Observing Interaction, pp. 10–25 (2010)
25. Schervish, M.: Theory of statistics. Springer, New York, New York, NY (1995)
26. Graham, A.: Statistics. Hodder Education, London (2011)
27. Fitzgerald, B.: Open source software adoption. Int. J. Open Sour. Softw. Process. **1**, 1–23 (2009). https://doi.org/10.4018/jossp.2009010101

Part III
Industry and Transport Green IT Systems Applications

Green Assurance Case: Applications for Internet of Things

Vladimir Sklyar and Vyacheslav Kharchenko

Abstract The article discusses Assurance Case methodology for its implementation for Green Information Technology (IT) domain related with Internet of Things (IoT). We name this approach as "Green Assurance Case". The fundamentals of Assurance Case are explained as the initial part of the article. Safety and security requirements are considered as a part of typical Assurance Case. Green Assurance Case for IoT includes: sustainability assurance part, safety and security assurance part, Green IT principles, power consumption parameters part as well as requirements to green features of each of the four IoT layers. An example of the proposed Green Assurance Case methodology application for medical IoT system is represented.

Keywords Assurance case · Green IT · Internet of things

1 Introduction

Information Technologies (IT) and different computers application around the world consume 3% of the produced electrical energy as well as gives 2% of carbon emissions. Analysis of future trends in IT shows that power consumption and carbon emissions very probably will increase. So it becomes more and more important to assess computer systems and IT requirements compliance from the point of view of energy consumption. Such assessment shall also consider functional requirements as well as safety and security requirements. Also there are social, economic and environmental parts of sustainability which should be taken into account as parts of Green IT sustainability.

V. Sklyar (✉) · V. Kharchenko
National Aerospace University "KhAI", 17, Chkalova Street, Kharkiv, Ukraine
e-mail: v.sklyar@csn.khai.edu

V. Kharchenko
e-mail: v.kharchenko@csn.khai.edu

© Springer Nature Switzerland AG 2019

V. Kharchenko et al. (eds.), *Green IT Engineering: Social, Business and Industrial Applications*, Studies in Systems, Decision and Control 171,
https://doi.org/10.1007/978-3-030-00253-4_15

1.1 Motivation

In this article we apply Assurance Case [1] as a promising methodology, which is able to provide a basis for requirements compliance assessment of sustainable Green IT [2]. Assurance Case provides arguments that software or a system is acceptable for its intended use with respect to security and safety requirements. To be more specific we consider Internet of Thing (IoT) which is dramatically important to implement green IT approach [3]. Figure 1 provides general relations between three domains which give inputs to be analyzed. The coincidence of three sets is a subject of the article.

1.2 State-of-the-Arts

Available in the web information presents state-of-the-art researches and resources in area of Assurance Case. A challenge is to monitor and analyze relevant information by a systematic way [4]. The above analysis is based on the main results obtained in [4].

We perform such analysis in three directions. Firstly, we analyze achievement in standardization area, when Assurance Case or Safety Case implementation is required by some normative documents. Secondly we take into account companies and organizations which perform research in Assurance Case area. Finally, we investigate software tools applicable to build Assurance Case in semi-formal notations. Below we list the main sources available in the web.

The following standards and regulations are applicable for Assurance Case implementation:

- ISO/IEC 15026 “Systems and software engineering—Systems and software assurance” is a standard which contains four parts; this standards describes a notation for Assurance Case representing; the ISO/IEC 15026 notation is a bit different from Claim-Arguments-Evidence (CAE) and Goal Structured Notation (GSN) notations supported by software tools;
- ISO 26262:2011 “Road vehicles—Functional safety” standard in ten parts requires Safety Case implementation for automotive systems;

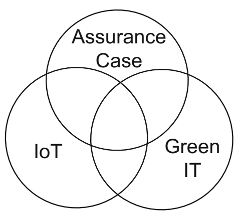


Fig. 1 Relations between domains analyzed in the article

- Object Modeling Group Structured Assurance Case Metamodel (OMG SACM) describes appropriate objects and relations [5];
- Goal Structured Notation (GSN) Community Standard [6] is closely related with OMG SACM providing the GSN description;
- The document "Common position of international nuclear regulators and authorised technical support organisations—Licensing of safety critical software for nuclear reactors" [7] describes software Assurance Case applicability in nuclear industry recognized by such countries as Belgium, Canada, Germany, Finland, Spain, Sweden, and UK;
- The document "European Organization for Safety of Air Navigation (EUROCONTROL)—Safety Case Development Manual" [8] describes Safety Case applicability for European Air Traffic Management Systems.

Some universities and research organizations pay efforts to development of Assurance Case methodologies for different domains at horizontal and vertical markets. Two pioneers in the Assurance Case semi-formal notations started their activities in the UK in 1990s. There is Adelard that developed CAE notation and designed software tools to implement computer added Safety and Assurance Case [9]. Researchers from High-Integrity Systems Engineering (HISE) Research Group in University of York proposed the GSN which after becomes a worldwide used notation for Assurance Case [10].

The last years VTT (Technical Research Centre of Finland) considers opportunities to implement Assurance Case for safety approval in nuclear industry [11].

In the U.S. two huge government organizations, such as National Aeronautics and Space Administration (NASA) and Department of Homeland Security (DHS) have already implemented Assurance Case approach for their products and services. NASA established the Robust Software Engineering Group in the Intelligent Systems Division for support of independent Verification and Validation (V&V) [12]. Independent V&V is implemented by NASA for space programs as well as for Unmanned Aircraft Vehicles (UAV). The DHS Cyber Emergency Response Team (US-CERT) implements Assurance Case methodology to establish security assurance ecosystem. The last activity, including lecture courses providing, is widely supported in Software Engineering University (SEI) which is a part of Carnegie Mellon University (CMU) [13].

Survey of software tools for Assurance Case support provides the three following candidates [14]:

- ASCE is designed by Adelard company (UK) in 1990s; it supports both CAE and GSN notations;
- Astah GSN is designed by Change Vision company (Japan) some years ago as a part of a large system modeling package; it supports only GSN;
- NOR-STA is designed in 2000s by Argevide (Poland) which is an university spin-off company; it supports its own notation TRUST-IT which complies with IEC/ISO 15026 "Systems and software assurance"; Assurance Case can be represented in both graphical and hyper-textual view.

All three companies provide tool trial versions as well as academic and commercial licenses. The last time NASA employers have developed the tool named AdvoCATE to support Assurance Case in GSN. Advantages of AdvoCATE include hierarchical and modular abstraction of Assurance Case as well as support of argument patterns [12].

Article [15] proposes a methodology called Assurance Case Driven Design (AC DD) which formalizes the life cycle of a system or software in terms of assessment of safety and security. Assurance Case was adapted to assess the safety of IoT systems in [16]. We can conclude from the above, that Assurance Case is advanced methodology which mature enough and supported with semiformal methods (CAE and GSN) and appropriated computer tools [5, 11, 12]. A positive experience of Assurance Case application in safety critical and security critical domains gives evidence that methodology can be extended for similar areas like Green IT. Since green features become more and more important, it can be considered as one from additional critical requirements. So Assurance Case approach can be applied also for Green IT domain. In this paper we introduce the term “Green Assurance Case” as Assurance Case which consider green features of systems or software.

From the point of view of Green IT, it is worth to consider already performed study of existing articles. For example, article [17] propose the following categories of the existing researches in green software:

- Energy Efficiency: Papers that present energy consumption calculation methods, energy-efficient systems, and/or metrics for optimizing system performance with lower costs;
- Development Methodology: Papers that present improved algorithms or frameworks for software development;
- Process Enhancement: Papers that focus on specific sub-section of processes used in the life cycle of a software product;
- Organizational Metrics: Papers that are directed towards administrations and present evaluation of current procedures or papers that aim to improve these;
- Life-Cycle Thinking: Papers that cover all processes in the software product lifespan.

Taking into account safety critical and security critical domain, energy efficiency has to be the main focus of Green Assurance Case. Another important issue is Safety and Security Life Cycle (SSLC). As it is shown below (see Sect. 3) development of Green systems and software shall follow a rigorous staged process which can be the same with SSLC. In this paper we discuss the joint Green SSLC.

Since energy efficiency is the main feature of Green IT it is true also for green IoT. IoT is a heterogeneous system which contains four layers including device layer, network layer, service layer, and application layer. Different approaches and models can be applied for different layers of IoT. For device layer it is important to consider power consumption on the level of electronic components [18]. A direct measurement of power consumption depending on hardware and software modes is also applicable at the device layer [3, 19]. For device and network layers an approach based on additive calculation of power consumption given in manufacture datasheets can be applied [20]. Service layer is usually implemented on the base of cloud technology, so models for power consumption measurement for data centers can be applied [21, 22].

There is the biggest uncertainty for power measurement at the IoT application layer that can be estimated mostly with some indirect methods [23, 24].

1.3 Goal and Structure

A goal of this article is to investigate a potential of Assurance Case methodology to be implemented in Green IT domain in respect to IoT systems. To achieve this goal the following sections are included in the chapter.

The main feathers of Assurance Case Driven Design approach are considered in Sect. 2 including Assurance Case levels, Claim-Argument-Evidence-Criteria (CAEC) Notation and Development-V&V-Assurance Case (DVA) Notation. Section is based on the main results obtained in [16].

After that we discuss framework of safety and security requirement that is a mandatory part of Assurance Case. This framework contains requirements to safety and security management, requirements to safety and security life cycle as well as requirements to failures and cyber-attacks avoidance countermeasures. Such requirements are discussed in Sect. 3. Section is based on the main results obtained in [15].

Green IT principles are discussed in Sect. 4. Some known principles are adapted for IoT with addition of new proposed principles.

Based on the above results, a structure and a content of Green Assurance Case is developed in Sect. 5 as "a big picture" for the proposed methodology. Green Assurance Case contains Sustainability Assurance part, Safety and Security Assurance part (based on Sect. 3 results), Green IT principles part (based on Sect. 4 results), Power consumption parameters part, and parts associated with each of the IoT architecture layer (device layer, network layer, service layer, application layer).

An example of the proposed Green Assurance Case methodology application for medical IoT system is represented in Sect. 6. General conclusions are discussed in Sect. 7.

2 Assurance Case Driven Design

As result of our latest research [4, 15, 16], we propose to apply Assurance Case Driven Design to meet stakeholders expectations from sides of developers, utilities and regulators. Section 2.1 discusses possible levels of Assurance Case depending on life cycle stages and product evolution. The discussed AC DD approach includes Claim-Argument-Evidence-Criteria Notation (see Sect. 2.2) and Development-V&V-Assurance Case Notation (see Sect. 2.3).

2.1 Assurance Case Levels

Assurance Case may be developed for different stages of life cycle as well as for different levels of the product implementation. From the point of view of life cycle evolution the following levels can be considered:

- Life Cycle Level 0 (L0)—Conceptual or Preliminary Assurance Case which is developed on the earliest stages of system or software life cycle, when Product Concept Document or Requirements Specification is developed;
- Life Cycle Level 1 (L1)—Architecture or Intermediate Assurance Case which is developed at the middle of system or software life cycle, when Product Architecture Design or Detailed Design is developed;
- Life Cycle Level 2 (L2)—Implementation Assurance Case which is developed at the end of system or software life cycle, when a product is developed and is ready to be installed and commissioned;
- Life Cycle Level 3 (L3)—Operational Assurance Case which is developed when system or software is in operation, and during long time operation, Operational Assurance Case shall be updated if any condition is changed.

From the point of view of product evolution the following levels can be considered:

- Product Level 0 (P0)—Platform Assurance Case which is developed for a basic system or software functionality;
- Product Level 1 (P1)—Application Family Assurance Case which is developed for some set of applications which can be installed at different sites;
- Product Level 2 (P2)—Specific Application Assurance Case which is developed for an application which is installed at some specific site.

Taking into account two types of Assurance Case categories, it possible to implement 2D-identification, depending on Life Cycle Level and Product Level (see Fig. 2). For example, (L2, P1) means Implementation Assurance Case for Application Family. All or some parts of the twelve identified types of Assurance Case can be requested by Regulatory Authority.

Assurance Case has two sides of description and implementation:

(1) A static part which describes an approach to combine arguments for assurance support (see Sect. 2.2);
(2) A dynamic part to support a static part movement between stages of analyzed system life cycle (see Sect. 2.3).

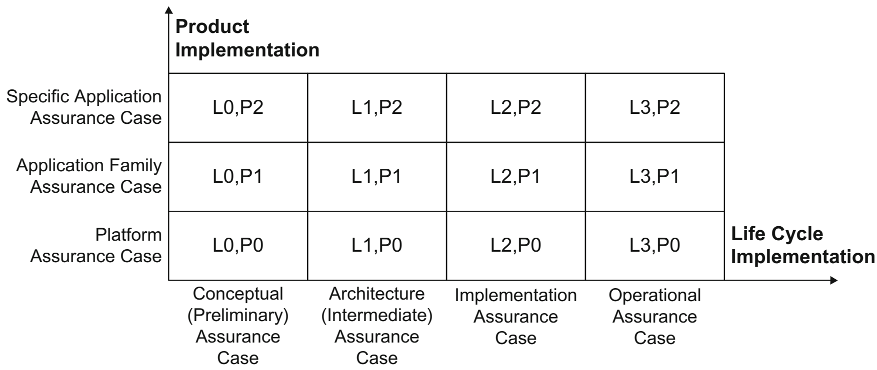


Fig. 2 Assurance case levels for life cycle evolution and product evolution

2.2 *Claim-Argument-Evidence-Criteria (CAEC) Notation*

In the AC DD framework we propose some addition for Assurance Case CAE notations to be able assess specific features of critical systems. Acceptance criteria and coverage criteria are two additional entities which have to be taken into account for support arguments and evidences. Acceptance criteria are the conditions when stated requirements are met. From the point view of Assurance Case, acceptance criteria provide us ability to state the right arguments which are consistent with the claim and to provide the evidences which are consistent with the arguments.

Coverage criteria describe how completely the claim is met. From the point view of Assurance Case, coverage criteria provide us ability to state multiple arguments to completely cover all claim features and to provide multiple evidences which completely cover the arguments. Acceptance criteria for a claim can be extracted from both argument and/or evidence. In general case acceptance criteria provide a quantitative and qualitative description of a situation when the claim is met. A coverage criterion is a measure used to describe the degree to which evidence for specific arguments is provided. A graphical view of the modified CAE notation which we name Claim-Argument-Evidence-Criteria (CAEC) notation is given on Fig. 3.

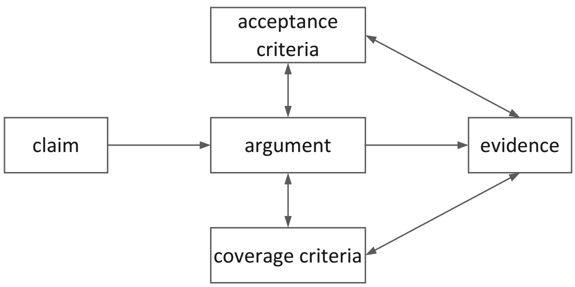


Fig. 3 Claim-Argument-Evidence-Criteria (CAEC) notation

The next step of CAE/CAEC notation development is to support activities of SSLC stages with implementation of Assurance Case. Specification and design requirements are the inputs for each of the SSLC stage. After any stage fulfillment, requirements implementation assessment has to be performed.

2.3 Development-V&V-Assurance Case (DVA) Notation

The following activities are mandatory for each of the SSLC stage (see Fig. 4):

- Development targeted to move an implemented product representation stage by stage through SSLC;
- V&V targeted to check conformance of the SSLC stage development outputs to the SSLC stage development inputs;
- Assurance Case update based on assessment of performed development and V&V activities.

The proposed Development-V&V-Assurance Case (DVA) notation bases on the following fundamentals:

- Safety and Security Life Cycle can be represented in a view of three components: Development, Verification and Validation and Assurance Case;
- Development activities are staged implementation of requirements in design description of system, hardware and software, and after that implementation of requirements in a physical system, hardware and software;
- Development also covers processes implementation to support development of the product; processes also are described in a view of requirements which are collected in project plans;
- Typically requirements are represented and handled as database records; from this point of view the main operation with requirements are CREATE (to add),

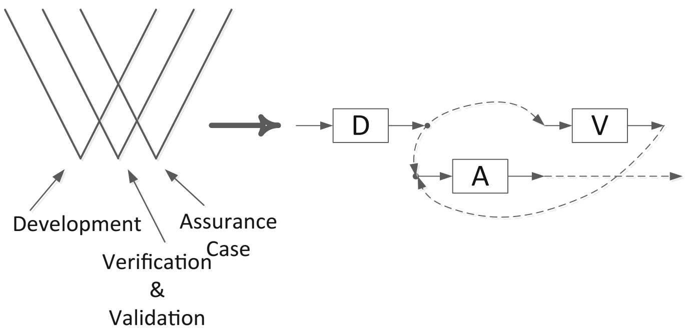


Fig. 4 Transition from V-shape life cycle to Development-V&V-Assurance case (DVA) notation

DELETE, MODIFY (if requirement needs some sense correction), EDIT (if requirement needs only editorial correction without changing of a sense);

- Forward and backward requirement tracing shall be implemented at each of SSLC stage to assure: (1) all previous stage requirements are implemented into the next stage documents; (2) no new requirements appear in the next stage documents; (3) all the requirements are verified or validated;
- Compliance of the product of next SSLC stage with the product of the previous SSLC stage is checked by implementation of V&V process;
- Compliance of processes implementation (including development and V&V processes) is checked by audits when processes implementation evidences are investigated against the project plans requirements; these audits can be a part of Assurance Case activities;
- All three D, V and A components of SSLC have specific inputs and outputs for each of the stage; so a diagram on Fig. 4 represents DVA relations for some single stage.

To develop a graph and theoretical-set based model for DVA notation a diagram on Fig. 4 should be elaborated to reflect feedback relations after V&V and Assurance Case performance (see Fig. 5). Direct data transmission and feedback data are highlighted with different templates of lines. It is clear for Fig. 5, that input and output sets have some overlapping, so sets of DVA data flows ate described in terms of inputs. The data sets transmitted between components of DVA are described in [16].

3 Framework of Safety and Security Requirements

Result of industrial standards considering [15] allows representing existing security requirements to Industrial Control Systems (ICS) and IoT related with a restricted set of categories. This conceptual security requirements taxonomy includes four the main parts:

- Risk management and assessment as a corner stone for definition of acceptable risks levels and countermeasures for risks reduction;

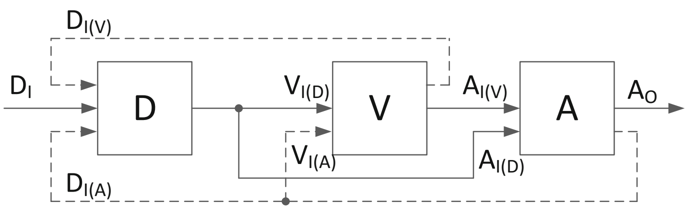


Fig. 5 Graph and theoretical-set based description of DVA notation

- Categories of security features implementation which include triad "People-Process-Technologies";
- ICS context which drive to define requirement taking into account specifics of ICS; this concept includes three types of models (reference, physical architecture and zone models) as well as functionality, components, assets and other definitions, as well as security and safety coordination issues;
- ICS security levels concept which grades risk levels for ICS separated parts and establishes different life cycle processes and countermeasures for different security levels.

A set of ICS functional safety requirement can be found in series of industrial standards, for example, IEC 61508 "Functional safety of electrical/electronic/programmable electronic safety-related systems" or IEC 61511 "Functional safety —Safety instrumented systems for the process industry sector". These functional safety requirements can be divided in some following categories [15] (see Fig. 6):

- Requirements to functional safety management;
- Requirements to functional safety life cycle;
- Requirements to systematic (system and software design) failures avoidance;
- Requirements to random (hardware) failures avoidance.

A scope of the above requirements is highly dependent from as named Safety Integrity Level (SIL) which establishes relation between system risk level and a

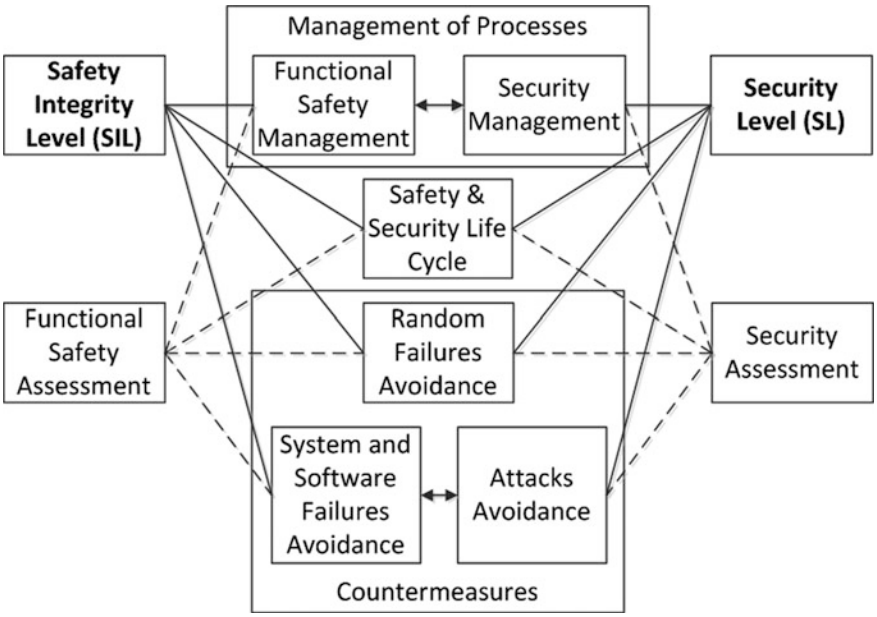


Fig. 6 A concept of harmonized safety and security requirements

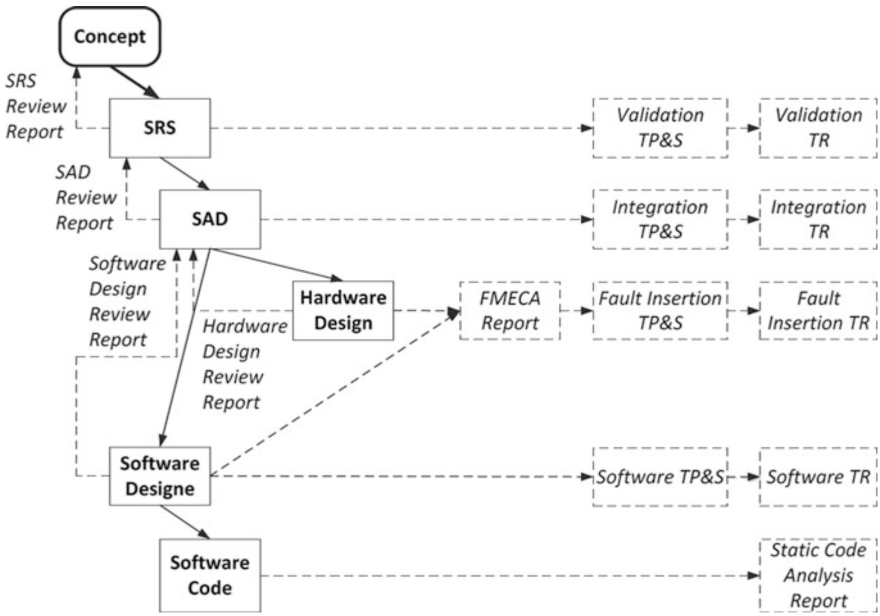


Fig. 7 V-shape green safety and security life cycle

scope of the related safety assurance countermeasures. The above approach can be applied for security concept (see Fig. 6).

At the present requirements related with countermeasures are rapidly developed, especially from the prospective of new appeared types of cyber-attacks. Parts of requirements related with safety and security management and with SSLC are well defined in the standards and can be formalized for common use. Figure 7 provides a structure of SSLC which is complied with the IEC 61508 and many safety and security standards. It worth to note, that SSLC structure and content is applied as Green Life Cycle [23]. This position is discussed as one of the Green IT principles in Sect. 4.

4 Green IT Principles for Internet of Things

Today IT consume up to 3% of general power consumption of humanity. Important trends in IoT industry are miniaturization, longevity, performance increasing and other similar issues related with power efficiency.

General Green IT principles shall be implemented in IoT area. Such principles include the following [20]:

1. Power modes management (PMM) and power off all the equipment when it is not needed.

2. Advanced communications technologies and communication optimization including data volume and communications routing optimization.
3. Improvement efficiency of batteries and other power storages.
4. Using renewable power sources and power harvesting from motion, vibration, radio frequency and electromagnetic emission, thermal sources, radiation, etc.
5. Implementation of best practices for algorithms and software development from the point of view of resources allocation, such as time duration, memory, frequency, computations support etc.
6. Trade off with implementation of safety and security features; since safety and security features require additional resource consuming, optimization criteria have to be implemented during IoT systems design.
7. Implementation Green life cycle which is coincident with SSLC (see Fig. 7) and can be applied as common Green SSLC.
8. Separation of requirements and metrics between four IoT layers (device layer, network layer, service layer, application layer). For example, green metrics, such as Power metrics, Efficiency metrics, and Productivity metrics can be interpreted separately for each of the IoT layer [21]. The same is right for a power consumption model which is an integral model collecting all four IoT layers.

The above Green IT principles implementation is defined by IoT system architecture. Requirements for IoT components have been identified by different vendors, system integrators, consortia etc. IoT Reference Architecture (IoT-RA) is a subject of standardization, what is developing now by International Electrotechnical Commission (IEC) and Institute of Electrical and Electronics Engineers (IEEE). The IoT-RA should describe the system or systems for the IoT, the major components involved the relationships between them, and their externally visible properties [3]. Existing layers and interfaces are points to implement standardized green IoT solutions. Green IoT metrics also can be a part of the future standards [21]. Device layer is represented by sensor networks which are connected with minicomputers or controllers.

The modern researches discuss concerning high immaturity of the IoT market which is still be dynamical from the point of view of appearance and disappearance of the main market players. At the present, the device layer is the most predictable in IoT–RA. The device layer has a typical structure of the Computer Control System (CCS), like, for example, embedded systems. Control systems fundamentals lay in interaction with processes of the real world via three the main parts which are sensors, controllers and actuators.

PMM is the most important approach for sensors and on-board computers at the device layer. Also servers and data centers at the service layer can save a lot of energy, if PMM is implemented. Advanced communication technologies implementation is obviously important at the network layer as well as for the communications associated with clouds at the service layer. Application of power storages including advanced batteries is critical for device layer, network layer and service layer. For autonomous parts of device layer and network layer it is applicable renewable energy using as well power harvesting from motion, vibration, thermal sources, electromagnetic emitting, etc.

5 Structure of IoT Green Assurance Case

Structured Green Assurance Case should include some main parts when it is implemented for IoT area. All parts shall be specific for any IoT application, but we focus on the patterns which are common for the most part of IoT-based systems.

5.1 *Green Assurance Case Diagram*

Figure 8 represents "a big picture" for IoT Green Assurance Case by joining all elements of assessment.

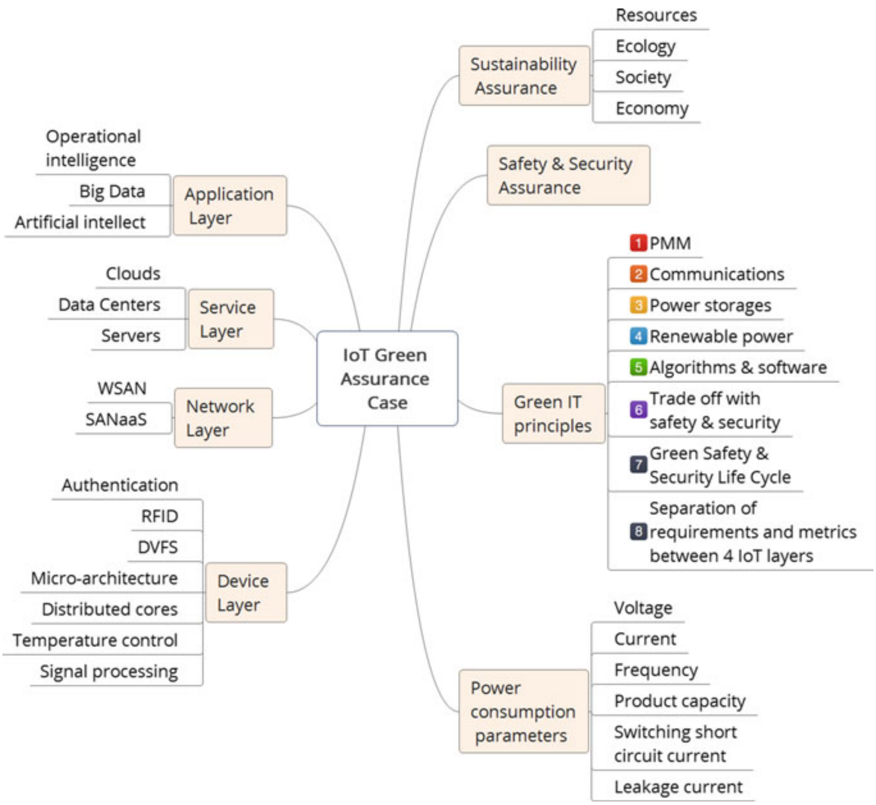


Fig. 8 IoT green assurance case

Firstly, Green ITs are directed to support sustainable development. For that sustainability assurance part is included in Assurance Case to check an influence to resources, ecology, society, and economy [2].

Secondly, the main issues related with safety and security requirements assurance and assessment shall be incorporated to Assurance Case. This part of Assurance Case is considered above in Sect. 3 (see Fig. 6). After that it makes a sense to submit above six Green IT principles adopted for IoT.

Power consumption parameters are also included as a part of Assurance Case since they provide a well-defined part of requirements to IoT-based system. For example, processor power consumption can be calculated in accordance with the following equation [19]:

$$\mathrm{P = A \cdot C \cdot V^2 \cdot f + A \cdot t \cdot V \cdot Ishort \cdot f + V \cdot Ileakage},$$

where the first item measures the dynamic power consumption caused by the charge and discharge of the capacitive load at the output of each key, which is equal to the product of the capacity—C, the square of supply voltage—V, the processor's frequency—f, and the coefficient A, which characterizes the activity of the keys in the system; the second item is the power expended as a result of short-circuit current, which takes place at the time of switching the logical element; the third item are losses due to current leakage.

The main specific part of Assurance Case includes specific requirement to green features which should be implemented for each of the four IoT layers.

5.2 *Green Features at the Device Layer*

Device layer includes mainly field sensors and actuators as well as programmable controllers mainly in a view of on-board computers. Device layer is relatively simple and well defined, so for it we can describe common green features with higher degree of certainty [19].

One of the key technologies applied at the device layer is technique of identification and authentication. For example, Radio Frequency Identification (RFID) is widely used at the present in IoT systems. It is important to specify device identification functions to meet Green IT principles.

After that PMM with power-aware scheduling based techniques shall be implemented at the device layer. For example, Dynamic Voltage and Frequency Scaling (DVFS) is widely used for embedded systems to alter the voltage and/or frequency of a programmable component based on performance and power requirements.

Micro-architecture solutions support saving energy in specific components with dynamical reconfiguration. For example, there are different techniques for buffering, memory compression, memory size adjusting, cash providing for simultaneous reading/writhing access etc.

Use of distributed cores for calculation allows to manage multitask environment and assign task to alternative programmable components, such as Digital Signal Processors (DSP), Field Programmable Gates Arrays (FPGA) and other. Tasks are assigned depending which component is more appropriate from the point of view of energy consumption.

Temperature control techniques are used for devices since a temperature mode affects longevity of components operation.

Finally, signals processing techniques are widely used at the device layer, so power efficiency of signals processing algorithms and software can noticeably decrease power consumption.

5.3 Green Features at the Network Layer

Wireless Sensor Networks (WSN) with different protocols and topologies should be based on advanced communication technology such as, for example, cognitive radio with autocorrecting of power efficiency parameters or Multiple Input Multiple Output (MIMO) with enforcement the capacity of communication channel.

Since the same sensors and networks can be used for different application, service providers operate global and local WSN infrastructure which can be considered as "Sensor Network as a Service" (SNaaS).

5.4 Green Features at the Service Layer

There are a lot of features to be implemented for green clouds which serve as a platform for the service layer. Modern green data centers and green servers are based on techniques using PMM, advanced communications, and advanced power storages.

For example, [22] propose to define energy consumption of a data center as

$$E = E\,\mathrm{const} + E\,\mathrm{storage} + E\,\mathrm{computation} + E\,\mathrm{communication},$$

where E const—constant or fixed part of energy consumption, E storage—energy consumption of data storage resources, E computation—energy consumption of computation resources, E communication—energy consumption of communication resources.

5.5 Green Features at the Application Layer

Successful operation of application layer depends on implementation of power efficient algorithms in software. From this prospective a business can be supported with operational intelligence based on advanced technique of big data and artificial intellect.

For example, GREENSOFT model [23] states sustainability for software based on consideration of life cycle, sustainability criteria and metrics, procedure model, as well as recommendations for actions and tools (see Fig. 9).

GREENSOFT model is highly related with Green SSLC (see Fig. 7). From this point of view application layer implementation is based on Green IT principle #7 "Implementation Green life cycle which is coincident with SSLC" (see Fig. 8).

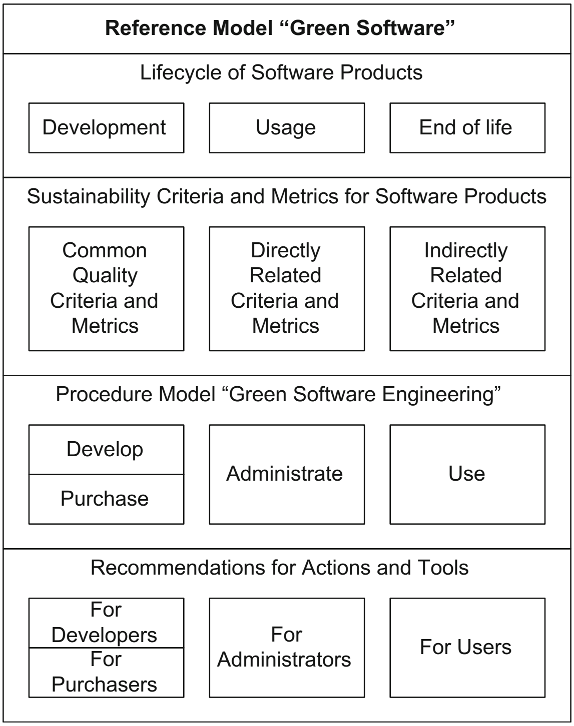


Fig. 9 GREENSOFT model of sustainable software

6 Example of Green Assurance Case Application for Medical IoT System

Let's consider medical IoT based system which measure blood pressure of some hospital patients and collect these data at the cloud. User application provides access to the collected data as well as support of analysis of data and data-based decision making (see Fig. 10). From the point of view of IoT layers, device layer include a set of blood pressure devices, network layer includes local network routers and medical service gateway, service layer includes an appropriate part of cloud data center, and application level includes user terminals with installed user application. Blood pressure devices, local network routers and user terminals can be multiple.

A general Green Assurance case diagram provides inputs for methodology application. Implementation of this methodology can be provided in a table view (see Table 1). Table 1 proposes a top-level framework for combined sustainability-safety-security assessment. This framework can specified for each of the part of the Green Assurance Case.

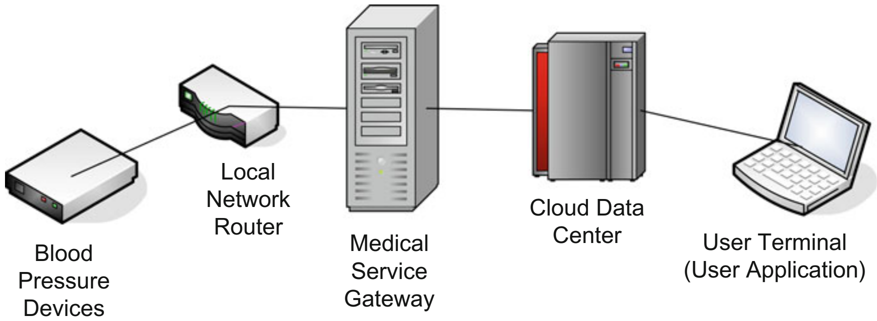


Fig. 10 An example of IoT based blood pressure system

Table 1 Results of green assurance case methodology application for IoT based blood pressure system

Parts of green assurance case	Evidence providing
Sustainability assurance: resources, ecology, society, economy	Consistency with sustainability policy of the hospital
Safety and security assurance	See Fig. 6
Safety integrity level (SIL)	To establish SIL in accordance with the IoT system risk assessment
Security level (SL)	To establish SL in accordance with the IoT system risk assessment
Functional safety management	To establish and to assess functional safety management process in accordance with the standard IEC 62304, medical device software
Security management	To establish and to assess security management process in accordance with the standards series ISO/IEC 27000, information security management systems
Green SSLC	To establish system development in accordance with V-shape life cycle (see Fig. 7) and GREENSOFT model (see Fig. 9)
Random failures avoidance	To implement protective measure (redundancy, self-diagnostics, etc.) in accordance with the standard IEC 61508
Systematic failures	To implement protective measure (project management, semi-formal methods, etc.) in accordance with the standard IEC 61508
Cyber-attacks avoidance	To implement protective measure in accordance with some cyber security framework (ISO/IEC 15408, ISA/IEC 62443, etc.)
Green IT principles	See Sect. 4
Power consumption parameters	Shall be specified in accordance with devices which are applied for four IoT layers
Device layer	Blood pressure devices are used. Power consumption parameters shall be analyzed and approved. Authentication features shall be analyzed and verified. Used wireless protocol shall be analyzed and approved
Network layer	Local network routers and medical service gateway are used. Power consumption parameters shall be analyzed and approved. Network topology shall be analyzed and approved
Service layer	Appropriate equipment of cloud data center is used. Power consumption parameters shall be analyzed and approved
Application layer	User terminals are used. Power consumption parameters shall be analyzed and approved Application software shall be developed in accordance with Green SSLC and/or GREENSOFT model. Software sustainability criteria shall be analyzed and verified

7 Conclusions: The Future Researches

Assurance Case Driven Design fundamentals are represented in the article with the objective to apply Assurance Case approach for Green IT. Notations and supporting tools for Assurance Case have been represented. We also consider in general the main structure of safety and security requirements including management and life cycle requirements. A specific part of Assurance Case has been developed to establish an interconnection between requirements to power consumption and approach to produce argument for evidence of compliance with the stated requirements. The proposed methodology of integrated safety, security and power consumption assessment is named as "Green Assurance Case".

The following Green Assurance Case issues shall be discussed in frameworks of the proposed methodology. Firstly, there are issues related with power consumption restrictions or energy efficiency requirements for safety and security critical systems (embedded components for IoT-based applications) which shall be introduced into general Assurance Case template elements. Secondly, there are issues related with specific Assurance Case targeting on green features. In such case requirements and constraints regarding safety and security must be taken into account. For modernized systems pre-development of Green Assurance Case could be based on technique of Green Gap Analysis. Thirdly, we consider specific requirements, indicators and constraints for power consumption features of IoT based systems. It is worth to notice that today we do not have any finalized standard which would provide a sufficient list of requirement on this issue, so some best practices are analyzed and implemented.

Finally, Green Assurance Case patterns for IoT based systems have been proposed. These patterns contain typical sets of requirements, arguments and evidences related with safety, security and energy consumption issues. Green IT features for each of the four IoT Reference Architecture levels are considered. Relations between Green IT principles and appropriate IoT layers are established. Table 1 proposes a top-level framework for combined sustainability-safety-security assessment, which can be detailed for any specific IoT system

References

1. Kelly, T.: Arguing safety: a systematic approach to managing safety cases. Ph.D. thesis, University of York, 152 p (1998)
2. Kharchenko, V., Illiashenko, O.: Concepts of green IT engineering: taxonomy, principles and implementation. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Concepts, Models, Complex Systems Architectures. Studies in Systems, Decision and Control, vol. 74, pp. 3–19. Springer International Publishing, Berlin, Heidelberg (2017). https://doi.org/10.1007/978-3-319-44162-7_1

3. Sklyar, V.: Vedic mathematics as fast algorithms in green computing for internet of things. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Components, Networks and Systems Implementation. Studies in Systems, Decision and Control, vol. 105, pp. 3–21. Springer International Publishing, Berlin, Heidelberg (2017). https://doi.org/10.1007/978-3-319-55595-9_1
4. Sklyar, V., Kharchenko, V.: Challenges in assurance case application in industrial IoT. In: Proceedings of the 9th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS 2017), Bucharest, pp. 736–739, Sept 2017
5. Structured Assurance Case Metamodel, v2.0, Object Management Group (2016)
6. GSN Community Standard, Version 1, Origin Consulting (York) Limited (2011)
7. Common Position of International Nuclear Regulators and Authorised Technical Support Organisations—Licensing of Safety Critical Software for Nuclear Reactors (2015)
8. Safety Case Development Manual, European Organization for Safety of Air Navigation (EUROCONTROL) (2006)
9. Bishop, P., Bloomfield, R.: A methodology for safety case development. In: Proceeding of the 6th Safety-Critical Systems Symposium, Birmingham, UK, Feb 1998
10. Hawkins, R., Kelly, T., Habli, I.: Developing assurance cases for D-MILS systems. In: Proceedings of International Workshop on MILS: Architecture and Assurance for Secure Systems, Amsterdam, The Netherlands, Jan 2015
11. Valkonen, J., Tommila, T., Linnosmaa, J., Varkoi, T.: Safety demonstration of nuclear I&C—an introduction. SAUNA task 3.1 report. Research report VTT-R-00167-16, VTT (2016)
12. Denney, E., Pai, G.: Safety case patterns: theory and applications, research report NASA/TM–2015–218492, NASA (2015)
13. Weinstock, C., Goodenough, J.: Towards an assurance case practice for medical devices, technical note CMU/SEI-2009-TN-018, SEI (2009)
14. Linnosmaa, J.: Structured safety case tools for nuclear facility automation. MS Thesis, Tampere University of Technology (2016)
15. Sklyar, V., Kharchenko, V.: Assurance case driven design based on the harmonized framework of safety and security requirements. In: Proceedings of the 13th International Conference on ICT in Education, Research and Industrial Applications (ICTERI 2017). Kyiv, pp. 670–685, May 2017
16. Sklyar, V., Kharchenko, V.: Assurance case driven design for internet of things. WSEAS Trans. Comput. Res. **4**, 173–182 (2016)
17. Berntsen, K., Olsen, M., Limbu, N., Tran, A., Colomo-Palacios, R.: Sustainability in software engineering—a systematic mapping. In: Mejia, J., Muñoz, M., Rocha, Á., San Feliu, T., Peña, A. (eds.) Trends and Applications in Software Engineering, pp. 23–32. Springer International Publishing, Berlin, Heidelberg (2017). https://doi.org/10.1007/978-3-319-48523-2_3
18. Mittal, S.: A survey of techniques for improving energy efficiency in embedded computing systems. Int. J. Comput. Aided Eng. Technol. **6**(204), 440–459
19. Sklyar, V., Odarushchenko, O., Bulba, E., Horbenko, R., Ivasyuk, A., Kotov, D.: Assessment of energy consumption for safety-related PLC-based systems. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Concepts, Models, Complex Systems Architectures, Studies in Systems, Decision and Control, vol. 74, pp. 269–281. Springer International Publishing, Berlin, Heidelberg (2017). https://doi.org/10.1007/978-3-319-44162-7_14
20. Sheng, Z., Mahapatra, C., Zhu, C., Leung, V.: Recent advances in industrial wireless sensor networks toward efficient management in IoT. IEEE Access **3**, 622–637 (2015)
21. Ardito, L., Morisio, M.: Green IT—available data and guidelines for reducing energy consumption in IT systems. Sustain. Comput. Inf. Syst. **4**, 24–32 (2014)
22. Chen, F., Schneider, J.-G., Yang, Y., Grundy, J., He, Q.: An energy consumption model and analysis tool for cloud computing environments. In: Proceedings of the First International Workshop on Green and Sustainable Software. Piscataway, NJ, USA, pp. 45–50, June 2012

23. Naumann, S., Dick, M., Kern, E., Johann, T.: The GREENSOFT model: a reference model for green and sustainable software and its engineering. Sustain. Comput. Inf. Syst. **1**, 294–300 (2011)
24. Kondratenko, Y., Korobko, O., Kozlov, O.: PLC-based systems for data acquisition and supervisory control of environment-friendly energy-saving technologies. In: Kharchenko, V., Kondratenko, Y., Kacprzyk J. (eds.) Green IT Engineering: Concepts, Models, Complex Systems Architectures, Studies in Systems, Decision and Control, vol. 74, pp. 247–267. Springer International Publishing, Berlin, Heidelberg (2017). https://doi.org/10.1007/978-3-319-44162-7_13

Synthesis and Optimization of Green Fuzzy Controllers for the Reactors of the Specialized Pyrolysis Plants

Oleksiy Kozlov, Galyna Kondratenko, Zbigniew Gomolka and Yuriy Kondratenko

Abstract This paper presents the developed by the authors generalized step-by-step method of synthesis and optimization of green fuzzy controllers (FC) for the automatic control systems (ACS) of the reactor's temperature of the specialized pyrolysis plants (SPP). The proposed method gives the opportunity to synthesize and optimize Mamdani type green FCs of the temperature modes of the SPPs reactors that provide (a) high accuracy and quality indicators of temperature control, (b) low energy consumption in the process of functioning as well as (c) relatively simple software and hardware implementation. The initial synthesis of the structure and parameters of green FCs is implemented on the basis of expert assessments and recommendations. Their further optimization for improving the quality indicators, reducing energy consumption and simplification of soft/hardware realization is carried out using specific optimization procedures by means of mathematical programming methods. In order to study and validate the effectiveness of the developed method the design of the Mamdani type green FC for the temperature ACS of the pyrolysis reactor of the experimental SPP has been carried out in this work. The developed green FC has a relatively simple hardware and software implementation as well as allows to achieve high quality indicators of

O. Kozlov · G. Kondratenko · Y. Kondratenko (✉)
Department of Computer-Aided Control Systems, Admiral Makarov National University of Shipbuilding, Mykolaiv, Ukraine
e-mail: y_kondrat2002@yahoo.com; yuriy.kondratenko@chmnu.edu.ua

O. Kozlov
e-mail: kozlov_ov@ukr.net

G. Kondratenko
e-mail: galvlad09@rambler.ru

G. Kondratenko · Y. Kondratenko
Department of Intelligent Information Systems, Petro Mohyla Black Sea National University, Mykolaiv, Ukraine

Z. Gomolka
Department of Computer Engineering, University of Rzeszow, Rzeszow, Poland
e-mail: zgomolka@ur.edu.pl

© Springer Nature Switzerland AG 2019

V. Kharchenko et al. (eds.), *Green IT Engineering: Social, Business and Industrial Applications*, Studies in Systems, Decision and Control 171,
https://doi.org/10.1007/978-3-030-00253-4_16

temperature modes control at a sufficiently low energy consumption, that confirms the high efficiency of the proposed method.

Keywords Green fuzzy controller · Synthesis and optimization
Automatic control system · Specialized pyrolysis plant · Pyrolysis reactor

1 Introduction

The multi-loop circulatory pyrolysis is an advanced green technology that is widely used for solving the environmental problem of the thermal utilization of the polymeric solid waste [1–3]. This technology allows complete utilization of the whole scope of the organic part of waste with further obtaining of alternative liquid and gaseous fuel in the environmentally friendly and energy-saving modes. For realization of the given technology specialized pyrolysis plants are used, which are, in turn, complex multi-component technical objects, that should be automated on the basis of highly effective automatic control systems [3–5].

One of the most important tasks of automation of the SPPs control processes is the implementation of proper control and stabilization of temperature modes of their pyrolysis reactors under conditions of various types of disturbing influences for providing the SPPs green operating mode, which is characterized by low energy consumption, high economic and operational indicators, as well as quality of the alternative fuel obtained at the outlet [3, 5, 6]. The complexity of this task increases due to the fact that the SPP pyrolysis reactor is a complex control object with substantially nonlinear and non-stationary parameters, and the value of its heating temperature may change under the influence of uncertain parametric and coordinate disturbances in rather wide limits [5–7]. Thus, to solve this problem it is expedient to develop and use specialized intelligent green controllers in the pyrolysis reactors ACSs, which would allow to achieve enough high quality indicators of temperature control at a sufficiently low energy consumption, that, accordingly, would significantly increase the energy and economic efficiency of the SPPs generally [8–10]. Also, such green controllers should have a relatively simple software and hardware implementation [11–13].

Analysis of approaches, methods and algorithms for the design and implementation of specialized intelligent green controllers for the temperature ACSs of the SPPs reactors shows the expediency of fuzzy control principles [14–16] using based on the theory of fuzzy sets and fuzzy logic [17–19]. Automatic control systems with Mamdani and Sugeno fuzzy controllers are developed and successfully used for controlling of complex technical objects with uncertain or non-linear parameters in various technical fields to achieve their high quality indicators of control, low energy consumption as well as high economic and operational efficiency [20–22].

An important feature of the ACSs based on FCs is that they are developed mainly on the basis of expert assessments, and their performance indicators significantly depend on the qualifications and experience of the developers (operators or experts), as well as on the number of subjective factors [23–25]. In particular, FCs of Mamdani type, developed on the basis of expert knowledge, have low efficiency at controlling of objects, for which there is no significant manual control experience gained by their operators [26–28]. Thus, the urgent problem is the development of effective approaches, methods and algorithms for fuzzy control devices designing, which will reduce the influence of subjective factors on the design process, increase in general the quality indicators of control, as well as energy and economic efficiency of the fuzzy ACSs of complex little-studied control objects.

2 Related Works and Problem Statement

A lot of approaches and methods of FCs synthesis for ACSs of technical objects of various purposes, which are based on the use of expert assessments in determining their structure and parameters, are developed for now [29–31]. The subjective factor in this case has a significant impact on the FC design process, and when making false decisions in general, the effectiveness of fuzzy ACSs may decrease, or their functioning will be carried out with understated (in terms of potential FC capabilities) quality indicators. For example, in the work [27] the application of the Mamdani type FC, developed on the basis of expert assessments, in the temperature ACS of the multi-circuit pyrolysis reactor did not improve the performance in comparison with the classical controller.

In order to reduce the negative influence of subjective factors on the FCs design process and to increase the efficiency of their operation it is expedient to use mathematically formulated methods and algorithms for FC synthesis [32–34] that are based on certain optimization procedures [35–37]. Thus, methods and algorithms for FCs synthesis are presented in papers [38–40], which include optimization procedures of the parameters of the linguistic terms membership functions (LTMF) of their input and output variables on the basis of the ACSs desired transients. In turn, papers [41–43] present the approach of FCs designing with the application of the structural optimization of their rule bases (RB), which allows to significantly reduce the number of rules in the FC RB without quality indicators deterioration of the ACS. These approaches and algorithms allow FCs designing for ACSs of complex objects in order to increase their energy, economic and operational indicators as well as can become the basis for creating of an effective, formalized from a mathematical point of view, method of synthesis and optimization of green fuzzy controllers for the temperature ACSs of the SPPs reactors.

The aim of this work is development and efficiency research of the integrated method of synthesis and optimization of green fuzzy controllers for the temperature ACSs of the SPPs reactors for ensuring the SPPs green operating mode, which implies high economic and operational indicators, low energy consumption, as well as high quality of obtained at the outlet alternative fuel.

3 Design Features of the Automatic Control Systems of the Temperature Modes of the SPP Reactors

The functional structure of the generalized ACS of the temperature modes of the SPP pyrolysis reactor is shown in Fig. 1 [27], where the following notations are adopted: UCL is the upper control level; SD is the setting device; TC is the temperature controller; RHS is the reactor heating system; LG is the liquefied gas; A is the air; GLFR and ALFR are the linear flow regulators of the liquefied gas and air, respectively; GB is the gas burner of the reactor RHS; TS is the temperature sensor; T_S, T_R are set and real values of the reactor heating temperature; u_{SD}, u_{TC}, u_{TS} are the output signals of SD, TC and TS, respectively; Q_{LG}, Q_A are the values of liquefied gas and air flow, respectively; P_{GB} is the heating power value of the RHS GB; F_D are the disturbances that effect on the SPP reactor.

The TC includes the Mamdani type fuzzy controller and certain auxiliary blocks. The TC structure is presented in Fig. 2, where the following notations are adopted: FB is the fuzzification block of the FC; FIB is the fuzzy inference block of the FC; DFB is the defuzzification block of the FC; MB is the multiplication block; LB is the limiting block; u_F and u_{FC} are the output signals of the function converter and fuzzy controller, respectively; K_P, K_D, K_I and K_u are the coefficients of proportionality for the input and output variables of the FC.

The TC implements automatic control of the temperature modes of the SPP reactor on the basis of feedback and feedforward. The function converter implements the dependency $u_F = f(u_{SD}) = f(T_S)$, which corresponds to the inverse static characteristic of the open-loop control system of the reactor temperature.

The signals of error of temperature control ε_T, its derivative $\frac{d\varepsilon_T}{dt}$ and/or integral $\int \varepsilon_T dt$ can be the FC input variables. The output variable of the FC, in turn, is the signal u_{FC}. In order to normalize the input signals of the FC and to convert them into relative units, according to the maximum values, the coefficients of proportionality K_P, K_D and K_I are applied. To form the control signal u_{TC} of the TC, the output signal u_{FC} of the FC is multiplied by the output signal u_F of the function converter with the help of the MB. For the implementation of its further adjustment and limiting the proportional coefficient K_u and the LB are applied, respectively.

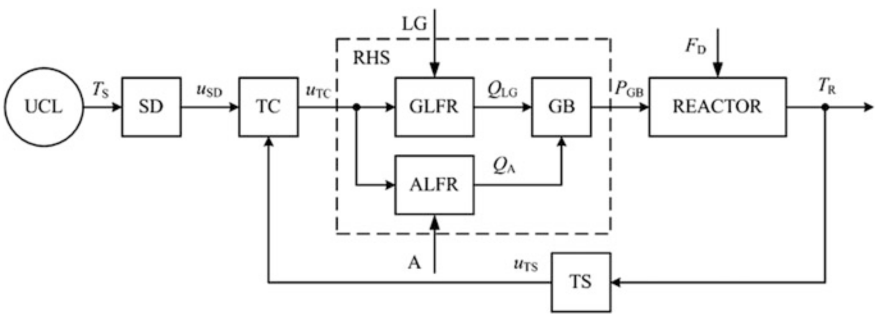


Fig. 1 Structure of the generalized ACS of the temperature modes of the SPP reactor

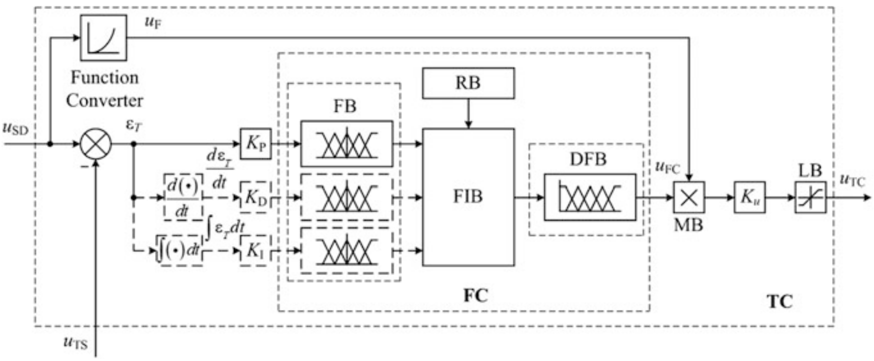


Fig. 2 Structure of the temperature controller

At designing of the green FC of the Mamdani type, which is part of the TC (Fig. 2) of the ACS of the temperature modes of the SPP reactor, it is advisable to use the integrated method of synthesis and optimization developed by the authors.

4 Integrated Method of Synthesis and Optimization of Green FC for the ACS of the Reactors Temperature of SPP

At designing the effective green FC for controlling of the complex technical objects at the initial stage, it is advisable to synthesize their structure and parameters based on expert assessments and recommendations of operators [28–31]. After that, for further improvement of the control quality, energy and economic efficiency of the ASC, it is necessary to perform consistent procedures of optimization of the FC parameters and structure on the basis of the corresponding methods of mathematical programming [44–46]. Herewith, it is advisable to carry out parametric optimization of the green FC based on the desired transients of the ACS [38, 47, 48], that provide the green operating mode of the SPP reactor, and structural optimization of the FC RB on the basis of calculating the influence of each rule on the control process [41, 49, 50]. Application of such an approach allows to provide rather high economic and operational indicators as well as low energy consumption of ACSs of the complex technical objects at a relatively simple implementation of their FC.

The proposed integrated method of synthesis and optimization of green fuzzy controllers for the temperature ACSs of the SPPs reactors takes into account the above recommendations and consists of the following steps.

Step 1. Selection of the vector of input and output variables of the green FC. The vector of input and output variables of the green FC **X** is selected at this stage, which in general can be represented as follows

$$\mathbf{X} = \{\mathbf{X}_i\}, \quad i = \{1, \ldots, n\}, \tag{1}$$

where i is the number of input or output variable of the FC; n is the total number of input and output variables of the FC.

Depending on the set quality indicators of control of the temperature modes of the SPP reactor, the vector of the input and output variables of the green FC can be represented by the expressions (2), (3) or (4)

$$\mathbf{X} = \left\{ \varepsilon_T, \frac{d\varepsilon_T}{dt}, u_{\mathrm{FC}} \right\}; \tag{2}$$

$$\mathbf{X} = \left\{ \varepsilon_T, \int \varepsilon_T dt, u_{\mathrm{FC}} \right\}; \tag{3}$$

$$\mathbf{X} = \left\{ \varepsilon_T, \frac{d\varepsilon_T}{dt}, \int \varepsilon_T dt, u_{\mathrm{FC}} \right\}. \tag{4}$$

Step 2. Preliminary determination of the vector of coefficients of proportionality for the input and output variables of the green FC. The vector of coefficients of proportionality **K** for the input and output variables of the temperature FC and its initial values are determined at this stage. In general, the vector of coefficients **K** can be represented as follows

$$\mathbf{K} = \{\mathbf{K}_i\}, \quad i = \{1, \ldots, n\}. \tag{5}$$

For $n = 3$ (depending on the vector of the FC variables **X**, selected at the *Step 1*) the vector **K** can be represented as

$$\mathbf{K} = \{K_{\mathrm{P}}, K_{\mathrm{D}}, K_u\}; \tag{6}$$

$$\mathbf{K} = \{K_{\mathrm{P}}, K_{\mathrm{I}}, K_u\}. \tag{7}$$

In turn, for $n = 4$ the vector of coefficients **K** has the form

$$\mathbf{K} = \{K_{\mathrm{P}}, K_{\mathrm{D}}, K_{\mathrm{I}}, K_u\}. \tag{8}$$

The initial values of the proportionality coefficients $\mathbf{K}_{i0}$ for the input variables of the green FC $\mathbf{X}_i$ ($i = 1,\ldots, n - 1$) are determined on the basis of their maximum allowable values $\mathbf{X}_{i\max}$

$$\mathbf{K}_{i0} = \frac{1}{\mathbf{X}_{i\max}}, \quad i = \{1, \ldots, n-1\}. \tag{9}$$

In turn, it is expedient to set the initial value of the coefficient K_u for the output variable u_{FC} equal to one ($K_{u0} = 1$) at this stage.

Step 3. Setting of the operating ranges of the input and output variables of the green FC. The operating ranges, within which the input and output variables **X** of the FC may change during its operation, are set at this stage. Since the input variables $\mathbf{X}_i$ (i = 1,…, n − 1) with the previous multiplication by the corresponding coefficients $\mathbf{K}_i$ (i = 1,…, n − 1) arrive to the FC input in relative units from the maximum value and can have both positive and negative values, then their operating ranges in the FC should be set from −1 to 1. In turn, the values of the flow of liquefied gas Q_{LG} and air Q_{A} as well as the heating power P_{GB} of the RHS gas burner can be only positive, therefore it is advisable to set the operating range of the output variable u_{FC} of the FC from 0 to 2.

Step 4. Selection of the number of linguistic terms for the input and output variables of the green FC. The values of the number of linguistic terms m_i for each i-th input and output variable of the FC (i = 1,…, n) are selected at this stage. It is expedient to select the number of linguistic terms for the input variables of the FC of the temperature ACSs of the pyrolysis reactors in the range from 3 to 7 [27]. In turn, for the FC output variable—in the range from 5 to 9.

Step 5. Selection of the LTMF types for the input and output variables of the green FC. The types of the LTMF for each i-th input and output variable of the FC are selected at this stage. The triangular, trapezoidal, Gaussian of 1st and 2nd types, π-like and others types of LTMF can be selected for the input and output variables **X** of the FC of the temperature ACS of the pyrolysis reactor [19, 26].

Step 6. Preliminary determination of the parameters vector of the LTMF for the input and output variables of the green FC. The vector **P** of the parameters (vertices) of the LTMF of the input and output variables of the FC and its initial values $\mathbf{P}_0$ are determined on the basis of knowledge of experts or operators at this stage. In turn, the vector of the LTMF parameters has the form

$$\mathbf{P} = \left\{\mathbf{P}^i_{j,k}\right\}, \quad i = \{1,\ldots,n\}, j = \{1,\ldots,m\}, k = \{1,\ldots,l\}, \tag{10}$$

where j is the term number of the i-th input or output FC variable; k is the parameter (vertex) number of the j-th term; l is the total amount of parameters (vertices) of the j-th term.

At the absence of expert knowledge the linguistic terms of the input and output variables of the green FC can be evenly distributed in their operating ranges.

Step 7. Preliminary synthesis of the rule base of the green FC. The preliminary synthesis of the RB of the green FC is carried out on the basis of the input and output variables **X**, as well as their linguistic terms selected at *Steps 1* and *4*, respectively, at this stage. The total number of rules s is determined by the number of all possible combinations of the linguistic terms of the input variables of the FC

$$s = \prod_{i=1}^{n-1} m_i. \tag{11}$$

The consequents for each r-th rule (r = 1, …, s) are selected by the expert or operator from a set of linguistic terms of the output variable u_{FC} of the FC, the number of which m_n is chosen at *Step 4*.

Step 8. Selection of computational procedures of fuzzy inference of the green FC. The selection of operations of aggregation, activation and accumulation of fuzzy inference of the green FC is carried out at this stage. As an aggregation operation, for example, operations “min” or “prod” can be selected. As an activation operation the operation “max” can be selected and as an accumulation operation—“max” or “sum” [14, 18].

Step 9. Selection of the defuzzification method of the green FC. The selection of the defuzzification method of the green FC is carried out at this stage. One of the following methods can be selected as the defuzzification method of the FC of the temperature modes ACS: the gravity center method, the square center method, the left modal value method, the right modal value method and the other [15, 26].

Step 10. Optimization of the parameters vector of the LTMF for the input and output variables of the green FC. The optimization procedures of the vector **P** of the LTMF parameters of the input and output variables of the FC are carried out at this stage on the basis of desired transients of the reactor ACS and methods of gradient descent of numerical optimization.

The maximum amount of liquefied gas and, accordingly, the SPP energy is expended by the gas burner of the RHS during the transient mode of the reactor temperature before achieving its steady mode. Thus, the desired transients of the ACS of the reactor temperature should have the minimum duration and aperiodic character, that corresponds to a high quality indicators of temperature control, low energy consumption and, consequently, green operating mode of the SPP. In turn, the desired transient characteristic $T_D(t)$ for the optimization procedures of the green FC can be obtained on the basis of the reference model (RM) of the ACS of the reactor temperature, which has the transfer function $W_{RM}(s)$ [6]

$$W_{RM}(s) = \frac{T_D(s)}{T_S(s)} = \frac{1}{(T_{RM}s + 1)^{\nu}}, \tag{12}$$

where T_{RM} and ν are the time constant and order of the transfer function of the RM.

To optimize the vector **P** of the LTMF parameters the goal function I and its optimal value I_{opt}, the initial values (initial hypothesis $\mathbf{P}_0$) and the constraints of the given vector [$\mathbf{P}_{min}$, $\mathbf{P}_{max}$] are selected. Also, the type of the procedure of the iterative optimization is chosen [35, 48].

At optimizing the vector **P** of the LTMF parameters of the green FC on the basis of desired transients of the reactor temperature ACS in many cases it is advisable to choose the mean integral quadratic deviation $I(t, \mathbf{P})$ of the real transient characteristic of the ACS $T_R(t, \mathbf{P})$ from the desired $T_D(t)$ as the goal function

$$I(t, \mathbf{P}) = \frac{1}{t_{\max}} \int_{0}^{t_{\max}} (T_{\mathrm{D}}(t) - T_{\mathrm{R}}(t, \mathbf{P}))^2 dt, \tag{13}$$

where t_{max} is the total time of the ACS transient of the reactor temperature.

The values of the time constant T_{RM} and order ν of the RM as well as the optimal value of the goal function I_{opt} are determined experimentally for the reactor of each particular SPP.

As the initial values (initial hypothesis) of the vector **P** of the LTMF parameters of the input and output variables of the FC the values of vector $\mathbf{P}_0$ are selected, which are previously determined at *Step 6*.

As the constraints for the possible values of the vector **P** of the LTMF parameters it is advisable to select the boundaries of the operating ranges of the input and output variables of the green FC, which are set at *Step 3*.

It is expedient to conduct the iterative procedures of the vector **P** of the LTMF parameters on the basis of the vector equation (14) according to the method of gradient descent

$$\mathbf{P}^{i}_{j,k}[\tau + 1] = \mathbf{P}^{i}_{j,k}[\tau] - \gamma[\tau] \frac{\partial I\left(\mathbf{P}^{i}_{j,k}\right)}{\partial \mathbf{P}^{i}_{j,k}} \Bigg|_{\mathbf{P}^{i}_{j,k}[\tau]}, \tag{14}$$

where γ is the vector of steps of the gradient descent, τ is the iteration number. In turn, the vector γ can be determined in different ways depending on the chosen method of gradient descent (with a fixed step, with a crushed step, the fastest descent, etc.).

These iterative procedures are carried out before fulfilling the condition of completion of the optimization of the LTMF parameters of the input and output variables of the green FC. The condition of completion of the optimization procedures of the LTMF parameters can be considered to be fulfilled in the following cases: (a) if the optimal value of the goal function is reached ($I = I_{opt}$); (b) if the maximum number of iterations is performed ($\tau = \tau_{max}$); (c) in the case when for a certain number of iterations τ the value of the goal function I has not decreased. If the condition (a) is fulfilled, the transition to *Step 12* is performed. In turn, if the conditions (b) or (c) are fulfilled, the transition to *Step 11* is carried out.

Step 11. Optimization of the vector of coefficients of proportionality for the input and output variables of the green FC. The optimization procedures of the vector **K** of the coefficients of proportionality for the input and output variables of the FC are carried out at this stage on the basis of desired transients of the reactor ACS and methods of gradient descent of numerical optimization. For optimization of the vector **K** of the coefficients of proportionality the goal function I and its optimal value I_{opt}, the initial values (initial hypothesis $\mathbf{K}_0$) and the constraints of the given vector [$\mathbf{K}_{min}$, $\mathbf{K}_{max}$] are selected. Also, the type of the procedure of the iterative optimization is chosen [35, 49].

In this case, it is also advisable to choose the mean integral quadratic deviation $I(t, \mathbf{K})$ of the real transient characteristic of the ACS $T_R(t, \mathbf{K})$ from the desired $T_D(t)$ as the goal function

$$I(t, \mathbf{K}) = \frac{1}{t_{\max}} \int_0^{t_{\max}} (T_D(t) - T_R(t, \mathbf{K}))^2 dt. \quad (15)$$

The desired transient characteristic $T_D(t)$ for the optimization procedures of the green FC, in this case, is also obtained on the basis of the RM with the transfer function (12).

As the initial values (initial hypothesis) of the vector $\mathbf{K}$ of the coefficients of proportionality the values of vector $\mathbf{K}_0$ are selected, which are previously determined at *Step 2*.

The constraints for the possible values of the vector $\mathbf{K}$ of the coefficients of proportionality are determined in the form

$$\mathbf{K}_i \in [\mathbf{K}_{i\min}, \mathbf{K}_{i\max}], \quad i = \{1, \ldots, n\}, \quad (16)$$

where $\mathbf{K}_{i\min}$ and $\mathbf{K}_{i\max}$ are the corresponding minimum and maximum values of the vector of coefficients of proportionality.

It is expedient to conduct the iterative procedures of the vector $\mathbf{K}$ of the coefficients of proportionality on the basis of the vector equation (17) according to the method of gradient descent

$$\mathbf{K}_i[\tau + 1] = \mathbf{K}_i[\tau] - \gamma[\tau] \frac{\partial I(\mathbf{K}_i)}{\partial \mathbf{K}_i} \Big|_{\mathbf{K}_i[\tau]}. \quad (17)$$

These iterative procedures are also carried out before fulfilling the condition of completion of the optimization, presented at *Step 10*. If the condition (a) is fulfilled, the transition to *Step 12* is performed. In turn, if the conditions (b) or (c) are fulfilled, the transition to *Steps 10*, *9*, *8*, *7*, *5*, *4* or *1* can be carried out.

Step 12. Structural optimization of the RB of the green FC. The structural optimization of the RB of the green FC is carried out at this stage on the basis of the evaluation of the influence of each r-th rule ($r = 1,\ldots, s$) on the control process and further exclusion of those rules whose influence is negligible. The procedure of structural optimization of the FC RB consists of identifying the influence degree of the RB rules on the output signal u_{FC} of the FC, forming of the corresponding ranking series of rules $\mathbf{R}$ by the reduction of this parameter, as well as the exclusion of those rules from the RB, whose influence on the formation of the signal u_{FC} is insignificant [49, 50].

The identification of the influence degree of the RB rules on the output signal u_{FC} of the FC should be based on the calculation of changing of the truth degree of rules $\mu^R(t)$ in the control process and on the evaluation functional $G[\mu^R(t)]$, which argument is $\mu^R(t)$[49].

The changing of the truth degree $\mu^R(t)$ of the r-th rule in the control process is presented by the function from time t

$$\mu_r^R(t) = \bigcap_{i=1}^{n-1} \mu_i^r(X_i(t)) = \inf_{i=1}^{n-1} \mu_i^r(X_i(t)), \tag{18}$$

where μ_i^r is the fuzzification result of the i-th input variable of the FC X_i (i = 1, …, $n - 1$) by the corresponding linguistic term of the r-th rule.

The evaluation functional $G[\mu^R(t)]$ of the influence of the r-th rule on the control process, in turn, is calculated according to the equation [49]

$$G_r\left[\mu_r^R(t)\right] = \frac{1}{t_{\max}} \int_0^{t_{\max}} \mu_r^R(t)dt = \frac{1}{t_{\max}} \int_0^{t_{\max}} \inf_{i=1}^{n-1} \mu_i^r(X_i(t))dt. \tag{19}$$

At the same time, (for an adequate evaluation of the influence of the RB rules of the FC on the control process) it is expedient to carry out the simulation of the temperature ACS of the pyrolysis reactor at calculating of changing of the truth degree $\mu^R(t)$ and evaluation functional $G[\mu^R(t)]$ of the RB rules on the time interval from 0 to $t_{\max}$ in all possible functioning modes (with different types of stepped and smoothly increasing input and disturbing influences).

The ranking series of rules **R** consists of all the rules of the FC RB, located in order of decreasing of their values of the calculated evaluation functional $G[\mu^R(t)]$.

The exclusion of the RB rules, whose influence on the formation of the signal u_{FC} is insignificant, should be carried out consistently one by one rule (beginning from the end of the formed ranking series **R**) to a certain optimal value of their number s_{opt}, at which the value of the goal function I does not exceed the optimal value I_{opt}, and the quality indicators of the ACS of the reactor temperature remain in the acceptable limits [49, 50].

Step 13. Implementation of the developed green FC using the appropriate element base. The structure of the green FC will be simplified by the reduction the RB conducted at *Step 12*.

Step 14. Usage and experimental research of the developed green FC in the ACS of the reactor temperature of the SPP. The using and researching of the developed and optimized green FC in the ACS of the reactor temperature of the SPP is performed at this stage.

The effectiveness study of the given above integrated method of synthesis and optimization of green fuzzy controllers is conducted in the design process of the Mamdani type green FC for the temperature ACS of the pyrolysis reactor of the experimental SPP.

5 Synthesis and Optimization of the Green FC for the ACS of the Reactor Temperature of the Experimental SPP

The operating volume of the pyrolysis reactor and the maximum power of the gas burner of this experimental SPP are 100 L and 25 kW, respectively [5]. The temperature ACS of the reactor of this SPP and its TC have the functional structures presented in Figs. 1 and 2, respectively. The mathematical model of this pyrolysis reactor as a temperature control object is developed on the basis of the equations of heat energy processes and given in [4].

The vector **X** is selected as a vector of input and output variables of this green FC at *Step 1* of the proposed method, which is represented by the expression (2). The vector of coefficients of proportionality **K** for the input and output variables **X**, presented by the expression (6), and its initial values $\mathbf{K}_0$ are determined at *Step 2*. In turn, the initial values of the coefficients K_{P0} and K_{D0} for the input variables of the FC ε_T and $\frac{d\varepsilon_T}{dt}$ are calculated by the formula (9) and are equal to 0,1 and 0,2, respectively (K_{P0} = 0,1; K_{D0} = 0,2). The initial value of the coefficient K_{u0} for the output variable u_{FC} is set equal to one (K_{u0} = 1).

The operating ranges of the input and output variables **X** of the green FC are set at *Step 3* of the developed method. For the input variables ε_T and $\frac{d\varepsilon_T}{dt}$ the operating ranges are set from −1 to 1, for the output variable u_{FC}—from 0 to 2.

5 linguistic terms ($m_i = 5$, $i = 1, 2, 3$) are selected for each input and output variables of the given FC at *Step 4*. For the input variables ε_T and $\frac{d\varepsilon_T}{dt}$ the following linguistic terms are selected: BN—big negative; SN—small negative; Z—zero; SP—small positive; BP—big positive. In turn, for the output variable u_{FC}, the given linguistic terms are selected: Z—zero; S—small; M—middle; B—big; VB—very big. For all of the above terms the triangular type of membership functions is selected at *Step 5*.

The vector **P** of the parameters (vertices) of the LTMF of the input and output variables of the FC, that is presented in general form by the expression (10), and its initial values $\mathbf{P}_0$ are determined at *Step 6*. In this case, the expression (10) takes the form

$$\mathbf{P} = \left\{\mathbf{P}^i_{j,k}\right\}, \quad i = \{1,2,3\}, j = \{1,\ldots,5\}, k = \{1,2,3\}.$$

In turn, the initial values of the vector of the LTMF parameters $\mathbf{P}_0$ are defined in such a way that the linguistic terms of the input and output variables **X** are evenly distributed in their operating ranges. The appearance of the LTMF of the input and output variables of the green FC with the parameters specified at this stage is shown in Fig. 3.

The preliminary synthesis of RB of the green FC is carried out at *Step 7*, on the basis of the input and output variables **X**, as well as their linguistic terms selected at *Step 1* and *Step 4*, respectively. The total number of rules s, according to the

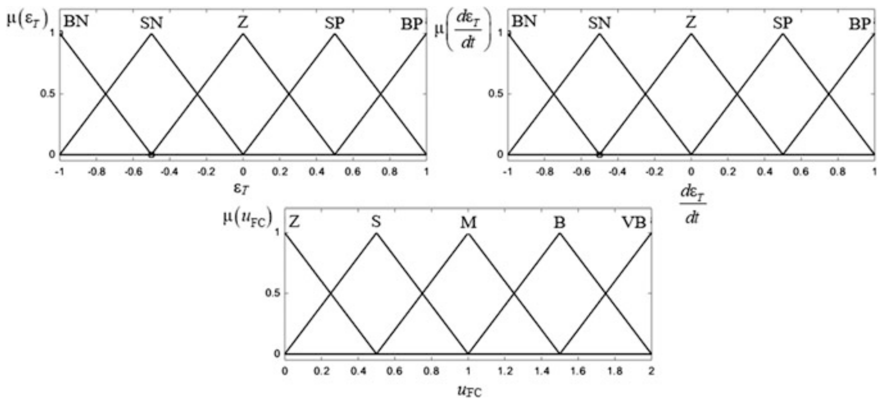


Fig. 3 Green FC linguistic terms and their parameters

Eq. (11), in this case is 25. Each rule of the synthesized RB is the linguistic statement, represented by the expression (20)

$$\text{IF ``}\varepsilon_T = x\text{'' AND ``}\frac{d\varepsilon_T}{dt} = y\text{'' THEN ``}u_{\mathrm{FC}} = z\text{'',} \tag{20}$$

where x, y, z are the corresponding values of the linguistic terms.

The RB of the green FC, synthesized at this stage, is presented in Table 1.

The computational procedures of the fuzzy inference of this green FC are selected at *Step 8*. In this case, the operation “min” is selected as an aggregation operation, the operation “max” is selected for both activation and accumulation operations. In turn, the gravity center method is chosen as the defuzzification method of the given green FC at *Step 9*.

The computer simulation of transients and comparative analysis of the quality indicators of the ACS of the reactor temperature with the developed green FC and optimally tuned conventional PID-controller are carried out to verify the efficiency of the synthesized at this stage green FC.

For providing the green operating mode of the given experimental SPP the transient of the pyrolysis reactor heating should have a duration of about 400 s and aperiodic character. The given desired transient characteristic can be obtained on

Table 1 Rule base of the green FC

		Rate of error change, $d\varepsilon_T/dt$				
		BN	SN	Z	SP	BP
Error, ε_T	BN	Z	Z	Z	Z	S
	SN	Z	Z	Z	S	M
	Z	Z	S	M	B	VB
	SP	M	B	VB	VB	VB
	BP	B	VB	VB	VB	VB

the basis of the reference model of the ACS of the reactor temperature with the transfer function (12), parameters of which are defined and have the following values: T_{RM} = 64 s, ν = 3. Also, the optimal value of the goal function I_{opt} is determined to be equal 500 (I_{opt} = 500) for the given experimental SPP. In this case, if the value of the goal function I does not exceed 500, the difference between the transients of the RM and real ACS of the reactor temperature remains insignificant, and the given experimental SPP remains in its green operating mode.

The transfer function $W_{PID}(s)$ of the conventional PID-controller is represented by the expression (21)

$$W_{PID}(s) = K_{P1} + \frac{K_{I1}}{s} + \frac{K_{D1}s}{T_{F}s + 1}, \quad (21)$$

where K_{P1}, K_{I1}, K_{D1} and T_F are the coefficients and time constant of the filter of the conventional PID-controller, which are found in the process of parametric optimization based on the desired transients using the gradient method of the fastest descent and the goal function, represented by the expression (15). In turn, K_{P1} = 2.189; K_{I1} = 0.0094, K_{D1} = 127.1; T_F = 1.035. In addition, to eliminate the effect of integral windup, the function of the algorithmic prohibition of integration is applied in this PID controller, when the control signal reaches the maximum possible value.

The quality indicators comparative analysis of the temperature ACS of the SPP reactor with the developed green FC and optimally tuned conventional PID-controller are presented in Table 2, where σ is an overshoot, $\sigma = \frac{T_{R\max} - T_R}{T_R} \cdot 100\%$; t_t is the transient time; Δ_T is the static error, $\Delta_T = \frac{T_S - T_R}{T_S} \cdot 100\%$; ξ is the oscillation (a number of steady value T_R crossing for the time t_t), W_{GB} is the heating energy, spent by the gas burner of the RHS during the transient mode of the reactor temperature before achieving its steady mode.

The transients of the temperature ACS of the SPP reactor are presented in Fig. 4.

As can be seen from Table 2 and Fig. 4 the ACS with non-optimized green FC, synthesized at this stage, has worse quality indicators than RM and ACS with optimally tuned conventional PID-controller. In turn, the values of the goal function

Table 2 Quality indicators comparative analysis of the temperature ACS of the SPP reactor

Quality indicators	ACS quality indicators values		
	RM	Optimally tuned conventional PID-controller	Non-optimized green FC
σ (%)	0	11	25.3
t_t (s)	400	655	760
Δ_T (%)	0	0	0
ξ	0	1	1
I	0	1582	6861
W_{GB} (MJ)	–	11.86	14.24

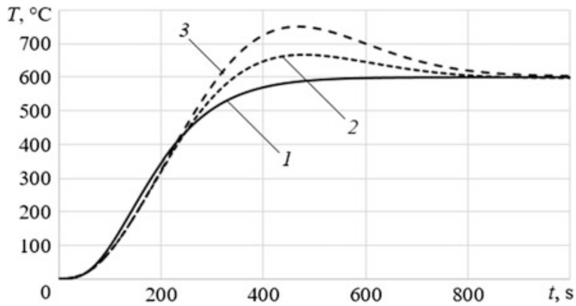


Fig. 4 ACS transients: *1* RM; *2* with optimally tuned conventional PID-controller; *3* with non-optimized green FC

I for both ACSs with non-optimized green FC and optimally tuned conventional PID-controller are much larger than the given optimal value I_{opt}. Also, the value of the heating energy W_{GB}, spent by the gas burner during the transient mode of the reactor heating is bigger for the ACS with non-optimized green FC than for the ACS with optimally tuned conventional PID-controller. Thus, the temperature ACS of the SPP reactor with developed at this stage FC does not provide the green operating mode of the given experimental SPP.

Therefore, the optimization procedures of the vector **P** of the LTMF parameters of the input and output variables of the FC are carried out at *Step 10* in order to increase the quality indicators as well as to reduce the energy consumption of the temperature ACS of the SPP reactor. Equation (13) is chosen as the goal function in this case. The values of vector $\mathbf{P}_0$, which are previously determined at *Step 6*, are selected as the initial values (initial hypothesis) of the vector **P** of the LTMF parameters of the input and output variables of the FC. In turn, the boundaries of the operating ranges of the input and output variables of the green FC, which are set at *Step 3*, are selected as the constraints for the possible values of the vector **P** of the LTMF parameters. The iterative procedures of optimization of the vector **P** of the LTMF parameters are carried out on the basis of the vector equation (14) according to the gradient method of the fastest descent.

The appearance of the LTMF of the input and output variables of the green FC with the parameters optimized at *Step 10* is shown in Fig. 5.

Transients and comparative analysis of the quality indicators of the temperature ACS with the developed green FC with optimized LTMF parameters and optimally tuned conventional PID-controller are presented in Fig. 6 and Table 3, respectively.

The value of the goal function is reduced to I = 1166 and the quality indicators of the temperature ACS of the SPP reactor are somewhat improved by means of the optimization of the LTMF parameters, carried out at this step. As can be seen from Table 3 and Fig. 6 the ACS with green FC with LTMF parameters, optimized at this stage, has better quality indicators than ACS with optimally tuned conventional PID-controller.

Also, the value of the heating energy W_{GB}, spent by the gas burner during the transient mode of the reactor heating is lower for the ACS with green FC with optimized LTMF parameters than for the ACS with optimally tuned conventional

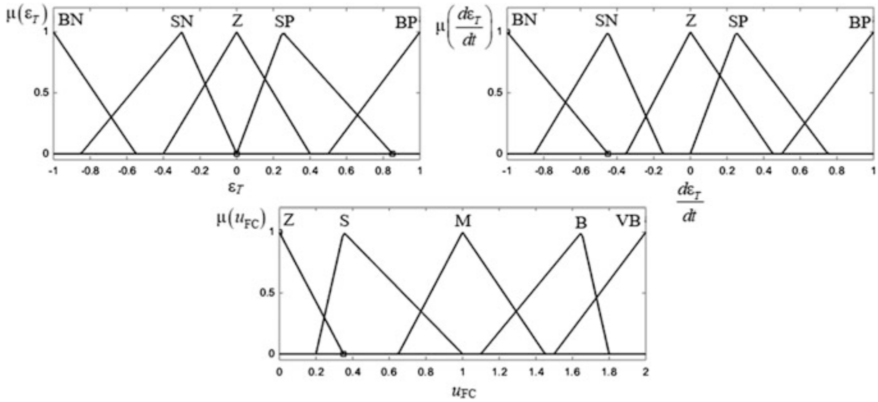


Fig. 5 Green FC linguistic terms and their optimized parameters

Fig. 6 ACS transients: *1* RM; *2* with optimally tuned conventional PID-controller; *3* with green FC with optimized LTMF parameters

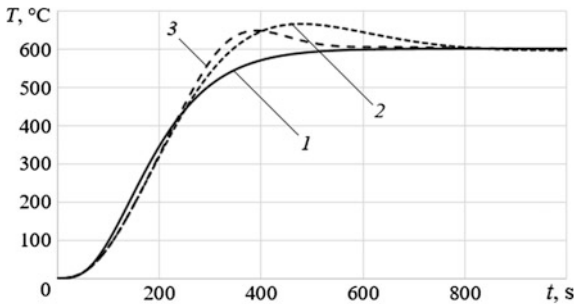


Table 3 Quality indicators comparative analysis of the temperature ACS with green FC with optimized LTMF parameters and optimally tuned conventional PID-controller

Quality indicators	ACS quality indicators values		
	RM	Optimally tuned conventional PID-controller	Green FC with optimized LTMF parameters
σ (%)	0	11	8.16
t_t (s)	400	655	459
Δ_T (%)	0	0	0
ξ	0	1	1
I	0	1582	1166
W_{GB} (MJ)	–	11.86	8.65

PID-controller. However, the value of its goal function I is still greater than the given optimal value I_{opt}. Thus, the temperature ACS of the SPP reactor with developed and optimized at this stage FC also does not provide the green operating mode of the given experimental SPP.

Therefore, the optimization procedures of the vector **K** of the coefficients of proportionality for the input and output variables of the developed FC are carried out at *Step 11* for further increasing the quality indicators as well as reducing the energy consumption of the temperature ACS of the SPP reactor. Equation (15) is chosen as the goal function in this case. The values of vector $\mathbf{K}_0$, which are previously determined at *Step 2*, are selected as the initial values (initial hypothesis) of the vector **K** of the coefficients of proportionality. In turn, the given boundaries are selected as the constraints for the possible values of the vector **K**

$$K_P \in [0.05; 0.2];$$
$$K_D \in [0.1; 0.35];$$
$$K_u \in [0.5; 2].$$

The iterative procedures of optimization of the vector **K** of the coefficients of proportionality for the input and output variables of the developed FC are carried out on the basis of the vector equation (17) according to the gradient method of the fastest descent.

The optimized values of the vector **K** of the coefficients of proportionality are: $K_P = 0.142$; $K_D = 0.285$; $K_u = 1.127$.

In turn, transients and comparative analysis of the quality indicators of the temperature ACS with the developed green FC with optimized LTMF parameters and coefficients **K** as well as with optimally tuned conventional PID-controller are presented in Fig. 7 and Table 4, respectively.

The value of the goal function is reduced to $I = 236$, that is less than the given optimal value I_{opt}, and the quality indicators of the temperature ACS of the SPP reactor are considerably improved by means of the optimization of the vector **K** of the coefficients of proportionality, carried out at this stage. As can be seen from Table 4 and Fig. 7 the ACS with green FC with coefficients of proportionality, optimized at this stage, has high enough quality indicators and the value of its goal function lies within the optimum limits. Also, the value of the heating energy W_{GB}, spent by the gas burner during the transient mode of the reactor heating is much lower for the ACS with green FC with optimized LTMF parameters and coefficients of proportionality than for the ACS with optimally tuned conventional

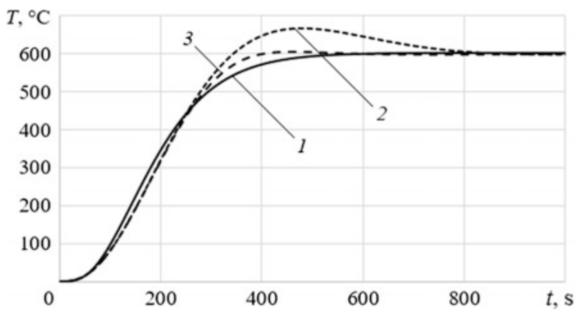


Fig. 7 ACS transients: *1* RM; *2* with optimally tuned conventional PID-controller; *3* with green FC with optimized LTMF parameters and coefficients **K**

Table 4 Quality indicators comparative analysis of the temperature ACS with green FC with optimized LTMF parameters and coefficients **K** as well as with conventional PID-controller

Quality indicators	ACS quality indicators values		
	RM	Optimally tuned conventional PID-controller	Green FC with optimized LTMF parameters and coefficients K
σ (%)	0	11	0.83
t_t (s)	400	655	343
Δ_T (%)	0	0	0
ξ	0	1	0
I	0	1582	236
W_{GB} (MJ)	–	11.86	6.78

PID-controller. Thus, the temperature ACS of the SPP reactor with developed and optimized at this stage green FC provides the green operating mode of the given experimental SPP.

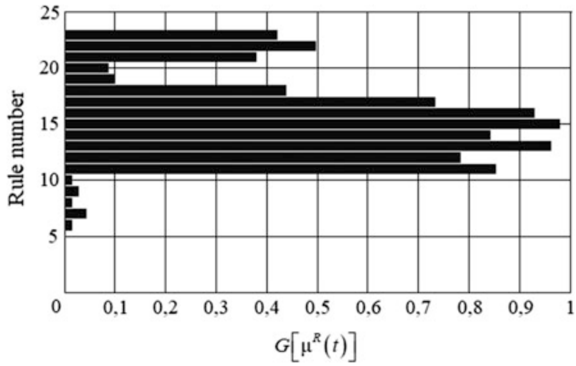
After that, the transition to *Step 12* is carried out for further structural optimization and reducing the RB of the developed green FC for simplifying of its hardware and software implementation.

The calculating of changing of the truth degree $\mu^R(t)$ and evaluation functional $G[\mu^R(t)]$ of the influence of the RB rules of the green FC on the control process is carried out at *Step 12* by means of Eqs. (18) and (19) at the simulation of the temperature ACS of the pyrolysis reactor in all possible functioning modes (with different types of stepped and smoothly increasing input and disturbing influences). The diagram of the evaluation functional $G[\mu^R(t)]$ of the influence of all the rules of the green FC RB on the control process is presented in Fig. 8.

The ranking series of rules **R**, that include all the rules of the green FC RB located in order of decreasing of their values of the calculated evaluation functional $G[\mu^R(t)]$, has the form

$$\mathbf{R} = \{15, 13, 16, 11, 14, 12, 17, 22, 18, 23, 21, 19, 20, 7, 9, 6, 8, 10, 1, 2, 3, 4, 5, 24, 25\}.$$

Fig. 8 Diagram of the evaluation functional of the influence of RB rules of the green FC on the control process



The exclusion of the RB rules, whose influence on the formation of the signal u_{FC} is insignificant, is carried out consistently one by one rule (beginning from the end of the formed ranking series **R**) to a certain optimal value of their number s_{opt}, at which the value of the goal function I does not exceed the optimal value I_{opt}, and the quality indicators of the ACS of the reactor temperature remain in the acceptable limits.

The graph of changing of the goal function value from the total number of the RB rules of the developed green FC is shown in Fig. 9.

As can be seen from Fig. 9 the optimal value of the rules number s_{opt} = 14.

The transients and comparative analysis of the quality indicators of the temperature ACS with the developed and optimized green FC with full RB as well as with reduced (optimized) RB are presented in Fig. 10 and Table 5, respectively.

As can be seen from Table 5 and Fig. 10 the ACS with optimized green FC with reduced at this stage RB has high enough quality indicators and the value of its goal function lies within the optimum limits. Also, the value of the heating energy W_{GB}, spent by the gas burner during the transient mode of the reactor heating is not much more than for the ACS with optimized green FC with full RB. Thus, the

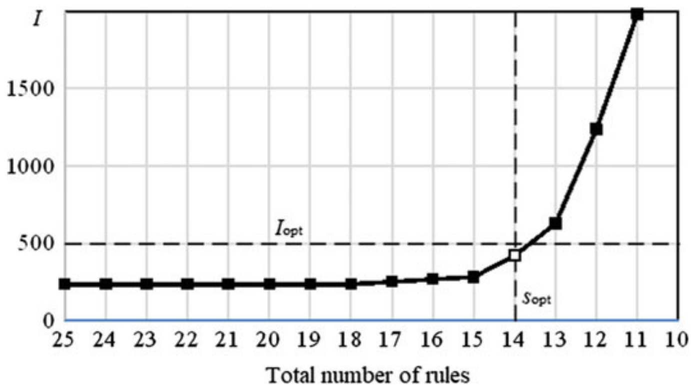


Fig. 9 Graph of changing of the goal function value from the total number of the RB rules of the developed green FC

Fig. 10 ACS transients: *1* RM; *2* with optimized green FC with full RB; *3* with optimized green FC with reduced RB

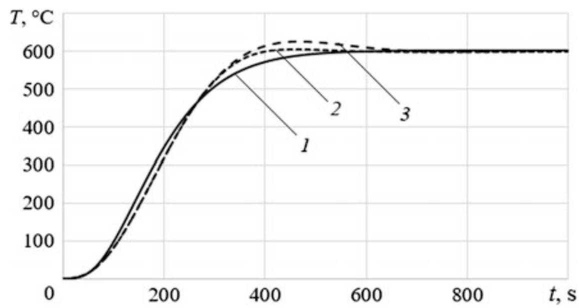


Table 5 Quality indicators comparative analysis of the temperature ACS with optimized green FC with full and reduced RB

Quality indicators	ACS quality indicators values		
	RM	Optimized green FC with full RB	Optimized green FC with reduced RB
σ (%)	0	0.83	4
t_t (s)	400	343	349
Δ_T (%)	0	0	0
ξ	0	0	0
I	0	236	422
W_{GB} (MJ)	–	6.78	6.88

temperature ACS of the SPP reactor with developed and optimized at this stage green FC provides the green operating mode of the given experimental SPP.

Also, the software and hardware implementation of the developed and optimized green FC at *Step 13* will be simplified by the reduction of the RB, conducted at *Step 12*, and it can be used for further experimental research in the ACS of the reactor temperature of the given experimental SPP at *Step 14*.

The application of the optimization procedures of the LTMF parameters **P** and the proportional coefficients **K** of the input and output variables of the green FC at *Step 10* and *Step 11* allowed to significantly increase the quality indicators of the temperature ACS of the reactor and to reduce the heating energy of the experimental SPP as well as to provide its green operating mode. In turn, further structural optimization of the RB of the given green FC at *Step 12* allowed to reduce the total number of the RB rules up to 14 rules without degrading the quality indicators of the temperature ACS and significant increasing of the heating energy consumption of the experimental SPP. This gives the opportunity to significantly simplify the software and hardware implementation of the synthesized green FC in the temperature ACS of the reactor of this experimental SPP.

The above studies confirm the high efficiency of the proposed integrated method of synthesis and optimization of the green fuzzy controllers for the ACSs of the reactors temperature of the SPPs. The application of the developed on the basis of this method green FCs will provide high quality indicators of control of the reactors temperature, low energy consumption and, consequently, green operating mode of the SPPs at their relatively simple software and hardware implementation.

6 Conclusions

The developed by the authors integrated method of synthesis and optimization of green FCs for the ACSs of the reactors temperature of the SPPs is presented in this work.

The proposed integrated method allows to carry out the preliminary synthesis of the structure and parameters of the green FCs of Mamdani type for the temperature ACSs of the pyrolysis reactors of the SPPs on the basis of expert assessments as well as further sequential procedures of optimization of the parameters and RB structure of the given FCs on the basis of mathematical programming methods. In particular, the parametric optimization of the green FCs is performed on the basis of the desired transients and methods of gradient descent of numerical optimization, that allows to provide high quality indicators of control of the temperature modes of the pyrolysis reactors at a sufficiently low energy consumption, that corresponds to the SPPs green operating mode. In turn, the structural optimization of the RBs of the given FCs is carried out on the basis of the evaluation of the influence of each rule on the control process and allows to reduce the total number of rules of the RBs without deterioration of the quality indicators of temperature control and significant increasing of the heating energy consumption of the SPPs. This allows to significantly simplify the software and hardware implementation of the synthesized and optimized green FCs in the temperature ACSs of the reactors of the SPPs.

For studying the effectiveness of the presented method the design of the Mamdani type green FC for the temperature ACS of the pyrolysis reactor of the experimental SPP is carried out in this work. The developed green FC has a relatively simple hardware and software implementation as well as allows to achieve high quality indicators of temperature modes control, low energy consumption and, consequently, green operating mode of the given experimental SPP, that confirms the high efficiency of the proposed method.

Thus, the application of the green FCs, developed on the basis of the given method, in the temperature ACSs of the pyrolysis reactors of the SPPs of different types and capacities allows to provide their green operating mode, which implies high economic and operational indicators, low energy consumption, as well as high quality of obtained at the outlet alternative fuel. Also, such green controllers have a relatively simple software and hardware implementation.

References

1. Ryzhkov, S.S., Markina, L.M.: Experimental researches of organic waste recycling method of multiloop circulating pyrolysis. J. Collected Works NUS **5**, 100–106 (2007). (in Russian)
2. Markina, L.M.: Development of new energy-saving and environmental safety technology at the organic waste disposal by ecopyrogenesis. J. Collected Works NUS **4**, 8 (2011). (in Ukrainian)
3. Kondratenko, Y.P., Korobko, O.V., Kozlov, O.V.: PLC-based systems for data acquisition and supervisory control of environment-friendly energy-saving technologies. In: Kharchenko, V., Kondratenko, Y., Kacprzyk J. (eds.) Green IT Engineering: Concepts, Models, Complex Systems Architectures, Studies in Systems, Decision and Control, vol. 74, pp. 247–267. Springer International Publishing, Berlin, Heidelberg (2017). https://doi.org/10.1007/978-3-319-44162-7_13

4. Kondratenko, Y.P., Kozlov, O.V.: Mathematic modeling of reactor's temperature mode of multiloop pyrolysis plant. In: Modeling and Simulation in Engineering, Economics and Management. Lecture Notes in Business Information Processing, vol. 115, pp. 178–187 (2012). https://doi.org/10.1007/978-3-642-30433-0_18
5. Kondratenko, Y.P., Kozlov, O.V., Kondratenko, G.V., Atamanyuk, I.P.: Mathematical model and parametrical identification of ecopyrogenesis plant based on soft computing techniques. In: Berger-Vachon, C., Lafuente, A.M.G., Kacprzyk, J., Kondratenko, Y., Merigó, J.M., Morabito, C.F. (eds.) Complex Systems: Solutions and Challenges in Economics, Management and Engineering. Studies in Systems, Decision and Control, vol. 125, pp. 201–233. Springer International Publishing, Berlin, Heidelberg (2018). https://doi.org/10.1007/978-3-319-69989-9_13
6. Kondratenko, Y.P., Kozlov, O.V.: Mathematical model of ecopyrogenesis reactor with fuzzy parametrical identification. In: Zadeh, L.A., et al. (eds.) Recent Developments and New Direction in Soft-Computing Foundations and Applications. Studies in Fuzziness and Soft Computing, vol. 342, pp. 439–451. Springer-Verlag, Berlin, Heidelberg (2016). https://doi.org/10.1007/978-3-319-32229-2_30
7. Kharchenko, V., et al. (eds.).: Green IT engineering: concepts, models, complex systems architectures. In: Decision and Control, vol. 74. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-44162-7
8. Kharchenko, V., et al. (eds.).: Green IT engineering: components, networks and systems implementation. In: Studies in Systems, Decision and Control, vol. 105. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-55595-9
9. Drozd, J., Drozd, A., Antoshchuk, S.: Green IT engineering in the view of resource-based approach. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Concepts, Models, Complex Systems Architectures, Studies in Systems, Decision and Control, vol. 74, pp. 43–65. Springer International Publishing, Berlin, Heidelberg (2017). https://doi.org/10.1007/978-3-319-44162-7_3
10. Palagin, A.V., Opanasenko, V.N.: Design and application of the PLD-based reconfigurable devices. In: Adamski, M., Barkalov, A., Wegrzyn, M. (eds.) Design of Digital Systems and Devices. Lecture Notes in Electrical Engineering, vol. 79, pp. 59–91. Verlag, Springer, Berlin, Heidelberg (2011)
11. Kharchenko, V., Illiashenko, O.: Concepts of green IT engineering: taxonomy, principles and implementation. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Concepts, Models, Complex Systems Architectures. Studies in Systems, Decision and Control, vol. 74, pp. 3–19. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-44162-7_1
12. Kuchuk, G., Kovalenko, A., Kharchenko, V., Shamraev, A.: Resource-oriented approaches to implementation of traffic control technologies in safety-critical I&C systems. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Components, Networks and Systems Implementation, Studies in Systems, Decision and Control, vol. 105, pp. 313–337. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-55595-9_15
13. Kondratenko, Y., Gordienko, E.: Implementation of the neural networks for adaptive control system on FPGA. In: Katalinic, B. (ed.) Annals of DAAAM for 2012, Proceeding of the 23th International DAAAM Symposium on "Intelligent Manufacturing and Automation", vol. 23, no. 1, pp. 0389–0392. DAAAM International, Vienna, Austria, EU (2012)
14. Zadeh, L.A., Abbasov, A.M., Yager, R.R., Shahbazova, S.N., Reformat, M.Z. (eds.): Recent Developments and New Directions in Soft Computing. In: STUDFUZ, vol. 317. Springer, Cham (2014)
15. Jamshidi, M., Kreinovich, V., Kacprzyk, J. (eds.): Advance Trends in Soft Computing. Springer-Verlag, Cham (2013)
16. Jang, J.-S.R., Sun, C.-T., Mizutani, E.: Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence. Prentice Hall (1996)
17. Zadeh, L.A.: Fuzzy sets. Inf. Control **8**, 338–353 (1965)

18. Zadeh, L.A.: The role of fuzzy logic in modeling, identification and control. Model. Ident. Control **15**(3), 191–203 (1994)
19. Piegat, A.: Fuzzy Modeling and Control, vol. 69. Physica (2013)
20. He, X., He, S.: Research of energy consumption pattern classification based on fuzzy logic and RBF networks in hydraulic systems. Sens. Transducers **205**(10), 58–62 (2016)
21. Xiao, Z., Guo, J., Zeng, H., Zhou, P., Wang, S.: Application of fuzzy neural network controller in hydropower generator unit. J. Kybernetes **38**(10), 1709–1717 (2009)
22. Kondratenko, Y., Gerasin, O., Topalov, A.: A simulation model for robot's slip displacement sensors. Int. J. Comput. **15**(4), 224–236 (2016). Retrieved from http://computingonline.net/computing/article/view/854
23. Pasieka, M., Grzesik, N., Kuźma, K.: Simulation modeling of fuzzy logic controller for aircraft engines. Int. J. Comput. **16**(1), 27–33 (2017). Retrieved from http://computingonline.net/computing/article/view/868
24. Gomolka, Z., Dudek-Dyduch, E., Kondratenko, Y.P.: From homogeneous network to neural nets with fractional derivative mechanism. In: Rutkowski, L., Korytkowski, M., Scherer, R., Tadeusiewicz, R., Zadeh, L.A., Zurada, J.M. (eds.) Artificial Intelligence and Soft Computing. Lecture Notes in Artificial Intelligence 10245, 16th International Conference ICAISC 2017, Zakopane, Poland, 11–15 June 2017. Proceedings, Part 1, pp. 52–63. Springer International Publishing, Cham (2017). https://doi.org/10.1007/978-3-319-59063-9_5
25. Tanaka, K., Wang, H.O.: Fuzzy Control Systems Design and Analysis: A Linear Matrix Inequality Approach. Wiley, New York, USA (2001)
26. Hampel, R., Wagenknecht, M., Chaker N.: Fuzzy Control: Theory and Practice, p. 410. Physika-Verlag, Heidelberg, New York (2000)
27. Kondratenko, Y.P., Kozlov, O.V., Kondratenko, G.V., Atamanyuk, I.P.: Mathematical model and parametrical identification of ecopyrogenesis plant based on soft computing techniques. In: Berger-Vachon, C. et al. (eds.) Complex Systems: Solutions and Challenges in Economics, Management and Engineering, Studies in Systems, Decision and Control, vol. 125, pp. 201–233. Springer, Berlin, Heidelberg (2018). https://doi.org/10.1007/978-3-319-69989-9_13
28. Driankov, D., Hellendoorn, H., Reinfrank, M.: An Introduction to Fuzzy Control. Springer Science & Business Media (2013)
29. Merigo, J.M., Gil-Lafuente, A.M., Yager, R.R.: An overview of fuzzy research with bibliometric indicators. Appl. Soft Comput. **27**, 420–433 (2015)
30. Rotshtein, A.P., Rakytyanska, H.B.: Fuzzy Evidence in Identification, Forecasting and Diagnosis, vol. 275. Springer, Heidelberg (2012)
31. Von Altrock, C.: Applying fuzzy logic to business and finance. Optimus **2**, 38–39 (2002)
32. Suna, Q., Li, R., Zhang, P.: Stable and optimal adaptive fuzzy control of complex systems using fuzzy dynamic model. J. Fuzzy Sets Syst. **133**, 1–17 (2003)
33. Oh, S.K., Pedrycz, W.: The design of hybrid fuzzy controllers based on genetic algorithms and estimation techniques. J. Kybernetes **31**(6), 909–917 (2002)
34. Lodwick, W.A., Kacprzhyk, J. (eds.).: Fuzzy optimization. In: STUDFUZ, vol. 254. Springer-Verlag, Berlin, Heidelberg (2010)
35. Kondratenko, Y.P., Simon, D.: Structural and parametric optimization of fuzzy control and decision making systems. In: Zadeh, L. et al. (eds.) Recent Developments and the New Direction in Soft-Computing Foundations and Applications. Studies in Fuzziness and Soft Computing, vol. 361, pp. 273–289 (2018). Springer, Cham. https://doi.org/10.1007/978-3-319-75408-6_22
36. Jayaram, B.: Rule reduction for efficient inferencing in similarity based reasoning. Int. J. Approximate Reasoning **48**(1), 156–173 (2008)
37. Yam, Y., Baranyi, P., Yang, C.-T.: Reduction of fuzzy rule base via singular value decomposition. IEEE Trans. Fuzzy Syst. **7**(2), 120–132 (1999)
38. Kondratenko, Y.P., Al Zubi, E.Y.M.: The optimization approach for increasing efficiency of digital fuzzy controllers. In: Annals of DAAAM for 2009, Proceeding of the 20th

International DAAAM Symposium “Intelligent Manufacturing and Automation”, pp. 1589–1591. DAAAM International, Vienna, Austria (2009)
39. Simon, D.: H∞ estimation for fuzzy membership function optimization. Int. J. Approximate Reasoning **40**, 224–242 (2005)
40. Simon, D.: Evolutionary Optimization Algorithms: Biologically Inspired and Population-Based Approaches to Computer Intelligence. Wiley (2013)
41. Kondratenko, Y.P., Klymenko, L.P., Al Zu’bi, E.Y.M.: Structural optimization of fuzzy systems’ rules base and aggregation models. Kybernetes **42**(5), 831–843 (2013). http://dx.doi.org/10.1108/K-03-2013-0053
42. Simon, D.: Design and rule base reduction of a fuzzy filter for the estimation of motor currents. Int. J. Approximate Reasoning **25**, 145–167 (2000)
43. Ishibuchi, H., Yamamoto, T.: Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining. Fuzzy Sets Syst. **141**(1), 59–88 (2004)
44. Koczy, L.T., Hirota, K.: Size reduction by interpolation in fuzzy rule bases. IEEE Trans. Syst. Man Cybern. B Cybern. **27**(1), 14–25 (1997)
45. Alcalá, R., Alcalá-Fdez, J., Gacto, M.J., Herrera, F.: Rule base reduction and genetic tuning of fuzzy systems based on the linguistic 3-tuples representation. Soft. Comput. **11**(5), 401–419 (2007)
46. Pedrycz, W., Li, K., Reformat, M.: Evolutionary reduction of fuzzy rule-based models. In: Fifty Years of Fuzzy Logic and Its Applications, STUDFUZ, vol. 326, pp. 459–481. Springer, Cham (2015)
47. Simon, D.: Sum normal optimization of fuzzy membership functions. Int. J. Uncertainty Fuzziness Knowl. Based Syst. **10**, 363–384 (2002)
48. Simon, D.: Training fuzzy systems with the extended Kalman filter. Fuzzy Sets Syst. **132**, 189–199 (2002)
49. Kondratenko, Y., Korobko, V., Korobko, O., Kondratenko, G., Kozlov, O.: Green-IT approach to design and optimization of thermoacoustic waste heat utilization plant based on soft computing. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Components, Networks and Systems Implementation. Studies in Systems, Decision and Control, vol. 105, pp. 287–311. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-55595-9_14
50. Kondratenko, Y.P., Kozlov, O.V., Korobko, O.V., Topalov, A.M.: Synthesis and optimization of fuzzy control system for floating Dock’s docking operations. In: Santos, W. (ed.) Fuzzy Control Systems: Design, Analysis and Performance Evaluation, pp. 141–215. Nova Science Publishers, Hauppauge, NY (2017)

Models and Algorithms for Prediction of Electrical Energy Consumption Based on Canonical Expansions of Random Sequences

Igor Atamanyuk, Volodymyr Kondratenko, Yuriy Kondratenko, Vyacheslav Shebanin and Marina Solesvik

Abstract The given chapter is devoted to the development of the mathematical support, in particular, mathematical models and algorithms, which can be successfully used for solving prediction tasks in various areas of human activity, including energetic and ecological management. The development peculiarities and the use of models and algorithms as elements of green technology to predict electric energy consumption based on mathematical apparatus of canonical expansions of random sequences are currently being discussed. Developed calculation method doesn't impose any limitations on the qualities of random sequences of the change of electric energy consumption (requirement of linearity, Markovian behavior, monotony, stationarity etc.) and has maximal accuracy characteristics in this connection. Block diagrams of algorithms and results of the applied realization of the developed models and algorithms, for example the prediction of electric energy consumption by one of the local neighborhoods in Mykolaiv, Ukraine are introduced in the work. Comparative analysis of the results of a numerical experiment with the use of a Kalman filter and the linear prediction method confirms the high efficiency of the developed models and algorithms (relative error of prediction of electric energy consumption is 2–3%).

Keywords Models · Algorithms · Green information technology
Extrapolation · Electrical energy · Consumption · Prediction

I. Atamanyuk (✉) · V. Shebanin
Mykolaiv National Agrarian University, Mykolaiv, Ukraine
e-mail: atamanyuk@mnau.edu.ua

V. Kondratenko
Charter Communications, Greenwood Village, CO, USA
e-mail: vkondrat87@gmail.com

Y. Kondratenko
Petro Mohyla Black Sea National University, Mykolaiv, Ukraine
e-mail: yuriy.kondratenko@chmnu.edu.ua

M. Solesvik
Nord University Business School, Nord University, Bodø, Norway
e-mail: marina.solesvik@hvl.no

© Springer Nature Switzerland AG 2019

V. Kharchenko et al. (eds.), *Green IT Engineering: Social, Business and Industrial Applications*, Studies in Systems, Decision and Control 171,
https://doi.org/10.1007/978-3-030-00253-4_17

1 Introduction

Electric energy is the basis of modern civilization. In the entire world it is considered as the most preferable intermediate kind of energy, universal (transformed easily in any quantities into heat, light, mechanical energy etc.) and is transmitted into considerable distances. Electric energy is the basis for the development of basic industries, determining the progress of social production (the overwhelming majority of the machinery and equipment which the people use contains electric circuits and corresponding nodes the work of which is impossible without electric energy). At the same time the electric-power industry is one of the most unfavorable influence sources on the environment.

The influence of production systems, transmission and electric energy consumption on the environment reveals itself in such processes and as phenomenal as [1–3]:

(1) the withdrawal of territories for fuel production, the placement of electric power stations and transmission facilities and waste disposal;
(2) pollution of atmosphere and lithosphere with combustion products (atmospheric emissions, waste products, radioactive wastes etc.);
(3) heat (thermal) pollution—heat energy dumping of electric power station into the environment and the rise of environmental temperature;
(4) electromagnetic pollution—the creation of electric, magnetic and electromagnetic fields creating the threat for humankind and the biosphere;
(5) radioactive pollution;
(6) flood of usable territories (in case of hydroelectric power station);
(7) influence on climate;
(8) influence on flora and fauna;
(9) induced seismicity—the earthquake occurrence during the plants creation power, in the first place of hydroelectric power stations.

Thereupon the problem of creating efficient green IT to predict and determine the necessary volumes of consumed electric energy to increase effective management of electric-power industry objects, that will allow the reduction of the influence on the environment, is important and actual.

Peculiarity of solving the problem of electric energy consumption prediction is a stochastic character of values change of a given index considering the different factors' influence: type of day (workday, day off, holiday); weather conditions (air temperature, rain, fog, snow); dependence on time of day; duration of daylight hours etc. A non-stationary character of the factors change that influence on the processes of electric energy consumption determines the necessity of the effective green IT development (for the solving of corresponding prediction problems) on the basis of computational methods and extrapolation models of random sequences. The most general extrapolation form for solving the nonlinear prediction problem is the Kolmogorov-Gabor polynomial [4] but the determination of its parameters for the highest number of known values and the used order of nonlinear connections is very difficult and laborious process (with thus 11 already known values and the order of nonlinearity 4 it is necessary to obtain and solve 1819 equations of partial derivatives of extrapolations mean-square error). Thereupon during the formation of realizable in practice algorithms to predict different simplifications and limitations on the qualities of random sequence are used. For example, a range of suboptimal nonlinear extrapolation methods [5] with a limited stochastic relation order on the approximation basis of a posteriori probability density to estimated vector by orthogonal Hermite polynomial expansion or in the form of the Edgeworth series is offered by V. S. Pugachev. A solution to the non-stationary Kolmogorov equation [6] (in particular, the case of Stratanovich differs an equation to describe the Markovian processes) is obtained and provided that a drift coefficient is a linear function of condition, and diffusion the coefficient of which is equal to constant. An exhaustive solution to the optimal linear extrapolation problem for different classes of random sequences [7, 8] and different levels of informational support to the prediction problem (the Kolmogorov equation for random stationary sequences measured without errors; the Kalman method [9, 10] used for Markovian noisy random sequences; the Wiener-Hopf filter-extrapolator [11] for noisy stationary sequences etc.) das exist. But the use of simplifying assumptions considerably limits the extrapolation accuracy.

Thus, to accurately predict this, developed green IT must include processing methods on the basis of complete information about random sequences qualities of change of electric energy consumption.

Today, the studies in the sphere of creation and improvement of green information technologies (green IT) are carried out by scientists from different countries. Whereupon special attention is paid to the development of the concepts, models, components, networks, complex systems architectures and systems implementation in green IT engineering [12, 13]. Among the main directions of green IT in which theoretical and applied studies are actively carried out, the following ones should be named: energy saving [14, 15], resource and energy optimization and forecasting [16–18], data acquisition and supervisory control of environment-friendly energy-saving technologies [19], green-IT techniques for design and optimization of new technological processes and modern industrial systems [20] and others [12, 13]. Some articles are devoted to the theoretical aspects of green-IT implementation in energy saving and energy management processes, the stochastic approach in particular, is considered in [15] as a model-based evaluation of energy saving systems, the paper [21] is devoted to the development of the Markov models of smart grid digital substations availability for the investigation of multi-level degradation and recovery of power resources issues, etc. Special attention should be paid to the application of modern intelligent methods and technologies for solving tasks in power and electric energy sectors of the economy. For example, the authors of the paper [22] describe the hybrid adaptive systems of computational intelligence for IT in energy management tasks based on the on-line prediction of non-stationary nonlinear chaotic and stochastic time series, which describe electrical load producing and consuming processes.

The main goal of this chapter is to describe the developed models and algorithms (concept, theoretical base and step-by-step procedure for implementation) to predict regional electricity consumption to increase the efficiency of the energy management processes with verification and testing of developed software complex in the Mykolaiv region, Ukraine.

2 Mathematical Models and Computational Procedures Used to Predict Electric Energy Consumption

2.1 Polynomial Canonical Model of the Random Sequence of the Electric Energy Consumption Change

The studied random sequence $\{X\} = X(i), i = \overline{1, I}$ describing the change of the values $x(i), i = \overline{1, I}$ of the electric energies consumption in the moments of time $t_i, i = \overline{1, I}$ can be introduced in the form of a canonical model [23–25]:

$$X(i) = M[X(i)] + \sum_{\nu=1}^{i} V_\nu \varphi_\nu(i), i = \overline{1, I}. \tag{1}$$

Random coefficients $V_i, i = \overline{1, I}$ and coordinative functions $\varphi_\nu(i), i = \overline{1, I}$ are determined as

$$V_i = X(i) - \sum_{\nu=1}^{i-1} V_\nu \varphi_\nu(i), i = \overline{1, I}; \tag{2}$$

$$D_i = M\left[\left\{V_i\right\}^2\right] = D_x(i) - \sum_{\nu=1}^{i-1} D_\nu \varphi_\nu^2(i), i = \overline{1, I}; \tag{3}$$

$$\begin{aligned}\varphi_\nu(i) &= \frac{1}{D_\nu} M[V_\nu(X(i) - M[X(i)])] \\ &= \frac{1}{D_\nu}\Bigg[M[X(\nu)X(i)] - M[X(\nu)]M[X(i)] \\ &\quad - \sum_{\mu=1}^{\nu-1} D_\mu \varphi_\mu(\nu)\varphi_\mu(i)\Bigg], \nu = \overline{1, I}, i = \overline{\nu, I}.\end{aligned}$$

where $M[\cdot]$ is a mathematical expectation; $D_x(i)$ is a variance of the sequence $\{X\}$ in the sampling point t_i; D_ν is a random variance with coefficient V_ν.

In expression (1) the peculiarities (non-stationarity, non-Markovian behavior, non-monotony and etc.) of a studied random sequence $\{X\} = X(i), i = \overline{1, I}$ are adequately taken into account, but correlational functions were considerably drawn back from model (1).

To construct a canonical expansion while taking nonlinear stochastic relations into account $M\left[X^{\lambda}(i)X^{h}(j)\right], (\lambda \geq 1, h \geq 2) \vee (\lambda \geq 2, h \geq 1)$ of a random sequence of the electric energy change consumption let's consider the array of random values

$$X = \begin{bmatrix} X(1) & X(2) & \ldots & X(i) & \ldots & X(I-1) & X(I) \\ X^2(1) & X^2(2) & \ldots & X^2(i) & \ldots & X^2(I-1) & X^2(I) \\ \ldots & \ldots & \ldots & & \ldots & & \ldots \\ X^{N-1}(1) & X^{N-1}(2) & \ldots & X^{N-1}(i) & \ldots & X^{N-1}(I-1) & X^{N-1}(I) \\ X^{N}(1) & X^{N}(2) & \ldots & X^{N}(i) & \ldots & X^{N}(I-1) & X^{N}(I) \end{bmatrix} \quad (4)$$

where $X^2(i), \ldots, X^N(i), i = \overline{1, I}$ is the result of raising a corresponding power of elements $X(i), i = \overline{1, I}$ in a random sequence $\{X\}$.

Prior information about the sequence $\{X\} = X(i), \mathrm{i} = \overline{1, I}$ can be obtained by determining the cross correlation of the elements of the array X. Let's consider a given array as a vector random sequence $\overline{X}(i) = \left\{X(i), X^2(i), \ldots X^{N-1}(i), X^N(i)\right\}$, $i = \overline{1, I}$, the components of which are lines of the array, and the ordinal number of a sampling point $i = \overline{1, I}$ is unambiguously connected with the ordinal number of a column. The application of a linear vector canonical expansion to the sequence $\overline{X}(i), i = \overline{1, I}$ gives the following expression for the first component [26, 27]

$$X(i) = M[X(i)] + \sum_{\nu=1}^{i} \sum_{\lambda=1}^{N} W_{\nu}^{(\lambda)} \beta_{1\nu}^{(\lambda)}(i), i = \overline{1, I} \quad (5)$$

Elements $W_{\nu}^{(\lambda)}, \beta_{h\nu}^{(\lambda)}(i)$ of the model are determined by recurrent relations:

$$W_{\nu}^{(\lambda)} = X^{\lambda}(\nu) - M\left[X^{\lambda}(\nu)\right] - \sum_{\mu=1}^{\nu-1} \sum_{j=1}^{N} W_{\mu}^{(j)} \beta_{\lambda\mu}^{(j)}(\nu) - \sum_{j=1}^{\lambda-1} W_{\nu}^{(j)} \beta_{\lambda\nu}^{(j)}(\nu), \nu = \overline{1, I}; \quad (6)$$

$$\begin{aligned} \beta_{h\nu}^{(\lambda)}(i) = & \frac{M\left[W_{\nu}^{(\lambda)}\left(X^{h}(i) - M\left[X^{h}(i)\right]\right)\right]}{M\left[\left\{W_{\nu}^{(\lambda)}\right\}^{2}\right]} \\ & - \frac{1}{D_{\lambda}(\nu)}\{M\left[X^{\lambda}(\nu)X^{h}(i)\right] \\ & - M\left[X^{\lambda}(\nu)\right]M\left[X^{h}(i)\right] - \sum_{\mu=1}^{\nu-1}\sum_{j=1}^{N} D_{j}(\mu)\beta_{\lambda\mu}^{(j)}(\nu)\beta_{h\mu}^{(j)}(i) \\ & - \sum_{j=1}^{\lambda-1} D_{j}(\nu)\beta_{\lambda\nu}^{(j)}(\nu)\beta_{h\nu}^{(j)}(i)\}, \lambda = \overline{1, h}, \nu = \overline{1, i}, h = \overline{1, N}, i = \overline{1, I}. \end{aligned} \quad (7)$$

$$D_{\lambda}(v) = M\left[\left\{W_v^{(\lambda)}\right\}^2\right] = M\left[X^{2\lambda}(v)\right] - M^2\left[X^{\lambda}(v)\right] \\ - \sum_{\mu=1}^{v-1}\sum_{j=1}^{N} D_j(\mu)\left\{\beta_{\lambda\mu}^{(j)}(v)\right\}^2 - \sum_{j=1}^{\lambda-1} D_j(v)\left\{\beta_{\lambda v}^{(j)}(v)\right\}^2, \lambda = \overline{1,N}, v = \overline{1,I}. \quad (8)$$

where $D_{\lambda}(v)$ is a random variance coefficient of $W_v^{(\lambda)}$.

The coordinate functions $\beta_{hv}^{(\lambda)}(i), v = \overline{1,i}; \lambda, h = \overline{1,N}; i = \overline{1,I}$ are characterized by the qualities:

$$\beta_{hv}^{(\lambda)}(i) = \begin{cases} 1, & \text{if } (h = \lambda) \wedge (v = i); \\ 0, & \text{if } (i < v) \vee ((h < \lambda) \wedge (v = i)). \end{cases}$$

The model (5) of the sequence $\{X\} = X(i), \mathrm{i} = \overline{1,I}$ contains N arrays $\left\{W^{(\lambda)}\right\}, \lambda = \overline{1,N}$ of uncorrelated centered random coefficients $W_i^{(\lambda)}, \lambda = \overline{1,N}, i = \overline{1,I}$. Each of the given coefficients contain the information about the corresponding value $X^{\lambda}(i), \lambda = \overline{1,N}, i = \overline{1,I}$, and the coordinate functions $\beta_{hv}^{\lambda}(i), \lambda, h = \overline{1,N}, v, i = \overline{1,I}$ describe probabilistic relations of the order $\lambda + h$ between two different sections v and i, $\left(v, i = \overline{1,I}\right)$.

A canonical expansion (5) provides the mean-square error minimum of a random sequence presentation and allows the implementation of modeling random (depending only on the possibilities of a computer) intervals.

An essential advantage of model (5) in comparison with model (1) is the use of nonlinear stochastic relations $M[X^{\lambda}(i)X^h(j)], (\lambda \geq 1, h \geq 2) \vee (\lambda \geq 2, h \geq 1)$ of a random sequence of the change of electric energy consumption.

2.2 *Optimal Parameters Identification of Nonlinear Canonical Model of a Random Sequence of the Electric Energy Consumption Change*

The use of canonical presentations of random sequences has certain peculiarities. In particular, during the formation of a canonical expansion it is necessary to use such values of its parameters by which the model (5) describes as accurately as possible the studied random sequence: (a) the interval of random sequence modelling within the limits of which significant stochastic relations exist; (b) the number of sampling points on the interval of modelling, and also (c) higher orders of nonlinear stochastic relations which are necessary to use in the expansion.

The mechanism of aftereffect calculations and estimations of stochastic relations in the canonical expression (5) of a random sequence is laid in the coordinate functions:

$$\beta_{1\nu}^{(\lambda)}(i) = \frac{M\left[W_\nu^{(\lambda)}(X(i) - M[X(i)])\right]}{D_\lambda(\nu)} = \frac{M\left[W_\nu^{(\lambda)} \overset{o}{X}(i)\right]}{D_\lambda(\nu)}, i > \nu, \tag{9}$$

which by determination reflect the degree of stochastic relation between coefficients $W_\nu^{(\lambda)}$ and consequent sections of a random sequence $X(i), i > \nu$ ($\overset{o}{X}(i) = X(i) - M[X(i)]$ is centered random value). Thus it is possible to assume that the influence of the coefficients $W_\nu^{(\lambda)}$ in section t_ν on the consequent values of a random sequence equals to zero if starting from a certain $i_{k_\nu} > \nu$ for arbitrary λ it is true $\beta_{1\nu}^{(\lambda)}(i) \equiv 0$. Correspondingly the interval of aftereffect k_ν from the sampling point t_ν is determined at that as

$$k_\nu = i_{k_\nu} - \nu. \tag{10}$$

In a considered situation the elements of canonical expansions formation is carried out by a finite sample $x_l(i), \mathrm{i} = \overline{1, I}; l = \overline{1, L}$ of the volume L. This is why the determination of k_ν is reduced to the solution of the consequent problem. By the data of sample, the values of coordinate functions estimations $\beta_{1\nu}^{(\lambda)(L)}(i), \lambda = \overline{1, N}, i = \overline{\nu, I}$. are obtained. It is necessary to determine the following value $i_{k_\nu}, \nu < i_k$. The identity for this value $\beta_{1\nu}^{(\lambda)}(i) \equiv 0, i > \nu$ is carried out with a given degree of trust.

The mechanism of linearization laid in the model (5) allows the reduction of the problem of aftereffect estimation and the determination of the order of nonlinearity to the analysis of linear relation between $W_\nu^{(\lambda)}$ and $X(i), \nu < \mathrm{i}$, by standard numerical characteristics which is a normalized correlation coefficient

$$r^{(\lambda)}(\nu, i) = \frac{M\left[W_\nu^{(\lambda)} \overset{o}{X}(i)\right]}{\sqrt{D_\lambda(\nu)}\sqrt{D_x(i)}}, \lambda = \overline{1, N}, \nu = \overline{1, I}, i = \overline{\nu, I}. \tag{11}$$

Taking (9) into account, the last expression is reduced to its final form

$$r^{(\lambda)}(\nu, i) = \frac{\sqrt{D_\lambda(\nu)}\beta_{1\nu}^{(\lambda)}(i)}{\sqrt{D_x(i)}}, i > \nu, \tag{12}$$

where the coefficient of correlation $r^{(\lambda)}(\nu, i)$ is expressed through the elements $D_\lambda(\nu), \beta_{1\nu}^{(\lambda)}(i)$ of a canonical expansion. As the estimations of all these elements are obtained during the process of processing original statistical data, their application in formula (12) allows obtaining the estimation $r^{(\lambda)(L)}(\nu, i)$ of normalized correlation coefficient for any values ν, λ and i. Using this information, the given estimation problem of correlation relations significance of the coefficient $W_\nu^{(\lambda)}$ with

i-the section of a studied random sequence can be formulated as a testing problem of statistical hypothesis

$$r^{(\lambda)}(v,i) = 0 \tag{13}$$

and the alternative

$$r^{(\lambda)}(v,i) \neq 0.$$

As shown in work [17] a random value

$$\frac{1}{2}\ln\left[\frac{1 + r^{(\lambda)(L)}(v,i)}{1 - r^{(\lambda)(L)}(v,i)}\right]$$

should be considered as the one which is normally distributed, with mathematical expectation

$$m = \frac{1}{2}\ln\left[\frac{1 + r^{(\lambda)}(v,i)}{1 - r^{(\lambda)}(v,i)}\right] + \frac{r^{(\lambda)}(v,i)}{2(L-1)}$$

and variance $D = \frac{1}{L-3}$, which is why value

$$a^{(\lambda)}(v,i) = \sqrt{L-3}\left[\frac{1}{2}\ln\left[\frac{1 + r^{(\lambda)(L)}(v,i)}{1 - r^{(\lambda)(L)}(v,i)}\right] - \left(\frac{1}{2}\ln\left[\frac{1 + r^{(\lambda)}(v,i)}{1 - r^{(\lambda)}(v,i)}\right] + \frac{r^{(\lambda)}(v,i)}{2(L-1)}\right)\right]$$

has a standard normal distribution (0, 1).

Thus, data should be considered to be compatible with the hypothetical value $r^{(\lambda)}(v,i)$ with the significance level α if value

$$\frac{1}{2}\ln\left[\frac{1 + r^{(\lambda)}(v,i)}{1 - r^{(\lambda)}(v,i)}\right] + \frac{r^{(\lambda)}(v,i)}{2(L-1)}$$

lies within the limits

$$\left[\frac{1}{2}\ln\left[\frac{1 + r^{(\lambda)(L)}(v,i)}{1 - r^{(\lambda)(L)}(v,i)}\right] - \frac{z_\alpha}{\sqrt{L-3}}, \frac{1}{2}\ln\left[\frac{1 + r^{(\lambda)(L)}(v,i)}{1 - r^{(\lambda)(L)}(v,i)}\right] + \frac{z_\alpha}{\sqrt{L-3}}\right],$$

where z_α is a value of standard normal deviation that corresponds to a confidence probability of α. In other cases, the hypothetical value $r^{(\lambda)}(v,i)$ doesn't correspond to statistical data.

Taking the equality testing of correlation into account the coefficient to zero $r^{(\lambda)}(v,i) = 0$ is carried out, a random value $a^{(\lambda)}(v,i)$ takes the form

$$a^{(\lambda)}(v,i) = \frac{\sqrt{L-3}}{2}\ln\left[\frac{1+r^{(\lambda)(L)}(v,i)}{1-r^{(\lambda)(L)}(v,i)}\right]$$

and hypothesis (13) is accepted if the condition

$$-z_\alpha < a^{(\lambda)}(v,i) < z_\alpha \qquad (14)$$

is carried out.

Hypothesis (14) is tested out many times by the growing parameter λ till a certain boundary of value $N^{(v,i)}$, by which condition (14) is true ($N^{(v,i)}$ is higher order of nonlinear relation between sections t_v and t_i). After that the interval $i = i+1$ grows and the search of a higher order of nonlinearity is repeated for a new interval.

If for a certain i_{k_v} from the area $i > v$ and arbitrary λ (as a rule, testing at $\lambda = 1$ is enough: nonlinear relations decay quicker than linear ones) the assertion $r^{(\lambda)}(v,i) = 0, \mathrm{i} > i_{k_v}^{(\lambda)}$ turns out to be true, this means that the aftereffect interval for the sampling point t_v is equal to $i_{k_v} - v$ and for all $i > i_{k_v}$ the value of coordinate function $\beta_{1v}^{(\lambda)}(i)$ must be assumed to be zero.

The testing of the aftereffects for all sampling points $t_v, v = \overline{1, I^*}$ in which random sequence $\{X\}$ is studied, allows the values formation of the parameters N and I of canonical expansion (5) as

$$I = \max_v (i_{k_v} - v),$$
$$N = \max_{v,i} N^{(v,i)}.$$

The problem of choice of sampling points t_i (their quantity and interval between points) is connected with the principle of construction of a canonical expansion itself according to which each subsequent coefficient $W_i^{(1)}$ of a canonical expression (5) determines that only information which a subsequent section of a random sequence $\{X\}$ is relative to the previously accumulated one carries:

$$W_i^{(1)} = X(i) - M[X(i)] - \sum_{\mu=1}^{i-1}\sum_{j=1}^{N} W_\mu^{(j)} \beta_{1\mu}^{(j)}(i), i = \overline{1, I}. \qquad (15)$$

While using expression (15), the following situation is possible when the subsequent coefficient $W_i^{(1)}$ is determined as a difference of close numbers in case of an error. The results of this error grow sharply. In consequence of this starting from a certain moment according to model (5) a canonical expansion of the noise of calculations is formed instead of a canonical expansion of a studied random sequence. There are two reasons why this situation occurred [23]:

(a) a degenerate character of an investigated random sequence, when this sequence is described by the expression with random parameters;
(b) the sample interval is too small and as a result the neighboring sections of the sequence $\overset{o}{X}(i)$ and $\overset{o}{X}(i+1)$ start to faintly distinguish between themselves.

Another common error to these situations is that that consequent coefficient $W_i^{(1)}$ carries a small amount (in the limit—zero) of new information about a studied random sequence.

To keep the rule of taking a decision about informational significance of sampling point t_i it is appropriate to use approach [23] which is based on the analysis of a mean-square error of the linear canonical expression in a random sequence in the given point t_i provided that $i-1$ sampling points are used in the expansion. For a canonical expansion (5) of a random sequence the subject for analysis is

$$\Delta^{(i-1)}(i) = M\left[\left(\overset{o}{X}(i) - \sum_{\mu=1}^{i-1}\sum_{j=1}^{N} W_\mu^{(j)}\beta_{1\mu}^{(j)}(i)\right)^2\right].$$

The mean-square error of the expression of the random sequence $\{X\}$ in the point t_i is equal to the variance of a random coefficient of $W_i^{(1)}$:

$$D_1(i) = M\left[\left\{W_i^{(1)}\right\}^2\right] = M\left[X^2(i)\right] - M^2[X(i)] - \sum_{\mu=1}^{i-1}\sum_{j=1}^{N} D_j(\mu)\left\{\beta_{\lambda\mu}^{(j)}(i)\right\}^2.$$

Thus, the estimation of informational significance $S(i)$ for the i-th coefficient of an expansion can be determined as

$$S(i) = \frac{D(i)}{D_1(i)}, i = \overline{1, I}, \tag{16}$$

where $D(i)$ is a variance of a studied random sequence $\{X\}$ in the i-th section.

The larger value $S(i)$ is, the smaller informational contribution of the i-th coefficient $W_i^{(1)}$ which is in the description of a random sequence, and in a such way a certain boundary value can always be determined on reaching on the basis of heuristic arguments where a corresponding coefficient is considered to be informationally insignificant and is excluded from the following study.

Practically during the determination of coefficient $S(i)$ using formula (16) estimations $D^{(L)}(i)$ and $D_1^{(L)}(i)$ of the corresponding variances have obtained by the sample of a limited volume L. This is why the estimation problem of the i-th coefficient $W_i^{(1)}$ significance must be considered as a problem of testing statistical hypotheses formulating in the following way.

The estimation of significance is obtained based on data from a sample taken from volume L

$$S^{(L)}(i) = D^{(L)}(i) \Big/ D_1^{(L)}(i).$$

Based on this information it is necessary to test out the hypothesis

$$H_0 : S(i) \geq S_0 \tag{17}$$

concerning the alternative hypothesis H_1 according to which the condition is not carried out. At that the level of significance $\alpha_Д$ is assumed to be given (conditional probability of a hypothesis deviation H_0 in case it is not true).

The comparisons problem of the random values $X(i)$ and $W_i^{(1)}$ variances to which the formulated problem is reduced and is considered in the work [24] where its strict solution is given on the assumption that all samples on the basis of which variances $D^{(L)}(i)$ and $D_1^{(L)}(i)$ that are obtained are independent and distributed according to Gaussian law at that variance relation submits to F—Fisher distribution.

But the centered value $\overset{o}{X}(i)$ of a random sequence and coefficient $W_i^{(1)}$ are connected through corresponding coordinate functions and in a such way samples on based on $S(i)$ are calculated and are dependent. The given problem is solved if instead of the value $\overset{o}{X}(i)$ of a random sequence in the i-th sampling point its estimation in the form of a canonical expansion with $(i-1)N$ independent random coefficients

$$\overset{o}{X}(i) = \sum_{\nu=1}^{i-1} \sum_{\lambda=1}^{N} W_\nu^{(\lambda)} \beta_{1\nu}^{(\lambda)}(i) \tag{18}$$

is used.

Thus, the problem of dependence is removed if variance $D(i)$ is substituted by its estimation in the formula (16)

$$D_{i-1}(i) = \sum_{\mu=1}^{i-1} \sum_{j=1}^{N} D_j(\mu) \left\{ \beta_{\lambda\mu}^{(j)}(\nu) \right\}^2. \tag{19}$$

At the same time assumptions about the Gaussian distribution of samples based on variances $D^{(L)}(i)$ and $D_1^{(L)}(i)$ are calculated and are not removed by this. Under given conditions it is reasonable to reduce the problem of significance estimation to the determination of a lower confidence limit $S_Д(i)$ for a true value of $S(i)$ at a given value of $\alpha_Д$ assuming a sample normality that corresponds to the fact finding $S(i) > S_Д(i)$with the probability of $1-\alpha_Д$.

A comparison of the obtained value $S_Д(i)$ with the given value S_0 allows decision-making about the significance of coefficient $W_i^{(1)}$.

At a given value of $\alpha_{Д}$ a lower confidence limit is determined as

$$S_{Д}(i) = \frac{D_{i-1}^{(L)}(i)}{D_1^{(L)}(i)F(\alpha_{Д}, L, L)}, \tag{20}$$

where $F(\alpha_{Д}, L, L)$ is F for Fisher distribution.

If the next random coefficient $W_i^{(1)}$ has inessential information, given the i-th section is excluded from consideration and discrete step is increased. If all subsequent sampling points in which the values of a random sequence are measured, the decision of informational insignificance of random coefficients is taken. This means that a studied random sequence is degenerate-able and its accuracy is maximally reachable to describe a given volume of sample that is provided by $i-1$ members of expansion.

Thus, on the basis of expressions (12)–(20) for the model of a random sequence the changes of electric energy consumption are determined by: (a) a modelling interval of a random sequence in which significant stochastic relations exist, (b) a number of sampling points on the modelling interval, as well as (b) a higher order of nonlinear stochastic relations which are necessary to be used in the expansion.

2.3 Algorithm of a Random Sequence Prediction of the Electric Energy Consumption Change

Let's assume that as a result of measuring the first value $x(1)$ of the sequence of the electric energy consumptions $\{X\}$ in the point t_1 is known. Consequently, the values of coefficients $W_1^{(\lambda)}, \lambda = \overline{1,N}$ are known as:

$$w_1^{(\lambda)} = x^{\lambda}(1) - M\left[X^{\lambda}(1)\right] - \sum_{j=1}^{\lambda-1} w_1^{(j)} \beta_{\lambda 1}^{(j)}(1), \lambda = \overline{1,I}. \tag{21}$$

The substitution of $w_1^{(1)}$ into (5) allows the obtain of a polynomial canonical expansion of the posteriori random sequence $\left\{X^{(1,1)}\right\} = X(i/x_1(1))$:

$$\begin{aligned} X^{(1,1)}(i) = X(i/x(1)) = M[X(i)] + (x(1) - M[X(1)])\beta_{11}^{(1)}(i) \\ + \sum_{\lambda=2}^{N} W_1^{(\lambda)} \beta_{11}^{(\lambda)}(i) + \sum_{\nu=2}^{i} \sum_{\lambda=1}^{N} W_\nu^{(\lambda)} \beta_{1\nu}^{(\lambda)}(i), i = \overline{1,I}. \end{aligned} \tag{22}$$

The application of the operation of mathematical expectation to expression (22) gives optimality by the criterion of the mean-square error minimum of the extrapolation estimation of future values of sequence $\{X\}$ provided that the value $x(1)$ is used to determine a given estimation:

$$m_x^{(1,1)}(1,i) = M[X(i/x(1))] = M[X(i)] + (x(1) - M[X(1)])\beta_{11}^{(1)}(i), i = \overline{1,I}. \quad (23)$$

Taking the coordinate functions $\beta_{hv}^{(\lambda)}(i), \lambda, h = \overline{1,N}, v, i = \overline{1,I}$ into consideration, these are determined from the condition of the minimum of mean-square error of approximation in the intervals between the arbitrary values $X^{\lambda}(v)$ and $X^h(i)$, expression (23) and can be generalized in the case of prediction $x^h(i), h = \overline{1,N}, i = \overline{2,I}$:

$$\begin{aligned} m_x^{(1,1)}(h,i) = M\left[X^h(i/x(1))\right] = M\left[X^h(i)\right] \\ + (x(1) - M[X(1)])\beta_{h1}^{(1)}(i), i = \overline{1,I}. \end{aligned} \quad (24)$$

where $m_x^{(1,1)}(h,i)$ is an optimal estimation for the future value $x^h(i)$ provided that the value $x(1)$ is used for prediction.

The concretization in (22) of the second value $w_1^{(2)}$ gives a canonical expansion of the posteriori sequence $\left\{X^{(1,2)}\right\} = X\left(i/x_1(1), x_1(1)^2\right)$:

$$\begin{aligned} X^{(1,2)}(i) = X\left(i/x(1), x(1)^2\right) = M[X(i)] + (x(1) - M[X(1)])\beta_{11}^{(1)}(i) \\ + \left[x^2(1) - (x(1) - M[X(1)])\beta_{21}^{(1)}(1)\right]\beta_{11}^{(2)}(i) + \sum_{\lambda=3}^{N} W_1^{(\lambda)}\beta_{11}^{(\lambda)}(i) + \sum_{v=2}^{i}\sum_{\lambda=1}^{N} W_v^{(\lambda)}\beta_{1v}^{(\lambda)}(i). \end{aligned} \quad (25)$$

The application of the operation of a mathematical expectation to (25) allows an obtain of an algorithm of extrapolation by two values $x(1), x(1)^2$ using expression (24):

$$\begin{aligned} m_x^{(1,2)}(h,i) = M\left[X^h\left(i/x(1), x(1)^2\right)\right] \\ = m_x^{(1,1)}(h,i) + \left[x^2(1) - m_x^{(1,1)}(2,1)\right]\beta_{h1}^{(2)}(i), i = \overline{1,I}. \end{aligned} \quad (26)$$

For the first section t_1 in the last recurrent cycle the value $x^N(1)$ is used as the posteriori information

$$\begin{aligned} m_x^{(1,N)}(h,i) = M\left[X^h\left(i/x(1), x(1)^2 \ldots x(1)^N\right)\right] \\ = m_x^{(1,N-1)}(h,i) + \left[x^N(1) - m_x^{(1,N-1)}(N,1)\right]\beta_{h1}^{(N)}(i), i = \overline{1,I}, \end{aligned} \quad (27)$$

after which, prediction is specified by the use of the value $x(2)$

$$m_x^{(2,1)}(h,i) = M\left[X^h\left(i/x(1), x(1)^2 \ldots x(1)^N, x(2)\right)\right] \\ = m_x^{(1,N)}(h,i) + \left[x(2) - m_x^{(1,N)}(1,2)\right]\beta_{h2}^{(1)}(i), i = \overline{1,I}. \tag{28}$$

For section t_2 the last recurrent cycle is the operation

$$m_x^{(2,N)}(h,i) = M\left[X^h\left(i/x(1),\ldots,x(1)^N, x(2),\ldots,x(2)^N\right)\right] \\ = m_x^{(2,N-1)}(h,i) + \left[x^N(2) - m_x^{(2,N-1)}(N,2)\right]\beta_{h2}^{(N)}(i), i = \overline{1,I}.$$

The generalization of the obtained regularity allows to write down an algorithm of the prediction for an arbitrary number of the known values [27, 28]:

$$m_x^{(\mu,l)}(h,i) = \begin{cases} M\left[X^h(i)\right] \text{ at } \mu = 0; \\ m_x^{(\mu,l-1)}(h,i) + \left(x^l(\mu) - m_x^{(\mu,l-1)}(l,\mu)\right)\beta_{h\mu}^{(l)}(i) \text{ at } l \neq 1, \\ m_x^{(\mu-1,N)}(h,i) + \left(x^l(\mu) - m_x^{(\mu-1,N)}(l,\mu)\right)\beta_{h\mu}^{(1)}(i) \text{ at } l = 1. \end{cases} \tag{29}$$

The expression $m_x^{(\mu,l)}(h,i) = M\left[X^h(i)/x^\nu(j), j = \overline{1,\mu-1}, \nu = \overline{1,N}; x^\nu(\mu), \nu = \overline{1,l}\right]$ for $h = 1, l = N, \mu = k$ is an optimal estimation of $m_x^{(k,N)}(1,i)$ of the future value $x(i), i = \overline{k+1,I}$ provided that the values $x^\nu(j), \nu = \overline{1,N}, j = \overline{1,k}$ are used for the calculation of the given estimation which are the results of measurements of the sequence $\{X\}$ in points $t_j, j = \overline{1,k}$ which are known.

An expression for the sought value $m_x^{(k,N)}(1,i)$ can be written down in the following explicit form [29, 30]:

$$m_x^{(k,N)}(1,i) = M[X(i)] + \sum_{j=1}^{k}\sum_{\nu=1}^{N}(x^\nu(j) - M[X^\nu(j)])S_{((j-1)N+\nu)}^{(kN)}((i-1)N+1), \tag{30}$$

where

$$S_\lambda^{(\alpha)}(\xi) = \begin{cases} S_\lambda^{(\alpha-1)}(\xi) - S_\lambda^{(\alpha-1)}(\alpha)\gamma_k(i), & \text{at } \lambda \leq \alpha - 1; \\ \gamma_\alpha(\xi), & \text{for } \lambda = \alpha; \end{cases} \tag{31}$$

$$\gamma_\alpha(\xi) = \begin{cases} \beta_{1,[\alpha/N]+1}^{(\mathrm{mod}_N(\alpha))}([\alpha/N]+1), & \text{for } \xi \leq kN; \\ \beta_{1,[\alpha/N]+1}^{(\mathrm{mod}_N(\alpha))}(i), & \text{if } \xi = (i-1)N+1. \end{cases} \tag{32}$$

In expression (32) [.] is the operation of rounding.

A canonical expansion of the posteriori random sequence $\{X\}$, provided that the first k of values are fixed, will be written as

$$X^{(k,N)}(i) = X\left(i/x^{v}(j), v = \overline{1,N}, j = \overline{1,k}\right) = m_{x}^{(k,N)}(1,i) + \sum_{\mu=k+1}^{i} \sum_{\lambda=1}^{N} W_{\mu}^{(\lambda)} \beta_{1\mu}^{(\lambda)}(i), i = \overline{k+1,I}. \quad (33)$$

The error of a single extrapolation at known values $x^{v}(j), v = \overline{1,N}, j = \overline{1,k}$ is determined by the formula [31]

$$\delta\left[i/x^{v}(j), v = \overline{1,N}, j = \overline{1,k}\right] = m_{x}^{(k,N)}(1,i) - x^{(k,N)}(i), i = \overline{k+1,I}. \quad (34)$$

where $x^{(k,N)}(i), \mathrm{i} = \overline{k+1,I}$ are the future values of extrapolated realization.

The realization $x^{(k,N)}(i), \mathrm{i} = \overline{k+1,I}$ in the area of prediction develops randomly which is why it is impossible to specify its exact values. However, the model (33) is an accurate description of a stochastic development of the realization $x^{(k,N)}(i), \mathrm{i} = \overline{k+1,I}$ that allows to reduce the expression (34) to the form

$$\Delta\left[i/x^{v}(j), v = \overline{1,N}, j = \overline{1,k}\right] = m_{x}^{(k,N)}(1,i) - X^{(k,N)}(i), i = \overline{k+1,I}. \quad (35)$$

Thus, a single extrapolations error has stochastic character determined by a random nature of the studied sequence.

The application of the operation of mathematical expectation taking the expression (33)–(35) into account. This allows to determine the systematic constituent of the error of single extrapolation

$$S\left[i/x^{v}(j), v = \overline{1,N}, j = \overline{1,k}\right] = M\left[\Delta\left(i/x^{v}(j), v = \overline{1,N}, j = \overline{1,k}\right)\right] = \sum_{\mu=k+1}^{i} \sum_{\lambda=1}^{N} M\left[W_{\mu}^{(\lambda)}\right] \beta_{1\mu}^{(\lambda)}(i) = 0, i = \overline{k+1,I}. \quad (36)$$

The last result shows that the predictive models (29), (30) which provide the receiving of unbiased estimates of future values for each predictable realization.

This also implies that the algorithm provides unbiasedness on an average of

$$S^{(k,N)}(i) = M\left[S\left(i/X^{v}(j), v = \overline{1,N}, j = \overline{1,k}\right)\right] = 0, \mathrm{i} = \overline{k+1,I} \quad (37)$$

The estimation of the significance of a random constituent of the single extrapolation error is its variance

$$\begin{aligned} D_\Delta\left[i/x^v(j), v=\overline{1,N}, j=\overline{1,k}\right] &= M\left[\left(\Delta\left[i/x^v(j), v=\overline{1,N}, j=\overline{1,k}\right]\right.\right.\\ &\qquad \left.\left.-S\left[i/x^v(j), v=\overline{1,N}, j=\overline{1,k}\right]\right)^2\right]\\ &= M\left[\left(\sum_{\mu=k+1}^{i}\sum_{\lambda=1}^{N} W_\mu^{(\lambda)}\beta_{1\mu}^{(\lambda)}(i)\right)^2\right]\\ &= \sum_{\mu=k+1}^{i}\sum_{\lambda=1}^{N} D_\lambda(\mu)\left\{\beta_{1\mu}^{(\lambda)}(i)\right\}^2, i=\overline{k+1,I}. \end{aligned} \quad (38)$$

As $D_\Delta\left[i/x^v(j), v=\overline{1,N}, j=\overline{1,k}\right]$ doesn't depend on the specific values of extrapolated realization its averaging by the multitude of extrapolations allows to write down the expression for variance of a posteriori random sequence of

$$D_\Delta^{(k,N)}(i) = M\left[D_\Delta\left[i/X^v(j), v=\overline{1,N}, j=\overline{1,k}\right]\right] = \sum_{\mu=k+1}^{i}\sum_{\lambda=1}^{N} D_\lambda(\mu)\beta_{1\mu}^{(\lambda)}(i). \quad (39)$$

Taking (37), (39) into account, the expression for the mean-square error of prediction will be written down as

$$\begin{aligned} E_x^{(k,N)}(i) &= \left[S^{(k,N)}(i)\right]^2 + D_\Delta^{(k,N)}(i)\\ &= D_\Delta^{(k,N)}(i) = \sum_{\mu=k+1}^{i}\sum_{\lambda=1}^{N} D_\lambda(\mu)\beta_{1\mu}^{(\lambda)}(i), i=\overline{k+1,I}. \end{aligned} \quad (40)$$

Thus, the mean-square error of the prediction with the help of algorithm (29), (30) is equal to the variance of the posteriori random sequence.

The optimality of the obtained method of extrapolation is proved by the following theorem.

Theorem If the random sequence $\{X\}$ is completely set by a discretized moment of functions $M\left[X^\lambda(v)\right]$, $M\left[X^\lambda(v)X^h(j)\right]$, $\lambda, h=\overline{1,N}, v, j=\overline{1,I}$, the application of models (29), (30) for the purposes of calculation of the estimation $m_x^{(k,N)}(1,i)$ of the future values $X(i), i=\overline{k+1,I}$ is necessary and a sufficient condition of its optimality by the criterion of minimum of the mean-square extrapolation error.

Proof Let's assume that there's another estimation of $\hat{X}^{(k,N)}(1,i)$ with a better quality index. Then, for the purposes of the multitude of random coefficients $W_i^{(\lambda)}, \lambda=\overline{1,N}, i=\overline{1,I}$ containing information about $X^\lambda(i), \lambda=\overline{1,N}, i=\overline{1,I}$, another set of coordinate functions $\psi_{1v}^{(\lambda)}(i)$ corresponds such that:

$$\hat{X}^{(k,N)}(1,i) = M[X(i)] + \sum_{\nu=1}^{\kappa}\sum_{\lambda=1}^{N} W_{\nu}^{(\lambda)}\psi_{1\nu}^{(\lambda)}(i), i = \overline{k+1,I}.$$

Let's calculate the mean-square error of the prediction for this case

$$E(i) = M\left[\left\{\overset{o}{X}(i) - \sum_{\nu=1}^{k}\sum_{\lambda=1}^{N} W_{\nu}^{(\lambda)}\psi_{1\nu}^{(\lambda)}(i)\right\}^2\right], i = \overline{1,I},$$

where $\overset{o}{X}(i) = X(i) - M[X(i)]$.

Considering the quality of uncorrelated coefficients and expression (7) for the determination of coordinate functions $\beta_{1\nu}^{(\lambda)}(i)$ direct calculation gives the following

$$E(i) = M\left[\left\{\overset{o}{X}(i)\right\}^2\right] + \sum_{\nu=1}^{k}\sum_{\lambda=1}^{N} D_{\lambda}(\nu)\left[\left\{\psi_{1\nu}^{(\lambda)}(i)\right\}^2 - 2\psi_{1\nu}^{(\lambda)}(i)\beta_{1\nu}^{(\lambda)}(i)\right], i = \overline{1,I},$$

however $\left\{\psi_{1\nu}^{(\lambda)}(i)\right\}^2 - 2\psi_{1\nu}^{(\lambda)}(i)\beta_{1\nu}^{(\lambda)}(i) = \left[\psi_{1\nu}^{(\lambda)}(i) - \beta_{1\nu}^{(\lambda)}(i)\right]^2 - \left\{\beta_{1\nu}^{(\lambda)}(i)\right\}^2$, therefore

$$E(i) = D_x(i) - \sum_{\nu=1}^{k}\sum_{\lambda=1}^{N} D_{\lambda}(\nu)\left\{\beta_{1\nu}^{(\lambda)}(i)\right\}^2 + \sum_{\nu=1}^{k}\sum_{\lambda=1}^{N} D_{\lambda}(\nu)\left\{\psi_{1\nu}^{(\lambda)}(i) - \beta_{1\nu}^{(\lambda)}(i)\right\}^2, i = \overline{1,I}.$$

It follows from the last expression that at any i mean-square error will be the least and equal to a posteriori variance of the random sequence $\{X\}$ if $\psi_{1\nu}^{(\lambda)}(i) = \beta_{1\nu}^{(\lambda)}(i)$.

Thus, the use of the parameters $\beta_{1\nu}^{(\lambda)}(i)$ in (29), (30) is a sufficient condition for optimality of the estimation of $m_x^{(k,N)}(1,i)$.

The essentiality of the application of models (29), (30) to receive optimal prediction is proved by the existence of the theorem about the uniqueness of such a decision and recurrent character of the determination of coordinate functions $\beta_{1\nu}^{(\lambda)}(i)$(at first the function $\beta_{11}^{(1)}(i)$ is calculated unambiguously, with the use of $\beta_{11}^{(1)}(i)$ function $\beta_{11}^{(2)}(i)$ is determined etc. This is given as $M\left[X^{\lambda}(\nu)\right]$, $M\left[X^{\lambda}(\nu)X^{h}(j)\right]$ a single set of coordinate functions $\beta_{1\nu}^{(\lambda)}(i)$ which currently exists and on their base weighting functions $S_{((j-1)N+\nu)}^{(kN)}((i-1)N+1)$ are determined unambiguously).

It should be mentioned that the parameters of algorithm (29) are calculated by the statistical sample volume L. This can be specified at the expense of new posteriori information $x_{L+1}(v), v = \overline{1, I}$. For solving the prediction problem in real time, it is possible to calculate the algorithm (29) parameters in such way:

$$M_{(L+1)}[X^n] = \frac{\sum_{l=1}^{L+1} x_l^n}{L+1} = \frac{\sum_{l=1}^{L} x_l^n L}{(L+1)L} + \frac{x_{L+1}^n}{L+1} = \frac{M_{(L)}[X^n]L + x_{L+1}^n}{(L+1)};$$

$$D_{n,(L+1)}(v) = \frac{\sum_{l=1}^{L+1} \left(v_{v,l}^{(n)}\right)^2}{L} = \frac{\sum_{l=1}^{L} \left(v_{v,l}^{(n)}\right)^2 (L-1)}{L(L-1)} + \frac{\left(v_{v,L+1}^{(n)}\right)^2}{L}$$

$$= \frac{D_{n,(L+1)}(v) + \left(x_{L+1}^n(v) - m^{(n-1,v)}(n,v)\right)}{L};$$

$$\beta_{\lambda v,(L+1)}^{(n)}(i) = \frac{\beta_{\lambda v,(L+1)}^{(n)}(i)(L-1) + \left(x_{L+1}^n(v) - m^{(n-1,v)}(n,v)\right)\left(x_{L+1}^{\lambda}(i) - M_{(L+1)}\left[X^{\lambda}(i)\right]\right)}{LD_{n,(L+1)}(v)}.$$

Thus, the use of the extrapolation results allows to essentially simplify (to some arithmetical operations) the procedure of recalculation of the parameters of algorithm (29) by the entry of new statistical data of a studied random sequence. This gives the possibility to use updated parameters to solve the prediction problem.

3 The Main Stages of the Developed Models and Algorithm Implementation Used to Predict Electric Energy Consumption

The application of models and algorithms, used to predict electric energy consumption, assumes the realization of the following stages:

Stage 1. Statistic data collection about a studied random sequence of electric energy consumption;

Stage 2. The estimation of moment functions $M\left[X^{\lambda}(i)X^{h}(j)\right]$ on the basis of accumulated realizations of electric energy consumption random sequence;

Stage 3. The forming of canonical expansion (5) based on the block diagram (Fig. 1) for studied sequence. Optimal parameters for model (5) (the number of sampling points and interval between them, the order of stochastic relation and the duration of the aftereffect) are determined in accordance with the block diagram introduced in Fig. 2;

Stage 4. The prediction of the future values of electric energy consumption based on model (29), (30) (the peculiarities of calculating the process are introduced in the form of the diagram in Fig. 3);

Stage 5. The estimation of the quality to solve the problem of predicting electric energy consumption with the help of the expression (40).

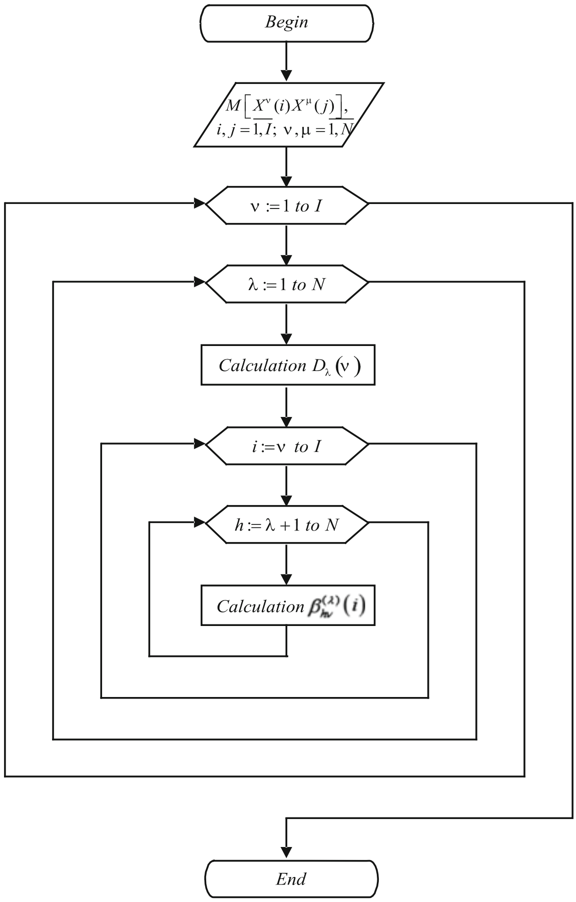


Fig. 1 A Block diagram to represent he calculations of the coordinate functions and the variances of random coefficients of canonical expansion Eq. (5)

4 The Results of a Numerical Experiment

Developed models and algorithms are approbated during the solving of the prediction problem of electric energy consumption for the Varvarovka district in Mykolaiv (in Fig. 4 graphs of electric energy consumption for the period 10.04.17–14.04.17 depending on times of a day).

To carry out comparative analysis based on the statistical data (period 1.04.17–31.05.17) a numerical experiment was carried out with the use of the linear method, the third-order Kalman method and with the help of an algorithm based on the predictive model (29), (30) of the third-order nonlinear relation. The results of the experiment (Fig. 5) show the high accuracy of prediction based on developed technology in comparison with a linear method at the expense of the use of nonlinear stochastic relations and in comparison with Kalman method at the expense of essential increase of posteriori information which is used for prediction.

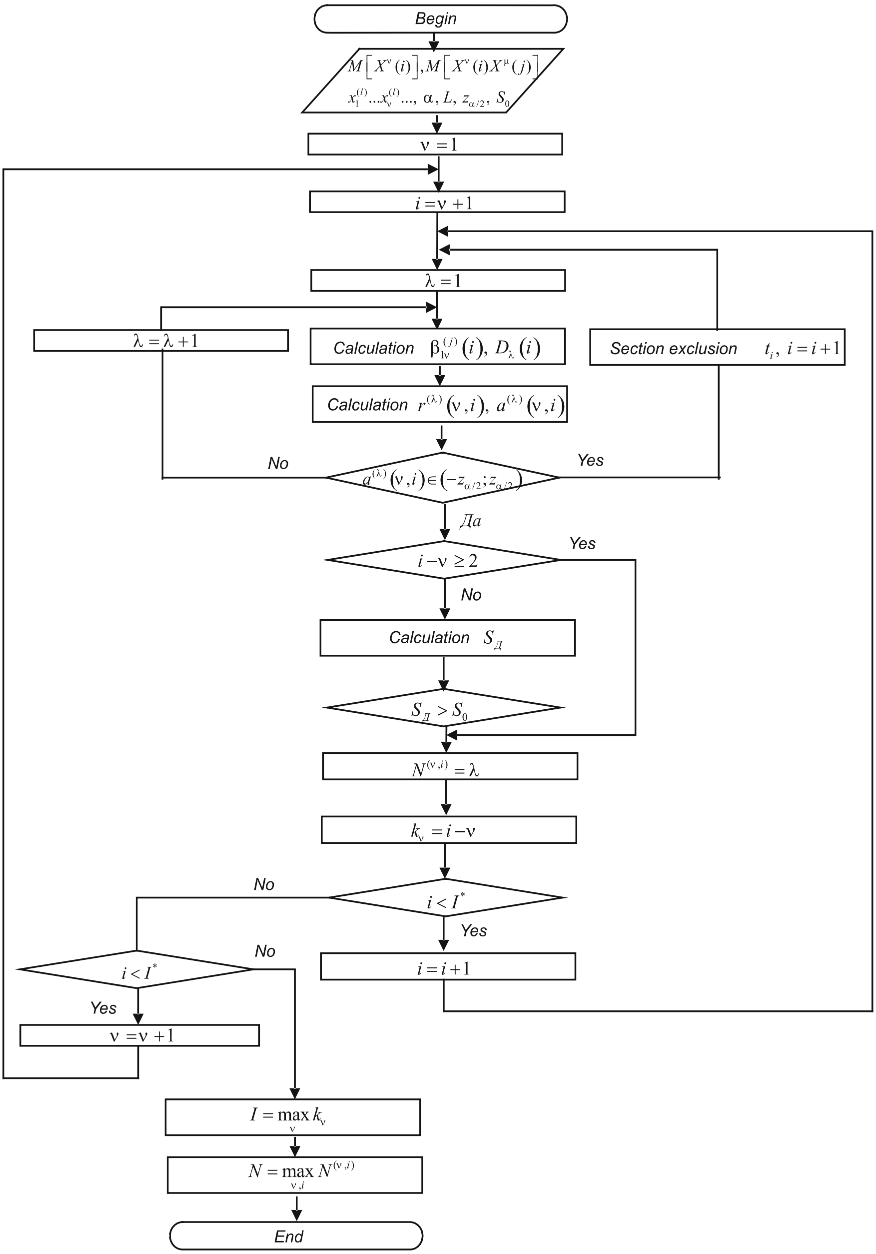


Fig. 2 A Block diagram to represent the calculation of optimal parameters of canonical expansion (5)

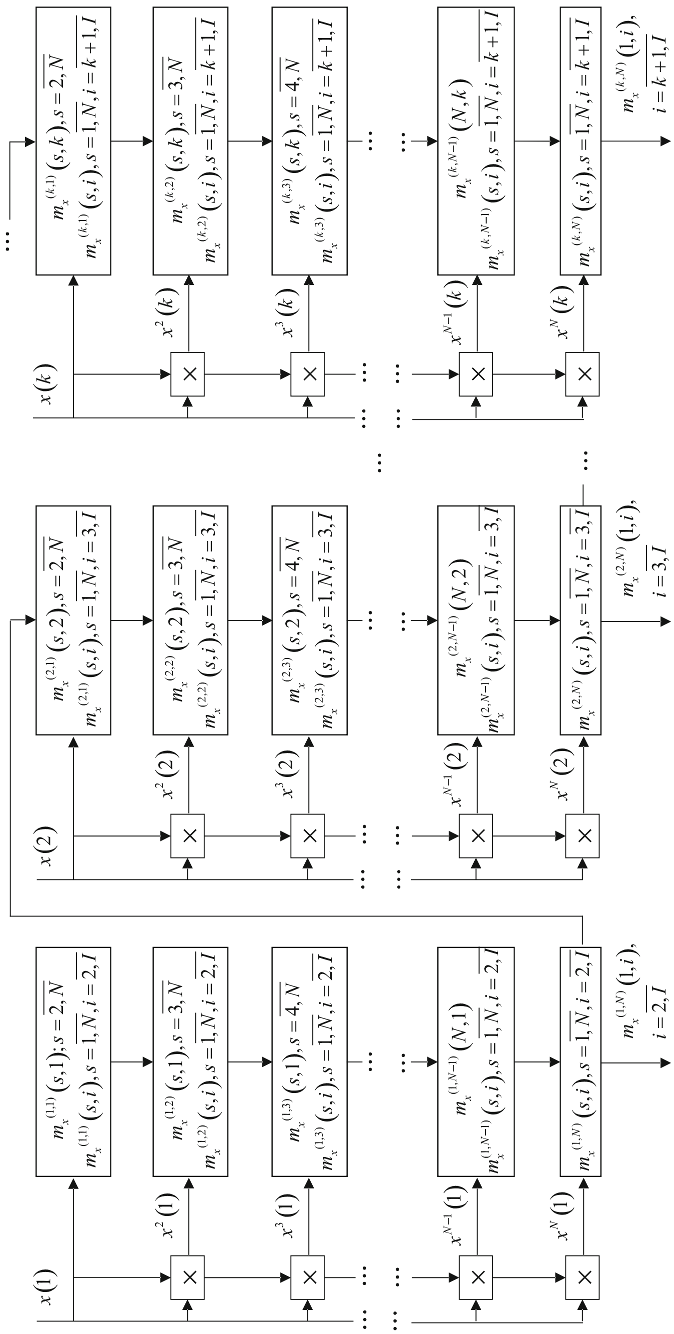


Fig. 3 A diagram to represent the forming of the future values of a random sequence with the help of model (29), (30)

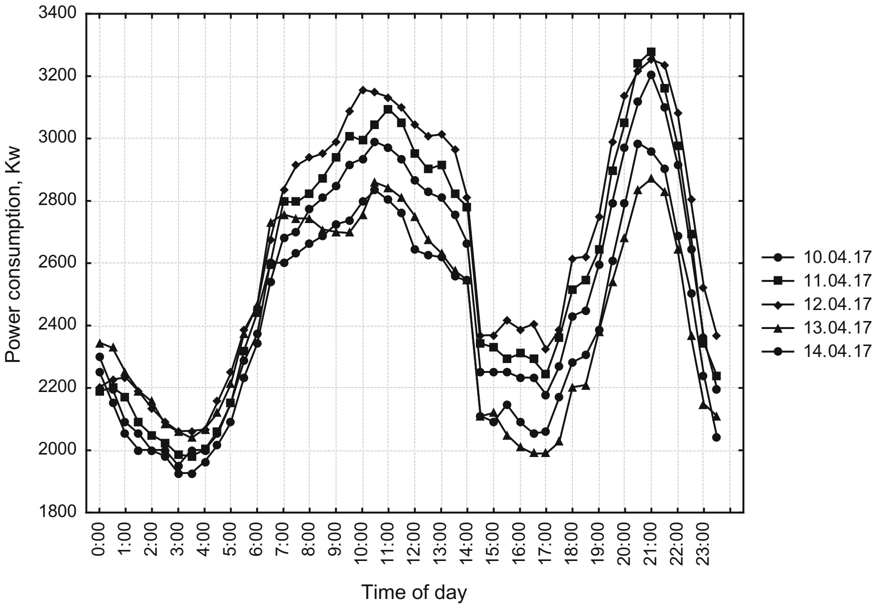


Fig. 4 A figure to represent electric energy consumption for the Varvarovka district, Mykolaiv for the period from 10.04.17 to 14.04.17

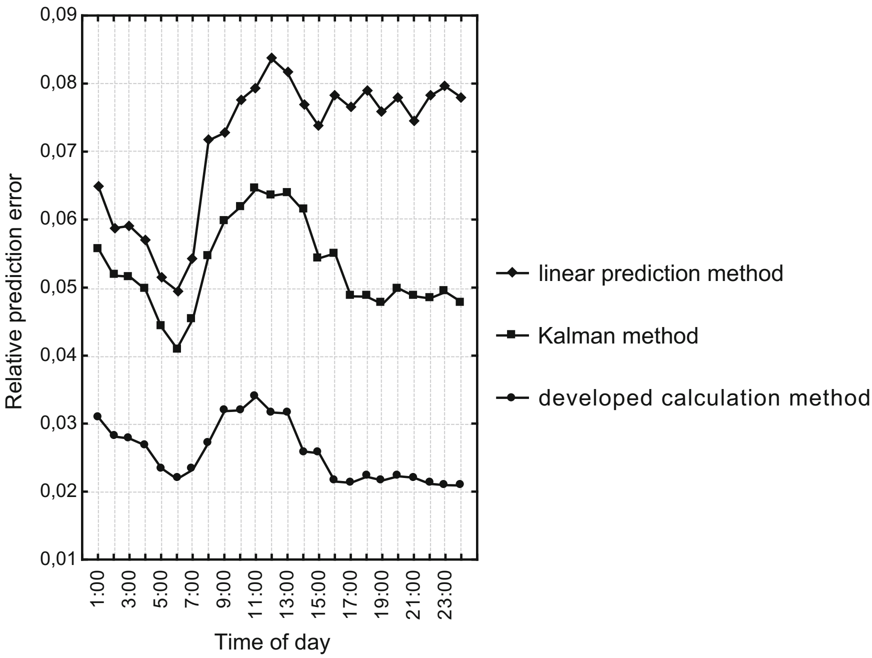


Fig. 5 The error of prediction of electric energy consumption for the Varvarovka district in Mykolaiv

5 Conclusion

The models and algorithms used in the prediction of future values of the process of electric energy consumption on the basis of a nonlinear canonical expansion of random sequences are offered. Models and algorithms don't impose any limitations on the qualities of random sequences of the electric energy consumptions change (requirement of linearity, Markovian behavior, monotony, stationarity etc.) that allow prediction to reach maximal accuracy. The results of the numerical experiment based on statistic data confirmed the high effectiveness of technology. The relative error of the prediction of electric energy consumption is 2–3%.

Increasing the accuracy of forecasting a reduction in the volume of electricity consumption will make it possible to more effectively regulate its production indicators and to carry out the redistribution of electricity between consumers. An accurate assessment of the consumption excess will enable (in addition to rational power generation) the mitigation of the consequences of electrical network congestion with the help of technical solutions and organizational measures and to avoid emergency situations associated with the cessation of electric power supplies to facilities that are classified as critical or dangerous.

The application of models and algorithms as green technology elements will allow an increase of effectiveness of planning electric energy generation, that will give the possibility to reduce unfavorable influences of the function of electric-power industry objects in the environment: the atmosphere (oxygen consumption, emissions of gases, moisture and solid particles), the hydrosphere (water consumption, the creation of artificial reservoirs, damping of polluted and heated water, liquid wastes) and the lithosphere (fossil fuels consumption, landscape change, emissions of toxic substances).

The proposed models and algorithms represent the core of the mathematical support of computer forecasting systems which can be used for solving different real-world prediction tasks in uncertainty.

References

1. Socolow, R.H. (ed.): Industrial Ecology and Global Change, 532 p. Cambridge University Press (2006)
2. Despeisse, M., Ball, P.D., Evans, S., Levers, A.: Industrial ecology at factory level—a conceptual model. J. Clean. Prod. **31**, 30–39 (2012)
3. Kalinichenko, A., Havrysh, V., Perebyynis, V.: Sensitivity analysis in investment project of biogas plant. J. Appl. Ecol. Environ. Res. **15**(4), 969–985 (2017)
4. Box, G.E.P., Jenkins, G.M.: Time–Series Analysis, Forecasting and Control, 575 p. Holden–Day, San Francisco (1970)
5. Pugachev, V.S.: Theory of Random Functions and its Application Control Problem, 833 p. Pergamon Press, London (2013)
6. Kolmogorov, A.N.: Foundations of the Theory of Probability. Chelsea Publishing Co., Oxford (1950)

7. Dorozhovets, M., Bubela, I.: Computing uncertainty of the extreme values in random samples. Int. J. Comput. **15**(2), 127–135 (2016). http://computingonline.net/computing/article/view/845
8. Gomolka, Z., Dudek-Dyduch, E., Kondratenko, Y.P.: From homogeneous network to neural nets with fractional derivative mechanism. In: Rutkowski, L. et al. (eds.) International Conference on Artificial Intelligence and Soft Computing, ICAISC-2017, Part I, Zakopane, Poland, June 11–15 2017, LNAI 10245, pp. 52–63. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59063-9_5
9. Kalman, R.E.: A new approach to linear filtering and prediction problems. Trans. ASME Ser. D, J. Basic Eng. **82**(Series D), 35–45 (1960)
10. Simon, D.: Training fuzzy systems with the extended Kalman filter. Fuzzy Sets Syst. **132**, 189–199 (2002)
11. Wiener, N.: Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications. MIT Press, New-York (1949)
12. Kharchenko, V., et al. (eds.): Green IT Engineering: Concepts, Models, Complex Systems Architectures. Studies in Systems, Decision and Control, vol. 74. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-44162-7
13. Kharchenko, V., et al. (eds.): Green IT Engineering: Components, Networks and Systems Implementation. Studies in Systems, Decision and Control, vol. 105. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-55595-9
14. Kharchenko, V., Illiashenko, O.: Concepts of green IT engineering: taxonomy, principles and implementation. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Concepts, Models, Complex Systems Architectures. Studies in Systems, Decision and Control, vol 74, pp. 3–19. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-44162-7_1
15. Basile, D., Di Giandomenico, F., Gnesi, S.: Model-based evaluation of energy saving systems. In: Kharchenko V., Kondratenko Y., Kacprzyk J. (eds.) Green IT Engineering: Concepts, Models, Complex Systems Architectures. Studies in Systems, Decision and Control, vol 74, pp. 187–208. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-44162-7_10
16. Palagin, A.V., Opanasenko, V.M., Kryvyi, S.I.: Resource and energy optimisation oriented development of FPGA-based adaptive logical networks for classification problem. In: Kharchenko V., Kondratenko Y., Kacprzyk J. (eds.) Green IT Engineering: Components, Networks and Systems Implementation. Studies in Systems, Decision and Control, vol. 105, pp. 195–218. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-55595-9_16
17. Drozd, J., Drozd, A., Maevsky, D., Shapa, L.: The levels of target resources development in computer systems. In: Proceedings of IEEE East-West Design & Test Symposium (EWDTS 2014), 26–29 Sept 2014, Kiev, Ukraine, pp. 1–5. https://doi.org/10.1109/EWDTS.2014.7027104
18. Huang, M., Li, X.: Optimal adaptive model for heating energy consumption forecasting. Revista de la Facultad de Ingeniería U.C.V. **32**(10), 252–260 (2017)
19. Kondratenko, Y., Korobko, O.V., Kozlov, O.V.: PLC-based systems for data acquisition and supervisory control of environment-friendly energy-saving technologies. In: Kharchenko V., Kondratenko Y., Kacprzyk J. (eds.), Green IT Engineering: Concepts, Models, Complex Systems Architectures. Studies in Systems, Decision and Control, vol 74, pp. 247–267. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-44162-7_13
20. Kondratenko, Y., Korobko, V., Korobko, O., Kondratenko, G., Kozlov, O.: Green-IT approach to design and optimization of thermoacoustic waste heat utilization plant based on soft computing. In: Kharchenko V., Kondratenko Y., Kacprzyk J. (eds.) Green IT Engineering: Components, Networks and Systems Implementation. Studies in Systems, Decision and Control, vol 105, pp. 287–311. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-55595-9_14

21. Fesenko, H., Kharchenko, V., Brezhnev, E., Zaitseva, E., Levashenko, V.: Markov model of smart grid digital substations availability: multi-level degradation and recovery of power resources issues. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Components, Networks and Systems Implementation. Studies in Systems, Decision and Control, vol 105, pp. 339–355. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-55595-9_16
22. Bodyanskiy, Y., Vynokurova, O., Pliss, I., Peleshko, D.: Hybrid adaptive systems of computational intelligence and their on-line learning for green IT in energy management tasks. In: Kharchenko V., Kondratenko Y., Kacprzyk J. (eds.) Green IT Engineering: Concepts, Models, Complex Systems Architectures. Studies in Systems, Decision and Control, vol. 74, pp. 229–244. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-44162-7_12
23. Kudritsky, V.D.: Filtering, Extrapolation and Recognition Realizations of Random Functions. FADA Ltd., Kyiv (2001). (in Russian)
24. Belan, V.V., Osadchiy, S.I.: Using canonical decomposition of spectral matrices to factor them. J. Autom. Inf. Sci. **27**(2), 57–62 (1995)
25. Atamanyuk, I., Kondratenko, Y.: Calculation method for a computer's diagnostics of cardiovascular diseases based on canonical decompositions of random sequences. In: Batsakis, S., et al. (eds.) ICT in Education, Research and Industrial Applications: Integration, Harmonization and Knowledge Transfer. Proceedings of the 11th International Conference ICTERI-2015, CEUR-WS, vol. 1356, pp. 108–120, Lviv, Ukraine, May 14–16 (2015)
26. Atamanyuk, I.P., Kondratenko, V.Y., Kozlov, O.V., Kondratenko, Y.P.: The algorithm of optimal polynomial extrapolation of random processes. In: Engemann, K.J., Gil-Lafuente, A. M., Merigo, J.L. (eds.) Modeling and Simulation in Engineering, Economics and Management. International Conference MS 2012, New Rochelle, NY, USA, 30 May–1 June 2012, Proceedings. Lecture Notes in Business Information Processing, vol. 115, pp. 78–87. Springer (2012). https://doi.org/10.1007/978-3-642-30433-0_9
27. Atamanyuk, I., Kondratenko, Y., Shebanin, V., Mirgorod, V.: Method of polynomial predictive control of fail-safe operation of technical systems. In: Proceedings of XIIIth International Conference "The Experience of Designing and Application of CAD Systems in Microelectronics", CADSM 2015, 19–23 Feb 2015, Polyana-Svalyava, Ukraine, pp. 248-251. https://doi.org/10.1109/CADSM.2015.7230848
28. Atamanyuk, I.P.: Algorithm of extrapolation of a nonlinear random process on the basis of its canonical decomposition. J. Kibernetika i Sistemnyj Analiz, №2, 131–138 (2005)
29. Atamanyuk, I., Kondratenko, Y.: Computer's analysis method and reliability assessment of fault-tolerance operation of information systems. In: Batsakis, S., et al. (eds.) ICT in Education, Research and Industrial Applications: Integration, Harmonization and Knowledge Transfer. Proceedings of the 11th International Conference ICTERI-2015, CEUR-WS, 14–16 May, vol. 1356, pp. 507–522, Lviv, Ukraine (2015)
30. Atamanyuk, I.P.: Optimal polynomial extrapolation of realization of a random process with a filtration of measurement errors. J. Autom. Inf. Sci. **41**(8), 38–48 (2009)
31. Atamanyuk, I., Kondratenko, Y., Sirenko, N.: Management system for agricultural enterprise on the basis of its economic state forecasting, complex systems: solutions and challenges in economics, management and engineering. Berger-Vachon, C., et al. (eds.) Studies in Systems, Decision and Control, vol. 125, pp. 453–470. Springer International Publishing, Berlin (2018). https://doi.org/10.1007/978-3-319-69989-9_27

On Quantitative Assessment of Reliability and Energy Consumption Indicators in Railway Systems

Davide Basile, Felicita Di Giandomenico and Stefania Gnesi

Abstract Stochastic model-based approaches are widely used for obtaining quantitative non-functional indicators of the analysed systems, as for example reliability, performance and energy consumption. However, a critical issue with models is their validation, in order to justifiably put reliance on the analysis results they provide. In this paper, we address cross-validation on a case study from the railway domain, by modelling and evaluating it with different formalisms and tools. Stochastic Activity Networks models and Stochastic Hybrid Automata models of rail road switch heaters, developed for the purpose of evaluating energy consumption and reliability indicators, will be evaluated with Mobius and Uppaal SMC. We will compare the obtained results, to improve their trustworthiness and to provide insights on the design and analysis of energy-saving cyber-physical systems.

Keywords Energy-saving · Reliability · Quality models · Stochastic analysis

1 Introduction

Recently, studies dedicated to reduce the energy consumption in cyber-physical systems (CPS) [1] are gaining increasing attention [2–7], aiming at saving in economic terms and reducing environmental impact. In CPS, digital control units interact with phenomena belonging to the system itself or the surrounding environment, whose nature is typically continuous (e.g., failure events, power energy

D. Basile (✉) · F. Di Giandomenico · S. Gnesi
Institute of Information Science and Technologies (ISTI), National Research Council (CNR), Pisa, Italy
e-mail: davide.basile@isti.cnr.it

F. Di Giandomenico
e-mail: felicita.digiandomenico@isti.cnr.it

S. Gnesi
e-mail: stefania.gnesi@isti.cnr.it

© Springer Nature Switzerland AG 2019
V. Kharchenko et al. (eds.), *Green IT Engineering: Social, Business and Industrial Applications*, Studies in Systems, Decision and Control 171,
https://doi.org/10.1007/978-3-030-00253-4_18

flow). Examples of CPS can be found in disparate application domains, including the transportation sector. The continuous dynamic nature of CPS is difficult to be expressed through discrete approaches, with proper control on the required approximations. Extensions of finite state automata [8–10], (extensions of) Petri Nets [11–13], have been adopted as formalisms for modelling them, where the evolution of the continuous variables can be described uniformly or by ordinary differential equations.

Dependability evaluations and formal verification are two separate research fields that in the last decade have been integrated to analyse critical CPS, where (i) measures as dependability, performance, reliability, energy consumption are formally assessed through rigorous approaches, and (ii) the correctness of the models is ascertained by proving that relevant properties hold in a suitable abstraction of the analysed system. These two aspects are complementary, and to cope with potential defects introduced in the modelling phase, validation of the developed models is paramount and highly recommended. Indeed, the introduced errors might compromise the accuracy of the results obtained through the analysis, which may lead to the delivery of flawed components, with both potential serious consequences for the components users and loss of time and money for industries, to recover from the late revealed deficiencies.

In this paper, we compare these two approaches for the analysis of CPS with reference to a case study from the railway domain. The aim of the investigation is to (i) emphasise their differences and provide considerations for modelling and evaluating CPS, and (ii) to verify that the results obtained through two separate formalisations of the case study are indeed in accordance.

Specifically, we consider rail road switch heaters, which are essential components for the correct functioning of railway stations, in absence of which possible disasters can occur (i.e. derailments, trains collision). In particularly cold regions, ice and snow can prevent the switches to work properly, hence heaters are used for guaranteeing their correct functioning. In particular, a central control unit is in charge of managing policies of energy consumption while satisfying reliability constraints, by communicating with the network of switches to manage the energy supply.

We will exploit two different methodologies to model and evaluate the system under analysis, chosen among popular ones in the research communities interested in verification and evaluation of complex critical systems. The first is based on the adoption of *Stochastic Activity Networks* (SAN) [13] to model rail road switch heaters for assessing the energy consumption and the probability of failure of these devices [14–16]. The quantitative properties (or measures of interest) are defined using Markov Reward Models [17] and evaluated in Möbius [18] through simulation.

The second adopted methodology is based on *Statistical Model Checking* (SMC) [19, 20]. The system is modelled as a network of *Stochastic Hybrid Automata* (SHA) [9, 10, 21] to uniformly deal with both continuous, discrete and stochastic aspects. The quantitative properties are defined in the Metric Interval Temporal Logic (MITL) [21] and evaluated with Statistical Model Checking [22].

The Uppaal SMC toolkit [21] is used for evaluating these measures. Moreover, the absence of deadlocks has been proved through standard model checking techniques and through refinement [23]. Those models are here extended to deal with different classes of priorities in the network of switches.

The main contributions of this paper are:

- modelling frameworks of rail road switch heaters through both SAN and SHA formalisms, in order to compare different strategies of energy consumption, considering failure events and different weather profiles;
- identification of indicators representative of energy consumption and reliability of the analysed system and their quantification through the developed frameworks in a variety of scenarios. The obtained results allow to gain insights on suitable trade-offs between energy consumption and reliability of the analysed system, to properly tune the parameters of the considered energy consumption strategies;
- comparing the results obtained through the two separate formalisations provides a further guarantee on the reliability of the obtained results.

Structure of the paper The paper is structured as follows. We will introduce dependability analysis and the two methodologies that we will adopt in Sect. 2. In Sect. 3 we will describe the system that we intend to model and analyse, with the goal of reducing the energy consumption and augmenting the reliability of the system. Generic guidelines for modelling energy-saving CPS are discussed in Sect. 4. The two different formalisations of the system are in Sect. 5, while the evaluation and validation of these models are in Sect. 6. Finally, related work and conclusions are respectively in Sects. 7 and 8.

2 Dependability Analysis

Dependability analysis of complex, critical systems is the topic of many research studies, since assurance on proper operation is typically among the requirements of the applications such systems are employed in. Depending on the specific aspects of interest in the dependability field, different approaches have been developed to accomplish analyses. The most active communities in dependability analyses have been and are: (i) the one focusing on fault tolerance and dependability, and (ii) the one focusing on formal methods. Although moving from rather different perspectives, both are progressively converging towards solution of common problems and in the last decade they influenced each other [24].

In particular, due to the emerging complexity and dependencies shown by modern systems, like many critical CPS around our everyday life, there is a need for expressing more complex system behaviours and measures of interest. This motivated growing relevance of quantitative measures in the field of verification, while in the past qualitative aspects were mainly tackled. Quantitative assessment of non-functional properties, especially dependability and performance related ones, consists in probabilistically estimate the occurrence of faults and their impact on the ability of the system to operate correctly.

To provide the context necessary to better understand the motivation of our study, in this section we briefly overview two main approaches to quantitative dependability analyses developed by the traditional dependability and the formal methods communities. Namely, quantitative model-based analysis and Statistical Model Checking are introduced in the following, together with the formalisms and tools that we will adopt in this paper.

2.1 Model-Based Dependability and Performance Analysis

Several approaches are available in the literature to perform assessment of dependable systems [25], mainly testing, fault injection and model-based evaluation. A wide range of modelling techniques has been developed for both dependability and security analysis, each focusing on particular levels of abstraction and/or system characteristics, as surveyed in [26].

For quantitative evaluation of dependability indicators, stochastic model-based analysis [27] has been proven to be particularly useful, versatile and cost-effective for manufacturers [28, 29]. The system under analysis is often described through stochastic processes, whilst measures of interest are generally obtained through mathematical analysis (closed-form expressions), numerical evaluation (linear programming techniques) and discrete-event simulation (statistical methods). This analysis approach is useful for expressing the stochastic nature of physical phenomena involved in CPS [1].

To keep the model manageable, the system needs to be represented at a properly identified abstraction level. Indeed, depending on the properties to be analysed the emphasis on the system representation is focused on those aspects that are relevant for analysis purposes, while irrelevant aspects are neglected. Therefore, a wide variety of models are used in practice, to tailor the right abstraction level for the system under analysis, in accordance with the properties to be assessed, the desired degree of accuracy and available resources to manage model development and solution. Formalisms such as (extensions of) Petri Nets [11–13] and (Non) Markov based models [13, 30] are used for modelling and evaluating CPS, where reward structures [31] are defined in order to evaluate measures of interest (e.g. reliability, performance, energy consumption) at the variation of relevant parameters, either analytically or through simulation. Stochastic Activity Networks [13] and Möbius [18] which we adopted in our study, are well-established formalism and tool, respectively, for modelling and evaluating these systems.

Stochastic Activity Network Stochastic Activity Networks (SAN) [13] is a formalism widely used for performance, dependability and performability evaluation of complex systems, given its high expressiveness and the powerful tools for modelling and evaluating them [18]. The SAN formalism is a variant of Stochastic

Petri Nets [30], and has similarities with Generalised Stochastic Petri Nets [11]. A SAN is composed of the following primitives: *places*, *activities*, *input gates* and *output gates*. Places and activities have the same interpretation as places and transitions of Petri Nets. Input gates control the enabling conditions of an activity and define the change of marking when an activity completes. Output gates define the change of marking upon completion of the activity. Activities are of two types: *instantaneous* and *timed*. Instantaneous activities complete once the enabling conditions are satisfied. Timed activities take an amount of time to complete following a temporal stochastic distribution function, which can be for example exponential or deterministic. When an activity completes, the following steps are executed: (1) one of the cases of the activity is chosen according to its marking depending probability; (2) the function of each input gate of the activity is executed; (3) the function of each output gate linked to the case selected at first step is executed. An enabled activity is aborted, i.e. it cannot complete, when the SAN moves into a new marking in which the enabling conditions of the activity no longer hold. Cases are associated to activities, and are used to represent probabilistic uncertainty about the action taken upon completion of the activity.

The primitives of the SAN models are defined using C++ code.

Möbius Möbius [18] is a multi-formalism multi-solver tool that can be used for defining and solving SAN models. Möbius supports various formalisms and different analytical and simulative solvers, and can be used for studying the reliability, availability, and performability of systems. It follows a modular modelling approach, where atomic models are building blocks that can be composed with proper operators *Rep* and *Join* to generate a composed model.

Notably, atomic models specified in different formalisms can be composed in this way. This allows to specify different aspects of a system under evaluation in the most suitable formalism. Along with an atomic or composed model, the user specifies a reward model, which defines a reward structure on the overall model. Rewards are the vehicle to define the measures for our case study. On top of a reward model, the tool provides support to define experiment series, called Studies, in which the user defines the set of input parameters for which the composed model should be evaluated. Each combination of input parameters defines a so-called experiment. Before analysing the model experiments, a solution method has to be selected: Möbius offers a powerful (distributed) discrete-event simulator, and, for Markovian models, explicit state-space generators and numerical solution algorithms. It is possible to analyze transient and steady-state reward models. The solver solves each experiment with parameters instantiation specified in the Study. Results can be managed by means of a database.

2.2 Formal Methods and Statistical Model Checking

Recent developments in probabilistic analysis using formal methods have improved the accuracy and reliability of dependability analysis, which was traditionally performed through proof methods not fully automatized and computer simulations [32]. In the literature, several approaches for the verification and validation of stochastic models have been proposed, as for example testing, theorem proving, model checking. In particular, model checking [33] is a widely-used and powerful approach for the verification of finite state systems, where a property φ, usually specified in a temporal logic, is automatically checked against a model of a system M, by performing an exhaustive exploration of the state-space of M, i.e. $M| = \varphi$, obtaining a counter-example in case $M| = \varphi$ does not hold.

The recently introduced Statistical Model Checking [19, 20] uses results from statistics on top of simulations of a system to decide whether a given property specified in a temporal logic is satisfied with some degree of confidence, and it represents a valid alternative to classical model checking and dependability evaluation, especially in the case of undecidability. PRISM [34] and Uppaal SMC [21] have been proposed as tools that implement the above techniques. An advantage is that quantitative properties are uniformly described through temporal logics, and hence have a well-defined semantics. Moreover, it is possible to assess qualitative as well as quantitative properties to evaluate and validate the proposed models.

Stochastic Hybrid Automata, analysed with Uppaal SMC is the specific model checking approach adopted in this paper to address the modelling and analysis of our case study.

Hybrid Automata Stochastic Hybrid Automata (SHA) [9, 10, 21] are a suitable formalism for describing cyber-physical systems, where both discrete and continuous dynamics and stochastic phenomena are involved. Timed automata [8] combine discrete systems with real-valued variables that evolve during the time a system spends in a state. These variables, called *clocks*, evolve with a uniform rate and they can be used for guarding transitions. Reachability and other key problems are decidable for timed automata, with algorithms supporting them implemented in tools such as Uppaal [35].

Hybrid automata [9, 10] generalise timed automata by including arbitrary dynamics for the real-valued variables (i.e. clocks), expressed through ordinary differential equations (ODEs). Stochastic Hybrid Automata include also probabilistic transitions and are used in tools such as Uppaal SMC [21]. A number of case-studies demonstrate their applications [21, 36].

For simplifying the presentation, we slightly elaborate the formal definition of hybrid automata in [9, 10]. We start by introducing some useful notation. In a hybrid automaton, the states progress according to both continuous and discrete clocks; in the first case this behaviour is called *continuous flow*, while in the second *jump*. A flow function $R^{|X|} \rightarrow R^{|X|}$ characterizes the flow (i.e. the dynamic) of the continuous variables in the set X through a system of ODEs $X^{\cdot} = F(X)$, where $X^{\cdot}$ is the first order derivatives of the variables in X, and as usual R is the set of real numbers. Moreover, let $\nu: X \rightarrow R$ be a valuation of the variables in X, $\pi \in$ pred (X) be a predicate over X and $[[\pi]] \in R^{|X|}$ be the set of valuations of X that satisfies the predicate π. Predicates are used to (*i*) guard transitions, (*ii*) specify the jumps of a system (i.e. how variables evolve in a discrete-time step) and (*iii*) define the invariants for each state of the automaton. A hybrid automaton H is defined as a tuple $H = \langle Q, Q_0, \Sigma, X, T, I, F, V_0 \rangle$ where:

- Q is a finite set of *states* including a distinguished initial singleton set $Q_0 \subseteq Q$,
- Σ is a finite set of *actions*,
- X is a finite set of real-valued *variables*, called *clocks*,
- $T \subseteq Q \times \text{pred}(X) \times \Sigma \times \text{pred}(X \cup X^0) \times Q$ is the *transition relation*,
- $I{:}Q \to \text{pred}(X)$ that assigns an invariant function to each state,
- $F{:}Q \to \left(R^{|X|} \to R^{|X|}\right)$ that assigns a flow function to each state $q \in Q$ as the set of ODEs $X = F(q)(X)$, and
- $V_0 \in \text{pred}(X)$ is the set of initial valuations.

It is assumed that for each state $q \in Q$ the flow function $F(q)$ has unique solution. We now briefly describe the semantics of hybrid automata. A configuration of a hybrid automaton is a tuple (q, ν) where $q \in Q$ is a state and $\nu \in \mathrm{R}^{|X|}$ is a variable valuation.

The initial configuration of a hybrid automaton is (q_0, ν_0), where $q_0 \in Q_0$, $\nu_0 = [[\pi]]$ such that $\pi \in V_0$ and $\nu_0 \in [[I(q_0)]]$ (the invariant constraints are satisfied). During the time t a system spends in a state q, the clocks in X are updated according to the flow function of q, and at each step the new valuation must respect the invariant constraints in q. A transition $\delta = (q, g, a, j, q_1)$ is enabled after t time when the guard $g \in \text{pred}(X)$ is satisfied. When δ is executed, the automaton jumps to a new configuration (q_1, ν_1) such that q_1 is the target state of δ, ν_1 is the valuation of the jump constraints $j \in \text{pred}(X \cup X^0)$, and $\nu_1 \in [[I(q_1)]]$.

Composing Hybrid Automata For modelling complex hybrid systems it is convenient to adopt a modular approach where systems are described by interacting entities. This allows to separately verify different smaller components more efficiently than verifying a bigger monolithic model. Hybrid automata can be composed through a synchronous product operator, and they interact through actions and shared variables. Let $I = \{1, \ldots, n\}$ be a set of indexes, the product of hybrid automata is denoted as $\otimes H_i \in C\ H_i$, where $C = \{H_i \mid i \in I\}$. The states of the product are composed by the product of the states of its components. Similarly, the alphabet and the variables are the union of those of its components. The invariants, flow function and initial valuations are defined homomorphically on their elements. Finally, the transitions are synchronous, i.e. all the components (satisfying the constraints on the corresponding transition) synchronise on an action a while the others stay idle (in the following we will also distinguish between input and output actions through broadcast channels).

Uppaal SMC Uppaal is a toolbox that has been adopted for verifying real-time systems, represented by (extended) timed automata, that interact through broadcast channels and shared variables. Uppaal SMC [21] is an extension of Uppaal that allows to express both stochastic and non-linear dynamic features, by adopting a stochastic extension of hybrid automata. The stochastic interpretation replaces the non-deterministic choices for multiple enabled transitions and time delays with, respectively, probabilistic choices and probability distributions (uniform for bounded time and exponential for unbounded time). By composing different automata through the product of SHA, arbitrary complex behaviours can be obtained, where it is possible to statically or dynamically generate new instances of automata, that are uniquely identified.

Uppaal SMC uses Statistical Model Checking to evaluate probabilistic properties of interest. SMC uses results from statistic area to decide, based on a given number of monitored simulations, whether the system under analysis satisfies the property of interest within a given degree of confidence. An advantage of SMC is that it avoids the exploration of the whole state-space of a model, which is a main drawback of standard model checking techniques.

Temporal Logic formulae. In addition to standard model checking techniques of properties as reachability, deadlock-freedom, in Uppaal SMC it is possible to evaluate the probability that a random run of a network M satisfies a property φ in a given amount of time t. Properties are defined using the Metric Interval Temporal Logic (MITL) [21]. A MITL formula φ is inductively defined by the following grammar:

$$\varphi ::= \mathrm{ap} | \neg\varphi | \varphi_1 \wedge \varphi_2 | \bigcirc\varphi | \varphi_1 \cup_{x \le t} \varphi_2$$

In the definition above, ap are atomic predicates over states of an automaton, and the logical operators are standard, except for $\varphi_1 \cup_{x \le t} \varphi_2$ that checks whether a formula φ_1 is satisfied in a run *until* a formula φ_2 is satisfied, and this must happen before the clock x exceeds the value t. As usual, it is possible to derive the operators *exists* and *forall* as $\Diamond_{x \le t}\ \varphi = \mathrm{true}\ \cup_{x \le t}\ \varphi$ and $\Box_{x \le t}\varphi = \neg\Diamond_{x \le t}\neg\varphi$, where both quantifiers are bounded by the time t for the clock x.

Generally, checking if a model M satisfies a property $P_M(\Diamond_{x \le t}\varphi) \ge p, p \in [0, 1]$ is undecidable [37]. Statistical algorithms are developed in Uppaal SMC for estimating the probability of cost-bounded reachability problems in a given interval of confidence. There are three types of queries: $P_M(\Diamond_{x \le t}\mathrm{ap})$ (probability estimation), $P_M(\Diamond_{x \le t}\mathrm{ap}) \ge p$, $p \in [0, 1]$ (hypothesis testing), $P_M(\Diamond_{x1 \le t1}\mathrm{ap}_1) \ge P_M(\Diamond_{x2 \le t2}\mathrm{ap}_2)$ (probability comparison). In Sect. 6 we will evaluate the measures of interest for the energy consumption and the probability of failure in Uppaal SMC through probability estimation, while in Sect. 7 we will compare the approaches based on Möbius and Uppaal SMC. In the next section, we will describe the case study.

3 Description of the Case Study

We propose a cyber-physical system from the railway domain as case study: a rail road switch heating system. In this section, we briefly describe the real-world devices that we want to model and the underlying logical system we have built for analysis purposes.

3.1 Description of the Network of Rail Road Switch Heaters

We consider a heating system composed of a series of tubular flat heaters along the rail road track, which warm up the rail road by induction heating. Sensors are used to communicate the temperatures of the air and of the rail road to the rail road switch heating system [38], to perform decisions. The central unit manages the maximum amount of power that can be delivered to the system, in order to prevent possible blackouts. In case of extremely cold conditions, the total amount of energy available may not be sufficient to heat the overall system of switches, hence it is important to duly choose the heaters that must be primarily turned on and those that may be later on. Indeed, in a railway station there are tracks which are less important than others, for example the side tracks, and the heating phase can be delayed for them if necessary. If the temperature cannot be kept above the freezing thresholds, the corresponding switch will experience a failure.

3.2 Logical Structure of the System

The two main logical components of our system are the *heater* and the *central coordinator*. The network of heaters is realised by replicating the heater component, and their activation/deactivation is controlled by the central coordinator. In the following we discuss the two main components.

Heater We based the policy employed to activate/deactivate the heating on two threshold temperatures:

– *warning threshold* (T_{wa}): this temperature represents the lowest temperature that the track should not exceed. If the temperature is lowest than T_{wa}, then the risk of ice or snow can lead to a failure of the rail road switch and therefore the heating system needs to be activated;
– *working threshold* (T_{wo}): this is the working temperature of the heating system. Once this temperature is reached, the heating system can be safely turned off in order to avoid an excessive waste of energy.

The energy consumption of the overall system depends on the value of T_{wa} and T_{wo}. A smaller gap between these thresholds will result in a frequent activation of the heating system, but for a shorter period of time. Alternatively, a wider gap between the thresholds will result in a less frequent activation, but it will be for longer periods of time. The time during which a single heater is active depends also on the weather conditions.

Coordinator The coordinator will collect the requests of activation from the pending heaters, and it will manage the energy supply according to a FIFO prioritised order. Indeed, the first heater that asks to be turned on will be the first to be activated. We will assign priorities to switches based on their criticality on the track and we will exploit the assigned priority in the performed analysis, so to guarantee higher reliability to those switches that are essential for the correct functioning of the overall station. The maximum amount of energy deliverable by the system that cannot be exceeded is represented by $NH_{\max}$, and it is measured as the percentage of heaters that can be turned on at the same time. If there is no energy available, each request will be enqueued in the queue of pending heaters. Below we describe in details the behaviour of the central coordinator.

Interactions We now give details about the protocol of communication between the network of heaters and the central coordinator:

- *Heater*: at starting time, each heater h_i is switched off and its internal temperature is set to T_{wo}. Once its internal temperature goes below T_{wa}, h_i asks the coordinator to be turned on and waits. Upon reception of the notification, h_i is turned on. After that, two events can happen:
  - the heater h_i reaches an internal temperature above T_{wo}, communicates to the central coordinator the termination of the heating phase and is switched off;
  - a second component h_j with a higher priority asks to be turned on. The energy delivered to h_i is turned off, even though it has not yet reached an internal temperature above T_{wo}. If the temperature is below T_{wa}, h_i will issue a new request of activation to the coordinator.
- *Coordinator*: at starting time the central coordinator is waiting for a message from one of the heaters h_i in the network. Two messages can be received:
  - h_i asks to be activated. This request is inserted in the queue of pending requests in case there is no energy available and the priority of h_i is not higher than that of the already activated switches. Otherwise, the request is accepted and we have two cases; (i) if there is energy available, h_i will be activated by issuing a notification; (ii) if no energy is available but h_i has a priority higher than one of the activated heaters, firstly the heater with lowest priority will be turned off with a notification, and then the activation is notified to h_i;
  - h_i asks to be deactivated. After the deactivation, if there are no heaters that are waiting for being activated then no action is performed. Otherwise, one of the pending heaters h_j (the first in the prioritized queue of pending heaters) is activated by issuing a notification to it.

Before presenting the different models of the system described above, in the next section we will discuss the generic guidelines for modelling energy-saving cyber-physical systems, that we have followed for our case study.

4 General Guidelines to Model Energy-Saving Cyber-Physical Systems

Guidelines on the analysis of reliability and energy consumption indicators of CPS systems are now discussed. Following them, in Sect. 5 the two analyses approaches are concretely applied to a case study representative of an energy-saving cyber-physical system, to emphasise pros and cons of the two alternatives and to compare the obtained results for cross-validation purposes.

Generally, in energy-saving CPS [1] the supervision of the cyber-control is in charge of strategies to supply energy to components of the physical system, necessary to keep them effective and reliable in the service they accomplish. Our interest is in assessing measures that are representative of the energy consumption, to be combined with other dependability-related properties dictated by the critical domain the CPS is employed in. It is then possible to study the benefits of different energy supply strategies to properly tune the parameters of these strategies toward most rewarding configurations. A diagram of the framework is depicted in Fig. 1. The proposed analysis framework is built around three major modules [2].

– *Physical-aspects module*: this module focuses on the physical components of the system and on their characterisation in terms of relevant aspects from the energy viewpoint. It includes models representing phenomena related with energy supply, which depends on the fabric of the supplied components and environmental conditions impacting on the energy consumption. Examples are: (i) internal and external temperatures, properly modelled taking into account their evolution in time, given the different means involved (such as iron or copper for the physical components, winter or summer days for the external air); (ii) fuel consumption, represented by properly considering the engine parameters, aerodynamic drag, weight, and other relevant parameters; (iii) supplied power, represented by properly considering the laws regulating the involved real process; (iv) battery charging and others.
– *Control-aspects module*: this module deals with the policies that dynamically regulate the energy consumption of the physical components. To manage potential complexity while assuring adequate accuracy of the analysis, the

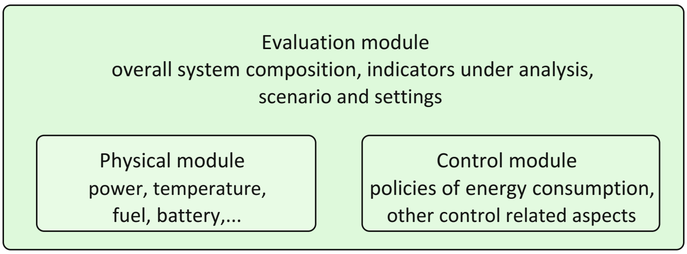


Fig. 1 The proposed analysis framework for energy saving CPS

representation of such policies is abstracted at the level of their impact on the energy parameters of the controlled physical components. As already mentioned, the primary objective of the proposed model-based approach is to assist the system designer in identifying the best policy to employ among several alternatives, in accordance with pre-established dependability and cost requirements. Trade-offs between energy consumption and dependability requirements are mandatory in critical domains, where, e.g., the energy should not be reduced in safety critical situations. For example, the dynamic power management can be of kind on-off, where energy is supplied or turned off on the basis of values assumed by parameters that depend on physical conditions of the system and of the environment.

– *Evaluation module*: Finally, the third module deals with the composition of the several models from the previous two modules, to end up with the overall evaluation framework. Exercising the overall composed model, energy supply policies trading energy saving and dependability properties can be quantitatively evaluated and compared in terms of properly defined indicators.

The generality of the above outlined approach allows assessing a variety of measures of interest to final customers, service providers and operators, in accordance with the specific application domain where the CPS system under analysis is utilised. Given the aim of trading energy consumption with dependability, typical indicators are energy supplied to individual system components or to the overall system in a certain time interval, as well as failure probability of an individual component or of the overall system due to lack of supplied energy.

A benefit of the proposed approach in Fig. 1 is its modularity and compositionality, that allows to be applied to a wide variety of scenarios relevant to CPS, and it has been adopted for our case study. In Sects. 5 and 6, an instantiation of the discussed guidelines is presented, following two different approaches based on SAN and SHA models.

5 Description of the Models

In this section, we will describe the models of the rail road switch heating system. We will follow the general guidelines discussed in Sect. 4. Firstly, we will discuss the physical-aspects module of the system under analysis, then we will discuss the control aspects module, that will be modelled both with SAN models and SHA models. Finally, the evaluation module will be discussed in Sect. 6.

The models are parameterized based on the two temperature thresholds T_{wa}, T_{wo} and $NH_{\max}$ that we recall to be the maximum power that the system can provide at every instant of time, expressed in percentage of heaters that can be turned on at the same time.

5.1 Physical-Aspects Module

The continuous physical behaviour concerning the increment and decrement of the temperature of the rail road track, respectively when the heater is turned on or off, is modelled by an ODE representing the balance of energy [14].

The heater is represented by a resistance that passes through the rail road in different points in order to warm up the iron. The set-up for the heating device is based on patents of heating switches [39], which contain data about the power consumed by a single heater and about the increment of the temperature of the track in cold winter nights. We assume that the power used by the heater is constant, in order to estimate the kilowatt per hours consumed during the time interval that we consider.

Assuming that the values of the temperature of the surrounding area T_e and the previous internal temperature T are known, the internal temperature T after time t is (we adopt the Newton's notation for differentiation) $T' = (-uA(T - T_e) + Q)/mc$, where u is the coefficient of convective exchange; c, the heat capacity of iron; A, the surface area exposed to the external temperature; m, the mass of the iron bar; Q, the power used when the heater is turned on, if the heater is turned off this value will be zero. We now discuss the stochastic aspects that have been considered as part of the physical module.

Stochastic aspects Stochastic aspects concern the possibility of experiencing a failure in a switch, that will be used for measuring the reliability of the system, and the influence of the weather forecast.

Switch Failures When the temperature of the rail road track is below the freezing threshold (i.e. 0 °C in our experiments), a switch may experience a failure. In this case, the time-to-failure is modelled with an exponential distribution with fixed rate, that will be based on the temperature of the rail road track.

Weather forecast To model the external weather conditions, our model takes in input data structures containing profiles of average temperatures in those days for

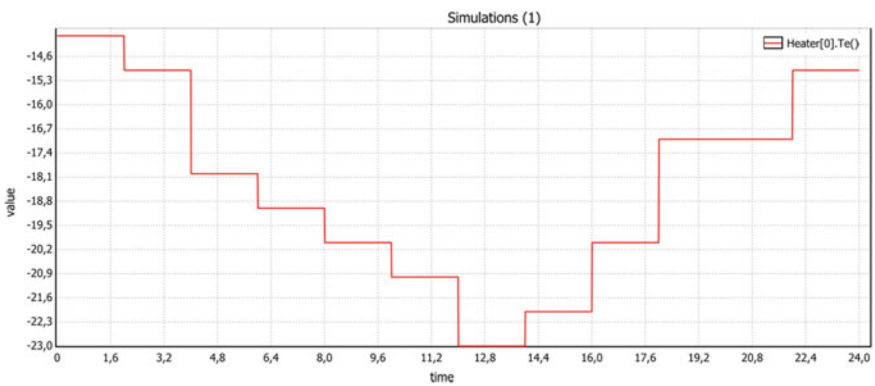


Fig. 2 A weather profile corresponding to the temperatures for coldest winter nights in a northern city, retrieved from [40]. The simulation starts at 6:pm and terminates after 24 h

which the analysis is relevant (e.g. winter days), as depicted in Fig. 2. For our experiments, we have five different daily weather profiles retrieved from the internet [40]. The time window under analysis is divided into intervals to which an average reference temperature is assigned. The current instance of the model concentrates on a whole day, divided into intervals of 2 h. However, the model can be easily modified to consider longer (or shorter) periods, as well as different number of intervals. A probability is assigned to each weather profile, based on weather statistics.

5.2 Control-Aspects Module

The control-aspects module has been modelled both in SAN and SHA models. Comparisons between these two different approaches will be derived in Sect. 7.

SAN Models We now describe the SAN based model realising the rail road switch heating system logic and the protocol described in Sect. 3, built through the functionalities provided by Möbius.

The overall model is obtained by the composition of the atomic models, using the join and rep operators, as shown in Fig. 3 (where the atomic models are the leaves of the tree while the overall composed model is the root). The atomic model *Coordinator* represents the central coordinator and it will interact with the network of switch heaters. The atomic SAN models *LocalitySelector*, *ProfileSelector* and *SwitchIDSelector* represent, respectively, the selector for the weather profile, the location of the switch and the unique identifier of each switch. The submodel *HeaterModuleM* represents an instance of a single heater module, obtained by the composition, using the join operator, of the four atomic SAN models. Those atomic models share the places relative to the locality of the device, its weather profile and

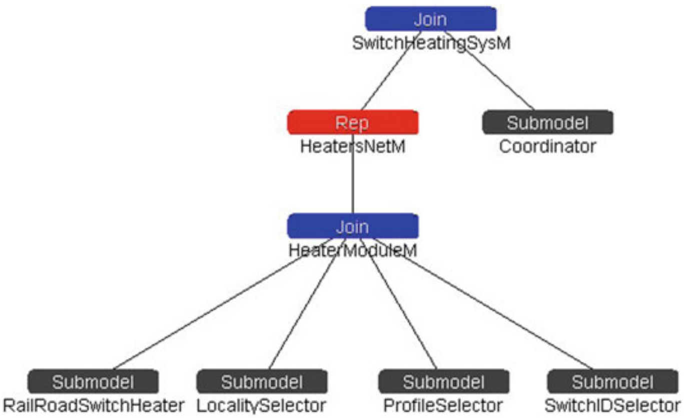


Fig. 3 The composed model

the unique ID. The submodel *HeatersNetM*, obtained by replicating *numRep* times the model *HeaterModuleM* using the Rep operator, represents the network of heaters, where the parameter *numRep* identifies the number of devices composing the network. Finally, the model *SwitchHeatingSysM*, obtained using the join operator, represents the overall system.

All these SAN models interact through *shared places*, a feature available in Möbius [18] for joining different SAN models thus allowing modularity.

Note that in Möbius different replicas of SAN models are anonymous, and it is not possible to distinguish between them. Hence, the *SwitchIDSelector* SAN model is used for assigning a unique ID to each heater module.

The main SAN model concerning the rail road switch heater is depicted in Fig. 4. It is partitioned into three logical components: the *init sub-net*, the *clock sub-net* and the *heater sub-net*. The *init* subnet initialises the data structures used by the SAN model. The clock sub-net models the evolution of time (during one day in our analyses), and it is used to update the environment temperature and the temperature of the rail road track. In this paper, we have considered as unit of time 1 h. The activity *clock* has a deterministic distribution of time (non-Markovian), and completes each hour. When *clock* completes, the place *Temperature* is updated: if the heater is turned on then the temperature increases, otherwise the temperature will be updated according to the temperature of the environment, as in the equation in Sect. 5.1.

The heater sub-net represents the status of the rail road switch heater. The heater can be activated (one token in the place *on*), waiting for being activated (one token in the place *ready*), turned off (one token in the place *off*), or failed (one token in the place *failure*). Indeed, according to the heating policy, once the system temperature falls below a pre-defined warning threshold (T_{wa}), the heating needs to be activated,

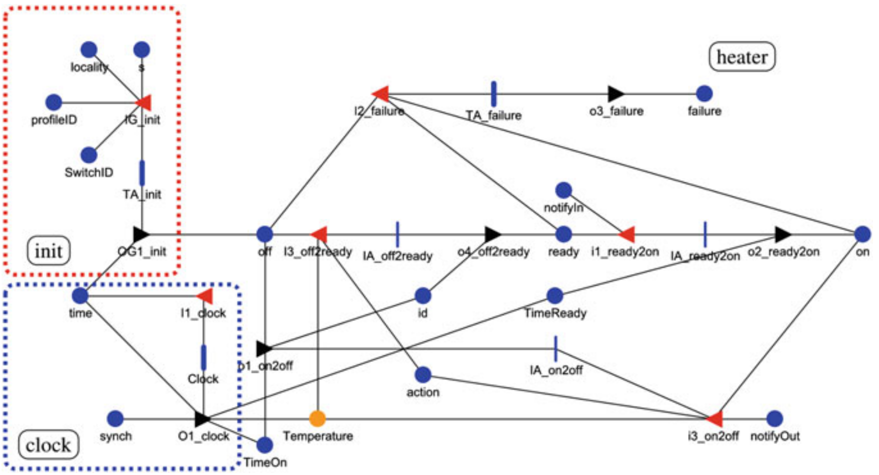


Fig. 4 The SAN model *RailRoadSwitchHeater*, logically divided into three sub-nets: the init sub-net, the clock sub-net and the heater sub-net

otherwise the associated switch fails. Then, once the temperature raises and reaches the working threshold (T_{wo}), the heating system can be safely turned off.

The heater sub-net interacts with the *Coordinator* SAN model through places shared among all the replicas of the heater model and the *Coordinator* model. For example, if the heater is in state ready, in order to be turned on, the input gate $i1_{ready2on}$ checks if the marking of the shared place *notifyIn* is equal to the marking of the place *SwitchID*, which means that the coordinator has notified the heater to be turned on.

The function representing the heating exchange is defined in C++, and it is called by the output gate $O1_{clock}$ in Fig. 4 to update the temperature of the rail road each interval of time *t*. The activity $TA_{failure}$ models the failure of a component. It has an exponential distribution of time based on the temperature of the rail road track: the more the temperature is below the freezing threshold the more is probable that the activity will fire (the activity is not activated if the temperature is positive).

The SAN model Coordinator represents the central management unit and it interacts with all the heaters in the network by activating, deactivating or moving them in a waiting state. Below we will describe the network of stochastic hybrid automaton used for modelling the system of rail road switch heaters.

SHA Models The rail road switch heating system has also been modelled through stochastic hybrid automata, with the purpose of cross-validating the two distinct formalisations and improve the trustworthiness of the obtained results.

Indeed, this formalism allows to capture both discrete, continuous and stochastic aspects in a single framework. We have been able to verify the correctness of the interactions, as well as energy and reliability indicators, by using the Uppaal SMC toolkit. We briefly outline the formalisation of the system of (remotely controlled) rail road switch heaters as a product of hybrid automata.

The ODE in Sect. 5.1 is expressed in the stochastic hybrid model H in Fig. 5, where the temperature *T* is a *continuous clock* and the flow function *F* (i.e. the ODE) is similar in different states. Indeed, when H is in state on, *F* is adding the term Q (i.e. the power); this is not the case in states off and ready.

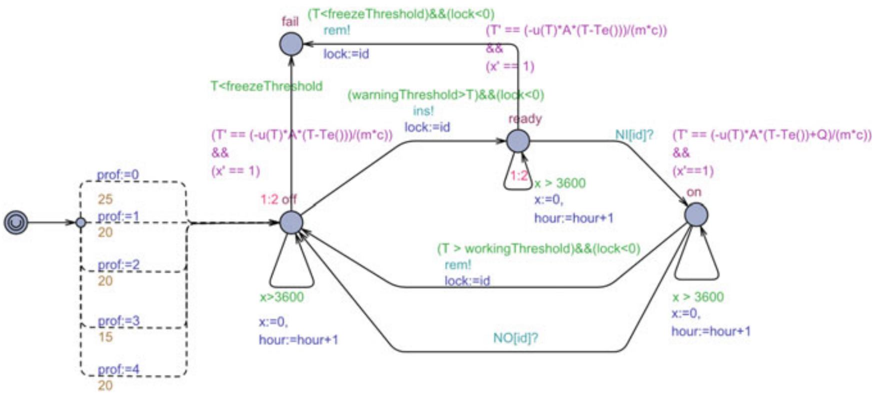


Fig. 5 The stochastic hybrid automaton H, modelling an instance of a rail road switch heater

The two main logical components describing the discrete cyber part of the analysed system are the *heater* H (depicted in Fig. 5) and the *central coordinator* K (depicted in Fig. 6). The network composed of n heaters and the coordinator is realised by making the product of K and the replicas of the stochastic hybrid automaton H_{id}, $id \in 1, \ldots, n$, where each heater is uniquely identified by its *id*, i.e. $(\otimes_{id \in 1, \ldots, n} H_{id}) \otimes K$. The heater model is in Fig. 5 and it implements the policy for activating and deactivating the heating phase, similarly to the SAN model in Fig. 4. In particular, the dotted transitions are urgent (i.e. instantaneous) probabilistic transitions used for selecting one of the available weather profiles. The main states are on, off, ready and fail, which correspond to the places of the SAN model *RailRoadSwitchHeater*. Here we note that each state has an inner cycle modelling the decrease and increase of the internal temperature according to the flow function, and that both the incoming transitions to state fail have an exponential distribution of time. During a simulation, the current time is stored in the clock x, and a variable hour stores the current hour.

The function Te(), used in the flow function of T, selects the actual external temperature based on the current hour, and it is implemented in Uppaal.

The coordinator is modelled as the hybrid automaton in Fig. 6. Its behaviour is similar to the one of the SAN model *Coordinator*. The queue of pending heaters is modelled with the array queue[] of length equal to NH_{max}, and the functions enqueue(id) and dequeue() are used for inserting and removing elements, while empty() returns true if the queue of pending heaters is empty.

The coordinator sends messages to the network of heaters through two arrays of channels NI[id] and NO[id] indexed by the identifiers of the heaters, to notify respectively the activation and deactivation of a heater (see Sect. 3 for the communication protocol). Note that Uppaal SMC only allows broadcast channels, hence an array of channels has been adopted in order to implement one-to-one communications. Following the standard notation, sending through a channel a is denoted

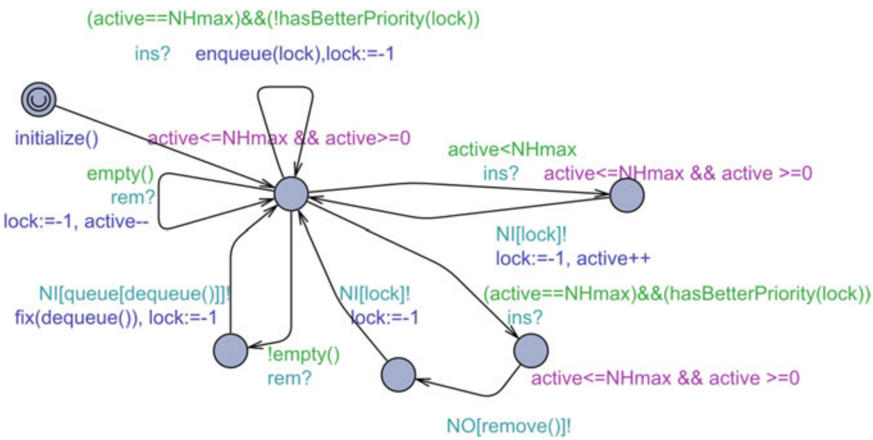


Fig. 6 The stochastic hybrid automaton K, modelling the coordinator

as a!, while reading as a?. Upon reception of the notification NI[id]?, the heater with identifier id switches from state ready to state on.

The heaters communicate to the coordinator their transition from off to ready through the channel ins, so asking for being activated, and their transition from on to off through the channel rem; both channels are many-to-one. All channels are *urgent*, which means that no delays will occur in case a synchronisation is available.

While the coordinator is in a busy state, a shared variable lock is used as a semaphore to prevent a heater from sending messages that cannot be elaborated, and it is used by the heaters for communicating their identifiers to the coordinator. In the next section, we will describe the evaluation module for both SAN and SHA models.

6 Evaluation Module

In the evaluation module, we will evaluate the energy consumption and the reliability of the system considering different thresholds-based policies of energy consumption. Concerning the scenarios and settings we have considered for our analysis, real world data are used. In particular, to keep the presentation simple, the layout of an average small-size railway station has been chosen. The cold winter days are considered for the temperatures and the parameters of the physical model are based on data available from real devices [38, 39].

6.1 Measures of Interest

We consider two different measures of interest. The first concerns the energy consumption while the second addresses the reliability of the system under analysis.

1. $CE(t, l)$: the time (in hours) a generic heater is activated in the time interval $[t, t + l]$. By multiplying $CE(t, l)$ for the power consumed (kilowatt per hour), that is the term Q in the heat exchange equation (see Sect. 5.1), it is possible to derive the energy consumed by the system;
2 $PFAIL(t, l)$: the probability that a generic switch fails (becomes frozen) at time $t + l$, given that at time t is not failed.

We remark that reliability is computed as the probability that no failure occurs in the interval of time under analysis [41], that is $1 - PFAIL(t, l)$.

Measures of Interest in the SAN models Concerning the SAN models, in Möbius reward structures have been used for evaluating the measures of interest. In particular, $CE(t, l)$ is defined by accumulating in the interval $[t, t + l]$ the time that each replica of the SAN model *RailRoadSwitchHeater* spends in markings with one token in place *on* (see Fig. 4), that is the time that each heater is activated. The measure *PFAIL*(t, l) is defined as the probability that at time $t + l$ there is one token in the place *failure* of the SAN model *RailRoadSwitchHeater*.

Measures of Interest in the SHA models In Uppaal SMC, the measures of interest are defined as formulae in MITL, enriched with quantification operators on the replicated models and expected values. We will consider a discrete clock energy that counts the hours H spends in state on. For enhancing readability, we have omitted this clock in Fig. 5. For the energy consumption, we estimate the number of hours in which the heaters are active as:

$$CE(t, l) = \mathrm{E}[< \, =24; 10{,}000]\left(\max : \sum_{\mathrm{i:idt}} \mathrm{H_i energy}\right)$$

where E stands for the expected value, 24 is the considered interval of time (24 h) and 10,000 are the simulations executed by the tool. The overall energy consumption is the sum for all H_i of all the clocks energy.

The probability of failure is estimated by Uppaal SMC with the formula:

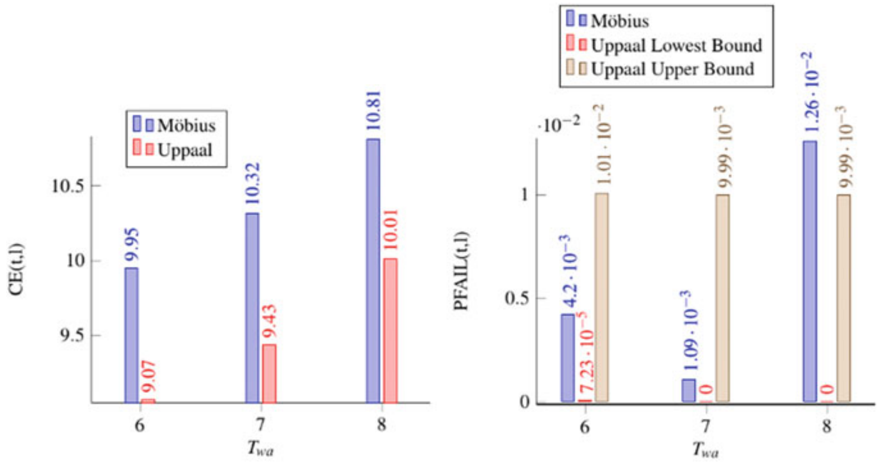
$$PFAIL(t, l) = \mathrm{P}(\Diamond_{\mathrm{h} \leq 24} \exists (\mathrm{i} : \mathrm{id_t})(\mathrm{H_i fail}))$$

The above formula evaluates the probability that in the interval [t, $t + l$] (24 h) there exists at least a switch H_i in the network which has failed, i.e. H_i is in state fail.

6.2 Evaluation

We now discuss the evaluation of *CE*(t, l) and *PFAIL*(t, l) on the two adopted tools. In our experiments, we assume to have a network with 10 switches, partitioned into 4 switches with high priority, 3 with medium priority and 3 with low priority. We report the results of the experiments performed with 9 pairs of T_{wa} and T_{wo}, that are those with a better trade-off between energy consumption and probability of failure. The amount of energy available to the system is set to $NH_{\max} = 50\%$, the values for T_{wa} are respectively 6, 7 and 8 °C, and the values for T_{wo} are 1 °C or 2 °C higher than the corresponding value of T_{wa}. We have considered cumulative values of *CE* (t, l) and *PFAIL*(t, l) for both heaters with high, medium and low priorities in the network. In Figs. 7 and 8 the results of the experiments are reported.

The measures of interest have been evaluated in Möbius through simulation, with minimum and maximum batches per experiment set to respectively 1000 and 10,000 and minimum interval of confidence level of 0.95. Concerning Uppaal SMC, the number of simulations used for evaluating *CE*(t, l) has been fixed to 10,000, while for *PFAIL*(t, l) from a minimum of 800 to a maximum of 10,000 simulations have been performed, that are those necessary for reaching a confidence interval of 0.995.



(a) The energy consumption (b) The probability of failure

Fig. 7 The measures of interests with $T_{wa} - T_{wo} = 1$ °C and $NH_{\max} = 50\%$

The estimated values of $CE(t, l)$ are displayed in Figs. 7a and 8a, and as expected both tools report similar results. In particular, by augmenting the value of T_{wa} the system consumes more energy for reaching a higher temperature. On the converse, we note that by augmenting the difference $T_{wo} - T_{wa}$, the values of $CE(t, l)$ slightly decrease (with the only exception of $T_{wa} = 6$ °C). Indeed, when augmenting $T_{wo} - T_{wa}$, the probability of failure increases, and the failed switches will no longer consume energy, resulting in an overall decrease of $CE(t, l)$.

In Figs. 7b and 8b the values of $PFAIL(t, l)$ are reported. Note that in this case Uppaal SMC only reports the intervals of the probability estimation. Also in this case the tools report similar results, with the main difference being the case $T_{wa} = 8$ °C and $T_{wo} = 10$ °C, which is the worst (i.e. highest) probability of failure. In particular, both tools identify as optimal pair of thresholds the values $T_{wa} = 7$ °C and $T_{wo} = 8$ °C. Indeed, in this case we have the lowest probability of failure, that is a better distribution of energy among all the heaters with different priorities. This is because with a tighter gap of $T_{wo}T_{wa} = 1$ °C the time needed for reaching T_{wo} is less than for the case of $T_{wo} - T_{wa} = 2$ °C. Concerning the performance of the experiments, it has been used a machine with CPU Intel Core i5-4570 at 3.20 GHZ with 8 GB of RAM, running 64-bit Windows 10, Uppaal SMC academic version 4.1.19 (rev. 5649) and Möbius 2.5.0. The elapsed running time is reported in Table 1. The longest running time in Möbius has been of 11.049 s. for computing both measures of interest with parameters $T_{wa} = 6$ °C and $T_{wo} = 7$ °C. On the converse, Uppaal SMC has shown worst performances. Indeed, with parameters $T_{wa} = 6$ °C and $T_{wo} = 7$ °C, we had a running time 295.922 s. for $CE(t, l)$ and 30.39 s. for $PFAIL(t, l)$.

However, we note that for evaluating $CE(t, l)$ Uppaal SMC is forced to perform 10,000 simulations, and for $PFAIL(t, l)$ we have a tight confidence level, i.e. 0.995.

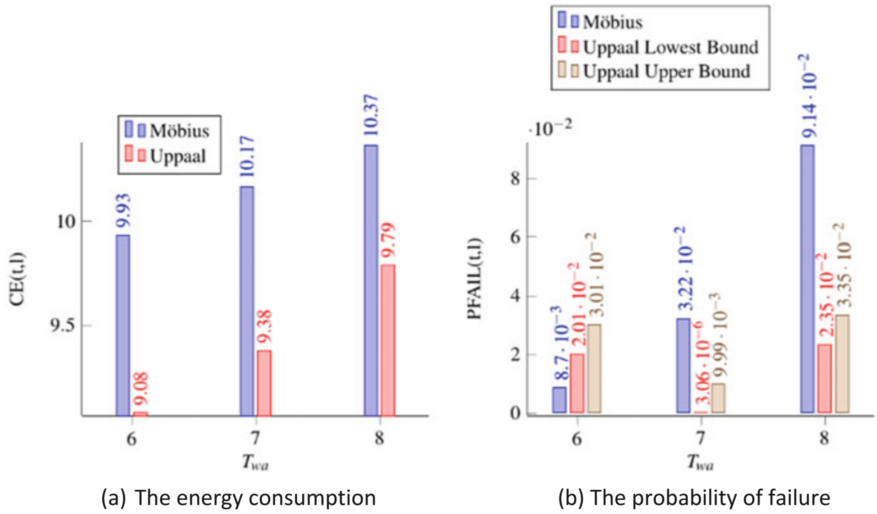


Fig. 8 The measures of interests with $T_{wa} - T_{wo}$ = 2 °C and $NH_{\max}$ = 50%

Nevertheless, in this case study the performances of Möbius overwhelm those of Uppaal SMC.

Finally, with both approaches it has been possible to model and evaluate the case study, and to cross-validate the obtained results.

7 Related Work

There is a wide literature concerning the analysis and optimisation of energy-saving systems in several application domains using formal approaches; in the following we discuss some of these recent efforts.

Table 1 The elapsed running time of the experiments for different settings of (T_{wa}, T_{wo}): for Möbius the total running time of each experiment is reported (i.e. both measures), while for Uppaal SMC the time needed for computing each measure is reported

	Uppaal (*PFAIL*(*t*, *l*)) (s)	Uppaal (*CE*(*t*,*l*)) (s)	Möbius total running time (s)
(6 °C, 7 °C)	30.39	295.922	9.657
(7 °C, 8 °C)	24.922	313.968	9.405
(8 °C, 9 °C)	18.422	316.344	3.868
(6 °C, 8 °C)	233.516	293.875	5.387
(7 °C, 9 °C)	25.015	294.797	2.093
(8 °C, 10 °C)	259.219	290.704	1.775

An example of combining stochastic simulation and model checking is in [42], where a tool chain comprising Uppaal and Möbius is used for the proactive schedule generation for manufacturing scenarios with resource competition, stochastic resources breakdowns, and earliness/tardiness penalties. The system is modelled with Modest [43], the optimal schedule is synthesised with the tool Uppaal Cora [21, 35] and simulations carried on with Möbius. Optimising energy consumption for energy aware buildings, represented as stochastic hybrid automata, is the selected case study in [36]. Statistical Model Checking and analysis of variance has been used to identify Pareto-optimal configurations in terms of both discomfort and energy consumption. In our analysis, we identify the best trade-off between energy consumption and reliability through simulations. An approach for estimating the energy consumption of mobile apps is proposed in [44]. Similarly, we build an abstract model to predict the energy consumption of the rail road switch heating system at the variation of temperature thresholds and available energy, to find the optimal setup of the parameters. Dynamic Power Management (DPV) and Dynamic Voltage and Frequency Scaling (DVFS) are adopted in [45] by using Statistical Model Checking. The system is modelled as Stochastic Hybrid Games. The tool Uppaal Stratego [46] is used to synthesise the safe and near optimal strategy. Similarly, we adopted Statistical Model Checking and DPM, to turn off the energy consumption of heaters when a given temperature is reached. Hybrid automata have been used in [47] to study the dynamic power management control problem. We do not consider a power adjustment mechanism in the rail road switch heating system. A prediction mechanism for minimising the power supplied could be added for example in case of warmer nights.

The applicability of self-organizing systems for different fields of power system control is discussed in [48]. In our case, we consider a central unit that manages the different heaters. The demand of energy is adjusted according to the maximum energy that can be delivered by the central unit. In case of failure of a heater, the energy is automatically shared among the remaining active heaters. We show that by managing the temperature thresholds it is possible to improve reliability even in case of low energy demand. The survivability of a smart house is analysed in [49]. Hybrid Petri Nets [12] are used for modelling this scenario. The authors consider a randomly chosen probability of failure and fixed thresholds, while in our case the probability of failure is derived from the model and we instantiate the thresholds to improve the energy consumption and reliability. The trade-off between energy saving and reliability is studied in [50], by managing frequencies and voltage of the delivered energy. In our approach, different policies of energy consumption are based on thresholds temperatures, and we do not deal with frequencies and voltage of the delivered energy and assume a fixed amount of power. Services negotiation of energy and reliability requirements is the selected case study in [51]. The entities and their protocol of communication have been modelled with the Remes Hdcl language [52] and rendered as timed automata [8], through which the absence of deadlock has been certified with Uppaal. Their reliability and energy requirements are given parameters that are negotiated between the parties; instead in our approach we estimate those values based on policies of energy consumption.

8 Conclusion

We have addressed cross-validation of reliability and energy consumption of a critical cyber-physical system belonging to the railway domain (a rail road switch heating system), by comparing the approach based on Stochastic Activity Network and Möbius with the approach based on Stochastic Hybrid Automata and Uppaal SMC.

We address some lines of future research concerning energy saving CPS. The behaviour of CPS is in general unpredictable, because of physic and environmental aspects that are involved in their design. A future line of research concerns the introduction of control techniques for restricting their possible behaviours in order to predict and avoid possible failures, improve and verify their dependability [53]. It is also interesting to study the formal specification of the energy requirements that a CPS must satisfy, verifying that the proposed model satisfies them or proving that such requirements are not satisfiable [54].

References

1. Lee, E.A.: Cyber physical systems: design challenges. In: ISORC'08. IEEE C.S. (2008)
2. Basile, D., Di Giandomenico, F., Gnesi, S.: Model-based evaluation of energy saving systems. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Concepts, Models, Complex Systems Architectures. Studies in Systems, Decision and Control, vol. 74, pp. 187–208. Springer International Publishing, Cham (2017)
3. Doukas, N.: Technologies for greener internet of things systems. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Components, Networks and Systems Implementation. Studies in Systems, Decision and Control, vol. 105, pp. 23–42. Springer International Publishing, Cham (2017)
4. Hahanov, V., Litvinova, E., Chumachenko, S.: Green cyber-physical computing as sustainable development model. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Components, Networks and Systems Implementation. Studies in Systems, Decision and Control, vol. 105, pp. 65–85. Springer International Publishing, Cham (2017)
5. Kharchenko, V., Illiashenko, O.: Concepts of green it engineering: taxonomy, principles and implementation. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Concepts, Models, Complex Systems Architectures. Studies in Systems, Decision and Control, vol. 74, pp. 3–19. Springer International Publishing, Cham (2017)
6. Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.): Green IT Engineering: Components, Networks and Systems Implementation, Studies in Systems, Decision and Control, vol. 105, 1st edn. Springer Publishing Company, Incorporated, Cham (2017)
7. Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.): Green IT Engineering: Concepts, Models, Complex Systems Architectures, Studies in Systems, Decision and Control, vol. 74, 1st edn. Springer Publishing Company, Incorporated, Cham (2017)
8. Alur, R., Dill, D.L.: A theory of timed automata. TCS **126**(2), 183–235 (1994)
9. Henzinger, T.A.: The theory of hybrid automata, pp. 278. LICS'96, IEEE C.S. (1996)
10. Krishna, S.N., Trivedi, A.: Hybrid automata for formal modeling and verification of cyberphysical systems. CoRR (2015). http://arxiv.org/abs/1503.04928

11. Balbo, G.: Introduction to generalized stochastic Petri nets. In: Bernardo, M., Hillston, J. (eds.) Formal Methods for Performance Evaluation, LNCS, vol. 4486. Springer (2007)
12. David, R., Alla, H.: On hybrid Petri nets. DEDS **11**(1–2), 9–40 (2001)
13. Sanders, W.H., Meyer, J.F.: Stochastic activity networks: formal definitions and concepts. In: Lectures on Formal Methods and Performance Analysis (2000)
14. Basile, D., Chiaradonna, S., Di Giandomenico, F., Gnesi, S.: A stochastic model-based approach to analyse reliable energy-saving rail road switch heating systems. JRTPM (2016). http://www.sciencedirect.com/science/article/pii/S2210970616300051
15. Basile, D., Chiaradonna, S., Di Giandomenico, F., Gnesi, S., Mazzanti, F.: Stochastic model based analysis of energy consumption in a rail road switch heating system. In: SERENE 2015. LNCS 9274
16. Basile, D., Di Giandomenico, F., Gnesi, S.: Tuning energy consumption strategies in the railway domain: a model-based approach. In: 7th International Symposium on Leveraging Applications of Formal Methods, Verification and Validation, ISOLA 2016 (2016)
17. Haverkort, B.R.: Lectures on formal methods and performance analysis. In: Markovian Models for Performance and Dependability Evaluation, pp. 38–83. Springer (2002)
18. Clark, G., Courtney, T., Daly, D., Deavours, D., Derisavi, S., Doyle, J.M., Sanders, W.H., Webster, P.: The Möbius modeling tool. In: PNPM. pp. 241–250 (2001)
19. Legay, A., Delahaye, B., Bensalem, S.: RV 2010. Proceedings. In: Statistical Model Checking: An Overview. Springer
20. Larsen, K.G., Legay, A.: Statistical model checking: past, present, and future. In: Leveraging Applications of Formal Methods, Verification and Validation: Foundational Techniques 7th International Symposium, ISoLA 2016, Imperial, Corfu, Greece, 10–14 Oct 2016, Proceedings, Part I. pp. 3–15 (2016)
21. David, A., Larsen, K.G., Legay, A., Mikuaionis, M., Poulsen, D.B.: Uppaal SMC tutorial. Int. J. Softw. Tools Technol. Transf. 17 (2015)
22. Basile, D., Giandomenico, F.D., Gnesi, S.: Statistical model checking of an energy-saving cyber-physical system in the railway domain. In: Proceedings of the Symposium on Applied Computing, SAC 2017, Marrakech, Morocco, 3–7 Apr 2017, pp. 1356–1363 (2017)
23. Basile, D., Giandomenico, F.D., Gnesi, S.: Enhancing models correctness through formal verification: a case study from the railway domain. In: MODELSWARD 2017, Porto, Portugal, 19–21 Feb 2017, pp. 679–686 (2017)
24. Baier, C., Haverkort, B.R., Hermanns, H., Katoen, J.P.: Performance evaluation and model checking join forces. Commun. ACM **53**(9), 76–85 (2010)
25. Avizienis, A., Laprie, J.C., Randell, B., Landwehr, C.: Basic concepts and taxonomy of dependable and secure computing. IEEE Trans. Dependable Secure Comput. **1**(1) (2014)
26. Sanders, W.H., Nicol, D.M., Trivedi, K.S.: Model-based evaluation: from dependability to security. IEEE Trans. Dependable and Secure Comput. **1**(undefined), 48–65 (2004)
27. Karlin, H.M.T. (ed.) An Introduction to Stochastic Modeling (Revised Edition), revised edn. Academic Press (1994). http://www.sciencedirect.com/science/article/pii/B978012684885450001X
28. Bernardi, S., Merseguer, J., Petriu, D.C.: Model-Driven Dependability Assessment of Software Systems. Springer (2013)
29. Diab, H.B., Zomaya, A.Y.: Dependable computing systems: paradigms, performance issues and applications. Wiley (2005)
30. Bause, F., Kritzinger, P.S.: Stochastic Petri nets: an introduction to the theory. SIGMETRICS Perform. Eval. Rev. **26**(2)
31. Reibman, A., Smith, R., Trivedi, K.: Markov and Markov reward model transient analysis: an overview of numerical approaches. Eur. J. Oper. Res. **40**(2) (1989)
32. Ahmed, W., Hasan, O., Tahar, S.: Formal dependability modeling and analysis: A survey. In: CICM, Bialystok, Poland, July 25–29. LNAI, vol. 9791, pp. 132–147. Springer (2016)
33. Clarke Jr., E.M., Grumberg, O., Peled, D.A.: Model Checking. MIT Press, Cambridge (1999)
34. Hinton, A., Kwiatkowska, M., Norman, G., Parker, D.: Prism: a tool for automatic verification of probabilistic systems. In: TACAS 2006, volume 3920 of LNCS, pp. 441–444. Springer

35. Larsen, K.G., Pettersson, P., Yi, W.: Uppaal in a nutshell. JSTTT 1 (1997)
36. David, A., Du, D., Guldstrand Larsen, K., Legay, A., Mikucionis, M.: NFM 2013, chap. Optimizing Control Strategy Using Statistical Model Checking, pp. 352–367. Springer (2013)
37. Henzinger, T.A., Ho, P.: Algorithmic analysis of nonlinear hybrid systems. In: CAV (1995)
38. http://www.railsco.com/electric_switch_heater_controls.htm. Accessed on June 2016
39. Brodowski, D., Komosa, K.: A railroad switch and a method of melting snow and ice in railroad switches (2013). https://data.epo.org/publication-server/rest/v1.0/publication-dates/20131225/patents/EP2677079NWA1/document.html
40. https://weatherspark.com/#!graphs;ws=27985. Accessed on Mar 2016
41. Trivedi, K.S.: Probability & statistics with reliability, queuing and computer science applications. Wiley (2008)
42. Mader, A., Bohnenkamp, H., Usenko, Y.S., Jansen, D.N., Hurink, J., Hermanns, H.: Synthesis and stochastic assessment of cost-optimal schedules. Int. J. Softw. Tools Technol. Transf. (STTT) **12**(5), 305–317 (2009). http://doc.utwente.nl/69344/
43. Hartmanns, A., Hermanns, H.: The Modest Toolset: An Integrated Environment for Quantitative Modelling and Verification, pp. 593–598. Springer, Berlin (2014)
44. Lu, Q., Wu, T., Yan, J., Yan, J., Ma, F., Zhang, F.: Lightweight method-level energy consumption estimation for android applications. In: 10th International Symposium on Theoretical Aspects of Software Engineering, TASE 2016, Shanghai, China, 17–19 July 2016. pp. 144–151. IEEE (2016). http://doi.ieeecomputersociety.org/10.1109/TASE.2016.27
45. Ahmad, W., van de Pol, J.: Synthesizing energy-optimal controllers for multiprocessor dataflow applications with uppaal stratego. In: ISOLA 2016 (2016)
46. David, A., Jensen, P.G., Larsen, K.G., Mikucionis, M., Taankvist, J.H.: Uppaal Stratego, pp. 206–211. Springer, Berlin (2015)
47. Erbes, T., Shukla, S.K., Kachroo, P.: Stochastic learning feedback hybrid automata for dynamic power management in embedded systems. In: SMCia/05 (2005)
48. Muller, S.C., Hager, U., Rehtanz, C., Wedde, H.F.: Application of self-organizing systems in power systems control. In: Dieste, O., Jedlitschka, A., Juzgado, N.J. (eds.) PROFES 2012 Proceedings, LNCS, vol. 7343, pp. 320–334. Springer (2012)
49. Ghasemieh, H., Haverkort, B.R., Jongerden, M.R., Remke, A.: Energy resilience modeling for smart houses. In: 45th Annual IEEE/IFIP, DSN 2015, pp. 275–286. IEEE C.S. (2015)
50. Zhu, D., Melhem, R., Mossè, D.: The effects of energy management on reliability in real-time embed DED systems. In: ICCAD, pp. 35–40, Nov 2004
51. Causevic, A., Seceleanu, C., Pettersson, P.: Distributed energy management case study: a formal approach to analyzing utility functions. In: ISoLA Proc.II, LNCS, vol. 8803, pp. 74–87. Springer (2014). https://weatherspark.com/#!graphs;ws=27985. Accessed on Mar 2016
52. Seceleanu, C., Vulgarakis, A., Pettersson, P.: Remes: a resource model for embedded systems. Eng. Complex Comput. Syst. **2009**, 84–94 (2009)
53. Antsaklis, P.: Goals and challenges in cyber-physical systems research editorial of the editor in chief. IEEE Trans. Autom. Control **59**(12), 3117–3119 (2014)
54. Banerjee, A., Venkatasubramanian, K.K., Mukherjee, T., Gupta, S.K.S.: Ensuring safety, security, and sustainability of mission-critical cyber-physical systems. Proc. IEEE **100**(1) (2012)
55. Kemper, P., Tepper, C.: Traviando—debugging simulation traces with message sequence charts. In: International Conference on Quantitative Evaluation of Systems, pp. 135–136 (2006)

Cooperative Ecology Human–Machine Interfaces for Safe Intelligent Transport Systems: Cloud-Based Software Case

Aleksandr Orekhov, Anastasia Stadnik and Vyacheslav Kharchenko

Abstract The analysis of the available solutions regarding the cooperative intelligent transport systems (ITS) and their human–machine interfaces (HMI) has been conducted. The standards and recommendations in the area of transport systems interfaces design is analysed. So-called ecology HMI and requirements for such interfaces and features of the application for ITS are discussed. Profiling of the requirements for the cooperative human–machine interface (CHMI) and the cooperative ecology human–machine interface (CEHMI) for such systems including the requirements to usability and safety based on a set of standards for ITSs has been performed. An approach and a design technique of human–machine interfaces for ITSs based on designing the cooperative human–machine interfaces for intelligent transport systems are suggested. The architecture of cloud-based CHMI for intelligent transport systems that consists of three main parts: client end, server end, and core-project and includes data models for the communication protocol and the common utility functions is developed. The prototype of the software system for CHMI for ITS examined in the laboratory conditions is described. The assessment using the GOMS method allowed to calculate the efficiency to execute the task of user interaction with the system.

Keywords Intelligent transport systems · Human–Machine interfaces
Cloud

1 Introduction

Human–Machine interfaces are an important part of information and control systems (I&Cs) for commercial and critical domains. The ability to control and manage systems and objects, for instance, reactor, aircraft or medical equipment, the

A. Orekhov (✉) · A. Stadnik · V. Kharchenko
National Aerospace University "KhAI", 17, Chkalova Street, Kharkiv, Ukraine
e-mail: a.orehov@csn.khai.edu

© Springer Nature Switzerland AG 2019

V. Kharchenko et al. (eds.), *Green IT Engineering: Social, Business and Industrial Applications*, Studies in Systems, Decision and Control 171,
https://doi.org/10.1007/978-3-030-00253-4_19

efficiency and reliability of the operator and I&Cs in whole, depend on the HMI quality.

The HMI development process is based on a variety of modern technologies and means, such as sensor monitors, high-performance computers and networks.

One of the today's challenges in the domain of creating the HMI is the development of the green human–machine interfaces (GHMI) [1, 2]. In contrast to the traditional HMIs, they have such properties as environmental friendliness, adaptability, safety, reliability, etc.

Scalable and flexible interfaces of the operator's panels allow to integrate different systems monitoring and control functions into the GHMI [3]. The hardware part of such GHMI includes the touch-screen, that gives access to the general infrastructure of automated house and Internet. The software part is presented by the smart-house control program, which communicates and operates with a variety of connected devices, such as an air conditioner, CCTV, intercom, lighting equipment, household devices, etc. GHMI can also offer some other Internet services [4].

GHMI for the automobile informational systems improves the traffic safety by decreasing the driver's informational overload, and thus minimizing the distractions [5]. The context-dependent interface adaptation, which is proposed in [6], can be achieved by user personalization.

GHMI should not only be safe and reliable but also adaptable and environmentally friendly.

The environmental interface is the interface, in which the information provided by the mental model closely matches the managed object, operated by the human. The environmental interface provides the information in a form of visual images, which makes its perception and execution easier for processing of cognitive operations. The modern environmental interface is created using the virtual reality technologies (immersion interface) and cognitive graphics [7].

Based on the definition of the environmental interface, we can conclude that environmental interface should be matched with the user's mental model, make the perception of the information easier and facilitate the implementation of the cognitive operations of the operator. We can expand the existing quality models by introducing new characteristics:

- *User Model Compatibility*—all aspects of the system should be compatible with the mental users' models.
- *Situation Awareness*—the information must be correct and easy to understand, and it should be supported on a high level to ensure users' awareness about system's status.
- *Cognitive Compatibility*—operator's work should consist of purposeful and meaningful tasks, that help staff in supporting the strong awareness and to keep the level of the load not too high to avoid bad effects on productivity, but high enough to maintain vigilance.
- *Cognitive Workload*—information should be quickly perceived and understood. The system should minimize the requirements to the in-mind calculations and

conversions, and use some hints. The background data should be presented in a convenient form.

- *Timeliness*—system design should take into consideration the users' cognitive capabilities and the time limits in connection with the process. The speed of the informational stream and performance monitoring requirements, which are too fast or too slow, may lead to productivity decline.

The adaptability is an additional GHMI factor. As a property of HMI, the adaptability is presented in several forms: change of the given informational input, maintaining of the dialog, distribution of the problems between human and machine, adaptation speed [8, 9].

One of the criteria of adaptability is the flexibility. Flexible HMI should provide the user with several ways to perform the action, and the display and control should be the most suitable for the problem.

In Sect. 2, the principles for the development of human–machine interfaces for vehicle information systems are reviewed.

In Sect. 3, the application of cloud computing technology for intelligent transport systems is given. The systems for driver state analysis are investigated.

In Sect. 4, the concept of the environmental cooperative HMI along with the architecture for its implementation are proposed.

In Sect. 5, the development of the cooperative human–machine interface prototype based on cloud computing is described.

In Sect. 6, the development of the assessment model for human–machine interface based on the classical GOMS method and the assessment method for vehicle navigation systems according to the J2365 standard are given.

In Sect. 7, the results of the experiment based on the emulation of the vehicles on the road moving from one point to another are given.

In Sect. 8, the general conclusion is communicated. Further directions of research in regards to the environmental cooperative HMI aim for the integration of the system with the onboard computer, enhanced efficiency and need to be based on testing the system in real conditions on a number of vehicles.

2 Principles of Human–Machine Interfaces Development for Vehicle Information Systems

In the European declaration on the principles of HMI functioning [10] In-Vehicle Information Systems (IVIS) designing foundations are offered. The systems should not distract a driver, and the information they convey to the driver has to be predictable and controlled. It is important that interaction with the informational systems neither overloads the driver of the vehicle nor distracts him. The systems should give the information in a concise and comprehensive way.

Systems that need to communicate with the driver should be easy to use and always provide for the driver to do the principal task, which is driving the vehicle safely. Good HMI system reduces the informational strain on the driver helping to select the most relevant and important information.

As noted in the documents of the European Commission, safe HMI design must take into account the need to integrate nomadic devices and ensure the safety of vulnerable traffic participants (e.g. aged people). Nomadic devices include information and communication equipment such as mobile phone, navigation system, PDA, etc. All these devices are typical examples of the vehicle information systems.

Using nomadic devices may be not matched with a car, especially if their HMI is designed poorly. It should be noted that in the future the increase in the new systems with different haptic, visual and auditory methods of communication with drivers is expected. Therefore, all the risks associated with the use of such systems should be estimated.

Recommendations on the design of safe HMI of the IVIS is suggested in the project Human–Machine Interface and the Safety of Traffic in Europe (HASTE) [11]. The aim of the HASTE project is to develop methodologies and guidelines for the assessment of In-Vehicle Information Systems. It is therefore implied that such systems should have a means of communicating with the driver. This could be through one or more sensory modalities, i.e. visual, auditory or tactile/haptic interfaces. This will also require some form of system input interface.

Among the objectives of the research program within the HASTE are the following:

- identify and explore scenarios in which safety issues are most important and relevant;
- explore the connection between the load and the risk in the context of these scenarios;
- conceive the mechanisms of risk increasing in terms of distraction and the driver situational awareness reducing;
- determine risk rates;
- apply existing risk assessment methods to real vehicles;
- consider possible causes of information systems threats related to safety and reliability.

3 Cloud-Based Intelligent Transportation System

Cloud computing (CC) is used to receive or transmit data over the Internet via a wireless connection. The idea of using cloud services in ITS is just beginning to gain popularity [12, 13]. The facilities of “clouds” can also affect the increase of transport safety.

Re-engineering the vehicle to a cloud-based technology is discussed in [14]. Vehicles with a global positioning system that are connected to the “cloud”, will always “know” their location and the road conditions. Today, some of the features of the vehicle are designed with innovative technology, using the Internet, for example, the communication function V2V (vehicle-to-vehicle) and V2R (vehicle-to-road).

The Volvo Car Group company is working on new car projects—exchange of information about the dangers on the road through the "cloud" and control of the driver's state [15]. The data about the slippy road parts generated based on the vehicle sensors are passed to the Volvo Cars database via the mobile network on a real-time basis. A warning is passed to other vehicles reaching this road part instantly, thus making it possible for the driver to take prompt measures in order to prevent the critical situation.

Another possible application of this technology is the remote diagnostics. Data can be transferred in advance, thus eliminating the problem in real-time [16].

The review of the systems for driver state analysis is given in Table 1.

Driver's exhaustion is assessed by processing multiple parameters:

- vehicle movement (speed, forward and side acceleration, rate of yaw);
- biometric indicators (heart rate, respiration rate, skin temperature);
- driver's vision (eyes opening rate and vision line);
- driver's actions (turning angle of the steering wheel, position of the foot and brake throttles);
- road condition (traffic density, road covering).

Table 1 Driver state analyzing systems

System	Sensors used	Implementation examples
DAS (driver attention support)—driver's exhaustion detection and preventing sleeping at the wheel	– IR sensor behind the steering wheel that controls the face temperature – piezoelectric sensor in the safety belt that monitors the breathing rate – patches at the rim of the steering wheel that measure the pulse – IR sensor behind the steering wheel that measures the temperature of the palms	(1) Attention Assist (Mercedes-Benz) (2) Driver Alert Control (Volvo) (3) Seeing Machines (General Motors)
Physical state assessment systems Assessment of the critical health indicators: – pulse – breathing rate – skin capacity – blood sugar level	– heart rate sensors installed in the seat – sensors at the rim of the steering wheel: electrodes that monitor the heart rhythm and optical sensors that assess palms capacity	(1) Driver load assessment system (Ford) (2) Aged driver's state control system (3) Vital indicators control system (Toyota) (4) Warning technology for the diabetic drivers (BMW)

4 Cooperative Human–Machine Interfaces

As noted above, the cooperative systems are such systems that communicate with other cars wirelessly. Therefore, under the term of a cooperative HMI we will consider the interface system, distributed among several vehicles [17]. Additional monitors are installed on each vehicle or a compact unit is embedded into the existing HMI to provide information about the safety in APH, which gives the information about the safety level. This information (risk matrix) is formed and dynamically adjusted based on the overall situation for each car (the state of the vehicle, driver and road conditions), which is in the danger zone.

It is clear that these must be adaptive HMI, which reflect not only information about the condition of the car, but of the driver as well. If a driver starts to doze off or falls asleep, it is necessary to wake him up and inform the drivers of motor vehicles that are nearby.

The property of adaptability in the HMI becomes apparent in several forms: changes in the content of the information provided, dialogue, sharing of tasks between man and machine, the speed of adaptation [8].

The project PRORETA is a research in the area of the cooperative HMIs. The research object is the prototype of the cooperative automobile HMI that implements the scenarios of preventing collisions at the crossroads. The PRORETA HMI system implements a huge number of use scenarios, it does not complicate or irritate and ensures the multimode support. The HMI provides 4 support levels—information messages, warnings, actions recommendations, automatic intervention.

One of the variants of cooperative HMI construction is to use the technology of CC. Figure 1 shows the proposed architecture of the cooperative HMI.

Cooperative HMI provides the measurement values of the parameters of vehicle and driver state in real time via the Internet into the “cloud.” Here, the data from all the cars are dynamically processed and transmitted to motoring public.

Information from the HMI of one vehicle (shown as a red dashed line, Fig. 1) passes through the “cloud” and is displayed on the HMI of another vehicle. In turn, the information from the HMI of another vehicle (shown as a green line, Fig. 1), is also transferred to the HMI of the first vehicle. This information is taken into account when the risk analysis of each vehicle is performed.

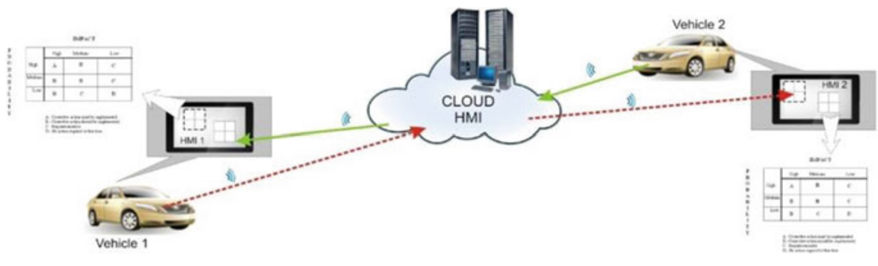


Fig. 1 Architecture of the cooperative HMI based on cloud computing

There are important issues in developing of HMI: optimization of the information necessary for the driver for the safe driving mode; determination of the information views, which stimulate the driver; control and prevention of the driver's detraction.

5 Case Study: Prototype of Cooperative Human–Machine Interface

The system consists of three projects combined in a single solution:

- server end—the decision support system (DSS);
- client end—the user HMI;
- core-project that includes data models for the communication protocol and the common utility functions.

The server end is the web-application, the core of which is the DSS. The web-application is managed by the Apache Tomcat server that supports the HTTP protocol. The protocol allows the interaction between the client and the server. The client end is implemented for the Android platform and it stands for the user interface. The ground map is the key element. The data exchange is performed wirelessly using the data types specified in the general Core-project. Java serves as the platform for creating the system in question. Figure 2 shows the architecture of the system.

The client and the server cooperate wirelessly through the communication module. The general convenience functions and data models for packetizing can be found in the Core-project that is used by both sides. Since the communication protocol should provide equal rights for the client and the server, it has been agreed to implement the communication protocol based on TCP from the specification Java EE—WebSocket.

The protocol ensures the free data exchange: two equal participants exchange data, each one working independently and sending data to the other one when necessary (Fig. 3).

The data is packetized. The packets stand for the data type from the core-project in the JSON format. The packets are formed and parsed on both sides in the communication module.

The human–machine interface provides the driver with the information about the road situation, the driver's state and the vehicle's state. At the first start of the client application the registration form is displayed (Fig. 4) where the driver needs to enter his personal data (nickname, age, sex).

The working area on the display is covered with the ground map (Fig. 5).

The current state and the direction of the vehicle are marked on the map with the help of the special arrow indicator. The map is to be centered according to the

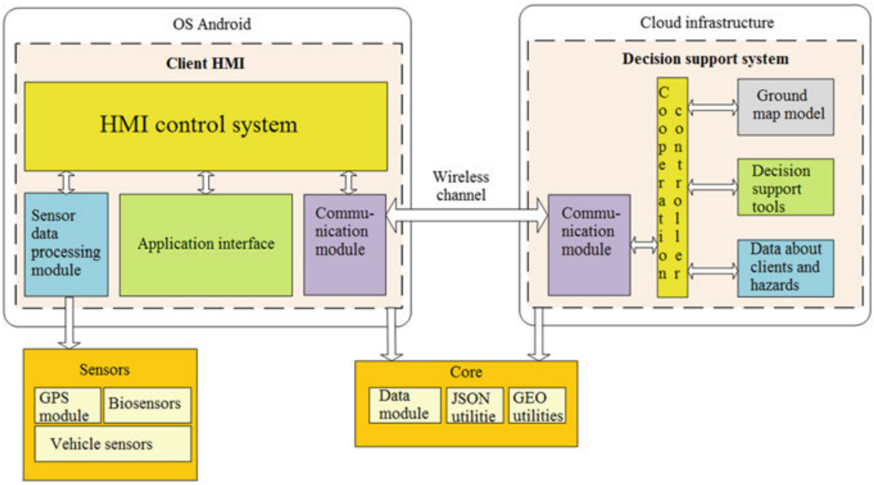
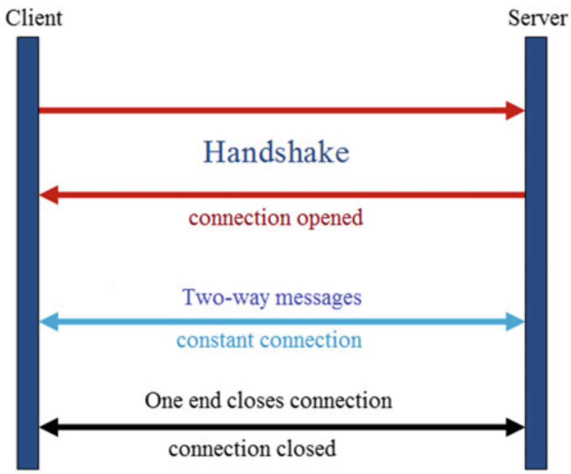


Fig. 2 Architecture of the system

Fig. 3 Mode of the communication protocol WebSocket



current position. The position of other vehicles is displayed by means of arrows having different colours.

The connection to the server is displayed by a special indicator. The indicator icon depending on whether there is a connection is shown in Fig. 6.

The driver's state is displayed by a special indicator. The indicator icon depending on the driver's state is given in Fig. 7.

The HMI provides for the feature of manual signals to other drivers about the dangerous road stretch by pressing the button with the schematic representation of hazards types on a special board (Fig. 8).

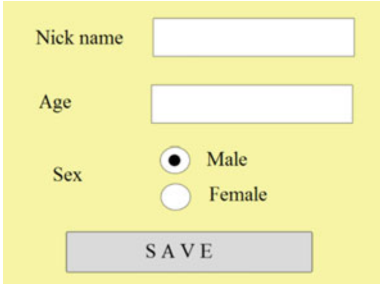


Fig. 4 Registration form

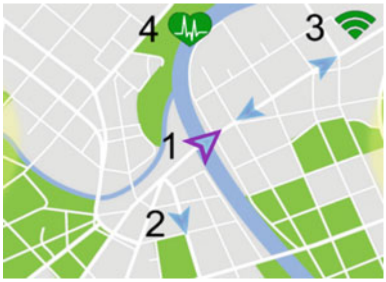


Fig. 5 Visual interface 1—current position; 2—another vehicle; 3—server connection indicator; 4—driver's state indicator

Fig. 6 Server connection indicator: 1—connection established, 2- no connection

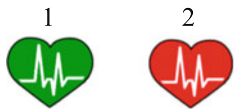


Fig. 7 Driver's state: 1—good, 2—poor

The speech recognition has been adopted in the HMI for the voice hazard signal transfer. The command for signal transfer consists of two fields: 1—key phrase, 2—hazard type. The key phrase should be brief and easy to pronounce. The possible key phrases are: "OK, motor", "Go, machine" or simply "Danger", "Danger ahead". According to the survey results, the majority of the drivers prefer to set their own key phrases for the control commands.

Hazards are indicated on the map with markers displaying the hazard type (Table 2). The marker is coloured according to the hazard level (low—yellow, middle—orange, high—red).

Map scale should be set according to the range of the lowest hazard level. When the hazard description is queried, the informative message with the enlarged hazard marker and the distance to the hazard object is displayed (Fig. 9).

Fig. 8 Hazards menu

Table 2 Driver state analyzing systems

Hazard type	Poor road	Ice condition	Fog	Caving	Reconditioning work	Poor driver's state	Aggressive driver
Marker							

Fig. 9 Displaying of the markers and informative messages

A new hazard occurred is accompanied by the short voice signals. If the hazard level is high, the driver is informed with the voice messages communicating the hazard, for example: «Aggressive driver ahead, distance 150 m, speed 90 kmph», «Fog in a 100 m». Voice messages should repeat at a 10-s interval. The driver is provided with the possibility to query the hazard description using voice commands like "Voice the hazard", "Describe the hazard". The map scale can be configured, the voice and sound messages and volume level can be set or disabled (Fig. 10).

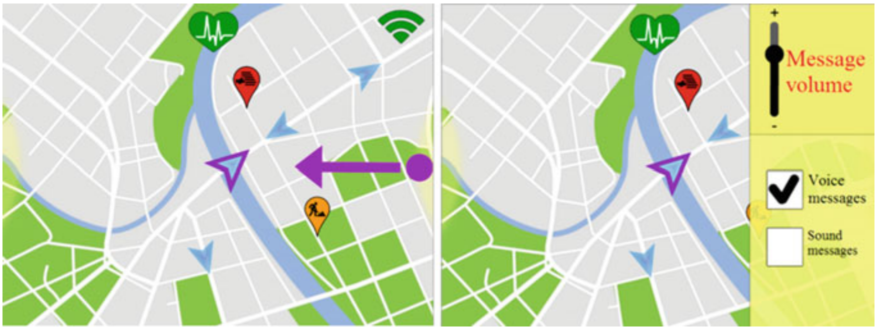


Fig. 10 Signal parameters setting

Display brightness and contrast should be adjusted to the daytime. The voice messages volume level is to be adjusted to the noise level in the car.

The overall picture of the road is at the driver's disposal. He can see the ground map, monitor other vehicles moving on a real-time basis. The driver's awareness is improved as the position of the cars undetected through the glass or by the mirror can be obtained. The blind spot issue is resolved. Due to the voice description of the hazards, the cognitive load is reduced, the probability of the driver's distraction of the display is lowered.

The client HMI subsystem is implemented on OS Android. It consists of several modules that interact via the HMI control system (Fig. 11).

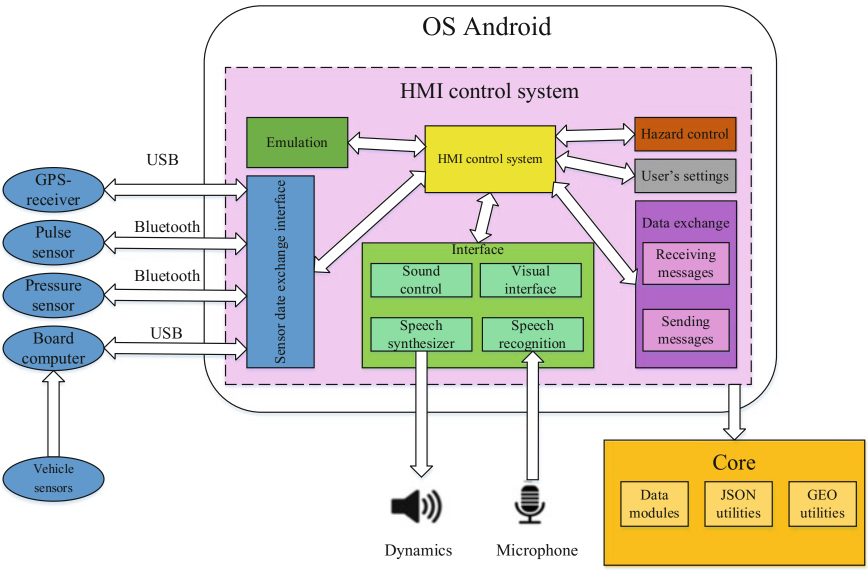


Fig. 11 Client HMI architecture

The vehicle's sensors data is obtained from the board computer through the wired interfaces. The requests to the biosensors can be done through wireless interfaces. The current coordinates are obtained from the GPS receiver through the wire communication channel.

The module for communication is responsible for receiving and transmitting the messages to the server. The packets are formed by the client subsystem using the data models from the Core-project.

The interface of the application includes the following modules:

1. Visual interface responsible for displaying the following elements on the monitor:
   - registration page (personal data filling);
   - ground map, current position and position of other participants, hazard objects on the ground map;
   - indicators of the driver's state and server connection;
   - hazards panel;
   - signals setup control panel.
2. Speech synthesizer responsible for voice warnings generation.
3. Speech recognition responsible for voice commands recognition.
4. Sounds management responsible for sound warnings.

The HMI control system is responsible for the interaction with other modules in the system. It obtains the data from the sensor interfaces and transmits it to the communication module where packetizing takes place and the packets are sent to the server.

The user setup control module is responsible for the configuring and storing the personal data and parameters of the signals. The hazard control module is responsible for refreshing of the hazards list provided by the server.

The emulation module is responsible for the emulation of the vehicle movement and the data obtained from the biosensors.

Figure 12 stands for the connection between the client, the server and the core modules.

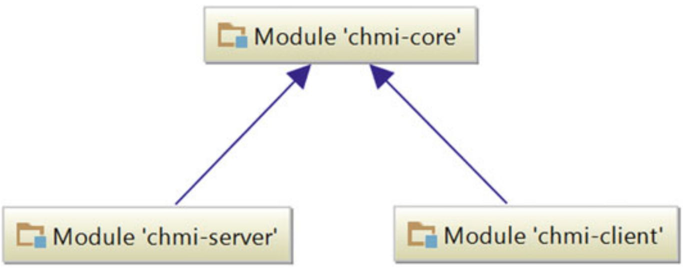


Fig. 12 Module structure of the system

The common classes for the client and the server applications are specified in the Core-project. It consists of the following packets:

1. Utils packet that contains 2 classes:
   - GeoUtils—the utilities for the geodata processing;
   - JsonUtils—the utilities for the work with JSON objects.
2. dto packet (Data Transfer Object) that includes the data models used to form the data transmitting packets.

The interface displays the road situation using the clear images. The driver can monitor the movement of the road users on the ground map (their current position, speed, direction). The ground map is associated with the environment, while the vehicles markers are associated with the real vehicles and the hazard markers are associated with the hazard objects. The colour of the markers allows the driver to identify the most dangerous objects. Using the voice warnings about the hazards and the voice commands for transmitting the signal the cognitive load is reduced as well as the driver's distraction of the monitor from the road is eliminated which in total reduces the risk of an accident on the road.

6 Human–Machine Interface Assessment

Let us specify the operators in order to do all the interaction tasks for the HMI and build the model to assess it based on the classical GOMS method and the assessment method for the automobile navigation systems according to the J2365 standard. The client HMI time indicators assessment model is given in Table 3. Table 4 displays the code of each operator and the duration for the young and elderly drivers.

Identification of the tasks for the work with the human–machine interface:

(1) Pass the signal about the dangerous road section via the monitor.
(2) Pass the signal about the dangerous road section using the voice command.
(3) Disable the sound signals.
(4) Request the hazard description.
(5) Learn the road situation.
(6) Listen to the message about the hazard.

Models development and the calculation of tasks execution.

Task 1. Pass the signal about the dangerous road section via the monitor.

(1) M—Mental psych-up—1.50/2.55.
(2) D—Reach the monitor with the hand—0.45/0.77.
(3) Y—Move the hand (finger) to the left hazard panel—0.80/1.36.
(4) H—Get the hazard panel out—1.2/2.04.

Table 3 GOMS method adaptation for the HMI

Operator	Classical GOMS method	GOMS method according to J2365, young/elderly	GOMS for CHMI code, time
Mental psych-up	Mental operation M = 1.35 s	Mental operation M = 1.50/2.55 s	Mental operation M = 1.50/ 2.55 s
Compare the hazard description with the position on the map			
Assess the most dangerous objects on the map			
Reach the monitor with the hand	Movement D = 0.4 s	Reach far Rf = 0.45/0.77 s	Reach D = 0.45/ 0.77 s
Mark the position on the monitor with the finger	Mark Y = 1.1 s	Cursor once s1 = 0.80/1.36 s	Mark Y = 0.80/ 1.36 s
Find the marker on the map	–	Search S = 2.30/3.91 s	Search P = 2.30/ 3.91 s
Find the control element on the monitor	–	Search S = 2.30/3.91 s	
Press the object on the map	Press the key К = 0.2 s	Enter E = 1.2/2.04 s	Press H = 1.2/ 2.04 s
Press the button	Press the key К = 0.2 s	Enter E = 1.2/2.04 s	
Wait for the response from the interface	System response R	Response time of system-new menu Rm = 0.50 s	Interface response O = 0.1 s
Say the key voice command aloud	–	–	Voice command G = 1.5 s
Say the signal voice command aloud	–	–	
Listen to the message about the hazard	–	–	Listen s = 3.5 s
React to the situation	–	–	Reaction PE

Table 4 Summary table of the operators

	Mental operation	Reach	Mark	Search	Press	Response	Voice command	Listen	Reaction
	M	D	Y	P	H	O	G	S	PE
Young	1.50	0.45	0.8	2.3	1.2	0.1	1.50	3.5	–
Elderly	2.55	0.77	1.36	3.91	2.04				–

(5) Y—Move the finger to the button of interest—0.80/1.36.
(6) H—Press the hazard button—1.2/2.04.
(7) O—Wait for the response from the system—0.1.

The sequence:

$$M + D + Y + H + P + Y + H + O$$

Time for the young people:

$$t1 = 1.50 + 0.45 + 0.80 + 1.2 + 1.2 + 0.1 = 5.25 \text{ s.}$$

Time for the elderly people:

$$t2 = 2.55 + 0.77 + 1.36 + 2.04 + 2.04 + 0.1 = 8.86 \text{ s.}$$

Task 2. Pass the signal about the dangerous road section using the voice command.

The sequence:

$$M + G + M + G + O$$
$$t1 = 1.50 + 1.50 + 1.50 + 0.1 = 4.6 \text{ s.}$$
$$t2 = 2.55 + 1.50 + 2.55 + 0.1 = 6.7 \text{ s.}$$

Task 3. Disable the sound signals.

The sequence:

$$M + D + Y + H + Y + H + O$$
$$t1 = 1.50 + 0.45 + 0.80 + 1.2 + 0.8 + 1.2 + 0.1 = 6.05 \text{ s}$$
$$t2 = 2.55 + 0.77 + 1.36 + 2.04 + 1.36 + 2.04 + 0.1 = 10.22 \text{ s}$$

The quantitative evaluation of the HMI shows that the hazard signal is passed more effectively using the voice commands. The execution of the signal setting task should be optimized through the voice commands, and thus less time will be spent by the driver. Additionally, we can conclude that the voice description requires more time compared to the situation of the driver executing the task of the ground map assessment. However, in this case, the driver pays his attention to the map far more quickly which allows him to react to the situation faster.

7 Experimental Research of Cooperative Human–Machine Interfaces

The experiment has been conducted in laboratory conditions basing on the emulation of the vehicle's movement on the road from one point to another.

Experiment 1. Driver's state indication.

(1) Starting state of the HMI. The identifier shows that the driver's state is good (Fig. 13).
(2) Set the parameters of the driver. Experimental profile 1 (Fig. 14):

Age = 25 years;
Pulse = 60;
Upper hypertension = 100;
Low hypertension = 50.

The expected result is poor driver's state, and the indicator coloured in red.

(3) The result of setting the parameters (Fig. 15).
The indicator shows that the driver has a poor state.

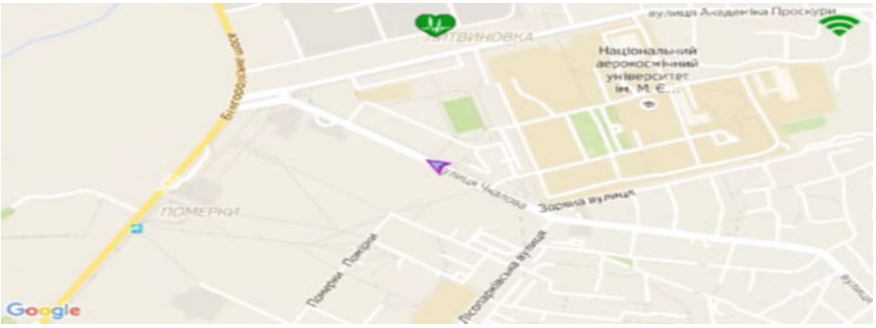


Fig. 13 Starting state of the HMI

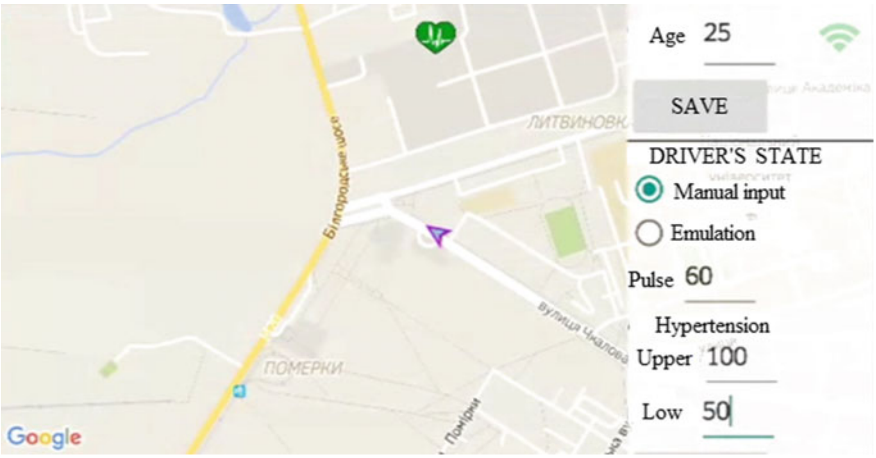


Fig. 14 Set the parameters of the driver. Profile 1

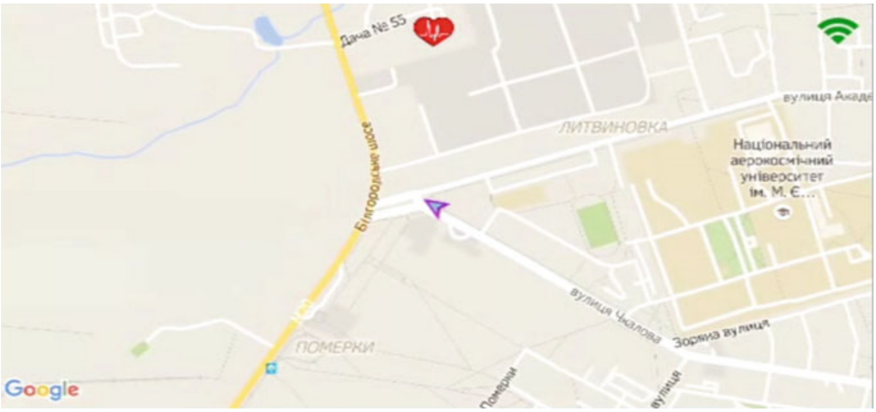


Fig. 15 Indication of the poor driver's state

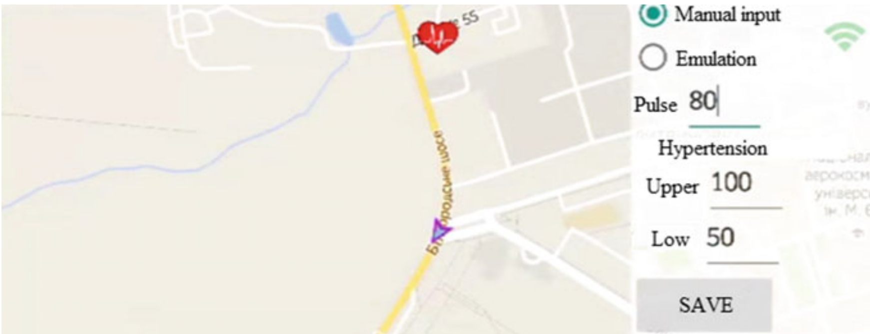


Fig. 16 Set the driver's parameters. Profile 2

(4) Set the parameters of the driver. Experimental profile 2 (Fig. 16):

Age = 25 years;
Pulse = 80;
Upper hypertension = 130;
Low hypertension = 80.

The expected result is the change of the driver's state indicator, the driver's state is good, and the state indicator gets coloured in green.

(5) The result of setting the parameters (Fig. 17).
The indicator shows that the driver has a good state.

Experiment 2. Emulation of the scenario «Dangerous road section».

When driving on the dangerous road section (ice condition, fog, poor road pavement, jam, road accident, reconditioning work) the driver can manually inform about the corresponding hazard using the tools of the HMI (the voice command or the menu) (Fig. 18). In this example, ice condition is reported.

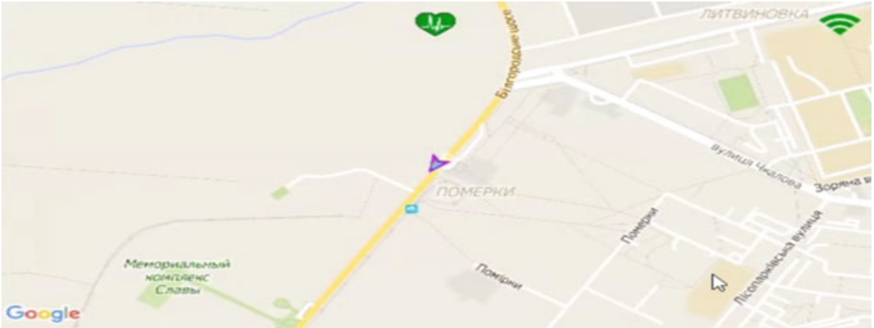


Fig. 17 Indication of good driver's state

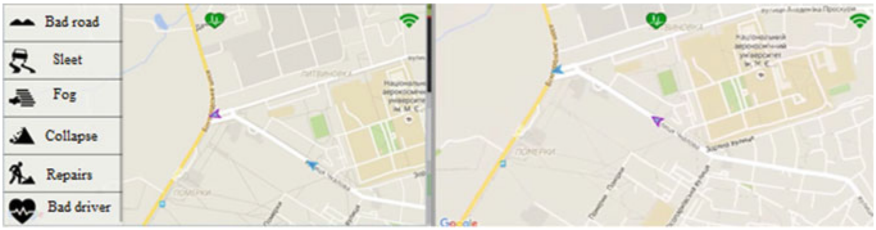


Fig. 18 Manual mode of hazard reporting

The information on the hazard is saved on the server and can be used to form the rod situation as to the other participants. Figure 19 shows that the hazard in question is relevant for the second road user and it is displayed on his HMI.

The road situation is formed on a real-time basis for each participant individually. The hazards are ranked by the server and displayed with the corresponding colours on the client HMI. Rank is assigned according to the distance and the hazard type.

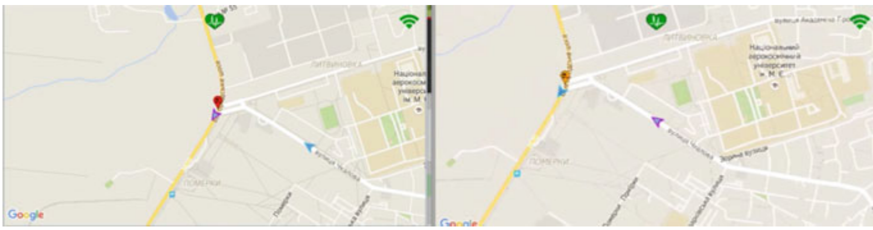


Fig. 19 Manual mode of hazard reporting

8 Conclusions: Future Research

Different vendors on IT market offer advanced driver assistance systems. Such systems as a collision warning system, parking assistant, are designed for improvement of safety when driving and reducing the driver's strain. One of the development lines of such systems is the improvement of the interaction between the driver and the vehicle control system "human–machine" and the provision of the information about the current situation on the road in real time to the driver.

The provision of such information leads to the increased situational awareness of the vehicle driver. The construction of secure dynamic human–machine interfaces has the great importance for enhancing of situational awareness.

The high amount of data used in the ITS leads to the necessity of improving the information access for all traffic participants. Improvement of situational awareness, risk assessment in the real-life improvement requires the use of large computing facilities to store, process and analyze the data. These facilities are not always available, even for modern onboard computing equipment of the vehicle.

Intelligent transport infrastructure includes the complex of equipment that secures obtaining almost full information about the road situation and the possibility of the quick response to the changing conditions. To use such systems effectively, one needs the HMI that maintains human-vehicle interaction and mitigates the negative errors that impact the safety, allows to avoid the misinterpretation of the information that the system provides.

The idea of using cloud services in ITS is just beginning to gain popularity. The facilities of "clouds" can also affect the increase of transport safety.

The cooperative transport systems are of the utmost interest nowadays, and the cooperative HMI stands for one of the trends in this area.

Cooperative HMI provides the measurement values of the parameters of vehicle and driver state in real time via the Internet into the "cloud." Here, the data from all the cars are dynamically processed and transmitted to motoring public.

The prototype of the cooperative human–machine interface for intelligent transport systems based on cloud computing has been developed. This system ensures the improved vehicle safety and the lower number of traffic incidents. The safety improvement is achieved through the driver's awareness of the road situation and possible hazards informed to him on a real-time basis through the human–machine interface. Further research development may include the integration with the onboard computer; improvement of the interface efficiency; testing of the system in real conditions on a number of vehicles.

References

1. Orekhova, N., Kharchenko, V., Orekhov, A.: Human-machine interface quality assessment techniques: green and safety issues. In: Proceeding of the 10th on Digital Technologies International Conference "DT 2014", pp. 259–264. Zilina, Slovakia (2014)

2. Kharchenko, V., Illiashenko, O.: Concepts of green IT engineering: taxonomy, principles and implementation. In: Kharchenko, V. et al. (eds.) Green IT Engineering: Concepts, Models, Complex Systems Architectures. Studies in Systems, Decision and Control, 74, Springer, pp. 3–20 (2017)
3. Green U-Home Terminal. [Online]. Available: http://pacificcontrols.net/products/green-u-home-terminal.html
4. Doukas, N.: Technologies for greener internet of things systems. In: Kharchenko, V. et al. (eds.) Green IT Engineering: Components, Networks and Systems implementation. Studies in Systems, Decision and Control, pp. 23–42. Springer (2017)
5. Safety Vehicle Adaptive Interface Technology (Save-IT). [Online]. Available: http://www.volpe.dot.gov/safety-management-and-human-factors/surface-transportation-human-factors/safety-vehicle-using
6. The Personal Adaptive In-Car HMI: Integration of External Applications for Personalized Use. [Online] Available: http://link.springer.com/chapter/10.1007%2F978-3-642-28509-7_5#
7. Anokhin, A., Ivkin, A.: Human-machine interface to support cognitive operator NPP. Nucl. Meas. Inf. Technol. **1**(41), 57–66 (2012)
8. Anokhin, A., Marshall, E.: Adaptive human-system interface for control of complex systems (in application to nuclear power plant), In: Proceedings of the 21st European Meeting on Cybernetics and System Researches “EMCSR 2012”, pp. 185–188. Vienna, Austria. 10–13 April (2012)
9. Bodyanskiy, Y., Vynokurova, O., Pliss, I., Peleshko, D.: Hybrid adaptive systems of computational intelligence and their on-line learning for green IT in energy management tasks. In: Kharchenko, V. et al. (eds.), Green IT Engineering: Concepts, Models, Complex Systems Architectures. Studies in Systems, Decision and Control, 74, pp. 229–246. Springer (2017)
10. European Statement of Principles, European Statement of Principles for in-vehicle information and communication systems (1999)
11. HASTE, Deliverable 4—Recommended Methodology for a preliminary safety analysis of the HMI of an IVIS (2005)
12. Stoltzfus, J.: Cloud Computing for Vehicles: Tomorrow’s High-Tech Car. [Online] (2012). Available: http://www.techopedia.com/2/28137/trends/cloudcomputing/cloud-computing-for-vehicles-tomorrows-hightech-car
13. Iqbal, A., Pattinson, C., Kor, Ah-Lian.: Introducing controlling features in cloud environment by using SNMP. In: Kharchenko, V. et al. (eds.), Green IT Engineering: Concepts, Models, Complex Systems Architectures. Studies in Systems, Decision and Control, 74, pp. 147–160. Springer (2017)
14. Cloud Computing can Reengineer the Car Interiors. [Online] (2012). Available: http://www.cbrdigital.com/2012/01/16/cloud-computing-canreengineer-the-car-interiors.html
15. Walford, L.: Volvo New Connected Car Features-Magnets, Real-Time Cloud Road Data and Driver Sensing. [Online]. Available: http://www.autoconnectedcar.com/2014/03/volvo-newconnected-car-features-magnets-real-time-cloud-road-datadriver-sensing/ (2014)
16. Sheehan, M Cloud Computing Cars and Mobile Devices. [Online] (2011). Available: http://scoop.intel.com/cloud-computing-cars-andmobile-devices/
17. Kharchenko, V., Orekhov, A., Brezhnev, E., Orekhova, A., Manulik, V.: The cooperative human-machine interfaces for cloud-based advanced driver assistance systems: dynamic analysis and assurance of vehicle safety. In: Proceedings of IEEE East-West Design and Test Symposium “EWDTS’2014”, pp. 82–86. IEEE Kyiv, Ukraine (2014)

Optimisation of Bi-directional Pulse Turbine for Waste Heat Utilization Plant Based on Green IT Paradigm

Yuriy Kondratenko, Serhiy Serbin, Volodymyr Korobko and Oleksiy Korobko

Abstract The problem of preservation of the environment is closely connected with issues of rational use of energy and material resources. In this regard, the development of new energy-saving technologies is a top priority. Analyzing the state of modern energy saving technologies, one can conclude that there is a lack of effective solutions for utilizing waste low-temperature and cryogenic heat resources (450–132) K. Thermoacoustic heat machines (TAHM), can effectively use such energy sources. TAHM are characterized by high reliability, relative low cost and environmental safety. The introduction of thermoacoustic technologies is hampered by the low power of TAHM and the difficulty of directly obtaining mechanical work from acoustic waves. There is the possibility of using bi-directional turbines as a converter of acoustic energy in mechanical work. Such approach will enable the creation of thermoacoustic turbine generators of high power. To determine the optimal design parameters of bidirectional thermoacoustic turbines it is necessary to conduct a complex of experimental and theoretical studies. Article presents the results of experimental research of the working processes of the "bi-directional turbine—electric generator—TAHM" complex. For the registration and analysis of experimental information, a special computer based control system was created.

Y. Kondratenko (✉)
Department of Intelligent Information Systems, Petro Mohyla Black Sea National University, Mykolayiv, Ukraine
e-mail: y_kondrat2002@yahoo.com

S. Serbin
Department of Turbines, Admiral Makarov National University of Shipbuilding, Mykolayiv, Ukraine
e-mail: serhiy.serbin@nuos.edu.ua

V. Korobko
Department of Marine and Stationary Power Plants, Admiral Makarov National University of Shipbuilding, Mykolayiv, Ukraine
e-mail: volodymyr.korobko@nuos.edu.ua

O. Korobko
Research Institute of Ecology and Energy Saving, Admiral Makarov National University of Shipbuilding, Mykolayiv, Ukraine
e-mail: oleksii.korobko@nuos.edu.ua

© Springer Nature Switzerland AG 2019

V. Kharchenko et al. (eds.), *Green IT Engineering: Social, Business and Industrial Applications*, Studies in Systems, Decision and Control 171,
https://doi.org/10.1007/978-3-030-00253-4_20

System includes MCU-based pulse-phase control subsystem for thermoacoustic engine heater, data acquisition subsystem and SCADA-based PC software for data storage and analysis. Designed with green IT approach paradigm control system is described in detail. Initial processing of measurement results was carried out in real time with the help of special software. The results of physical experiments allowed determining the boundary conditions for computational fluid dynamics (CFD) simulation of work processes in the bi-directional turbine. The peculiarities of the medium hydrodynamics in the TAHM resonator during the work of the bi-directional turbine at various load conditions were studied. Based on the results of CFD modeling the structural optimization of turbine is proposed.

Keywords Specialized computer system · Thermoacoustic processes
CFD modelling · Waste-heat utilization · Green IT approach

1 Introduction

The existence of modern society is based on strong industrial potential. Industry, energy and transport produce large volumes of thermal emissions that have a negative impact on the environment. The problem of rational use of energy resources remains valid for many years. Efficient heat engines and new energy saving processes are used to reduce energy consumption in different areas of modern society.

The technologies of thermal emissions re-usage, such as utilization or regeneration of thermal waste are widely developed. Usually, the highest attention is paid to the utilization of high-temperature thermal resources, the temperature of which exceeds 300 °C. In these cases it is possible to apply technologies of waste heat regeneration and implement cogeneration or trigeneration schemes. Currently, by improving the energy efficiency of thermal energy primary consumers, the temperature of thermal emissions decreases. With waste heat temperature below 250 °C the use of traditional energy saving schemes is complicated or becomes economically unreasonable. These circumstances determine the need to develop new technologies for energy conservation and environmental protection.

One of the promising ways of utilization of low-temperature heat energy is the usage of thermoacoustic heat machines (TAHM):

- thermoacoustic engines (TAE);
- thermoacoustic refrigerators or heat pumps (TAHP).

TAHM are characterized by high reliability, relative low cost, environmental neutrality [1, 2].

The introduction of thermoacoustic technologies is hampered by the low power of TAHM and the difficulty of the direct gain of mechanical work from acoustic waves. There is the possibility of the bi-directional turbines usage as a converter of acoustic energy in mechanical work. Such approach will enable the creation of thermoacoustic low-temperature turbine generators of high power.

To determine the optimal design parameters of bidirectional thermoacoustic turbines it is necessary to conduct a complex set of experimental and theoretical studies. An important role in these processes belongs to IT technologies. These technologies are useful for development of comprehensive solutions for technological and environmental issues. Particular attention deserves the green technologies area which is primarily focused on the implementation of environmentally friendly approaches and allows synthesizing effective integrated solutions based on various optimized and energy-efficient techniques.

The structure and main components of thermoacoustic experimental plant with consideration to stated approaches is discussed next.

2 Thermoacoustic Plant for the Low Temperature Heat Resources Utilization

There are different types of thermoacoustic plants that are driven by external heat sources or acoustic generators that are installed into resonator. TAHM with acoustic generator will not be considered in this article as their work is based on supplied electrical energy and thus they are less efficient for heat utilization problems.

2.1 Principal Scheme of TAHM Based Utilization Plant

Schematic diagram of thermoacoustic system for the use of waste heat resources is shown in Fig. 1. This plant is able to work with different sources of thermal energy, including emissions from industry and heat engines and thermal energy of solar systems. Recipient for waste thermal energy may be the environment or cryogenic substances, which is a significant advantage of thermoacoustic technology. It is promising to use TAHM in regenerative energy schemes, in particular in solar installations.

The proposed thermoacoustic installation consists of several subsystems:

- External subsystems, that provide energy exchange between the TAHM, the source and the recipient of the waste heat;
- Internal subsystem, the TAE itself, in which the transformation of thermal energy into acoustic is happening;
- Subsystem of control and monitoring of work processes in TAHM and external systems [3];
- The load of TAE, which can be a heat pump or a special converter of acoustic energy to mechanical work or electrical energy [1, 4, 5].

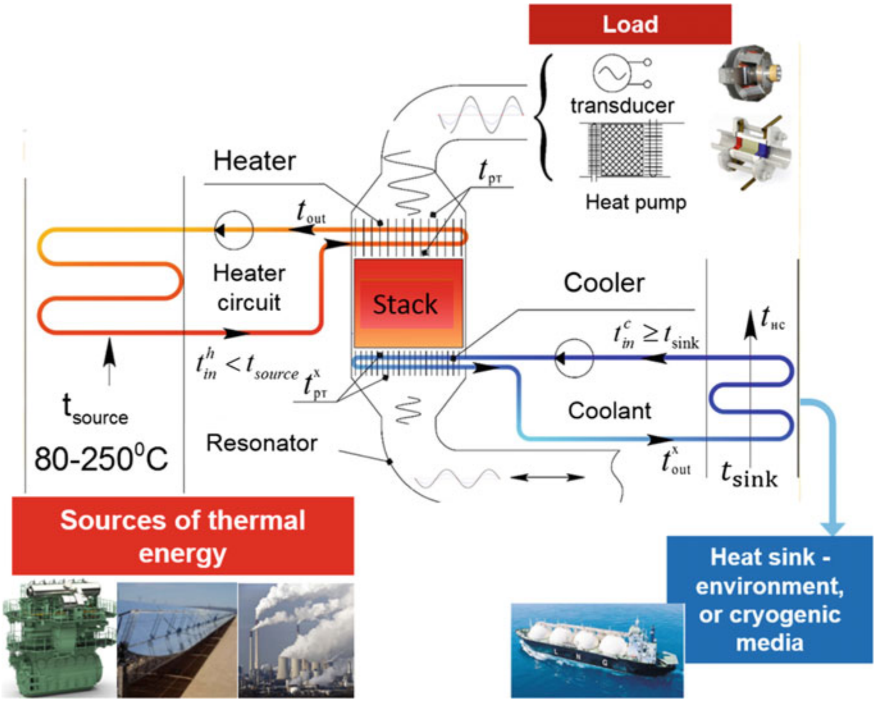


Fig. 1 Principal scheme of thermoacoustic plant for the use of low temperature heat resources

Examples of existing low-temperature thermoacoustic units with a thermal power up to 100 kW are already created. But the issue of the effective conversion of acoustic energy into electric energy has not yet been resolved. There is a number of known examples which are using piezoceramic [4] or electrodynamic transducers [5]. These transducers have details that are exposed to mechanical deformations. In practice, such structures have shown some significant disadvantages: low reliability and high cost.

According to [5], the cost of linear generator reaches 3,000 euros per 1 kW of electrical power. These circumstances require the development of new, more reliable types of load for TAHM.

2.2 *Bi-directional Turbine in Renewable Energetics and Thermoacoustic*

Possible variant of load for TAE can be bi-directional turbines that can transform longitudinal oscillations of the working environment into rotational motion.

There are several wave-triggered power stations (Fig. 2) with oscillating air columns OWC (Oscillating Water Column). In these installations, the fluctuations of the air pillar allow the operation of special types of turbines, such as the Welsh turbines and pulsed bi-directional turbines [6].

Since the TAHM has the same oscillating motion in the working environment, it is logical to use the bidirectional turbines for mechanical work [7–9]. It is important to take into account the significant differences in the parameters of the working environment. Comparison of parameters of working environment (1), (2) in these facilities is given in Fig. 3 [10].

OWC are open systems that operate under atmospheric pressure. The amplitude of the oscillating air ζ in the channels reaches several meters. In this case, the frequency of oscillations of the air is determined by the rapidity of the water column movement, which is usually not more than 0.2–1 Hz [6].

In TAHM, the situation is significantly different, it is a closed system in which the pressure of the working environment can reach 0.3–3.0 MPa.

Fig. 2 Wave power station with oscillating water column

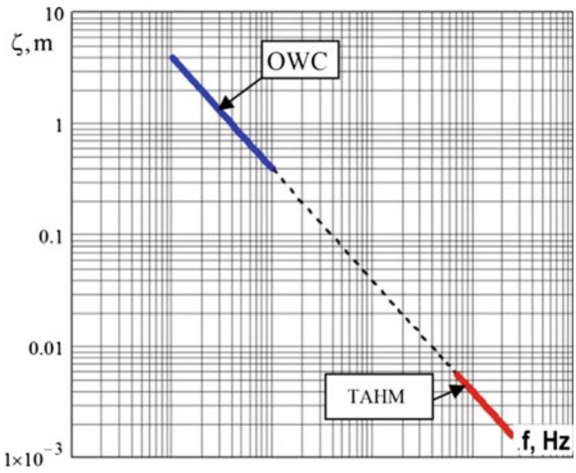


Fig. 3 Amplitude of the oscillating motion of the medium depending on the wave frequency

The frequency of oscillations is determined by the length of the resonator, and more often that is 50–150 Hz, respectively the amplitude of the oscillatory motion is much smaller.

$$u_1 = \frac{p_s}{\rho c}, \quad (1)$$

$$\zeta = u_1 / 2\pi f, \quad (2)$$

where p_s—the amplitude of the pressure of the acoustic wave (Pa), u_1—oscillatory velocity of the medium (m/s); ρ—density of the working environment (kg/m^3), c—sound speed in the working environment (m/s); f—frequency of acoustic oscillations (Hz).

At the moment, there is a lack of experimental data on the operation of bidirectional turbines under TAHM conditions.

2.3 Formulation of the Research Problem

It can be assumed that the interaction of the turbine and the pulsating medium forms a complex hydrodynamic situation at the entrance to the turbine nozzle devices, which has a significant effect on the characteristics of the turbine [11]. The operation of the turbine in the TAHM resonator is likely to result in intense secondary currents. These circumstances put some doubts on the possibility of using the linear theory of thermoacoustics [1] in this case.

These factors induce to carry out a complex of experimental studies which should solve a number of issues, namely:

- study the peculiarities of hydrodynamic and thermophysical processes that are taking place during the work of pulsed bi-directional turbines under TAHM conditions;
- determine the constructive factors that influence the characteristics of bidirectional turbines;
- investigate the work of "TAE—bidirectional turbine—electric generator" plant;
- determine the optimal design parameters of bidirectional turbines for thermoacoustic heat-generating machines;
- experimentally determine the boundary conditions required for CFD modeling of hydrodynamics of the oscillating medium in a bidirectional turbine.

Such information could be handy during the design process of efficient thermoacoustic turbine generators.

To suit the demands of stated tasks the MCU based data acquisition system with SCADA PC software for data storage and analysis was designed in accordance with the green IT approach paradigm [12, 13].

The main objective of the current research is to improve the existing control system and adapt it to study the working processes in thermoacoustic turbomachines.

3 Experimental Investigation of Bi-direction Turbine

A model of bi-directional turbine was developed for carrying out planned experimental research. This model was designed in accordance with the recommendations given in the papers [7, 8]. In the future, this approach should facilitate comparisons of experimental data.

3.1 Experimental Model of Turbine

Structural elements of a bidirectional turbine (a rotor with blades, nozzle grids and guide vanes) were made by the method of 3D-printing. In this case the PLS plastic was used as a fabric (Fig. 4).

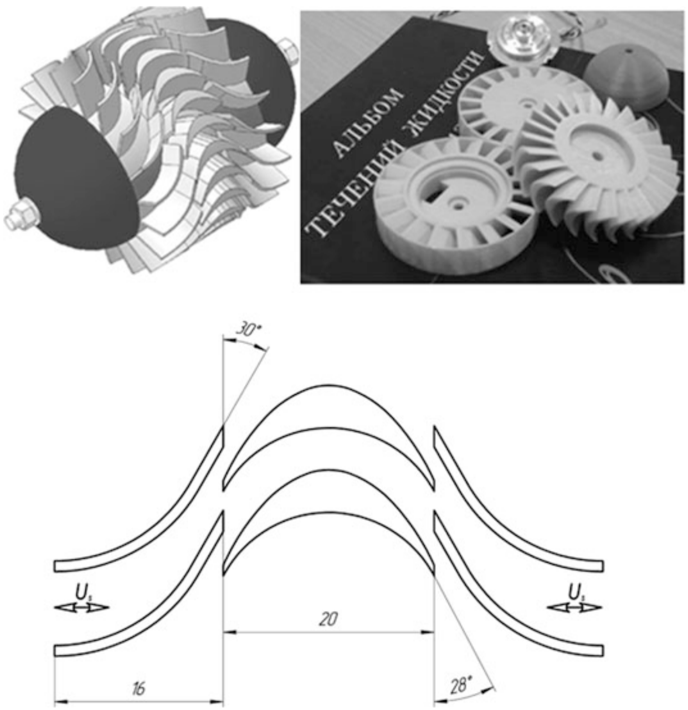


Fig. 4 Design of experimental bi-directional turbine

The nozzles of the turbine have 18 blades and the rotor has 24 blades. The outer diameter of the bi-directional turbine was 70 mm, the diameter of the rotor was 50 mm.

The length of the chord of the blades of the rotor was equal to the double amplitude of the oscillatory motion of the medium. This corresponds to the frequency of 20 Hz at amplitude of acoustic pressure at the level of 5000–6000 Pa [9].

In the rotor of a bi-directional turbine, a three-phase, full-cycle electric generator of the Sankyo F2JGL type was installed. The output voltage from the generator was applied to the rectifier, which was made by using Schottky diodes. Laboratory rheostat of type P 517-M was used as a load.

The turbine frequency was determined using the digital oscilloscope SDS 1074CFL which was connected to one of the generator windings, this connection allowed simultaneously to control the amplitude and shape of the signal at the output of the electric generator.

3.2 The Experimental Setup for Study of Bidirectional Turbine Performance

For the study of bidirectional turbine performance two experimental setups with digital data acquisition system were used. Their schemes are shown on Figs. 5 and Fig. 6.

The first setup (Fig. 5) is a quarter wave resonance tube, equipped with four high-speed pressure sensors (P1–P4) of type MPXV7007DP, low-frequency loudspeaker, and a hot-wire probe of thermoanemometer (DISA) with a coordinate mechanism. Acoustic oscillations were generated using a loudspeaker [14]. The control signal was generated by self-designed software and then tweaked to a high-precision amplifier, which allowed setting the signal frequency and acoustic power in the TAHM cavity.

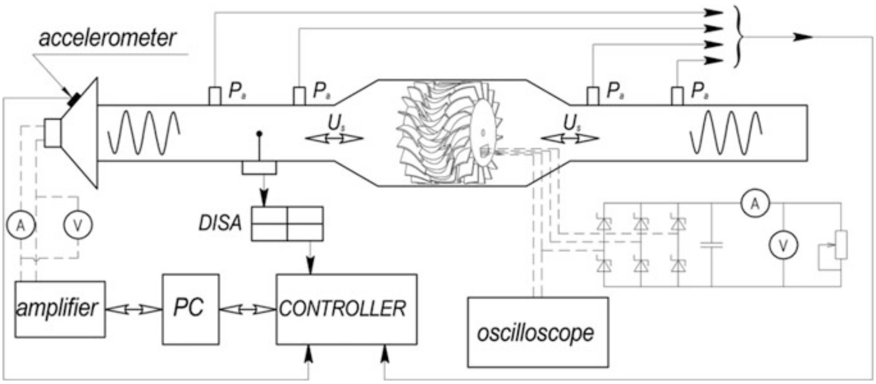


Fig. 5 The experimental setup on the basis of the resonance tube

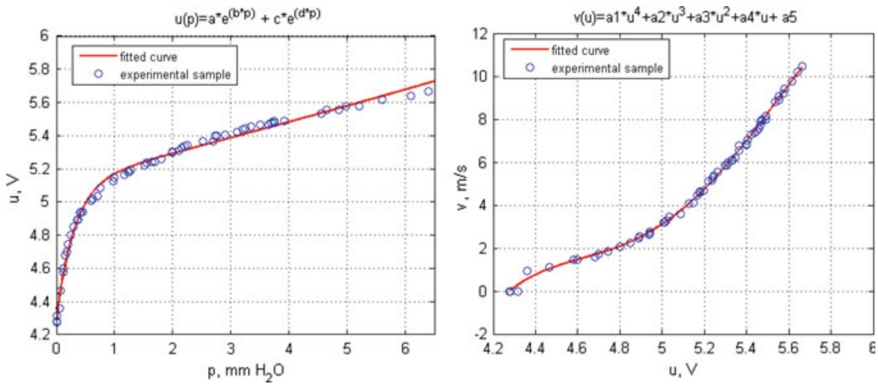


Fig. 6 DISA 5500 probe speed model fitting

Acoustic power, at different parts of the resonator, was determined by a modified two sensors method [15], and by direct measurement of oscillatory velocity and acoustic pressure. To measure the oscillatory speed of the medium, the hot wire anemometer DISA 5500 was used. Its output signal was gathered and processed by MCU. To measure the velocity field in the intersection plane, the probe of the thermoanemometer was able to change its position along the height of the channel thanks to a special coordinate mechanism. With the usage of data acquisition system thermoanemometer was calibrated to measure both the flow velocity and its temperature. Figure 6 shows the fitted model (3) of the probe speed as a function of output voltage $v = f(u)$ that was obtained using the regression methods.

$$v = -6.226u^4 + 126.8u^3 - 959.7u^2 + 3208u - 4002 \quad (3)$$

A bi-directional turbine is a complicated load for an acoustic system, its impedance can depend on the speed, load, and other factors. In this regard, the feedback between the speaker system and the loudspeaker can lead to a distortion of the form of the acoustic wave front, which will affect the turbine working efficiency.

To determine this possible effect and its prevention the system was enhanced by the control of the loudspeaker diaphragm movement with the help of an accelerometer sensor ADXL321 [15].

It should be noted that design allows the plant to change its overall length and location of the turbines along its resonator.

In the second experimental installation (Fig. 7), the TAE operated in a standing wave mode, was used, implementing the Brighton cycle [1]. The design of the TAE made it possible to change the length of the resonator and, accordingly, the frequency of the acoustic wave.

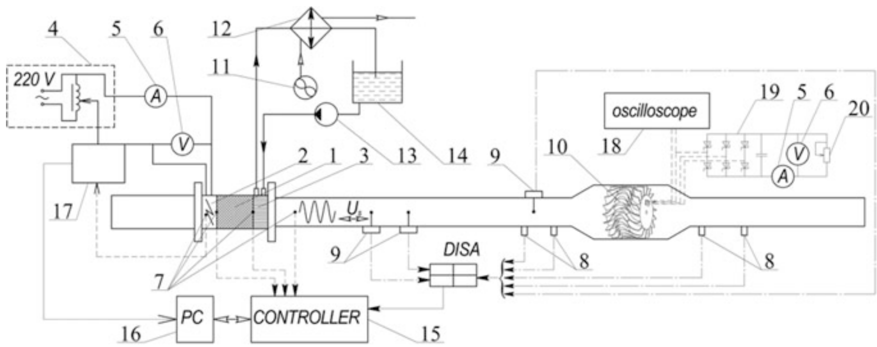


Fig. 7 Schematic of the experimental setup on the basis of TAE

The collection and processing were carried out by designed digital control system [16] that includes:

- nichrome heater that is used as a source of thermal energy and is driven by the supplied from the controlled power supply unit electrical energy (P_{SUPL});
- current sensor (CS) ACS712ELCTR-05B-T and the voltage divider (VD);
- two pressure sensors MPXV7007DP, which are situated at the Δx = 150 mm distance from each other and are used to calculate the acoustic parameters by means of modified two sensor method [15, 17];
- thermocouple (TS) for control of thermal data;
- STM32 Cortex M4 based microcontroller for data acquisition and analysis

It should be noted that for the determination of acoustic parameters system is using the complex form. In combination with implemented digital signal processing algorithms, it allows the intensity of the sound [1] wave to be calculated by the expression (4).

$$I = \frac{1}{2}\mathrm{Re}(\tilde{p}v), \tag{4}$$

where $\tilde{p}$ is the complex conjugate value of acoustic pressure *p(t)*, *v* is the particle velocity of the environment in which sound is distributed and Re() denotes the real part of complex number.

Such approach avoid the difficulties associated with the phase difference θ determination, because the oscillating acoustic pressure and velocity are determined by the expressions (5), (6), and their complex modulus and phase respectively calculated using expressions (6) and (7).

$$\begin{gathered} p(0) = \frac{p_A + p_B}{2\cos(k\Delta x/2)}, \\ k = k_0\left\{1 + \frac{1-j}{2}\frac{\delta_v}{r_0}\left(1 + \frac{\gamma-1}{\sqrt{\sigma}}\right)\right\}, \end{gathered} \tag{5}$$

$$v(0) = \frac{kF}{\omega\rho}\frac{p_A - p_B}{2j\,\sin(k\Delta x/2)},\quad F = 1 - \frac{1-j}{r_0/\delta_v}, \tag{6}$$

$$|x| = \sqrt{\frac{1}{N}\sum_{k=1}^{N} x_k^2}, \tag{7}$$

$$\varphi(x) = \sqrt{\frac{1}{N}\sum_{k=1}^{N} \left[arctg(\mathrm{Im}(x_k)/\mathrm{Re}(x_k))\right]^2}, \tag{8}$$

where k is the complex wave number, Δx is the distance between pressure sensors A and B, p_A, p_B are the A and B pressure sensors values, $k_0 = \omega/a$ is the wave number in free space, σ is the Prandtl number, δ_v is the viscous boundary layer thickness, r_0 is the duct radius, ρ is the average gas density, j is the imaginary unit, $\omega = 2\pi f$ is the sound wave angular frequency, γ is the adiabatic index of gas in resonator, N is the number of samples.

The designed research system allows the communication between all components via RS232 network, also to enable the transfer of large amounts of data, MCU can communicate with PC via USB. This grants maximum flexibility to designed distributed computer system. Usage of CortexM4 MCU module with built-in floating point processor allowed moving all complex computations from PC software to low-power MCU, thus lowering the overall energy consumption of the system.

Created software for PC allows storing the received from the controller information in the archive database and display of all available information on the user's screen.

Controlled voltage regulator of electric heating element system is based on a pulse-phase control (PPC) realized in the MCU. The triac BT137 is used as a managed element in the circuit. The triac [18] based system of pulse-phase control forms the rectangular pulses. The phase [19] of this pulses, in relation to the supplied to triac power, is changing proportionally to the voltage supplied to the control body PPC [14]. Control unit of PPC limits the minimum and maximum angles regulation, and set the initial angle adjustment. A detailed description of control system and software is provided in [3, 14, 15].

Indications on Fig. 7 are as follows: 1—stack; 2—heater; 3—cooler; 4—power supply unit; 5, 6—amperemeter and voltmeter; 7—thermocouples; 8—pressure sensors; 9—probes of thermoanemometer; 10—bi-directional turbine; 11—fan; 12—cooler; 13—pump; 14—cooler tank; 15—MCU; 16—PC; 17—controlled voltage regulator; 18—oscilloscope; 19—diode bridge; 20—resistive load.

Turbine has always been located in the nodes of acoustic pressure, where the velocity and amplitude of the oscillatory motion are highest.

It also should be noted that system can work in silent mode without human operator and without human–machine interface module, thus making it greener by shorten the power consumption. In future authors want to consider the IoT [13, 20] implementation of digital control system with cloud based server database and web based user interface. This approach will allow eliminating the need for dedicated computer on site thus decreasing the energy consumption and making system more flexible.

3.3 The Results of Experimental Studies

Figure 8 shows the dependency of the frequency of the turbine rotor rotation on the frequency of the acoustic wave and acoustic power (Fig. 8). During an experiment on a loudspeaker at constant power a signal with different frequencies was applied. The load on the generator of the bidirectional turbine was also constant.

During the experiments it was found that the bi-directional turbine rotational frequencies have several maxima at different acoustic frequencies. At the same time, there is also an increase in acoustic power, which confirms the presence of resonant effects.

The obtained results allow us to conclude that for a bi-directional turbine the amplitude of the oscillatory motion is not a decisive parameter. For such turbines, acoustic pressure and oscillatory speed are more important than other parameters. This makes it possible to create bi-directional turbines for higher frequencies, which will significantly reduce their size.

Unfortunately, the generator that was used in the turbine had a low power. Therefore, a study with greater acoustic power was not successful.

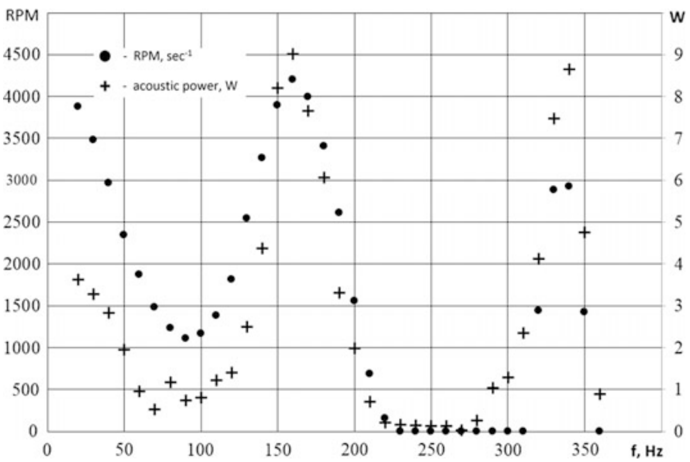


Fig. 8 The dependence of acoustic power turbines frequency and the frequency of the acoustic waves

This experimental setup had a fixed length and provided maximum acoustic power only at its own resonant frequencies. It shows the results of measurements, which clearly show the local maximum of acoustic pressure (acoustic power). At this points rotations speed of turbine have maximum values.

Figure 9 shows the results obtained during the operation of a bi-directional turbine in combination with TAE. In these experiments, a different installation was used (Fig. 7), due to which the work of the turbine at different frequencies was studied in details. In this installation, the turbine was mounted directly into the resonator of TAE, the length of which can be changed, thereby ensuring a change in the working frequency of the TAE.

The experiments were carried out at constant load of turbine generator. In addition, by the changes of the voltage supplied to the TAE heater the intensity of acoustic energy in the resonator was kept constant. By changing the length of the resonator, it was possible to investigate the effect of operational frequency on the characteristics of the turbine. The results of measurements showed (Fig. 9) that in the frequency range 75–250 Hz the revolutions of the bi-directional turbine were kept in the range of 4200–4500 rpm, which confirms the results of previous studies.

Consequently, a bi-directional turbine is capable of efficient operation at frequencies that are typical to TAHMs (50–150 Hz) and ensures the operation of an electric generator or other load.

The result obtained is very important in view of the possible practical use of TAHM. In turn, this confirms the expediency of in-depth study of the peculiarities of hydrodynamic processes in the system "TAE—Bi-Directional turbine".

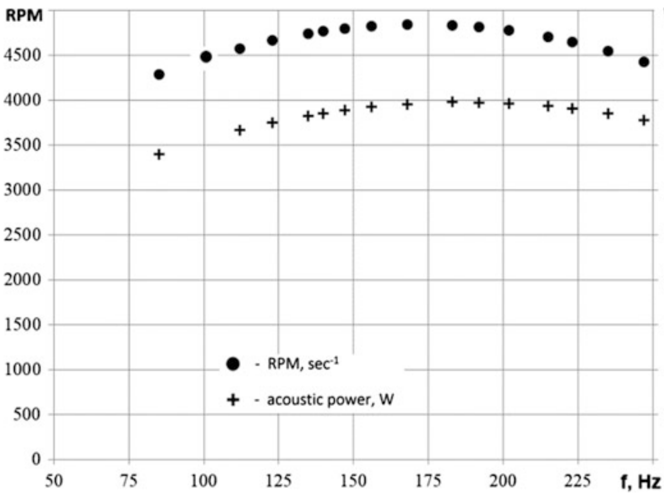


Fig. 9 Dependence of the speed of the turbine on the frequency of the acoustic wave

In addition to physical experiments, the research was conducted using CFD simulation methods [10, 21]. For this purpose, a 3D CFD model of a bidirectional turbine was built. The calculations were carried out in the 3D form. Experimental data was used to determine the boundary conditions.

The simulation results show (Fig. 10) that intense secondary currents are formed in the cavity region directly in the turbine area, which should affect the efficiency of the rectifying apparatus and the turbine as a whole.

At the same time, the physical experiment showed that in the TAE resonator at a distance of more than 2 diameters secondary circumferential currents lose their intensity and practically do not affect the structure of the environment. This can be evidenced by the absence of deformation of the front of oscillograms of acoustic pressure (Fig. 11).

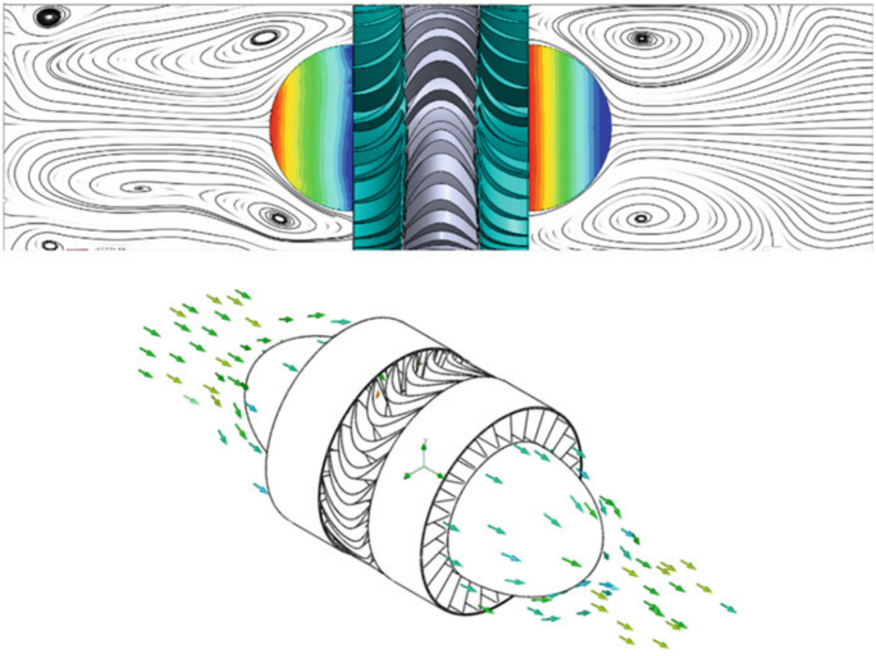


Fig. 10 The lines of flow in the TAE cavity at the area of turbine

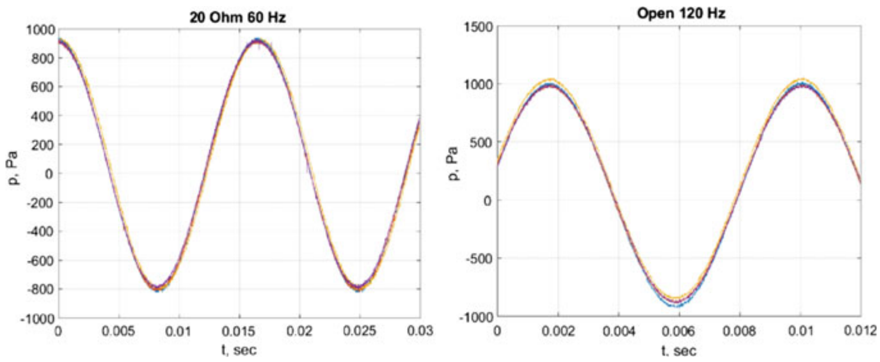


Fig. 11 Acoustic pressure oscillations at different cross sections of the resonator

4 Conclusions

Considering the taxonomy of green computing [12, 20, 22] it can be concluded that problem of thermoacoustic plants optimization applies to «greening» of IT-systems itself by increase of their efficiency. Also, the implementation of TAHM technologies into waste heat utilization of energy complexes leads to support of environmental sustainability by lowering overall heat emissions.

As a result of the research, the efficiency of bi-directional pulse turbine (BPT) usage in the thermoacoustic system has been confirmed. It is shown that turbine can work efficiently over a wide range of frequencies, enabling the creation of compact and powerful systems.

The control and measuring system is capable of conducting research on hydrodynamics and thermophysical processes in the complex "TAHM—Bidirectional Turbine—Generator". Furthermore, usage of digital signal processing and regression methods along with real time control algorithms implemented in MCU allowed creating low-powered scalable digital system that increases overall efficiency of TAHM.

The TAE complex with the BPT driven generator can become a promising technical solution for waste and reclamation energy resource utilization systems.

A classical model of a flat-front acoustic wave does not reflect all the peculiarities of a hydrodynamic picture in a TAHM resonator with bi-directional pulse turbine.

In future work authors should focus on the design of the rectifying apparatus that should prevent the appearance of secondary radial currents in the oscillating medium and thus increase the efficiency of waste heat utilization system.

References

1. Swift, G.W.: Thermoacoustics—A Unifying Perspective for Some Engines and Refrigerators. Acoustical Society of America (2002)
2. De Blok, K.: Low operating temperature integral thermo acoustic devices for solar cooling and waste heat recovery. In: Proceedings of Acoustic-2008 International conference, pp. 18–24, Paris (2008)
3. Kondratenko, Y., Korobko, V., Korobko, O.: Specialized computer system for control and parametric optimization of thermoacoustic processes. In: Computing in Science and Technology. Monographs in Applied Informatics, pp. 38–54. Warsaw University of Life Sciences & University of Rzeszow, Warsaw (2013)
4. Keolian, R. M.: Truck Thermoacoustic Generator and Chiller. Final Report, 190 p. Applied Research Laboratory, The Pennsylvania State University (2011)
5. De Blok, K., Owczarek, P., Francois, M.: Acoustic to electric power conversion. Aster Thermoacoustics (2014)
6. Thakker, A., Hourigan, F.: Modeling and scaling of the impulse turbine for wave power applications. Renew. Energy **29**(3), 305–317 (2004)
7. De Blok, K.: Bi-directional turbines for converting acoustic wave power into electricity. http://www.aster-hermoacoustics.com. Access December (2017)
8. Kloprogge, T.: Turbine design for thermo-acoustic generator: Master's thesis, 72 p. Hogeschool. Holland Delft, Aeronautical Engineering (2012)
9. Korobko, V.V., Moskovko, O.O., Voronov, I.V.: Study of features if bi-directional turbines in thermoacoustic machines. Scientific Bulletin of Kherson Sea State Academy **1**(14), 201–208 (2016)
10. Korobko, V.V.: Investigation of heat transfer and hydrodynamics in thermoacoustic engine components. Aerosp. Techn. Technol. **8**(105), 123–130 (2013). ISSN 1727-7337
11. Matveev, I.B., Serbin, S.I., Vilkul, V.V., Goncharova, N.A.: Synthesis gas afterburner based on an injector type plasma-assisted combustion system. IEEE Trans. Plasma Sci. **43**(12), 3974–3978 (2015)
12. Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.): Green IT engineering: concepts, models, complex systems architectures. In: Studies in Systems, Decision and Control, vol. 74. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-44162-7
13. Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.): Green IT Engineering: Components, Networks and Systems Implementation. Studies in Systems, Decision and Control, vol. 105. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-55595-9
14. Kondratenko, Y., Korobko, V., Korobko, O., Kondratenko, G., Kozlov, O.: Green-IT approach to design and optimization of thermoacoustic waste heat utilization plant based on soft computing. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Components, Networks and Systems Implementation. Studies in Systems, Decision and Control, vol. 105, pp. 287–311. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-55595-9_14
15. Kondratenko, Y., Korobko, V., Korobko, O., Gerasin, O.: Pulse-phase control system for temperature stabilization of thermoacoustic engine model driven by the waste heat energy. In: Proceedings of the 8th IEEE International Conference IDAACS'2015, Warsaw, vol. 1, pp. 58–61 (2015). https://doi.org/10.1109/IDAACS.2015.7340701
16. Kondratenko, Y.P.: Robotics, automation and information systems: future perspectives and correlation with culture, sport and life science. In: Gil-Lafuente, A.M., Zopounidis, C. (eds.) Decision Making and Knowledge Decision Support Systems. Lecture Notes in Economics and Mathematical Systems, vol. 675, pp. 43–56. Springer International Publishing, Switzerland (2015). https://doi.org/10.1007/978-3-319-03907-7_6
17. Fusco, A.M., Ward, W.C., Swift, G.W.: Two-sensor power measurements in lossy ducts. J. Acoust. Soc. Am. **91**, 2229–2235 (1992)
18. Fox, R.W.: Practical triac/SCR projects for the experimenter, G/L Tab Books, 191 p (1974)

19. Prokopenya, A.N.: Simulation of a quantum algorithm for phase estimation. Programming and Computer Software **41**(2), 98–104 (2015)
20. Bardis, N.: Secure, green implementation of modular arithmetic operations for IoT and cloud applications. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Components, Networks and Systems Implementation. Studies in Systems, Decision and Control, vol. 105, pp. 43–64. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-55595-9
21. Raghvendra, G., Fletcher, D.F., Haynes, B.S.: On the CFD modelling of Taylor flow in microchannels. Chem. Eng. Sci. **64**(12), 2941–2950 (2009). ISSN 0009-2509. https://doi.org/10.1016/j.ces.2009.03.018
22. Kharchenko, V., Illiashenko, O.: Concepts of Green IT engineering: taxonomy, principles and implementation. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Concepts, Models, Complex Systems Architectures. Studies in Systems, Decision and Control, vol. 74, pp. 3–19. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-44162-7_1

Part IV
Social, Educational and Business Aspects of Green IT Systems Application

Cyber-Social Computing

Vladimir Hahanov, Oleksandr Mishchenko, Tetiana Soklakova, Vugar Abdullayev, Svetlana Chumachenko and Eugenia Litvinova

Abstract The structures of cyber-social computing are proposed, which are considered as components of cloud-driven technologies for exact monitoring and moral governance of the society. The main trends of the development of the cyber-physical structure presented in Gartner's Hype Cycle 2017 are described to apply them in science, education, transport, industry and state structures. Prospective directions of the market-feasible technologies, related to green cyber-social monitoring and management of society, are proposed. An expanded description of technologies focused on the creation of the smart digital world, green cities and 5G telecommunications is performed. Recommendations are given for leveraging the top 10 technologies of 2017 in business, scientific and educational processes of higher education. Technologies of emotional-logical computing, a metric for measuring social relations along horizontal and vertical connections, rules of human behavior, focused on creating the emotional logic for modeling and simulation human behavior are proposed. A cyber-physical model of green statehood for the metric management of resources and citizens is introduced; it is based on digital monitoring and assessment of the needs of citizens, including compo-

V. Hahanov (✉) · O. Mishchenko · T. Soklakova · S. Chumachenko · E. Litvinova
Design Automation Department, Kharkov National University of Radio Electronics, Nauka Avenue 14, 61166 Kharkiv, Ukraine
e-mail: hahanov@icloud.com

O. Mishchenko
e-mail: santific@gmail.com

T. Soklakova
e-mail: tetiana.soklakova@gmail.com

S. Chumachenko
e-mail: svetachumachenko@icloud.com

E. Litvinova
e-mail: litvinova_eugenia@icloud.com

V. Abdullayev
Azerbaijan State Oil and Industry University, Baku, Azerbaijan
e-mail: abdulvugar@mail.ru

© Springer Nature Switzerland AG 2019

V. Kharchenko et al. (eds.), *Green IT Engineering: Social, Business and Industrial Applications*, Studies in Systems, Decision and Control 171,
https://doi.org/10.1007/978-3-030-00253-4_21

nents of social computing (relations, goals, management, personnel, infrastructure, resources), ordered by the degree of their influence on the market success.

Keywords Hype cycle emerging technologies · Digital green world
Cyber-physical computing · Human behavior simulation · Smart digital state
Cloud-driven computing · Emotional-logical computing · Cyber-social computing

1 Introduction

The key to the humanity future on a green planet contains five components: (1) The digital world. (2) New materials. (3) Genetic engineering. (4) Mutation of living. (5) Sun-energy. The Nature has a Genome or algorithm of the development that involves the existence of the beginning and the end. The genome of the development has all the components of nature: The Universe and its cosmological subjects, terrestrial nature, which includes living things, physical objects (natural and artificial), and humanity. To fulfill its own genome, humanity, as a part of green nature, has powerful self-regulation or immunity mechanisms or moral relations, which are intended to save life on the planet through combating numerous social viruses.

The motivation of the research is defined by high market feasibility of digital monitoring cyber-social management of the society, aimed at improving people's quality of life and preserving the planet's ecology. Humanity has learned to manage the physical processes and objects, but there are practically not moral technologies for metric management of human and social groups. Lack of competent decisions of cyber-social computing is the cause of social corruption, pollution of the environment, conflicts and local wars.

The state of the art in this field combines the historical experience of social groups management with the scientific and technological achievements of cyber-physical cloud computing. There are publications aimed at the formation of a new cyber culture of humanity, which is intensively implemented in all fields of human activity. Gartner's analytical reports, as well as the basic platforms for cyber-physical monitoring and management, such as IoT, cloud computing, embedded microsystems, big data analytics, blockchain, mobile computing [1–12], are important. There are publications, which consider digital monitoring and metric management of social processes and phenomena in the development of scalable cyber-physical systems and components, such as digital universities, electronic states, electronic document management, smart cities, electronic voting, cloud services for making socially significant decisions [13–21]. Practically all mentioned research has significant limitations related to the formation of data, which are further subjectively analyzed by a person to make a final decision. Subsequent researches are aimed at the maximum exclusion of the officials from monitoring and management cycle of resources, personnel, socially significant processes, focused on the destruction of wars and corruption.

The aim of the work is to develop logical components of cyber-social computing based on moral, social relationship for accurate cloud management of social processes, by leveraging big data, metric online monitoring of the social groups opinion in order to improve the quality of human life and preserve the green ecology of the planet.

Objectives are: (1) Introduction: definitions the structure of green management by social groups on the basis of moral public relations to prevent social conflicts and cataclysms. (2) The three main fields of the cyber culture for managing physical and social processes presented by Gartner. (3) The practice of leveraging green top technologies of cyberculture 2017 for universities, companies and states. (4) Emotional-logical computing: structure and components. (5) Cyber-physical statehood, as the future structure of the green social world. (6) Models of human relations and behavior.

Computing is a technological branch of knowledge focused on the theory and practice of computer exact metric management of virtual, physical and social processes and phenomena based on exhaustive big data and sensor-driven digital monitoring [13].

Cyber-physical computing is the theory and practice of cloud computing, metric actuative manage of virtual, physical and social processes and phenomena based on an exhaustive big data and sensor-driven digital monitoring [13].

Internet of Things is a global, scalable technological way or cyber culture of moral cloud computer metric actuative management of virtual, physical and social processes and phenomena based on an exhaustive big data and sensor-driven digital monitoring to ensure the quality of life and preserve the planet's ecology [13].

Cyber-Social Computing is the theory and practice of the moral, social relationship for precise cloud management of virtual and social processes and phenomena based on the exhaustive leverage of big data, metric online monitoring opinions of the social groups in order to improve the quality of people's life and preserve the planet's ecology.

There are certain trends in the green world, which provide the technological basis for creating cyber-social computing, as an analogue of cyber-physical computing, under umbrella of Internet of Things. Gartner Inc., which is creating a global technology, cyber-fashion, added eight new trends to its Hype Emerging Technologies Cycle brand in 2017 (Fig. 1): 5G, Artificial General Intelligence, Deep Learning, Deep Reinforcement Learning, Digital Twin, Edge Computing, Server-less PaaS, and Cognitive Computing [1, 2].

Edge computing is a technology aimed to improve the performance of cloud services by leveraging special intelligent computational procedures at the location of the mobile user or embedded microsystem. Digital Twin creates cyber images of physical processes and phenomena. As in the mirror, if there is no reflection of a digitized company, (university) in the cyberspace, then it is not in the physical space. Serverless PaaS is hardware-free architecture for organizing cloud-computing processes based on a platform as a service (Platform as a Service). The economy of the cloud platform is clearly advantageous in comparison with

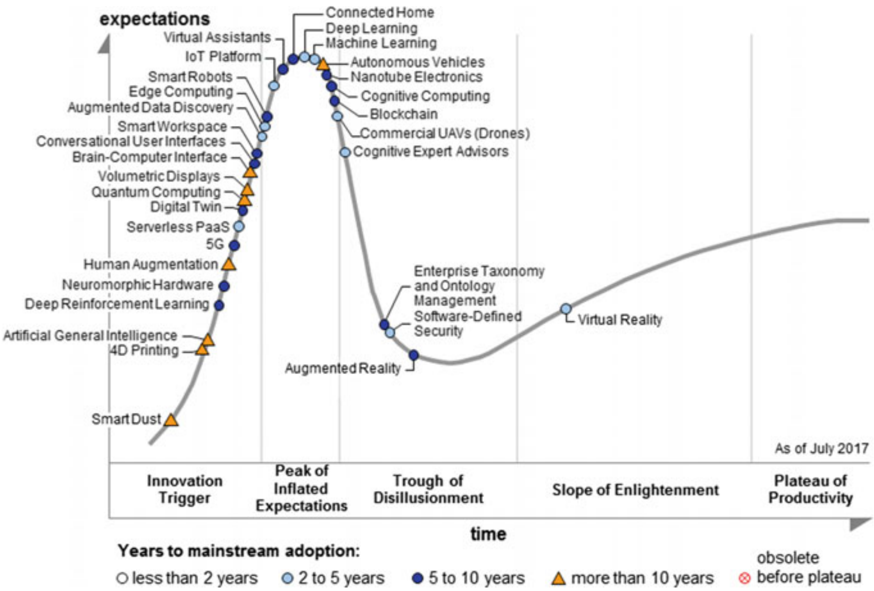


Fig. 1 Gartner's Hype Cycle for emerging technologies

server support of the activities of the companies. Therefore, all small and medium businesses within two years will move to cloud infrastructure and services.

How to understand the phases of the Gartner cycle? (1) Innovation trigger means the launch of innovation, where potential breakthrough technologies that are interesting for the market, with unproven commercial consistency, are replacing existing cyber-physical structures. (2) The peak of inflated expectations is the high level of bloated market expectations, where timely advertising creates successful precedents for implementing innovative technologies on the failure field. (3) A trough of disillusionment means the arrival of disappointment, when interest in technology goes down, experiments do not confirm the expected market attractiveness, individual developers improve their products and learn about investments. (4) Slope of Enlightenment means a slope of insight, when there are examples of technologies that benefit the enterprise; there are finances for pilot projects. (5) Plateau of Productivity means an area of sustainable productivity enhancement, when the technologies, products and services that are created find their consumers on the market.

2 The Three Main Fields of the Cyber Culture

Cyber culture is the development level of social and technological moral relations between society, physical green world and cyberspace, determined by the implementation of Internet services for precise digital monitoring and reliable metric management in all processes and areas of human activity, including education, science, life, production and transport, in order to improve people life quality and preserve the planet's ecosystems. Short definition: Cyber culture is a moral metric of computing excellence to achieve the social significance and recognition.

Hype-cycle 2017 forms the planet's cyber culture for the next 5–10+ years by expert analysis of more than 1800 possible technologies performed by leading research and consulting companies. A list of 33 + 2 top Gartner-table technologies creates a cyber culture, shown in Fig. 2, and also competitive advantages for subjects of the market in the fields of science, education, industry and transport.

The first three places in the Gartner-top-cycle are assigned to the following strategic directions: Artificial Intelligence Everywhere, Transparently Immersive Experiences and Digital Platforms. (1) Artificial intelligence everywhere becomes the most disruptive technology in the next 10 years due to the availability of computing power, infinite volumes of big data and achievements in the implementation of neural networks to adapt to the new situations that no one has ever

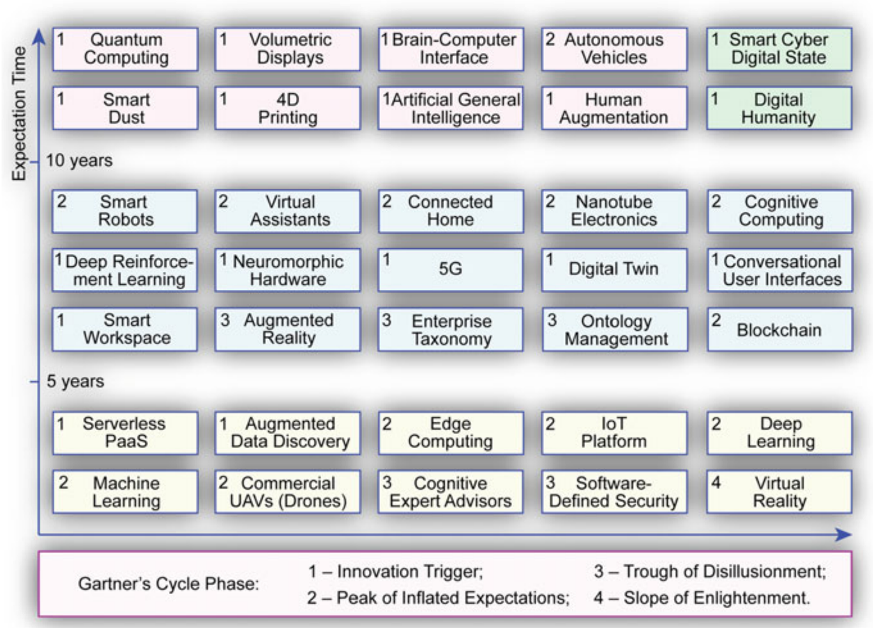


Fig. 2 Gartner's table for emerging technologies

encountered before. Enterprises, which are interested in leveraging artificial intelligence, consider the following technologies to be useful for themselves: Deep Learning, Deep Reinforcement Learning, Artificial General Intelligence, Autonomous Vehicles, Cognitive Computing, Commercial UAVs (Drones), Conversational User Interfaces, Enterprise Taxonomy and Ontology Management, Machine Learning, Smart Dust, Smart Robots and Smart Workspace. Thus, Artificial General Intelligence in the next 10 years will penetrate into all fields of human activity, as a technological service immersed in the cyber space, including 30% of high-tech and transport companies.

Smart workspace means to be connected to the infrastructure of solving production problems in space and time in the format of 24/7. At the same time, virtual private networks are used, the metric for measuring the potential and performance results, the presence of a certain cyber culture and the choice of the most convenient places for doing business. High self-motivation for the successful and efficient performance of the task stipulates the leverage of dynamically changing cyber-physical workspace for creativity, which is invariant to the office, home, transport, places of rest and sports.

(2) Transparently immersive experiences are becoming more human oriented and provide the following: transparency of relations between people, business and things; the flexibility and adaptability of the links between the workplace, home, enterprise and other people. Gartner-Inc. also predicts the introduction into practice of the following expecting critical technologies: Autonomous Vehicles, Brain-Computer Interfaces, Smart Dust, 4D Printing, Augmented Reality (AR), Connected Home, Human Augmentation, Nanotube Electronics, Virtual Reality (VR), and Volumetric Displays. Integration of cyber technologies is aimed at ensuring the quality of human life by creating: smart workspace, connected home, augmented reality, virtual reality and the growing brain-computer interface. For example, Human Augmentation technology aims to expand or supplement human capabilities to improve the health and quality of life through the harmonious use of cognitive and biotechnical improvements, as parts of the human body. Volumetric Displays visualize objects using 3D active voxels in three dimensions with a 360° spherical viewing angle, where the image of the phenomenon changes as the viewer moves. 4D Printing technology is an innovation of 3D printing, where structural materials can be transformed after manufacturing the product in order to adapt it to human needs and the environment.

(3) Digital Platforms of technological culture are formed by components: 5G, Digital Twin, Edge Computing, Blockchain, IoT, Neuromorphic Hardware, Quantum Computing, Serverless PaaS and Software-Defined Security. Technologies such as Quantum Computing and Blockchain will create the most unpredictable and disruptive breakthroughs for humans in the next 5–10 years. Neuromorphic Hardware is considered as the future of artificial intelligence, which is aimed at creating a neuromorphic-computing chip that can replace the cloud computing power of the Apple Siri Data Center in solving complex machine learning problems (Chris Eliasmith, a theoretical neuroscientist and co-CEO of the Canadian AI startup Applied Brain Research) [3]. Otherwise, inside the iPhone will

be a digital brain in the form of a neuromorphic IP-core, which solves all the tasks of interaction of a gadget with the outside world in real time. IBM's neuromorphic universal chip, due to spike asynchronism, consumes three orders of magnitude less energy with a fivefold increase in the number of transistors that exceed Intel's existing hardware solutions. For programming hardware-oriented algorithms, compilers are used: Nengo, Python. At present, the following chips are already implemented using the Nengo compiler: vision systems, speech systems, motion control, adaptive robotic controllers, and also Spaun-chip for offline interactive communication between the computer and the environment. Software-Defined Security (SDS) or Catbird is designed to protect system objects or logical structures in virtual space. This is due to the fact that network security no longer has physical boundaries within the framework of the logical architecture existence of the cloud service. Therefore, an accurate and flexible SDS is created as a complement to infrastructures and data centers without the presence of special security devices. Scaling SDS makes it possible to create or acquire the minimum necessary security conditions in a certain place and time, which significantly reduces the material costs of forming a quality SDS service.

3 The Practice of Leveraging Green Top Technologies of Cyberculture

High level costs for research and development from Amazon, Apple, Baidu, Google, IBM, Microsoft and Facebook stimulate the creation of original, patentable solutions in the field of Deep Learning and Machine Learning, among which are: Amazon Alexa, Apple Siri, Google Now, Microsoft Cortana. Gartner Inc. is sure that the tools for in-depth training will account for 80% of the standard funds for scientists by 2018. Today, on the websites of companies, technologies and data on scientific research are becoming available: Amazon Machine Learning, Apple Machine Learning Journal, Baidu Research, Google Research, IBM AI and Cognitive Computing, Facebook Research.

The introduction of telecommunication 5G-technology (Fig. 3) in the coming decade will provide the market with expected innovative solutions for the security, scalability and performance of global green networks and connections in transport, IoT, industry, healthcare.

Gartner Inc. predicts that by 2020, 3% of network providers of mobile services will launch commercial networks in 5G-format, which will provide qualitatively new conditions for the widespread introduction of telecommunications for scalable globalization: IoT, cloud-transport control, UHD-TV. Leaders of 5G-implementation in 2017–2018 are: AT&T, NTT Docomo, Sprint USA, Telstra, T-Mobile and Verizon. The 5G-technology is an ultra-wideband mobile connection in the millimeter range for Massive M2M real-time transactions with permissible delay control (1 ms), while simultaneously connecting about 10 million devices per

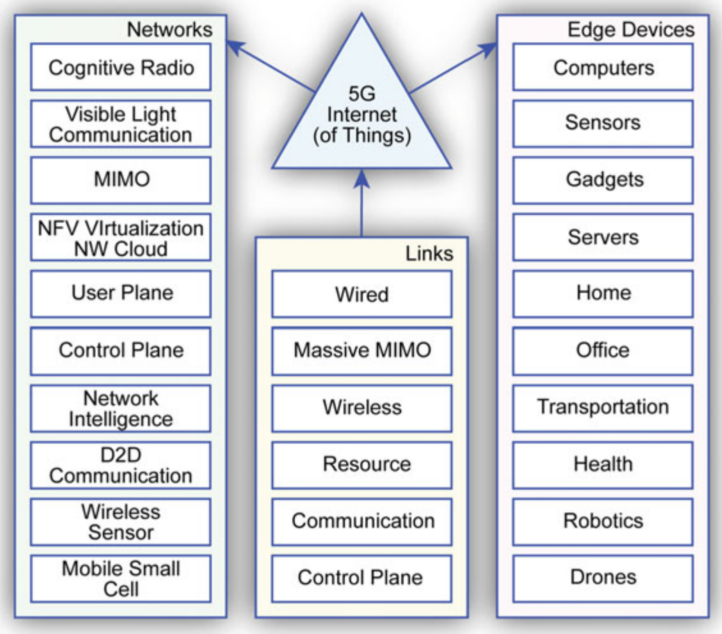


Fig. 3 5G-technology

1 km^2. The 5G uses Beam Division Multiple Access (BDMA) technology to interface the base station with mobile devices. The 5G wireless cellular architecture provides 10–50 Gbps throughput in the 30–300 GHz range of UHD video applications and virtual reality creation [4]. The innovative 5G technology features the use of: the Massive MIMO array, Cognitive Radio network, direct D2D connection for IoT, radio access network as a service and network function virtualization cloud —NFV.

The Hype cycle does not consider two technologies that are essential for humanity, related to cyber-social computing: Digital Humanity and Smart Cyber Digital State. They are added to the Gartner's Table of Emerging Technologies, as the launch of the innovation trigger, the implementation of which is expected in ten years.

Digital Humanity suggests an accurate digital identification of a person based on natural biometric parameters (fingerprints, scanning of the eye and DNA) excluding paper data carriers, plastic cards, certificates, diplomas and passports in a planetary scale. The digital identifier makes it possible to delegate the positioning of each person in time and space to the cloud service, which removes all the problems associated with the cyber-physical analysis of the human illegitimate actions. The consequence of the sustainable development of digital humanity technology within the green Internet of things to create a smart world [5] is a multi-billion saving in the costs of producing and using paper products, preserving forests and planetary ecology. The payment for receiving these dividends is the cost of creating an electronic infrastructure for digital authentication of each individual in time and space. Green IoT is a cyber-physical culture of human activity aimed at ensuring the quality of life of people and preserving the ecology of the planet, energy, materials, money and time. IoT components are: Identification, Sensing, Controlling, Communication, Computation Services, and Intelligent Infrastructure. The smart world provides each person with services from: smart devices (watches, mobile phones, computers), intelligent transport (aircrafts, cars, buses, trains), smart infrastructure (homes, hospitals, offices, factories, cities, states), and intelligent remote online education (school, university).

The world without mediator is an innovative technological way of direct telecommunication contacts between every person and any subject on the green planet, through the use of cyber physical services. If humanity decides to destroy corruption on a global scale, it will do this by using Blockchain [6] that is a decentralized cryptographic transparent technology for storing and exchanging data on completing transactions. Blockchain provides direct interaction of two businessmen without trusting legal mediators within the IoT-service. Without legitimate recognition of Blockchain by government structures, this technology is still a resource-consuming experimental anti-corruption enclave in the cyberspace of data and transactions legalized by governments. If the level of people's trust in Blockchain technology increases, then a cyber revolution may occur that will overthrow existing state institutions, which are now guarantors of mediators in relations between citizens.

IEEE Spectrum magazine [7] published a thematic collection of articles on Blockchain technology under the slogan: "In cryptography we trust". In essence, the technology removes all "trusted" intermediaries: bankers, a multi-million army of government officials in a deal between the two sides. The third party-intermediary has been always redundant, which reduces the profit of the business-producers, complicates transactions and, most importantly, is the only source of corruption on the green planet. "People's striving for secrecy makes them stupid: no bad or good results are seen". Julian Paul Assange.

Blockchain technology (in particular, bitcoin) is a cyber-physical metric cloud crypto protected computing for transparent monitoring and trust management of transactions in distributed blockchain data, Fig. 4.

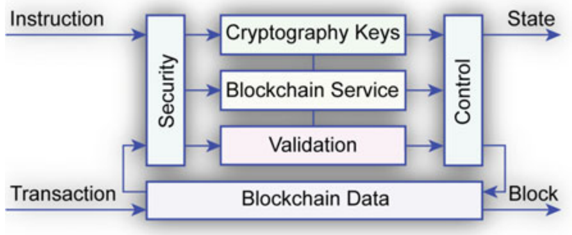


Fig. 4 Blockchain computing

The technology, originally proposed by Satoshi Nakamoto (2009), today sustainable evolves as trusted computing, distributed in space and time: Ethereum Virtual Machine (2013, Vitalik Buterin); Microsoft blockchain applications on the Azure cloud; IBM and Intel open Hyperledger resources. According to Blockchain.info, more than 375,000 people are already involved in the Bitcoin project (China, Eastern Europe, Iceland, Venezuela). The Crypto currency becomes a universal intermediary between the seller and the buyer for assessing the social importance of goods and services.

Therefore, substituting paper money for virtual (digital code) will be trivial, since the paper cannot stand up for itself. Difficulties arise when replacing millions of officials who act as intermediaries-trustees, on the cloud computing management systems. Officials have the power, the weapon and will stand to death in order to preserve mediation in a corruptive distribution of money and resources, Fig. 5.

The scope of the Blockchain technology is all cyber-physical processes and phenomena, often affected by corruption, thanks to the presence of intermediaries who cannot be really trustworthy, due to human weakness—to have someone else if there is no penalty. In fact, it is possible and necessary to build Blockchain trust systems for the open management of science, education, tourism, transport, finance, society, medicine and personnel. Ethereum is a blockchain (bitcoin) world computing that can replace Facebook, Twitter, Uber, Spotify, being invulnerable to censors and transparent regarding the processes that occur, and also to work long without the people who created them.

The disadvantages of Blockchain computing are: (1) multiple duplication of data in a distributed network; (2) the openness of valuable data, patents and codes, which will destroy some of the specific companies; (3) high costs of creating an infrastructure oriented to Blockchain; (4) stubborn resistance of state structures to the implementing the system of moral distribution of resources. “Money is not the root of all evil. Equity is the root of all evil”, Joel Monegro. In other words, equality in poverty is the cause of the society’s degradation.

Smart Cyber Digital State will certainly emerge as an innovative cyber-social culture, providing tolerant services based on digital monitoring and cloud

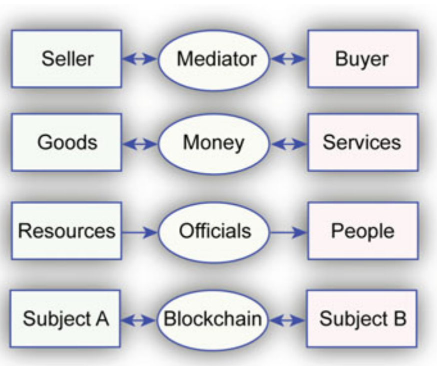


Fig. 5 The areas of application Blockchain computing

management of citizens. Why should a cyber state necessarily take place? (1) A citizen must have a choice of acceptable public services in exchange for taxes that he pays. There should be no monopoly on the provision of traditional state services to a citizen, especially if they become logically or emotionally unacceptable. (2) The care of citizens turns from residence (place) dependent into extraterritorial cloud service, invariant to the location, due to the presence of cyberspace, which online connects the citizen with state green organizations [7]. (3) Conflicts between the authorities and the people should be resolved by online monitoring of public opinion on the substantive issues. (4) Like social networks, new states will emerge, and old, not-competitive social entities will collapse. Who will be the leader in cyberspace that will rule in the physical world. (5) The moral and material consistency of the goal and program of the society's development, put forward by the political elite, as well as its real actions, becomes the main argument of the market attractiveness of the authorities in the struggle for the citizen (taxpayer) in the market of cyber-physical states. The formula of the interactive game in the transformation of society is simple and understandable: "The best citizens in the best countries". It is interesting that today a similar formula for the game between higher education institutions and students in some post-Soviet states has been practically implemented: "The best students in the best universities". (6) In the cyber world, the priorities between physical prototypes and digital images are gradually changing. The "Master-Slave" pair or who is the leader and who is the slave, is transformed into a sequence: "a cyber-image—a physical prototype". This is especially true for the states that existed throughout the history of mankind, while citizens supported the political doctrine, as a rule, the immaterial substance put forward by the political elite. Today, it becomes even more important to make personnel decisions not by a physical citizen, but by the metric cyber-image of its social significance, formed in cyberspace. (7) Statehood becomes an integral cyber-service in the market on the planet, which can be bought by a citizen in exchange for his taxes. (8) The main functionality of cyber-statehood is the formation of the moral emotional intelligence of a society through a tolerant integrating association of languages, histories, cultures, religions and traditions, which is dominant when making strategic decisions, in comparison with the logical intelligence of a social group. Ignorance and incompetence of the political elite associated with the intra-territorial constitutional division of citizens according to the metrics of languages, nationalities, cultures, histories and traditions inevitably leads to the collapse of the state. The policy of emotional and logical unification of all citizens through a tolerant attitude to the mentioned metric differences creates a powerful integrating impulse for the development of each individual, economy, culture, science and education.

The world is becoming more and more green, intelligent, digitized and strongly connected (networked) to the people, things and services [8, 9]. Figure 6 presents a picture consisting of top-ten strategic green trends in the IT industry in 2017 [10], which should be adopted by all companies and universities wishing to form new business projects on the NASDAQ market of goods and services.

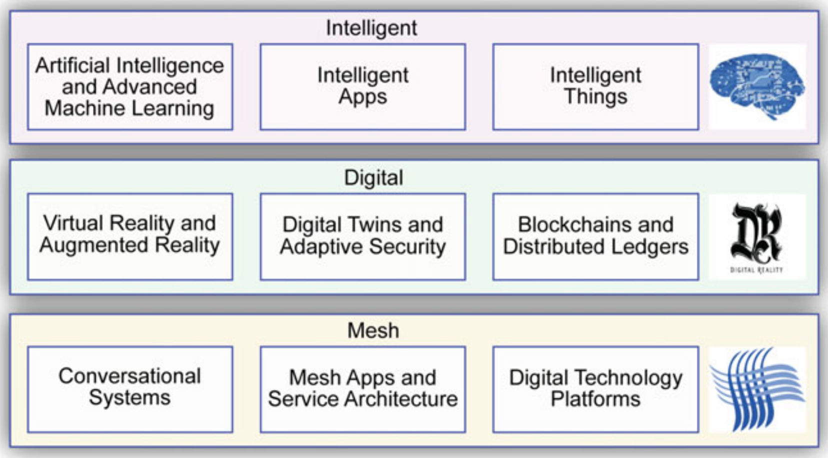


Fig. 6 Top-9 disruptive areas in the computing industry

Practically today, a new, more sophisticated smart digital intelligent cyber-physical world is created to share a harmonious existence of people in an environment of people-friendly goods and services. Thus, the general picture of the desired world consists in the desire: (1) to digitize all objects and processes on the planet (spatial, biological, technical, social, virtual); (2) to implement scalable artificial intelligence into all digitized cyber-physical processes and phenomena; (3) to connect all intelligent objects and processes into smart, scalable network within the framework of a digitized common cyberspace. The purpose of creating a digital intelligent smart cyber-physical world is the quality of life of a moral person, the elimination of social defects and the green ecology of the planet.

Gartner Inc. considered it necessary to withdraw from the umbrella of the current market fashion the following technologies, which did not meet the expectations of IT-business: Affective Computing, Micro Data Centers, Natural-Language Question Answering, Personal Analytics, Smart Data Discovery and Virtual Personal Assistants.

To create successful businesses and new educational courses, Gartner Inc. recommends taking into account assumptions about strategic planning, which include 10 points: (1) By 2020, 100 million consumers will buy into the expanded reality, including using Head-Mounted Displays (HMDs). (2) By 2020, 30% of web browsing sessions will be performed without using a screen. More than 5 of the 550 million Apple iPhone owners will use AirPods to exchange voice messages. Five percent of consumer-oriented websites will be equipped with audio interfaces (including voice chats with voice support). (3) By 2019, 20% of brands will abandon their mobile applications (in favor of the MASA—Mesh App and Service Architecture). (4) By 2020 smart algorithms will positively affect the behavior of more than 1 billion global workers. (5) By 2022, a business based on the leverage of detachments will cost 10 billion dollars. (6) By 2021, 20% of all human activities will be included, at least, in the services of one of the seven leading global companies (Google, Apple, Facebook, Amazon, Baidu, Alibaba and Tencent) (7) Until 2019, every dollar invested in innovation, will require an additional 7 dollars for the basic execution of the project. (8) During the course of 2020, Internet of Things (IoT) will increase the demand for data centers by 3%. Indoor display devices, such as Amazon Echo and Google Home, will be located in more than 10 million homes. (9) By 2022, IoT will save consumers and businesses $1 trillion a year, targeting services and supplies. By 2020, Android Auto will use about 40 million cars, and 37 million vehicles will leverage CarPlay. (10) By 2020, 40% of employees will be able to cut their spending on health using a fitness tracker.

4 Emotional-Logical Computing (EL-Computing)

"The calm and modest life brings more happiness than the race for success, which causes only permanent stress." Albert Einstein. "Everything is the word and the word is God." The Gospel of John. The word creates and destroys, uplifts and overthrows, heals and maims, notifies and governs, makes the rich and the poor, wise and foolish. For computing, the word is: a command, and a resource, a state and a goal, control and notification, logic and emotions creating the architecture. The Bible not only integrates the history or experience of humanity (warning signals), but also programs each person for decent behavior by developing operational, regulatory actions, having in mind the goal function—the moral development of mankind in order to preserve the planet. In other words, the Bible is the first and so far, the most successful attempt to create emotional and logical computing in the form of human artificial intelligence, where every person can find an almost exact answer to any question related to decision-making.

Emotions and logic are related to the functioning of two co-processors: the right and left hemispheres of the brain, which participate in monitoring the situation and/or the surrounding reality to make a decision in the direction of movement toward a pre-determined goal. The hemisphere of the brain is nothing more than two mutually complementary alternative co-processors, to which the person himself and/or other people have access. The interaction of the hemispheres in order to make a decision fits into the four forms shown in Fig. 7. Types of emotional and logical computing can be denoted by the symbols of the Cantor alphabet: $\{0, 1, X, \varnothing\}$ Here there are two hemispheres {0,1} of the brain, which create the following types of emotional-logical computing (EL-computing): (1) Logical (0), when the developed left hemisphere dominates in the decision-making process based on acquiring knowledge and rich life experience, typical for elderly and/or uncreative people. (2) Emotional (1), when the developed right hemisphere dominates in the decision-making process, based on heuristic insights and emotional experience, typical for young and/or creative people. (3) Emotional-logical

(X = {0,1}), when the equally developed hemispheres of the brain participate in the decision-making process, based on the acquiring knowledge and emotional experience, typical for educated and creatively gifted people. (4) Repetitive (∅), when the equally undeveloped hemispheres of the brain, unable to make independent decisions because of the lack of knowledge and experience, perform the function of a repeater, which is typical for uneducated and/or weak-willed people. It is interesting that practically any educational process (kindergarten, school, university, enterprise) is focused on raising the level of logical intelligence by reducing the level of emotional intelligence associated with creative activity. Knowledge helps to avoid trivial mistakes made by humanity but limits the flight of fantasy for the discovery of new technologies, materials, processes and phenomena.

Emotional logical (psychological) computing is the metric management of the emotional and logical co-processors of the brain based on the digital monitoring of the cyber-social environment for choosing the optimal decision on the way to achieving the goal (Fig. 8).

Objectively, it should be considered that such computing is accessible to the owner of the brain and for the people around him, with positive and negative goals. Such dualism of access to the brain co-processors has its advantages and disadvantages: (1) It is good that I have influence not only on myself, but also for other people, to heal them, to make them successful and happy. (2) It's bad that knowing the structure of co-processors and access codes, it is possible to reprogram a citizen to immoral behavior for harming him, social groups and the ecology of the planet. Programming human behavior means starting the functional modules of his emotional and/or logical intelligence (co-processor) by such warning signals, which

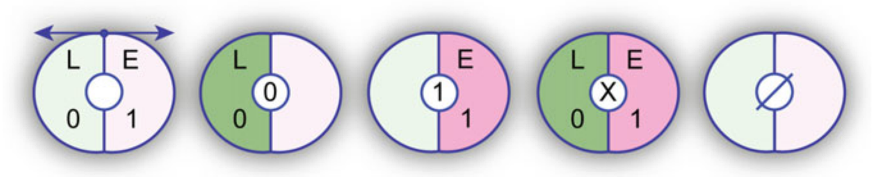
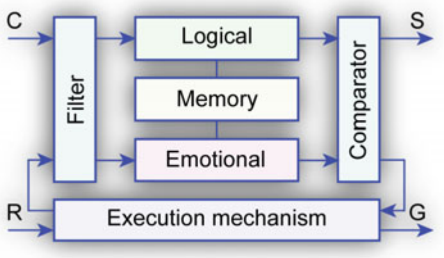


Fig. 7 Types of emotional and logical computing

Fig. 8 Model of emotional logical computing



lead to the desired, own or imposed by someone, decision. In other words, a person can program himself, and also other people can do this to achieve their goals.

Logical processors for decision-making are sufficiently described in the modern scientific works [13, 14], but the creation of adequate emotional processors is at the stage of development and verification. However, one noteworthy event attracts an attention: the award of the Nobel Prize for 2017 to Richard Thaler, a scientist working in the field of financial and economic behavior of citizens. He created "a theory of pushing or controlled choice" based on a gradual transition from emotion to the logic of decision-making. Nevertheless, in the majority, the mankind while is guided, just on the contrary, by emotions for development of a line of behavior. For example, the desire not to lose what you have for the masses is more preferable than the possible acquisition of more.

The emotional processor, in addition to memory for data storage, has a functionally complete basis for describing primitives, circuits and devices. The emotional processor operates with clear positive and negative axioms, statements and actions. Negative one is more popular in everyday speaking than positive. The negative is more popular in everyday speaking than positive. For example, the word "clever" has 18 synonyms, and "fool"—192 synonyms. However, when analyzing the 100 most popular printed words in Russian, it turned out that only one of them —"bad", has a clearly negative meaning. The word "logic" loses to the word "emotion" in the number of synonyms as: 4–27 (http://synonymonline.ru). In other words, decision-making based on the negative emotions of the crowd directed against specific people, processes or phenomena is a win-win technology for coming to power. In wisdom, we always find miserable shortcomings, and we decorate stupidity with bloated virtues.

To create a basis for E-computing primitives, an axiomatic metric of human emotional behavior is proposed, built on the dominance of minimalism (the shortest path) in any system of relationships that is inherent in social groups with no special logical education: for analytical defining the relations of an oriented graph, the Algebraic Form of Graph Presentation (AFGP) is used [13, 15]. The metric of emotional relations or E-computing is represented by three axioms:

(1) Reflectivity $(a \to a)$: The shortest path in a system of relations. A system, closed to itself, retains one's own and does not acquire someone else's.
(2) Symmetry $[(a \to b) \vee (b \to a)]$: Minimum of connections with the outside world. I give to you, you return to me. Tolerance in relations and the absence of mediators.
(3) Transitivity $[(a \to b) \vee (b \to c) \vee (a \to c)]$: The minimum path to the goal. Direct contact with the subject and excluding the mediator means the removal of the extra link in the chain of events.

The fulfillment of these axioms, which create the behavioral metric of minimalism, forms binary equivalence relations in the society, which is good for the system as a whole (Fig. 9):

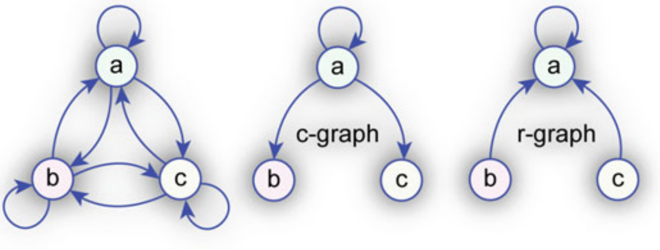


Fig. 9 Equivalent and selfish relations

$$Y = aa \vee bb \vee cc \vee ab \vee ba \vee ac \vee ca \vee bc \vee cb.$$

However, each individual has an egoistic desire to modify the system of tolerant equivalent relations for themselves, turning them into non-reflexive, non-symmetric and non-transitive (Fig. 9, c- and r-graph). Egoism is determined by a structure that minimizes the metric of connections in a system of relations for one node (subject): $Y^c = aa \vee ab \vee ac$—authoritarian management; $Y^r = aa \vee ba \vee ca$—ascending financial and resource flows. A scalable generalization of such modification generates nationalism—hatred of other nations, as well as Nazism, where equivalent relations are transformed into a pyramid with a titular nation on a top, which legislatively dominates the second-rate nations.

Nevertheless, selfish relations are dominant in the social groups, which are focused on the rapid, local and minimal success of each person without the risk of loss and the inability to receive larger resource gains after a long time. Naturally, equivalence relations by definition mean the legal equality of all citizens, giving the opportunity to participate fully in the production, economic, political and cultural processes of the social group with the subsequent receipt of remuneration in accordance with the measured social significance of the results of each person's work. Equivalent relations attract the best specialists in the social group, who produce competitive products and services for a successful sale in the market, which brings a high profit that provides larger salary and stimulates a person to more productive creative activity.

The metric of emotional relations along horizontal and vertical connections causes "rules" of the behavior of the average person:

(1) To save your own's is better than the possible acquisition of someone else's. (2) Violate the rules with guaranteed absence of punishment. (3) The small salary of this day will be better than the promised thousand tomorrow. (4) To elect and respect those who have real power. (5) Take care of yourself, and then about others. (6) Listen to yourself and obey the chief, ignore the opinions of subordinates. (7) To enjoy the shortcomings and mistakes of other people. (8) Discuss third parties, boast of successes and complain about fate. (9) Enter into emotional conflicts with others. (10) Do not keep promises and do not return debts. (11) To join the ranks of the majority, who united against a person or a subject. (12) Do not express your

own opinion, if it does not coincide with the opinion of the chief or the majority. (13) Confidence in state propaganda coming from power structures.

5 The Cyberphysical Statehood

"Where there is a will, there is a way." Albert Einstein. The cyber-physical statehood is intended for metric management of resources and citizens on the basis of digital monitoring and assessment of the social group needs in order to ensure the quality of their life and the prosperity of the country through the production and export of goods and services, Fig. 10. There are 6 main systems-forming components for the green statehood functionality, ordered by the degree of their influence on the project market success: (1) Relationship. (2) Goals. (3) Personnel. (4) Management. (5) Infrastructure. (6) Resources.

(1) Relationships. Statehood is created, first of all, by moral metric digital relations between citizens, which are formed by the constitution, laws, regulations, orders, careful and tolerant attitude to language culture, history and traditions uniting citizens. The political destructive elite often forms relations that separate society in languages, culture and history for the purpose of long-term and total domination over artificially disparate social groups that are not oriented toward the creation of market-demanded products and services.
(2) Goals. In one way or another, relations are formed by the political elite, proceeding from the goals or doctrine of the state, which in civilized countries is aimed at the moral development of the citizen, the growth of his well-being and the preservation of green ecology by leveraging resources obtained from the export of goods and services. The inability of the political elite to create a management mechanism for the production of goods and services can create an attractive doctrine of war and hatred for other peoples and states to improve the quality of titular nation life by destroying internal and external another nation.
(3) Personnel. Naturally, goods and services are created by the competent personnel, which must be correctly placed in the hierarchical structure of the

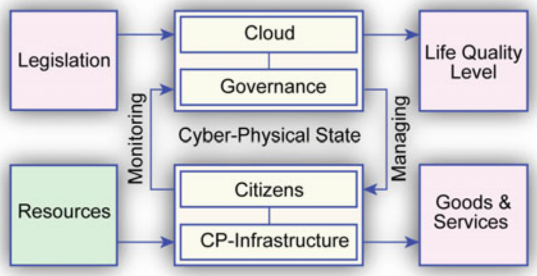


Fig. 10 The cyber-physical statehood

"management-execution" relationships to achieve maximum labor productivity in the creation of market-demanded goods and services. The attraction of the best personnel from all countries of the world should be elevated to the rank of state policy by creating the appropriate moral and material conditions for servicing specialists.

(4) Management. At the same time, the (cloud) management personnel (with MBA —Master Business Administration diplomas) as a key success component should not overlap with the executive mechanism and make up no more than 10% of the staff for the purpose of minimizing overhead costs at a similar level. Non-productive costs in today's computer occupy no more than 5%, which is a model for imitation of the processes of the state governance. In the past, the hierarchical administrative apparatus reached 50% in the absence of effective means of communication. Currently, direct cyber-physical contacts on the Internet provide the opportunity for cloud-based online management of all citizens of the country simultaneously. Officials should be only experts who received a special education in MBA. Equally, 90% of all digitized information flows must be transformed into electronic document flow on the basis of the cloud management.

(5) Infrastructure. The state's cyber-physical infrastructure is becoming increasingly dominant, which forces the political elite to leverage new cloud technologies for the paperless digital management by citizens. The role of purely physical state's infrastructure is transformed into secondary symbols of statehood associated with the border, roads and buildings that create comfortable conditions of traveling and living for every citizen of the green planet. Here, the dominant technology is IoT, which creates an e-infrastructure for authorized online access to the cyber-physical components of the statehood based on the leveraging the primary biometrics of a citizen (scan of the fingers, face or eye).

(6) Resources. They functionally depend on the above five components, even with the initial zero budget of the new statehood project. Free money in the world much more than interesting projects aimed at a positive future. To find sponsors, it is enough to demonstrate an interesting goal for potential investors, to create non-corrupt digitized relations, to collect a team for cloud-driven management and personnel for the implementation of the project, as well as to determine the territory in the cyber-physical space. Unfortunately, destructive projects aimed at destroying unwanted statehoods, easily will find sponsors.

What is the more important for success in the market, personnel or relationships? Often, the inefficiency of the social system is connected to the unsuccessful selection of personnel in key positions, where the stupid way out of which is the replacement of some specialists by others. We argue that relations have the highest priority over personnel when creating an effective company or statehood. "We are not looking for people who make mistakes, but we are correcting the structure of relations in which these mistakes are possible. Stanley Hyduke". Example 1. The premier companies on the planet simply create effective branches in any country of the world by intervening in it the relationship similar to the main office. Conclusion:

workers in the most developing territory can be technologically trained on a minimal timing frame, if the company pays substantially higher salary to its employees against the background of miserable salaries in the state enterprises of poor countries. The company's win consists in a substantially lower salary level for its employees compared to the employees of the head office. Example 2. In the case of equality of social relations within a company or a developed country, better production results will indeed be determined by the more qualified and creative personnel. Not the last place in the formation of market success takes the role of an idea-leader. So, there are successful projects: Apple, Google, Amazon, Microsoft, Alibaba in countries, where tolerant metric and non-corrupt relationships are created for work and creativity. To the question: what causes corruption, personnel or relations? There is only one answer: a relationship that allows two sources of corruption: (1) an intermediary official who "sells"; (2) the state resources to the people. Solution of the issue: (1) Cloud management of resources without the participation of official in accordance with predetermined rules. (2) The nullification of state ownership by replacing state structures (finance, science, education, medicine, transportation, law, court, energy) and enterprises to the private companies. The private companies do not have corruption. An example of digital cyber-physical creativity, strictly oriented to online service execution to all citizens' desires is Privatbank [16]. This is the prototype of a new type of moral cyber-physical green statehood, really caring for every person in 24/7 format.

What makes the state a loser in the market: personnel or relations? Exactly not the personnel, given that millions of fellow citizens successfully work throughout the world. Then the immoral relationship created by the political elite during the years of independence, including discrimination of citizens, constitutionally and legislatively created in terms of languages, history and culture. The moment of history, 26 years ago. The euphoria of high-speed authenticity of the country eclipsed the mind of self-educated politicians-deputies who in one night conceived and gave birth to the ugly system of relations (constitution) of the new independent state. The good solution is to change ten words in the constitution. It does not cost anything to get a successful country equal to Germany. "Where there is a will, there is a way." But the will of the elite for a long time is paralyzed by the insane idea of the title nation superiority. Conclusion: the degradation of statehood continues untimely.

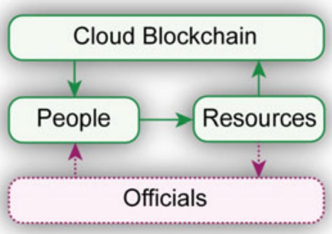
The identity and/or authenticity of statehood is determined not by propaganda bubbles of the imaginary superiority of one nation (language, history, culture) over others, but rather by a certain metric of consistency, which includes: (1) Real incomes of the population; (2) Export of products and services; (3) The level of development of science and technology, including the number of Nobel Prize laureates and the export of technology; (4) Life expectancy of citizens; (5) Tolerant relationship of state power to languages, history, culture of peoples. Surprisingly fact, the Republic of Crimea today has three official languages: Tatar, Ukrainian and Russian. And there is no person in the world suffers from such multilingualism.

Statehood, as an apparatus or service of violence against a person, is a thing of the past. States where the relationship of violence, hatred and Nazism dominate, in the 21st century will disappear from the face of the earth. The cyber-physical states with the moral, tolerant services for developing of every citizen come to replace them.

The basis of corruption in developing countries is determined by a model that contains three components interconnected by state relations: (1) the people; (2) officials; (3) resources. Eliminating just one of them will forever bury the phenomenon of corruption. It is easier to destroy the people that the authorities consider to be the source of corruption. It is more difficult to destruct a class of officials who has real unlimited power and resources. You can not eliminate resources. The most difficult thing for mankind to understand is that the source of corruption is not the components mentioned above, but the legal relations between them, always initiated and created by officials for themselves. In Fig. 11 there are two cycles of resource management: (1) <Resources, officials, people> are the state administration, where the mediator is the army of officials. (2) <Resources, a cloud blockchain service, the people> create a new structure of mutual green relationship, where the intermediary, which is the source of corruption, is excluded from the relations. The cost of implementing the second cycle of resource management is almost zero. It is necessary only the will of the first persons of the political elite and the people support, created by mass propaganda of the morality of relations. Thus, defeating corruption means replacing officials with a cloud-based resource management service, which automatically means minimizing their number. More difficult is the path of a long-term (50–300 years) morality education for officials who will refuse to sell resources that do not belong to them.

The model of emotional-logical computing operates with the discrete primitives that mathematically explain the real model of social processes and phenomena. For example, practically every state has two branches of power: dictatorship and democracy, which are projected onto the oligarchs and the people. For the people, democratic institutions are created that play fake games of making crucial decisions on the principle of the majority of voters. However, the voting people are 365 days a year dependent and governed by the oligarchs in power. Therefore, the people always do “vote” for the someone who has more money. You can not win poker from a partner who is not limited in finances. The following logical scheme, shown in Fig. 12, explains the interaction of democracy and dictatorship for the making fateful government decisions.

Fig. 11 Resource computing: two control loops



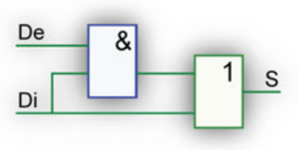


Fig. 12 The interaction of democracy and dictatorship

It seems that everything is honest, the people and the capital jointly make state decisions that formally depend on two variables:

$$S = f(Di, De) = (De \wedge Di) \vee Di.$$

However, in fact, minimization of the logical function defines the inessentiality or redundancy of the variable De to form an output state S, which forms responsible decisions for the statehood:

$$S = f(Di, De) = (De \wedge Di) \vee Di = Di \wedge (De \vee 1) = Di.$$

The similar result can be obtained by taking a Boolean derivative with respect to the variable De, which is equal to zero. It means that the variable is unessential when generating the output state of this function:

$$\frac{df}{d(De)} = [(1 \wedge Di) \vee Di] \oplus [(0 \wedge Di) \vee Di] = (Di \vee Di) \oplus Di = 0.$$

It should also be noted that the Boolean derivative with respect to the second variable Di is equal to one:

$$\begin{aligned}\frac{df}{d(Di)} &= [(De \wedge Di) \vee Di] \oplus [(De \wedge Di) \vee Di] \\ &= [(De \wedge 1) \vee 1] \oplus [(De \wedge 0) \vee 0] = 1 \oplus 0 = 1.\end{aligned}$$

This means not only the essentiality of the variable Di for the decision making, but also the complete absence of any other constraint-conditions that affect their acceptance.

Practical interest is the structure of digital social computing (Fig. 13), focused on solving problem of searching for optimal coverage.

In particular, this circuit can be used to determine the minimum number of competent developers (managers) in the execution of a social project. For example, there is a university management metric that includes eight parameters: (1) Legal culture for creating moral relationship in a team. (2) Management of science and education in digital format. (3) Personnel management based on the metric measurement of the social significance of each employee. (4) Planning the development of the university, taking into account global trends in science and education. (5) Management and development of the university's cyber-physical infrastructure.

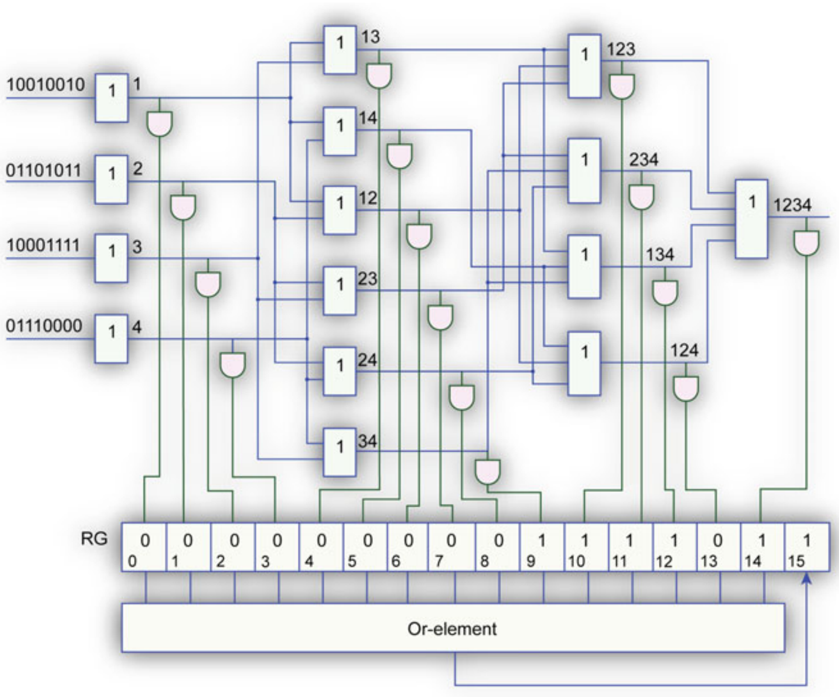


Fig. 13 Processor for searching an optimal coverage

(6) Managing and planning financial activities to attract public and private investment. (7) International cooperation based on close cyber-physical contacts of university scientists with leading companies and universities of the world for the implementation of joint projects in the field of science and education. (8) Education of the physical and aesthetic culture of students and employees, based on an individual approach to the disclosure of talents.

The following competency table is a possible coverage of eight parameters of the university management metric by four candidate experts:

C\P	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8
C_1	1	0	0	1	0	0	1	0
C_2	0	1	1	0	1	0	1	1
C_3	1	0	0	0	1	1	1	1
C_4	0	1	1	1	0	0	0	0

Solving the problem of defining the minimum number of experts able to manage the university is possible through the use of a circuit (see Fig. 13) that allows finding the optimal result in the form of two candidates C3 and C4, which cover all parameters of the metric management of the university by their competences.

6 Models of Human Relations and Behaviour

In developing humanity, local social viruses (collisions and cataclysms, revolutions and wars) mutate and acquire increasingly dangerous global forms (nationalism, Nazism, terrorism, corruptionism) capable of destroying living forms of life on the planet. The carrier of the social virus is a social group, in which there is no elementary hygiene of maintaining the moral relations prescribed in the bible as 10 God's Commitments. In the absence of immunity and vaccination, a social virus is now almost instantly able to hit countries and continents through Internet communications. What is a social virus? It is a negative attitude to the difference between the moral and material possessions of two social subjects, Fig. 14, where N and M are the neighbours, P(N) and P(M) are their social properties.

There are four scalable types of relationships between Me (M) and Neighbour (N) (people, companies, states): (1) I am Evolutionair. Evolution relations based on my P(M) perfecting targeted to the Neighbour P(N) destruction. Become strong and kill the weak. (2) I am Revolutionair. Revolution relations based on my P(M) destruction targeted to the Neighbour P(N) perfecting. Destroy your old and create another's new. (3) I am Moralist. Moral relations based on my P(M) perfecting targeted to the Neighbour P(N) perfecting. Create yourself and help neighbor. (4) I am Terrorist. Terror relations based on my P(M) destruction targeted to the Neighbour P(N) destruction. Kill myself and the world around. There is only one moral living way from point of view: «Fact—Comparison—Action», when the

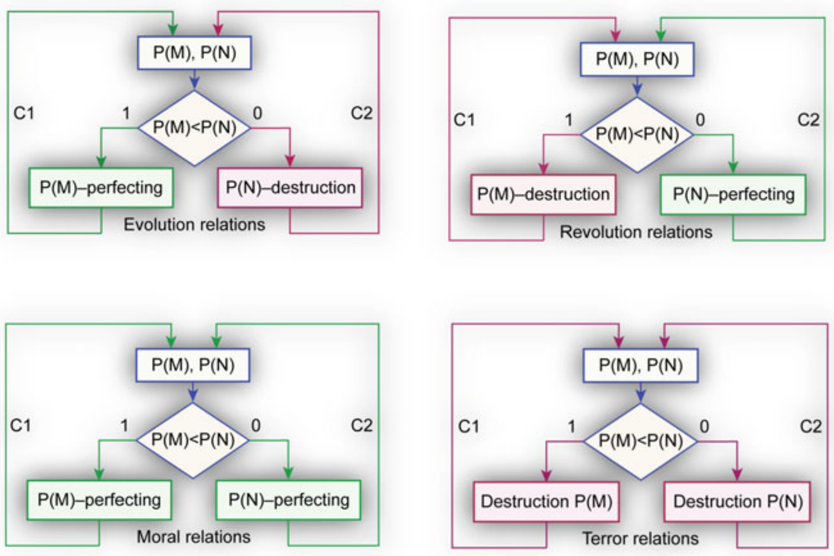


Fig. 14 Relations between two neighbour social subjects

actions are «My perfecting targeted to the Neighbour perfecting». The rest above mentioned three ways are Immoral ones. If the neighbor N lives better, then there are two ways of developing the subject M: (1) Positive way called P(M)-perfecting, which is long in time—to learn the experience of a neighbor and to surpass his achievements with honest creative work. (2) Negative way called P(N)-destruction, which is most effective in time—to destroy a neighbour and/or steal his property. What is social immunity? It is moral relations between the people, created by the constitution, culture and traditions, stimulating knowledge of the human experience and the dominance of the creative excellence of the citizen in the social significance of his work. Such morally healthy relations in a society can not be infected by social viruses of nationalism, terrorism and corruptionism. The alternative is the well-known doctrine of social equality of all citizens in income, which results in degradation of the social group: "We are happy in our equal poverty." Thus, the complete fulfillment of the human genome is entirely possible provided that the political elite of the planet creates social immunity—global moral relations aimed at ensuring the quality of life of each person and preserving the ecology of the planet. One possible variant of such immunity is cyber-social computing, designed to maintain the hygiene of social relations, preventing the penetration of social viruses.

7 Conclusion

1. Cyber-social computing as the theory and practice of moral, social relationship for precise cloud management of social processes based on big data and metric online monitoring of the social groups in order to improve the quality of people's life and preserve the green planet's ecology is proposed. The main idea is moral relationship between the people, created by the constitution, culture and traditions, stimulating knowledge of the human experience and the dominance of the creative excellence of the citizen in the social significance of his work. Such morally healthy human relations in a society can not be infected by social viruses of nationalism, terrorism and corruptionism.
2. The cyber trends from Gartner Inc. provide an opportunity for leaders of companies and university heads to keep up the digital business architecture in science, education and industry, timely respond to cyber threats, lead business innovation and determine an effective digital business strategy for sustainable development of the state.
3. In fact, Hype-cycle is a deep 4D-analysis, in time and space, the state of the modern market for sustainable cyber-physical development of smart cloud technologies for the next 10 years, which must be taken into account in business, science and education.
4. For universities, the Hype-cycle determines the vital need to invest in the knowledge of students the innovative technologies shown in the phases of the cycle, in order to obtain in 5–10 years the army of creative specialists capable of the state rising from the ruins of modern cyber ignorance. Otherwise, the

Gartner cycle for the university is a strategy of its cyber-physical sustainable development in time and space. Any strategy developed without knowledge of the direction of technological change has incorrect planning of actions, destruction of business, science and education. For example, it should be borne in mind that in 2018 robobosses will accurately monitor and remotely manage online 3 million workers in the world. The purpose of such management is to measure the potential of employees, the distribution of tasks that are invariant to the positions of the workers in the physical space, the evaluation of the quality of work, the salary calculation according to the metric results of achievements.

5. Hype-cycle implicitly differentiates all top technologies into master and slave (master-slave), which in fact mean that the development of Hardware (Physical Space) platforms towards compactness is always given priority. The rest of the virtual world (Cyber Space), striving for unlimited expansion of Software applications, will always be driven. Nevertheless, hardware and software technologies are represented in the Hype-cycle (on the market) in almost the same proportions (50:50): Hardware-driven technologies: 4D Printing, Volumetric Displays, Nanotube Electronics, Brain-Computer Interface, Human Augmentation, Autonomous Vehicles, Cognitive Computing, Commercial UAVs (Drones), Smart Dust, Smart Robots, Smart Workspace, Connected Home, 5G, IoT Platform, Edge Computing, Neuro-morphic Hardware, Quantum Computing. Software-driven technologies: Deep Learning, Deep Reinforcement Learning, Artificial General Intelligence, Enterprise Taxonomy, Ontology Management, Machine Learning, Virtual Assistants, Cognitive Expert Advisors, Digital Twin, Blockchain, Serverless PaaS, Software-Defined Security, Virtual Reality, Augmented Reality, Augmented Data Discovery, Conversational User Interfaces, Digital Humanity, Smart Cyber Digital State.
6. The same ratio of hardware and software technologies is shown in the Gartner-forecast, which means that the levels of their capitalization on the NASDAQ market tend to parity. A good example of parity is Apple ($800 billion—NASDAQ 2017) and Google ($570 billion). These manufacturers are significantly different in that they rely on the wisdom of their teams (experts) armed with the doctrine: “consumers can not predict their own needs” [11]. The alternative is the policy of Microsoft (503 billion), which conducts extensive research before launching the product, for example, such as Windows Phone. According to Gartner, Apple’s share in the global mobile phone market is 14.2% compared to 3.3% for Microsoft. Who do I trust, experts or consumers? The answer is unambiguous—to the experts, in 4D format (always, everywhere and on all issues).
7. Emotional-logical computing and its main components are proposed: the metric of measuring emotional relations on horizontal and vertical links, as well as the rules of behavior of the average person, focused on creating the emotional logic for modeling the behavior of the masses and green state power structures, as well as predicting their possible solutions.

8. A cyber-physical model of statehood is presented that is intended for metric management of resources and citizens based on digital monitoring and assessment of the social group needs in order to ensure the quality of their lives and the prosperity of the country through the production and export of goods and services. The structure of components of the green cyber-state computing, ordered by the degree of their influence on the market success is shown: (1) Relationship. (2) Goals. (3) Management. (4) Personnel. (5) Infrastructure. (6) Resources.

References

1. Columbus, L.: Gartner's Hype Cycle for Emerging Technologies, 2017 Adds 5G And Deep Learning for First Time. [Online] Available: https://www.forbes.com/sites/louiscolumbus/2017/08/15/gartners-hype-cycle-for-emerging-technologies-2017-adds-5g-and-deep-learning-for-first-time/#646a4cf34be2 (2017)
2. Press Release: Gartner Identifies Three Megatrends That Will Drive Digital Business Into The Next Decade. STAMFORD, Conn., 15 Aug 2017. [Online] Available: http://www.gartner.com/newsroom/id/3784363 (2017)
3. Frank, A.: The future of AI is neuromorphic. Meet the scientists building digital 'brains' for your phone. [Online] Available: http://www.wired.co.uk/article/ai-neuromorphic-chips-brains (2017)
4. Gupta, A., Jha, R.K.: A survey of 5G network: architecture and emerging technologies. IEEE Access. **3**, 1206–1232 (2015)
5. Zhu, C., Leung, V.C.M., Shu, L., Ngai, E.C.H.: Green internet of things for smart world. IEEE Access. **3**, 2151–2162 (2015)
6. Christidis, K., Devetsikiotis, M.: Blockchains and smart contracts for the internet of things. IEEE Access. **4**, 2292–2303 (2016)
7. Blockchains: How They Work and Why They'll Change the World. IEEE Spectrum. [Online] Available: https://spectrum.ieee.org/computing/networks/blockchains-how-they-work-and-why-theyll-change-the-world (2017)
8. Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.): Green IT Engineering: Concepts, Models, Complex Systems Architectures. Studies in Systems, Decision and Control (SSDC), vol. 1. Springer International Publishing, Berlin, Heidelberg (2017)
9. Kharchenko, V., Kondratenko, Y., Kacprzyk, J.: (eds.): Green IT Engineering: Components, Networks and Systems Implementation. In: Studies in Systems, Decision and Control (SSDC), vol. 2. Springer International Publishing, Berlin, Heidelberg (2017)
10. Zanella, A., Bui, N., Castellani, A., Vangelista, L., Zorzi, M.: Internet of things for smart cities. IEEE IoT J. **1**(1), 22–32 (2014)
11. Cearley, D.W., Walker, M.J., Burke, B.: Top 10 Strategic Technology Trends for 2017. 14 Oct 2016. [Online] Available: https://www.gartner.com/doc/3471559?srcId=1-7578984202&utm_campaign=RM_GB_2017_TRENDS_QC_E2_What&utm_medium=email&utm_source=Eloqua&cm_mmc=Eloqua-_-Email-_-LM_RM_GB_2017_TRENDS_QC_E2_What-_-0000 (2016)
12. Three Digital Marketing Habits to Break. 15 April 2015. [Online] Available: http://www.gartner.com/smarterwithgartner/three-digital-marketing-habits-to-break-2/ (2015)
13. Hahanov, V.: Cyber physical computing for IoT-driven services. Springer, New York (2018)

14. Hester, P.T., Adams, K.: Systemic Decision Making. Fundamentals for Addressing Problems and Messes. Springer Nature (2017)
15. Hahanov, V., Barkalov, A., Adamski, M.: Infrastructure intellectual property for SoC simulation and diagnosis service. In: Design of Digital Systems and Devices. Springer, Germany (2011)
16. Dubilet conceived a revolution. [Online] Available: http://dpchas.com.ua/politika/dubilet-zadumal-revolyuciyu (2017)
17. Wang, Y., Li, L., Yang, L.: Cyber-physical social systems. Intelligent human resource planning system in a large petrochemical enterprise. IEEE Intell. Syst. **28**(4), 102–106 (2013)
18. Liu, Z., Yang, D-Sh, Wen, D., Zhang, W.-M., Mao, W.: Cyber-physical-social systems for command and control. IEEE Intell. Syst. **26**(4), 92–96 (2011)
19. Hellinger, A., Seeger, H.: Cyber-physical systems. Driving force for innovation in mobility, health, energy and production. Acatech (2011)
20. Yaqoob, I., Ahmed, E., Hashem, I.A.T., Ahmed, A.I.A., Gani, A., Imran, M., Guizani, M.: Internet of things architecture: recent advances, taxonomy, requirements, and open challenges. IEEE Wirel. Commun. **24**(3), 10–16 (2017)
21. Song, A., Wang, L.: Higher education of measurement control and instrumentation specialty in China. IEEE Instrum. Meas. Mag. **19**(2), 39–42 (2016)

Architecture for Collaborative Digital Simulation for the Polar Regions

Marina Solesvik and Yuriy Kondratenko

Abstract Offshore oil and gas operations continuously go to the High North. The ecology of the Arctic area is very fragile. The maritime industry is mainly eco-destructive. In the Arctic, the problems of possible ship accidents and catastrophes are extremely dangerous for the nature. In this paper, we analyse recent literature on new digital simulation technology and propose new ways of the utilization of this novel technology in the maritime industry. We present fresh insights related to the architecture, software and hardware related to joint digital simulation platforms in planning safe maritime operations in the Arctic and effective search and rescue operations. For the facilitation of integrated systems, the project will work towards a cross industry, open platform ecosystem connecting operations with the help of cloud technology: from data collection to storage through analytics. The utilization of the joint digital simulation platforms will allow to minimize pollution effects during navigation and accidents. The study will be interesting for scholars, policy-makers and practitioners from the maritime industry.

Keywords Digital platform · Maritime industry · Green supply chain
Green IT

1 Introduction

Governments in many countries consider the development of renewable energy sources as one of the priority political tasks. Though alternative energy sources are actively being developed and implemented worldwide, the role of hydrocarbons (oil and gas resources) is still very important, and will continue to be so for many years

M. Solesvik (✉)
Nord University Business School, Nord University, 8049 Bodø, Norway
e-mail: mzs@hvl.no

Y. Kondratenko
Petro Mohyla Black Sea National University, Mykolaiv 54000, Ukraine
e-mail: y_kondrat2002@yahoo.com

© Springer Nature Switzerland AG 2019
V. Kharchenko et al. (eds.), *Green IT Engineering: Social, Business and Industrial Applications*, Studies in Systems, Decision and Control 171,
https://doi.org/10.1007/978-3-030-00253-4_22

in the future. The results of geological studies suggest that about 12% of the world's oil and 25% of gas resources are situated in the Arctic [1]. Five countries have a border with the Arctic (Canada, Denmark and Greenland, Norway, Russia, and the USA), and intend to use this opportunity and participate in oil and gas extractions in the Arctic. Taking into account the importance of the Arctic hydrocarbon resources, the governments of interested countries have established Arctic strategies; i.e. the Norwegian Government has developed long-term programmes and strategies for the Arctic, such as the Barents 2020 project; the Russian Government has worked out a Development Strategy of the Arctic Zone of the Russian Federation and National Security for the period up to 2020; and Canada has approved its Northern Strategy [2]. The oil and gas sector is very important for the economies of these countries. In Norway, the oil and gas sector and related industries (suppliers, shipping companies, etc.) are the major driving forces of prosperity and employment in the country. The same can be said about Russia. The depletion of traditional fields of hydrocarbons, such as in the North and Norwegian Seas in Norway, and Western Siberia in Russia, are driving oil and gas companies and other interested governments to look forward to exploiting rich hydrocarbon resources in the Arctic [3].

The increasingly growing number of maritime operations go in turbulent environments with high insecurity and difficult weather conditions. Turbulent environments are defined as environments with a high degree of volatility and complexity [4]. Logistics operations in turbulent maritime settings greatly influences on the entire value chain, as well as cooperation with search and rescue services and governmental bodies (such as coast guard, police, hospitals and others). Moreover, the imperfect search and rescue (SAR) facility in some regions of the High Arctic is a problem. Long distances and an absence of infrastructure demand, first, multi-purpose special ships with SAR, firefighting, oil-spill recovery, towing, and ice management functions in order to provide safety and back up. Simulations of operations and the orchestration of actions of the different services as well as coordination of information from different actors via the cloud-based technology would contribute to skill-development related to personnel. In real-life situations it will help to act quickly and correct which will result in minimal damages to people's lives and the environment. The analysis of prior research revealed a significant gap in the knowledge base related to collaborative simulation activities.

The objectives of this chapter are to contribute to increased safety and preparedness in polar regions through improving the capabilities of vessels to interlink through cloud-based platforms with main educational simulators. In this study, we offer a digital solution to connect different simulation facilities situated in remote places. Business and ICT innovations play an important role in the improvement of the maritime performance [5].

This chapter focuses on the ecology of the Arctic and shipping companies operating in the area. The Norwegian maritime sector and its shipping companies are innovative and are sustainably leading. The Norwegian Government pays a lot of attention and fund a lot of research projects related to green technology in the sea. The Green IT technology is one of the most important research areas.

Norwegian shipping companies were the first ones in the world to introduce electrical ferries, autonomous ships, hybrid vessels and etc. The ecological goals are often combined with economic targets. The introduction of green vessels allows achieving significant cost reduction benefits [6]. The research questions that guided our research are: (1) What is the role of ICTs during interorganizational collaboration? (2) How can joint collaborative digital platforms contribute to solving ecological issues in the Arctic area?

The paper will be interesting for policy-makers and scholars in different regions since the Arctic comprises Alaska, the North Eastern Part of Canada, the Greenland coast, the Barents Sea, North and East Russia, and the North East Passage (Northern Sea Route) through Northern Russia. The paper will also be interesting not only for specialists in the maritime business and shipping. The new simulation platform can be used in other industries as well.

The paper is organized as follows. In Sect. 2, we highlight the context of the study and draw attention to the specifics of the Arctic region and to challenges that face stakeholders for offshore operations. In the next section, we present the theoretical background of the study related to ICT and the digital simulation technology. In Sect. 4, we present the case on how the collaborative IT simulation technology contributes to the solution of green issues in the Polar region. Finally, we derive conclusions and present theoretical and practical implications.

2 Arctic Context and Challenges

Arctic areas can be defined as areas north of the longitudinal treeline [7]. This area has a mean July temperature of 10 °C [8]. The Arctic is very rich in oil and gas and other mineral resources [1]. However, there are a number of obvious and less obvious reasons that challenge extraction works in the Arctic. In this chapter, we will closely consider the challenges that are ecologically related to hydrocarbon extraction in the High North. A number of political, environmental, ecological, technological and economic issues will be taken into account regarding oil and gas firms' intents to explore hydrocarbons in the Arctic. First, the Arctic is still a disputed territory from the international law point of view. Several adjacent states claim that they have rights to the North Pole and its surrounding areas [9]. Even states that do not border the Arctic (e.g. China) assert that the Arctic is the shared property of all countries in the world. The border issues play an important role and will have an influence in the future. Border agreements between countries can also hinder hydrocarbon and other mineral resource extraction in the Arctic. For example, there is still a disputed offshore international boundary between the Yukon and Alaska. The maritime border agreement between Norway and Russia was only signed recently (in 2010), which opened up oil and gas exploration in the Barents Sea [10].

Second, the weather conditions in the Arctic are considerably more difficult than in the Northern Atlantic where major oil and gas companies have accumulated significant expertise. Severe weather conditions make offshore operations and subsea works exceptionally difficult [11–13]. Previous research has suggested that low temperatures (up to −40 to −50 °C), huge temperature differences, Arctic ice cover, drifting ice and floating icebergs, permafrost, waves, sea spray, wind, polar lows (unexpected changes of weather when a severe storm can start very quickly in 0.5–2 h), and darkness represent major climate challenges in the Arctic [8, 14]. This leads to additional challenges. For example, the existing technology allows offshore operations to only be carried out in some areas during periods when the water surface is free from ice. The offshore vessels are then only used several months per year. This means that they are hired at spot rates that are significantly higher than rates under long-term contracts for periods of up to 15 years.

Third, longer distances between the shore and the hydrocarbon fields are a drawback. The distances to the hydrocarbon reservoirs in the North and Norwegian Seas are relatively short. In some cases, it only takes three to four hours to reach the platform from the shore. In the Arctic, the distances between the mainland and the discovered oil and gas fields may be 500–600 km, and this is not the limit. This is a logistics challenge for oil and gas companies and their subcontractors who are supposed to deliver and evacuate people, and bring cargo to platforms. The emergency response time is also increased, when taking longer distances to the shore into account. In addition, the infrastructure is underdeveloped in areas adjacent to the Arctic seas. The above are major concerns of experts and practitioners since the development of appropriate infrastructures (e.g. railways, roads, bridges, telecommunication facilities, and so on) lead to significant capital costs, and increases the return-on-investment period [15]. Longer distances to the main supply bases increase the cost of supplies. Some offshore shipping companies that already have operations in the Norwegian part of the Barents Sea (Goliat and Snøhvit fields) established local warehouses in Hammerfest, while others deliver necessary parts and materials in almost a daily basis from their home bases (e.g. Esvagt shipping company delivers what is necessary from their base in Denmark by plane) [16]. Longer distances to the main consumers are also reflected in the higher costs for oil and gas extracted in the Arctic.

Fourth, the GPS, mobile connection, and telecommunications do not always work properly in the Arctic. This leads to difficulties communicating with personnel on ships and platforms. The dynamic positioning is complicated when the GPS does not work properly.

Fifth, the environment in the Arctic is very fragile. That is why the Technology Strategy of the Arctic in Norway [8] stipulates the strictest safety requirements for offshore operations. The governments of other Arctic states are also concerned with environmental issues. Oil and gas companies pay a lot of attention to the preservation of the unique nature of the Arctic. Possible oil spills there may have unpredictable consequences for the nature. Subsequently, oil spills would lead to a negative media image and large costs related to oil recovery. Oil and gas companies will pay higher prices to subcontractors who need to offer more expensive technical solutions that would work under low temperatures, ice and harsh conditions. For example, ships will additionally need winterization of the vessels, de-icing systems, metallization of the surfaces, ice-breaking capacities, thicker hulls, and etc. All these additional options are reflected in the price of the vessels. For example, the Vard shipyard in Norway estimates that the anti-icing equipment installed on a vessel costs about 600,000 Euros, and an ice-breaking capacity leads to a 200–250% increase in building costs, compared to light ice-class vessels [17]. The fuel consumption will increase as well. All these factors and the lack of established technology are responsible for much higher costs of hydrocarbon extraction in the Arctic.

Arctic offshore operations will need new technology, routines, and equipment (Research Council of Norway 2006). The Arctic can be roughly divided into three areas: (1) an ice-free zone; (2) a zone covered with the ice several months of the year; and (3) an area covered with the ice the whole year. For the areas without ice (e.g., the Barents Sea), the existing technology of oil extraction and offshore support is available with minor changes. This is reflected in the Norwegian Technology Strategy for the Arctic [8: p. 6]:

> From the Norwegian side it is now important to demonstrate the capability to adapt existing technology in this area together with developed concepts meeting the challenge in a region with ice and icing together with darkness, long distances and lacking infrastructure.

For areas covered with ice for some months of the year (e.g., the Kara Sea), incremental innovations need to be developed, and for the areas with surfaces covered with ice during the whole year, radical innovations need to be developed and implemented [18].

3 ICT and Digital Platforms Technology in Maritime Industry

In this section we will propose classification of ICT resources. We also will consider the issues related to the usage of ICT resources and competencies in enhancing collaboration between different actors aiming to support green IT activities.

3.1 Classification of ICT Resources

Prior research has applied different terms addressing issues relating to computer-aided solutions, notably information technology (IT) [19, 20]; information systems (IS) [21]; or a combination of both—IS/IT [22]; and information and communication technology (ICT) [23]. Recent developments on the Internet technologies allowed combining two different types of communication, notably, voice and data. For the purposes of this study, the term ICT is used as the most appropriate since it includes recent trends and takes account of the term ‘communication’, which is important in the interfirm collaboration context. The information and communication technology is defined as "computer and communication technologies (both hardware and software) used to process, store, retrieve, and transmit information in an electronic form" [24: p. 595].

The ability of firms and organizations to cooperate during operations in the Arctic has an important impact on a firm's ability to gain sustained competitive advantage [25]. ICT competences can enable a firm to pursue a range of strategies and they can enable collaborative strategies [26, 27]. Previous research [21] has classified three types of ICT resources, technical, human, and intangible. The classification of ICT resources is summarized in Table 1. Technical ICT resources consist of software, hardware, software, networks, and applications. Human ICT resources relate to skills in communication, information technology (IT) skills, and IT management skills. Intangible ICT resources are knowledge assets, collaboration between IT departments and other functional divisions, flexible and adaptable ICT culture, customer orientation and supplier relationships.

ICT competences are embedded in organizational processes and firm routines [28]. An information systems competence is created when processes and structures are utilized in inimitable ways to bundles of ICT resources, to build specific abilities for completing ICT-related organizational tasks [21]. Prior research [20, 29] has identified the following five potential characteristics of information technologies as possible sources of competitive advantages: the proprietary technology; managerial IT skills; technical IT skills; access to capitals; and customer switching costs. ICT competences may generate business values by cost reduction and differentiation of products and services [30, 31]. However, Mata et al. [20] have assumed that only managerial IT skills are sources of sustained competitive advantages.

Table 1 Classification of ICT resources and competences

ICT resource type	ICT resources	Sources
Technical	Hardware, software, applications, networks, databases, and applications	Armstrong and Sambamurthy [32], Keen [33], Weill and Broadbent [34]
Human	Skills of IS/IT personnel, IT management skills, skills in communication, understanding of business	Bharadwaj [35], Mata et al. [20], Peppard and Ward [36]
Intangible	Knowledge assets, a customer orientation, collaboration between IT department and other functional divisions, flexible IS culture, supplier relationships, an adaptable ICT culture	Bassellier and Benbasat [37], Feeny and Willcocks [19], Powell and Dent-Micallef [38]

3.2 The Role of ICT Resources and Competencies in Collaboration

Information and communication technologies develop rapidly. Managers need to take breath-taking changes into account and consider how to leverage new capabilities for their firms. Recent technological trends in ICT include the minimization of computing resources, the rapid penetration of broadband networks, the emergence of the Internet as a communication backbone, the appearance of the wireless technology, and the development of mobile devices [39]. At the same time, the rapid development of interfirm collaboration has been documented. The rate of collaboration agreements is growing at a rate of 25% since 1985 [40]. Business-process redesigns are important when firms engage into alliances and develop inter-organizational relationships. It has been argued that IT and business-process redesigns are natural partners [41].

Boland et al. [42] have suggested that the use of ICT in organizations leads to four types of changes: gaining large scale efficiencies, enhancing decision making and communication, changing the basis for competition and industry structure, and exploiting new business models. Nowadays, the fifth change is vitally important—the green vector and the growing importance of IT technologies in solving green issues [43, 44]. Gaining large-scale efficiencies is related to the introduction of different ICT tools (e.g. enterprise resource planning, computer-aided design, computer-aided manufacturing, cloud-based technologies, etc.), which allowed the decrease of the processing time from several days to several minutes [45]. Enhancing decision making and communication supports communication and collaboration processes in different firms. The introduction of group support systems (GSS), such as e-mail, computer conferencing systems, videoconferencing, and electronic meeting tools, in organizations added value to these organizations [46]. There are three ways on how GSS adds value: (1) the reduction of time and distance barriers related to face-to-face interactions, (2) increasing the scope of available information, (3) improving interaction processes within groups [39]. Emergence and the development of ICT enabled communication tools that lead to the appearance of the novel organization forms, notably virtual teams which are sometimes referred as "temporary, geographically dispersed, culturally diverse, and electronically communicating working groups" [39: pp. 600–601]. In the maritime context, virtual teams are a viable form of work. The use of ICT changes the industry structure in a number of ways. ICT might aggregate previously fragmented markets; bundle productive resources more effectively; change the nature of interfirm relations; and replace existing channels [39].

The first research question of this study is: what is the role of ICTs during collaboration? Collaborative management systems (which are also called groupware technologies) can add a firm's value in three ways: (1) significant cost savings, (2) faster cycle times [41], and (3) enhanced flexibility and responsiveness to customer requirements [47]. Collaborative management systems, such as Groove [48] assist collaboration by helping to ease communication, share knowledge, and manage projects. However, some researchers put that information forward and so the communication technology itself cannot be a source of competitive advantage [38]. The success or failure of ICT tool-applications depends on the competences of people who use them [49].

Significant research is done regarding how ICT can facilitate alliances. However, there is still a lacuna in our knowledge base. Notably, the research on ICT applications facilitating collaboration inside and outside maritime firms is scarce [50, 51]. The research in emerging collaborative ICT technologies also important because they "can help employees build in other's ideas, ease connections between people and develop shared knowledge. The result will be an organization with greater velocity and greater capacity to adapt to changing environs" [48: p. 222].

Recent developments in information and communication technologies (ICT) have enabled firms to engage more easily into various forms of interorganizational collaboration. New forms of collaboration are steadily getting more virtual. This is where organizational resources from different firms are pooled and linked together electronically [52]. ICTs help to overcome the spatial barrier between inputs from geographically remote firms. Furthermore, ICTs enable firms and organizations to share firm-specific and firm-addressable resources during collaboration processes [53]. ICT tools help the overcome of some shortcomings of external collaboration relating to the potential loss of flexibility and the time spent by managers negotiating and implementing a collaborative arrangement [54]. Grant and Baden-Fuller [55] argued that an interfirm collaborative arrangement is a vehicle to access and acquire knowledge. New knowledge and enhanced efficiency can be generated by an interfirm collaboration that uses appropriate ICTs [56].

Few studies have examined how ICTs might also facilitate interfirm collaboration in the production design phase. This study will empirically explore the role of ICT systems with regard to collaboration during operations in the Arctic. The role of ICTs in reducing coordination costs and improving knowledge management during the collaboration process will be explored in the next section.

4 The Case of a Collaborative Digital Simulator

Shipping companies need to have highly qualified, updated and skilled personnel on board of their vessels operating in the Arctic. Thus, the tight collaboration with educational institution is necessary to provide specific competencies. A novel digital platform that will enable linking separate simulation facilities are aimed to train different skills into the joint virtual simulation and training system. For example, a virtual campus will be the first step to real-life simulation. The virtual campus will be built upon the newly established simulator park at two maritime schools and one university in the Northern Norway. The University and maritime schools with unique maritime simulators share navigation and emergency modules. The first Maritime School has the oil spill preparedness simulator. The second Maritime School has a dynamic positioning simulator for operations in emergency situations and the University has a new part-task ship's bridge simulator with visual scenes, command and control systems as well as log systems, which are essential in crisis management education and training. These simulator systems are linked to cloud-based facilities for data collection and computation, and are connected to the navigation, safety, security and other decision-support systems of the vessels.

For the facilitation of integrated systems, the project will work towards a cross industry, open platform ecosystem connecting operations with IT: from data collection to storage through analytics. A centralized data collection and computation in the cloud are followed in a value chain by data generation and computation on the site through internet browser software and mobile device apps.

The virtual simulator platforms will facilitate:

(1) Safe navigation scheduling through the simulation of routes using integrated simulator-navigation scheduling systems.
(2) Efficient shore-vessel decision-making by connecting different subsystems including safety management systems, maintenance systems and administrative systems in the cloud.
(3) SAR and oil spill response mission coordinator instruments facilitating the simulation of emergency prevention, preparedness and response through the integration of command and control systems, log systems and vessel simulator systems.

For the present moment, we have achieved the connection of three simulators situated at three different locations via the cloud-based technology (Fig. 1). This allows cooperation between the three educational institutions, i.e., first, schools can share the expensive simulator facilities of each other through the Internet. Students do not need to commute between the campuses to use different types of simulators that schools posses. Second, they can drive joint exercises related to search and rescue operations in the sea.

The virtual campus digital platform may not only be used by students but also by vessel crews to learn new skills related to remote operations, and to simulate different operational solutions. The collaborative digital simulation platform will build upon the digital twin ship producing holistic and fully integrated digital twins including simulated data, real time data and visualizations.

The joint digital platform will promote the use of artificial intelligence for the digitalization of maritime operations. This, among others, will be important in integrated operations at sea such as search and rescue operations. So far, the parties involved in the search and rescue (SAR) operations do not have access to each other's data. This means that during critical situations on sea when each second counts for the rescue of people's lives and the prevention of environmental catastrophes, time is used to collect available information at other actors and coordinate different data streams. The creation of the joint digital platform gives all the involved parties instant access to real time data. Security and terror threats are also important issues. The development of the joint digital platform will be a building

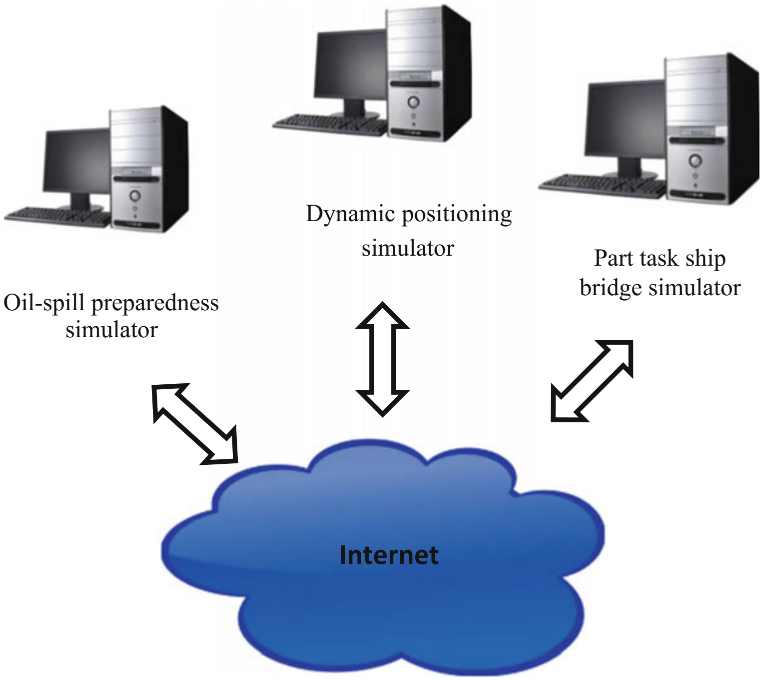


Fig. 1 The connection of functional structure of simulators' connection

block to a wider use of artificial intelligence in maritime operations, as well as for the transfer of steering and full control over the vessel to a remote operator.

In the future, we expect to connect simulators to the vessels in the sea and to carry out the training as well as search and rescue operations together with the experienced crews. Next, the information from different actors at the moment is not shared. The goal of the joint digital platform would be to serve as a platform for collecting and analysing the existing data coming from different actors (police, rescue company, shipping companies, coast guard, etc.) (Fig. 2).

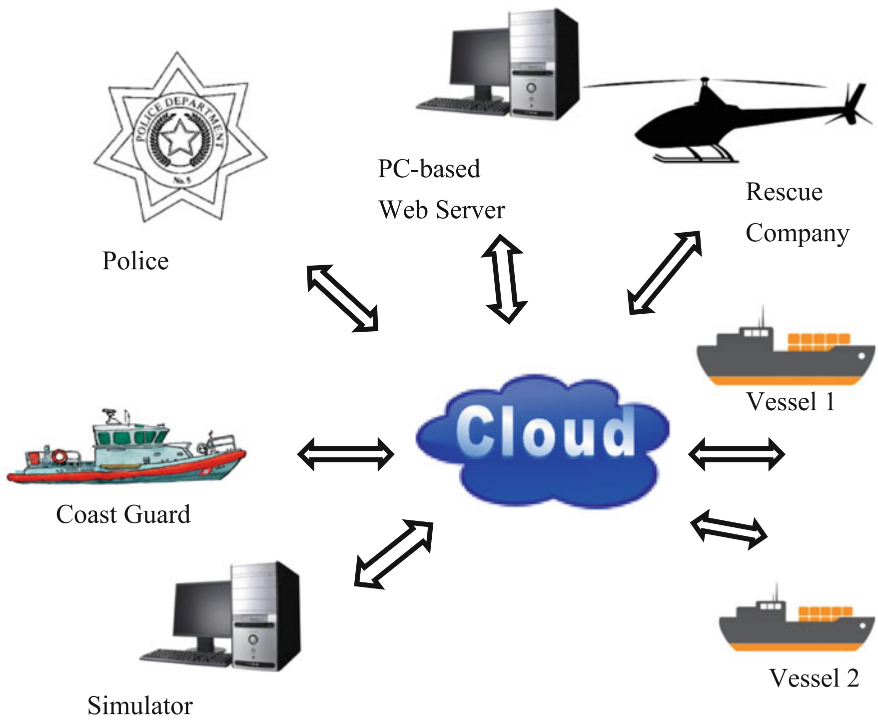


Fig. 2 Functional structure of the joint digital platform for the future

5 Conclusions

Governments and private companies have an interest in the development and exploitation of resources in Polar Regions. The depletion of oil and gas resources in the North Sea force and Norwegian petroleum activities move steadily to the North into the Norwegian Sea, the Barents Sea and further in Arctic waters. The region comprises about 25% of the outstanding oil and gas reserves in the world. The steadily growing exploration of polar waters and shelf, the increasing utilization of the Northern Sea Route, global warming in the world and other novel trends lead to the contamination of the environment. The High Arctic includes the northern regions of the Arctic Circle where cold climate may cause severe ice and icing conditions. The transportation into the Arctic may be alike in summer time to the North Sea. In winter, vessels may meet operational challenges related to polar lows, ice cover on the hull and superstructure, the mixture of waves, darkness and sea ice, together with fog and ice. These issues disturb scholars, practitioners and policy-makers. IT technologies made a significant progress during the last decades and has a wide range of instruments that can contribute to the solution of different green issues in order to preserve the environment.

Digitalization is one of the key priority areas of the Norwegian maritime industry. Modern ICT technologies open unique possibilities for many industries [57–66], including maritime sector [67, 68]. Green ICT technologies will be very important in a new era in the maritime sector from remotely operated vessels to autonomous ships. Maritime simulators can be excellent bridges from pure virtual simulations to real-time remote control over vessels. For shipping companies operating in Polar Regions, several challenges are present as to maintenance and the renewal of competence and operational planning. Long distances to education infrastructure together with extra demands both to planning, competences, and the renewal of certificates call for new educational platforms. Through the creation of an open, secure and collaborative platform and a new digital infrastructure, vessels may have continuous access to advanced simulators and may be able to use their own technology on board to provide both added competence and simulator opportunities to facilitate safe navigation and fast response in case of emergencies. A special interest for maritime industry specialists represents the perspectives of the utilization of the digital simulation technology in improving environmental efficiency of the maritime industry. The simulation technology allows the use of artificial intelligence for the training of sailors. It also makes it possible to coordinate rather big volumes of data from different actors using cloud-based technology.

References

1. Economist: Frozen conflict: Denmark claims the North Pole. 20 Dec 2014. http://www.economist.com/news/international/21636756-denmark-claims-north-pole-frozen-conflict?fsrc=scn%2Ffb%2Fte%2Fpe%2Fed%2Ffrozenconflict (2014)
2. Government of Canada: Canada's Northern Strategy. http://www.northernstrategy.gc.ca/index-eng.asp (2014)
3. Sundalsfoll, S., Bjerke, H.: Innovation in the Russian oil & gas sector-prepared for Arctic offshore development? Unpublished Master Thesis. University of Nordland, Bodø (2014)
4. Erikstad, S.O., Ehlers, S.: Decision-support framework for exploiting Northern Sea route transport opportunities. Ship Technol. Res. **59**(2), 34–43 (2012)
5. Jenssen, J.I., Randøy, T.: Factors that promote innovation in shipping companies. Marit. Policy & Manage. **29**(2), 119–133 (2002)
6. Solesvik, M., Gulbrandsen, M.: Partner selection for open innovation. Technol. Innov. Manage. Rev. **3**(4), 11–16 (2013)
7. Chapin III, F.S., Jefferies, R.L., Reynolds, J.F., Shaver, G.R., Svoboda, J.: Arctic plant physiological ecology in an ecosystem context. In: Chapin III, F.S., Jefferies, R.L., Reynolds, J.F., Shaver, G.R., Svoboda, J.: Arctic Ecosystems in a Changing Climate, pp. 441–452. Academic Press, San Diego (1992)
8. Research Council of Norway: Technology Strategy for the Arctic. www.forskningsradet.no (2006)
9. Sale, R., Potapov, E.: The Scramble for the Arctic. Ownership, Exploitation and Conflict in the Far North. Frances Lincoln, London (2010)
10. BBC: Russia and Norway Sign Maritime Border Agreement. 15 September. British Broadcasting Corporation. http://www.bbc.co.uk/news/business-11316430 (2010)

11. Borch, O.J., Batalden, B.M.: Business-process management in high-turbulence environments: the case of the offshore service vessel industry. Marit Policy Manage **42**(5), 481–498 (2015)
12. Masterson, D.M.: The Arctic Islands adventure and Panarctic Oils Ltd. Cold Reg. Sci. Technol. **85**(1), 1–14 (2013)
13. Staalesen, A.: Dramatic return from Dolginskoe. Barents Observer. 19 Nov. http://barentsobserver.com/en/energy/2014/11/dramatic-return-dolginskoye-19–11 (2014)
14. Pribytkov, E.A., Zolotukhin, A.B., Gudmestad, O.T.: Selection of subsea production systems for the field development in the Arctic environment. Paper presented at the SPE Arctic and Extreme Environments Technical Conference and Exhibition, Moscow, 15–17 Oct 2013
15. Harrison, C.: Industry perspectives on barriers, hurdles, and irritants preventing development of frontier energy in Canada's Arctic Islands. Arctic **7**(2), 238–242 (2006)
16. Falk-Petersen, M., Oterhals, T.: Entering the high tech offshore service vessel market. Unpublished Master's Thesis. Bodø, University of Nordland (2014)
17. Ulstein, S.: Vard: We are shipbuilders. Paper presented at the Oplog Workshop on Offshore Service Vessel Operations in the High North, on board of MS Richard With, 6–7 May 2014
18. Dynkin, A.: Changing Russia in the changing global energy landscape. Paper presented at the 12th Annual Forum, Club de Nice, Energie et Geopolitique, Nice, 28–30 Nov 2013. www.clubdenice.eu/2013/Alexandre_DYNKIN.pdf (2013)
19. Feeny, D.F., Willcocks, L.P.: Core IS capabilities for exploring information technology. Sloan Manage. Rev. **39**(3), 9–21 (1998)
20. Mata, F.J., Fuerst, W.L., Barney, J.B.: Information technology and sustained competitive advantage: a resource-based analysis. MIS Q. **19**(4), 487–505 (1995)
21. Tarafdar, M., Gordon, S.R.: Understanding the influence of information systems competencies on process innovation: a resource-based view. J. Strategic Inf. Syst. **16**(4), 353–392 (2007)
22. Pitt, M., MacVaugh, J.: Knowledge management for new product development. J. Knowl. Manage. **12**(4), 101–116 (2008)
23. Bouwman, H., van den Hooff, B., van de Wijngaret, L., van Dijk, J.: Information and Communication Technology in Organizations: Adoption, Implementation. Use and Effects. SAGE, London (2005)
24. Kogut, B.: Joint ventures: theoretical and empirical perspectives. Strategic Manage. J. **9**(4), 319–332 (1988)
25. Solesvik, M.: Interfirm collaboration in the shipbuilding industry: the shipbuilding cycle perspective. Int. J. Bus. Syst. Res. **5**(4), 388–405 (2011)
26. Barney, J.B.: Gaining and Sustaining Competitive Advantage. Prentice Hall, Upper Saddle River, NJ (2002)
27. Solesvik, M., Kondratenko, Y., Kondratenko, G., Sidenko, I., Kharchenko, V., Boyarchuk, A.: Fuzzy decision support systems in marine practice. In: IEEE International Conference on 2017 Fuzzy Systems (FUZZ-IEEE) 9 July, pp. 1–6. IEEE (2017)
28. Teece, D.J.: Strategies for managing knowledge assets: the role of firm structure and industrial context. Long Range Plan. **33**(1), 35–54 (2000)
29. McFarlan, F.W.: Information technology changes the way you compete. Harvard Bus. Rev. **62**(3), 98–103 (1984)
30. Bakos, J.Y., Treacy, M.E.: Information technology and corporate strategy: a research perspective. MIS Q. **10**(2), 107–119 (1986)
31. Porter, M.E., Millar, V.E.: How information gives you competitive advantage. Harvard Bus. Rev. **63**(4), 149–160 (1985)
32. Armstrong, C.P., Sambamurthy, V.: Information technology assimilation in aims: the influence of senior leadership and IT infrastructures. Inf. Syst. Res. **10**(4), 304–327 (1999)
33. Keen, P.G.W.: Information technology and the management difference: a fusion map. IBM Syst. J. **32**(1), 17–39 (1993)
34. Weill, P., Broadbent, M.: Leveraging the new infrastructure: how market leaders capitalize on information technology. Harvard Business School Press, Boston, MA (1998)

35. Bharadwaj, A.: A resource-based perspective on Information Technology and firm performance: an empirical investigation. MIS Q. **24**(1), 169–196 (2000)
36. Peppard, J., Ward, J.: Beyond strategic information systems: toward an IS capability. J. Strategic Inf. Syst. **13**(2), 167–194 (2004)
37. Bassellier, G., Benbasat, I.: Business competence of information technology professionals: conceptual development and influence on IT-business partnerships. MIS Q. **28**(4), 394–673 (2004)
38. Powell, T.C., Dent-Micallef, A.: Information Technology as competitive advantage: the role of human, business, and technology resources. Strategic Manage. J. **18**(5), 375–405 (1997)
39. Alavi, M., Yoo, Y.: Use information technology for organizational change. In: Locke, E.A. (ed.) Handbook of Principles of Organizational Behavior, pp. 595–614. Wiley, Chichester (2009)
40. Pekar, P., Allio, R.: Making alliance work. Long Range Plan. **27**(4), 54–65 (1994)
41. Attaran, M.: Information technology and business-process redesign. Bus. Process Manage. J. **9**(4), 440–458 (2003)
42. Boland, R.J., Lyytinen, K., Yoo, Y.: Wakes of innovation in project networks: the case of digital 3-D representations in architecture, engineering and construction. Organ. Sci. **18**(4), 631–647 (2007)
43. Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.): Green IT Engineering: Concepts, Models, Complex Systems Architectures. Studies in Systems, Decision and Control, vol. 74. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-44162-7
44. Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.): Green IT Engineering: Components, Networks and Systems Implementation. Studies in Systems, Decision and Control, vol. 105. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-55595-9
45. Davenport, T.H.: Putting the enterprise into the enterprise systems. Harvard Bus. Rev. **76**(4), 121–131 (1998)
46. Briggs, R.O.: On theory-driven design of collaboration technology and process. In: de Vreede, G.-J., Guerrero, L.A., Raventos, G.M. (eds.) Groupware: Design, Implementation and Use, pp. 1–15. Springer, Heidelberg (2004)
47. Choy, K.L., Lee, W.B., Lo, V.: An enterprise collaborative management system—a case study of supplier relationship management. J. Enterp. Inf. Manage. **17**(3), 191–207 (2004)
48. McKnight, B., Bontis, N.: E-Improvisation: collaborative groupware technology expands the reach and effectiveness of organizational improvisation. Knowl. Process Manage. **9**(4), 219–227 (2002)
49. Massa, S., Testa, S.: ICTs adoption and knowledge management: the case of an E-procurement system. Knowl. Process Manage. **14**(1), 26–36 (2007)
50. Bronsart, R., Grafe W., Koch, T.: A Collaborative Platform for Ship Design. In: Proceedings of the 12th International Conference on Computer Applications in Shipbuilding 23–26 Aug 2005, Busan, S. Korea, pp. 1–15 (2005)
51. Solesvik, M.Z., Encheva, S.: Partner selection for interfirm collaboration in ship design. Ind. Manage. Data Syst. **110**(5), 701–717 (2010)
52. Nagel, P., Dove, M.: The Virtual Corporation. Working Paper. Illinois, Le High University (1991)
53. Sanchez, R.: Quick-connect technologies for product creation: implications from competence-based competition. In: Sanchez, R., Heene, A., Thomas, H. (eds.) Dynamics of Competence-based Competition, pp. 299–322. Elsevier Pergamon, Oxford (1996)
54. Xie, F.T., Johnston, W.: E-business technological innovations. J. Bus. Ind. Mark. **19**(3), 208–222 (2004)
55. Grant, R.M., Baden-Fuller, C.: A knowledge accessing theory of strategic alliances. J. Manage. Stud. **41**(1), 61–84 (2004)
56. Papoutsakis, H.: How far can information systems support interfirm collaboration? J. Knowl. Manage. Pract. **7**(3), from http://www.tlainc.com/articl119.htm (2006)
57. Kharchenko, V., Illiashenko, O.: Concepts of green IT engineering: taxonomy, principles and implementation. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT

Engineering: Concepts, Models, Complex Systems Architectures. Studies in Systems, Decision and Control, vol. 74, pp. 3–19. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-44162-7_1
58. Kondratenko, Y., Korobko, O.V., Kozlov, O.V.: PLC-based systems for data acquisition and supervisory control of environment-friendly energy-saving technologies. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.), Green IT Engineering: Concepts, Models, Complex Systems Architectures. Studies in Systems, Decision and Control, vol. 74, pp. 247–267. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-44162-7_13
59. Kondratenko, Y., Korobko, V., Korobko, O., Kondratenko, G., Kozlov, O.: Green-IT approach to design and optimization of thermoacoustic waste heat utilization plant based on soft computing. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Components, Networks and Systems Implementation. Studies in Systems, Decision and Control, vol. 105, pp. 287–311. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-55595-9_14
60. Kuchuk, G., Kovalenko, A., Kharchenko, V., Shamraev, A.: Resource-oriented approaches to implementation of traffic control technologies in safety-critical I&C systems. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.), Green IT Engineering: Components, Networks and Systems Implementation. Studies in Systems, Decision and Control, vol. 105, pp. 313–337. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-55595-9_15
61. Atamanyuk, I., Kondratenko, Y.: Computer's analysis method and reliability assessment of fault-tolerance operation of information systems. In: Batsakis, S., et al. (eds.) ICT in Education, Research and Industrial Applications: Integration, Harmonization and Knowledge Transfer. Proceedings of the 11th International Conference ICTERI-2015, CEUR-WS, vol. 1356, pp. 507–522, Lviv, Ukraine, May 14–16 (2015)
62. Encheva, S., Kondratenko, Y., Solesvik, M.Z., Tumin, S.: Decision support systems in logistics. In: AIP Conference Proceedings, vol. 1060, no. 1, pp. 254–256. AIP (2008). https://doi.org/10.1063/1.3037065
63. Basile, D., Di Giandomenico, F., Gnesi, S.: Model-based evaluation of energy saving systems. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Concepts, Models, Complex Systems Architectures. Studies in Systems, Decision and Control, vol. 74, pp. 187–208. Springer, Cham (2017)
64. Palagin, A.V., Opanasenko, V.M., Kryvyi, S.I.: Resource and energy optimisation oriented development of FPGA-based adaptive logical networks for classification problem. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Components, Networks and Systems Implementation. Studies in Systems, Decision and Control, vol. 105, pp. 195–218. Springer, Cham (2017)
65. Drozd, J., Drozd, A., Antoshchuk, S.: Green IT engineering in the view of resource-based approach. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Concepts, Models, Complex Systems Architectures, vol. 74, pp. 43–65. Springer, Cham (2017)
66. Kondratenko, G., Kondratenko, Y., Sidenko, Ie.: Fuzzy Decision making system for model-oriented academia/industry cooperation: university preferences. In: Berger-Vachon, C., et al. (eds.) Complex Systems: Solutions and Challenges in Economics, Management and Engineering. Studies in Systems, Decision and Control, vol. 125, pp. 109–124. Springer International Publishing, Berlin, Heidelberg (2018). https://doi.org/10.1007/978-3-319-69989-9_7
67. Solesvik, M.: Partner selection in green innovation projects. In: Berger-Vachon, C., Gil Lafuente, A., Kacprzyk, J., Kondratenko, Y., Merigó, J., Morabito, C. (eds.) Complex Systems: Solutions and Challenges in Economics, Management and Engineering, vol. 125, pp. 471–480. Springer, Cham (2018)
68. Borch, O.J., Solesvik, M.: Collaborative design of advanced vessel technology for offshore operations in Arctic waters. In: Luo, Y. (ed.) International Conference on Cooperative Design, Visualization and Engineering, pp. 157–160. Springer, Berlin, Heidelberg (2013)

Computer Support for Decision-Making on Defining the Strategy of Green IT Development at the State Level

Igor Shostak, Igor Matyushenko, Yuri Romanenkov, Mariia Danova and Yuliia Kuznetsova

Abstract It is shown that in the conditions of the fourth industrial revolution and the formation of sixth technological order, Green IT technologies form one of the stable trend because they are classified as breakthrough technologies. The paper presents a complex estimation method of innovative potential of the national economy in order to determine the prospects of Green IT technology development. Methodological basis of the method is cluster analysis. As an example, the estimation of the innovation potential of Ukraine in the aspect of Green IT development is considered. Data analysis used for predicting the development of Green IT technologies on dynamic time series. As a predictive model is used Brown model and special structural parametric synthesis method that implements identification a time series by preliminary analysis and processing of initial data, identify the trend and formation the interval estimates. The results of forecasting are supposed to be used, along with other initial data, for foresight research on the definition of Green IT perspective directions development. A generalized structure of interactive decision support system for the development of Green IT at the state level is presented. In the future, it is supposed to create a generalized methodology of

I. Shostak · M. Danova (✉) · Y. Kuznetsova
Department of Engineering Software, National Aerospace University "KhAI", 17 Chkalova Street, Kharkiv, Ukraine
e-mail: danovamariya@gmail.com

I. Shostak
e-mail: iv.shostak@gmail.com

Y. Kuznetsova
e-mail: jk.25april@gmail.com

I. Matyushenko
Department of International Economic Relations, VN Karazin Kharkiv National University, 4 Svobody Square, Kharkiv, Ukraine
e-mail: imatyushenko@karazin.ua

Y. Romanenkov
Department of Management, National Aerospace University "KhAI", 17 Chkalova Street, Kharkiv, Ukraine
e-mail: KhAI.management@ukr.net

© Springer Nature Switzerland AG 2019

V. Kharchenko et al. (eds.), *Green IT Engineering: Social, Business and Industrial Applications*, Studies in Systems, Decision and Control 171,
https://doi.org/10.1007/978-3-030-00253-4_23

forecasting not only Green IT, but also other advanced technologies in countries with different levels of economic development.

Keywords Green IT technologies · Energy-saving green software
Green software engineering · Interactive decision support system
Foresight-project · Integration of forecast estimates · Interval data form
Brown model

1 Goal and Objectives of the Research

Today, when humanity faces a number of global problems, the activity of any state, which aims to improve the life quality of the population, should be aimed at solving these problems. One of such problems is the organization of efficient energy saving based on Green IT.

After the 2008–2009 crisis, most developed countries in the world consider convergent technologies as the main tool for solving global problems, based on the mutual influence and mutual penetration of the constituent parts of these technologies (first of all, nano-, bio-, informational and green), when the boundaries between individual technologies are erased, and the most significant results appear within the framework of interdisciplinary research. Green IT is a typical example of the convergence of these technologies. Since 2011 government policy of advanced countries aimed at developing key drivers of new industrial revolution converged technologies under the existing scientific and innovative potential. The above problem acute arises for modern Ukraine, taking into account its association with the EU.

The peculiarities of the modern, post-industrial society development and the prospects of a new industrial revolution are devoted to many works of famous scientists, in particular: C. Perez, J. Rifkin, Yu Yakovets, V. Heyts [1–4].

The problem of the convergent (in particular, Green IT) technologies development in Ukraine as a key factor in solving global problems in the context of a new industrial revolution needs further development. This is evidenced by the practice of determining the priorities of scientific and innovative activities in the countries of the world, and the available theoretical reserve of the above scientists. This led to the relevance of the topic of this study.

Thus, it is clear that at this stage, one of the sustainable development trends of humanity is the development and implementation of Green IT tools to solve energy saving problems. Green IT technologies are a typical result of the convergence of advanced technologies, especially information and knowledge. However, the definition of the Green IT development strategy is a fairly complex process, which is characterized by high levels of uncertainty. Based on this circumstance, it is appropriate to create a theoretical basis, and on its basis—special computer support tools for decision-making, through which the level of uncertainty in the process of formulating solutions for the development of Green IT will be reduced to an acceptable level.

The aim of the article is to substantiate the theoretical and methodological support for the implementation of the forecasting process of the Green IT development as one of the main trends in the conditions of the fourth industrial revolution and the formation of the sixth technological order.

To achieve this goal, the following tasks were formulated and solved in the research process:

1. To establish a link between the key factors of the development of Green IT, affecting the content of the new industrial revolution and the structure of the sixth technological order.
2. Identify areas of application of Green IT for the solution of global problems at the state level.
3. To assess the innovative potential as a prerequisite for the introduction of Green IT on the example of Ukraine.
4. To form scientific and methodical support of the forecasting processes of the Green IT development on the example of Ukraine.
5. To substantiate the methodical approach to carrying out foresight research on the selection of priority directions for Green IT development.
6. To propose a generalized structure of an interactive decision support system for defining Green IT development strategies at the state level.

2 The Influence of Green IT on the Content of the New Industrial Revolution and the Structure of the Sixth Technological Order

An analysis of the current global trends of economic development shows that for a number of industrialized countries, the leading areas of research were [5–7]: from the 50s and 60s of the XX century—information and communication technologies (ICTs); from the 70s and 80s—biotechnology; at the beginning of the XXI century—nanotechnology, and at the end of the first decade of the XXI century—cognotechnology (or technology of artificial intelligence) [9–11]. For today, one of the dominant areas of research is Green IT as an effective way to save energy. Figure 1 shows how the solution of global problems depends on the Green IT development.

Convergence is not only a mutual influence but also a mutual penetration of technologies, when the boundaries between individual technologies are erased, and the most interesting and unexpected results appear precisely within the framework of interdisciplinary work at the intersection of sciences.

With the development of the Green IT convergence for the first time in the history of mankind there is a parallel acceleration development of several scientific and technical areas that directly affect society. Accordingly, the Green IT development will lead to a leap in the capabilities of productive forces and to merge scientific and technological trends into a single scientific and technological knowledge industry [2–4, 8, 10, 11].

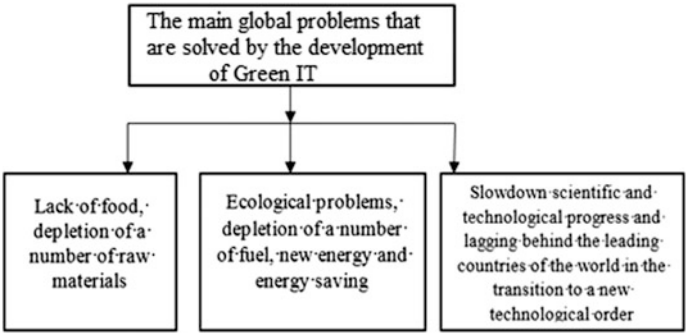


Fig. 1 Main global problems that are solved, in particular, through the Green IT development

Taking into account the interdisciplinary and objective nature of the convergence of scientific research and practical applications in the Green IT field, which allow qualitatively new opportunities for the main spheres of public life.

Table 1 presents the structure of the sixth technological order taking into account the Green IT development [12–15].

Since 2011 experts from the US, EU, Japan, China and Latin America in their regular meetings on the convergent technologies development discussed in detail possible ways of applying the convergence of these technologies that can improve human life in the next 10–20 years [9, 14], especially in a new industrial revolution [3, 4, 16]. They came to the conclusion that Green IT are one of the most influential factors of such convergence.

In Table 2 provides a generalized comparison of problems and opportunities that provides the Green IT development, namely the opportunities that can be realized in the next 20 years and the opportunities that can be realized in the distant future [9, 11, 13, 17].

3 Application of Green IT to Solve Global Problems at the State Level

Considering the interdisciplinary and objective nature of the convergence of scientific research and practical applications in the field of Green IT and the rapid development of a complex of advanced production technologies that provide qualitatively new opportunities for all spheres of public life, it is advisable to consider the conceptual foundations of the convergent technologies development in Ukraine as a key factor for solving global problems in a new industrial revolution

In the works of many scientists, including V. Heits [2], V. Khaustova [18] and others [16, 18, 19], whose engaged in nonlinear nature of the modern economy problems, has been proven the objective feasibility of using synergetic paradigm in

Table 1 The main characteristics of the sixth technological order

Long waves/cycles			State of education and science	Infrastructure		Preferred technology	Universal resource
Time frame	Leading countries	Characteristics of cycle		Transport and communications	Energy		
Sixth technological order 2010–2070 years (forecast)	USA, Japan, EU, China, South-East Asia, India, Brazil, Russia	Rapid development of cognitive science; information revolution deepening	Convergence of advanced technologies, network research and innovation systems	Integrated information systems and telecommunications, mobile Internet, broadband access	Renewable and thermonuclear energy	Green software, 3D-printing, Internet of things, artificial intelligence,	Green computing, green communication, green IT, nanoelectromechanical systems, bioprocessors

Table 2 Problems and opportunities that provide the Green IT development

Global problem	Possibilities that can be realized in the next 20 years	The opportunities that can be realized in a more distant future
Lack of food, depletion of a number of raw materials and fuel	–	Achieving a global material prosperity based on developed of Green IT
New energy and energy saving; ecological problems	Introduction of energy-saving technologies	Effective management of climate change and processes in the biosphere, global restoration of natural ecosystems
Slowdown scientific and technological progress	Expansion of intelligent human capabilities through the use of implanted or worn, sensor devices, computers, additional memory, communication devices	Introduction of noocomputing technologies

the study of revolutionary changes in scientific and technological development of any country [11, 18–20].

The analysis tendencies of the technological development in Ukraine on the basis of study the convergent technologies development processes (as a key factor of the new industrial revolution) has enabled, from the position of the synergetic paradigm, to offer the concept of the convergent technologies development, in particular Green IT, in Ukraine to solve global problems in the new industrial revolution. This concept is based on the following hypotheses:

Hypothesis 1. Use of Green IT, allows us to solve the main global problems of humanity and is a key factor in the new industrial revolution.
Hypothesis 2. It is expedient to implement Green IT only under the conditions and availability of a certain innovative potential of Ukraine.
Hypothesis 3. To manage the development of Ukraine's scientific, technical and innovation potential, it is necessary to forecast the consequences of Green IT implementation.
Hypothesis 4. The determination of priorities for scientific and technological development of Ukraine in connection with the complexity of the input information requires automation.

In the conditions of the new industrial revolution in modern terms, it is advisable to assume that technology Industry 4.0, combining factors Smart EMP (E (environment)—intelligent environment; M (manufacturing)—intelligent manufacturing; P (products)—smart products), create new markets and industries, promote the growth of productivity and competitiveness of certain sectors and national economies.

At the same time, the contours of the new technological system (TS) have already begun to be formed on the basis of the first convergent technologies and their explosive proliferation in 2010–2015. The fundamental abandonment of this TS from the previous ones is that human consciousness will become the same productive force as science has at one time become. The material basis and tool for implementing the convergent technologies of the new TS will be Industry 4.0 or Smart EMP.

To substantiate or refute the above hypotheses, it is advisable to propose a structural and logical scheme of scientific research into the development and implementation of Green IT in Ukraine as a key factor in solving global problems in the context of the new industrial revolution and association with the EU, as shown in Fig. 2.

In accordance with Fig. 3 the development and implementation of Green IT in Ukraine as a key factor in solving global problems in the new industrial revolution is carried out in five stages [21].

At the first stage, a comparative analysis of innovation development of Ukraine and EU countries is conducted on the basis of the: assessment of the creation conditions (CCIP) and the level of innovation potential (LVIP) of the countries with the definition of the integral index I_{CCL}; assessment of the implementation conditions (CRIP) and the results of implementation (RZIP) of the countries innovative potential with the definition of the integral index I_{CI}, as well as the comparison of the components of the specified integral indicators. At the same time, Ukraine is the main country under study, and calculations are carried out in comparison with other EU countries.

In the second stage, foresight research is carried out to determine the most promising areas of Green IT development in Ukraine, provided that they are further automated, taking into account formal methods of selecting experts, forming an initial list of thematic areas based on bibliometrics, science and patent analysis, and

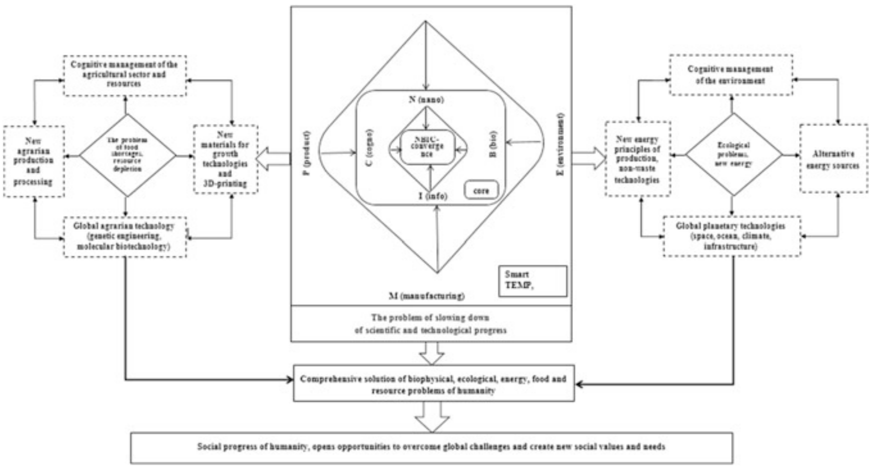


Fig. 2 The directions of solving global problems based on the introduction of Green IT as the core of the Smart EMP advanced production system development

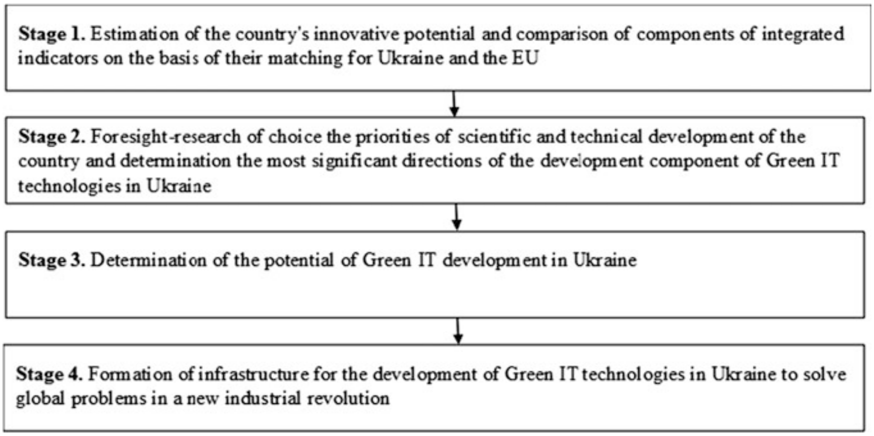


Fig. 3 The process of green IT developing at the state level (in Ukraine) to solve global problems in a new industrial revolution

evaluating and clarifying the thematic areas from using Pareto-optimality principles and t-ordering method. The scientific and technological priorities of the components of Green IT technologies are determined.

The third stage identifies the potential of Green IT technology development in Ukraine.

At the fourth stage, recommendations are being developed on the infrastructure supporting the scientific and technical development and implementation of Green IT in Ukraine.

4 Comparative Analysis of Ukraine Innovative Potential for Introduction of Green IT

In the context of increasing globalization processes and the rapid development of scientific and technological progress in the world, the development of technology innovation will contribute to direct and indirect solutions to the problems of social, economic and environmental situations in countries at all levels—from individual to global. Therefore, the paper proposes to conduct a comparative analysis of innovation development of Ukraine and the EU countries based on the indicators of the annual rating ōe Global Innovation Index of Cornell University, the French Business School and Research Institute (INSEAD), the World Intellectual Property Organization (WIPO) [22] and official EU statistics [23].

The aim of the study were the definition of creation conditions and the level of countries innovative potential based on the level of educational and institutional capacities development of the countries and its comparison with the conditions and results of implementation—the potential of R&D and commercial realization of the potential of countries.

An assessment of the creation conditions and the level of innovation potential of the countries under study is carried out through an analysis of the indicators that characterizing the financial and technological preconditions for obtaining education, proper institutional management, protection and stimulation of investments, the level and quality of the received education, the level of specialists qualifications, quality and ability to innovate. In Fig. 4 shows a structural scheme for assessing the creation conditions and the level of innovation potential of countries [21].

The assessment of the conditions and implementation of the innovative potential of the countries under study is carried out through the analysis of indicators that characterizing financial, technological and human resources for the implementation of the created innovative potential of the countries. The results of the created potential implementation are estimated through the level and quality of the patent activity of the country's agents and its revenues, the introduction of innovations into commercial activity and its results. In Fig. 5 shows a structural scheme for assessing the conditions and implementation of countries innovation potential [21].

To calculate the integral indexes an additive method of convolution the normalized partial indicators based on the simple average which used for the characterization of the hierarchical series of components is applied. Standardization of partial indicators was carried out according to the formula:

$$z_{ij} = \frac{x_{ij}}{x_{\max j}}, \quad (1)$$

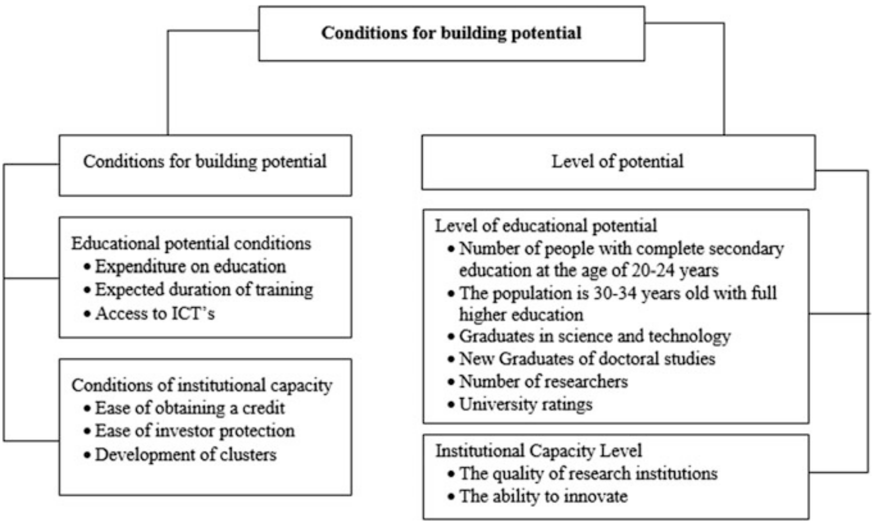


Fig. 4 Structural scheme for assessing the creation conditions and the level of innovation potential of countries

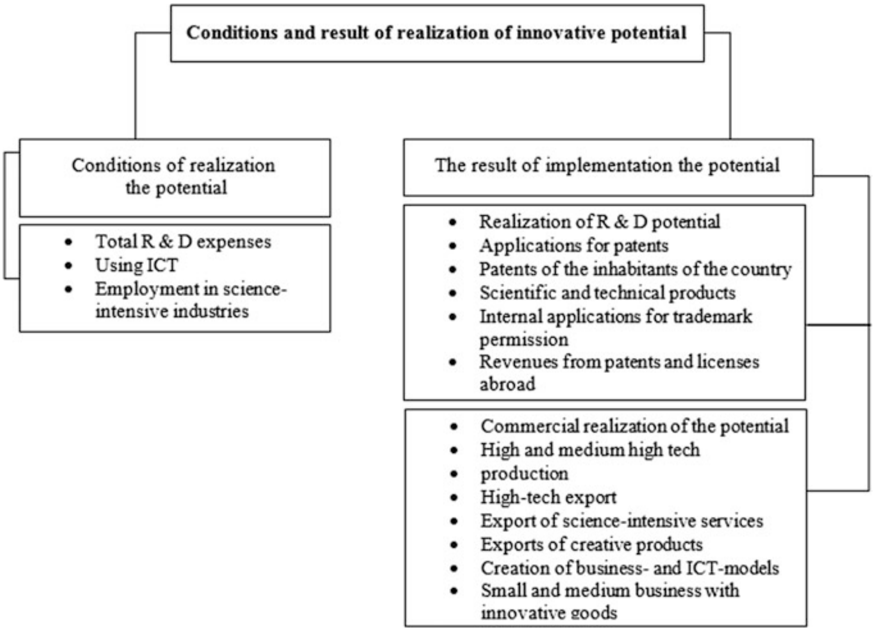


Fig. 5 Structural scheme of estimation the conditions and realization of innovative potential of countries

where

z_{ij} normalized j-th partial index of the i-th country;
x_{ij} value of j-th partial index of i-th country;
$x_{max\ j}$ is the maximum value of the j-th partial indicator.

The results of calculating the integral indicator of the creation conditions and the level of innovation potential of the countries in 2014 for the above method is shown in Fig. 6.

As can be seen from Fig. 6, among the analyzed countries, the leaders in terms of I_{CCL} were the United Kingdom, Sweden and Denmark. Bulgaria, Malta and Croatia were among the countries with the lowest levels of I_{CCL}. Ukraine, in terms of the integral indicator of the creation conditions and the level of innovation potential, occupied 19th place among the studied countries, ahead of Cyprus and yielding to Poland.

The clustering (Fig. 7) of the studied countries, depending on the components of the integral indicators of creation conditions (C_{CCIP}) and the level (C_{LVIP}) and conditions (C_{CRIP}) and realization (C_{RZIP}) of the innovation potential, makes it quite unambiguous to classify them: the first cluster (B)—a country with powerful conditions for creation, level and degree of implementation of innovative potential; second cluster (C)—a country with average conditions of creation, level and degree of implementation of innovative potential; third cluster (H)—with low conditions of creation, level and degree of implementation of innovation potential.

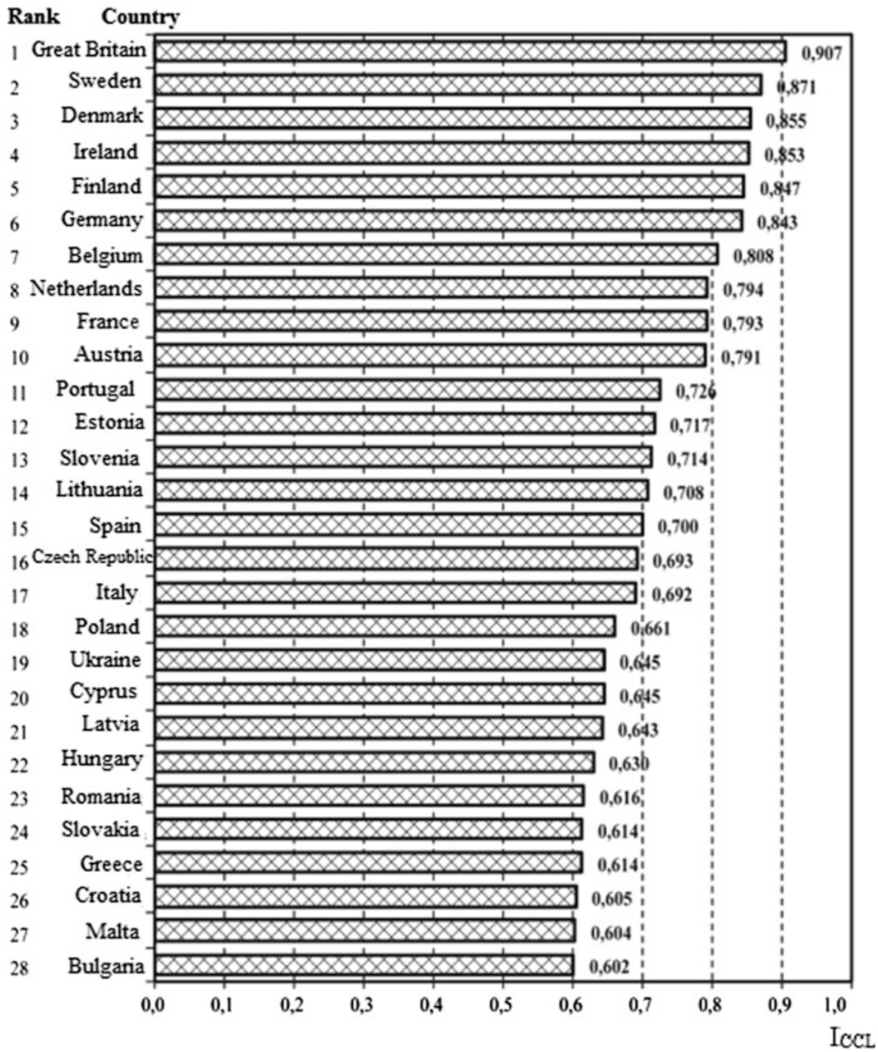


Fig. 6 The value of the integral indicator of the creation conditions and the level of innovative potential of countries in 2014

The classification of the studied countries, depending on the components of the integral indicators of the creation conditions and the level and conditions and the implementation of innovation potential, is given in the Table 3.

As can be seen from Table 3, the cluster structure of the studied countries by the integral indicators components level of the creation conditions of and the level and conditions and the implementation of the innovation potential was proportional to the structure: 10 countries (35.71%) belonged to the first cluster, 8 countries

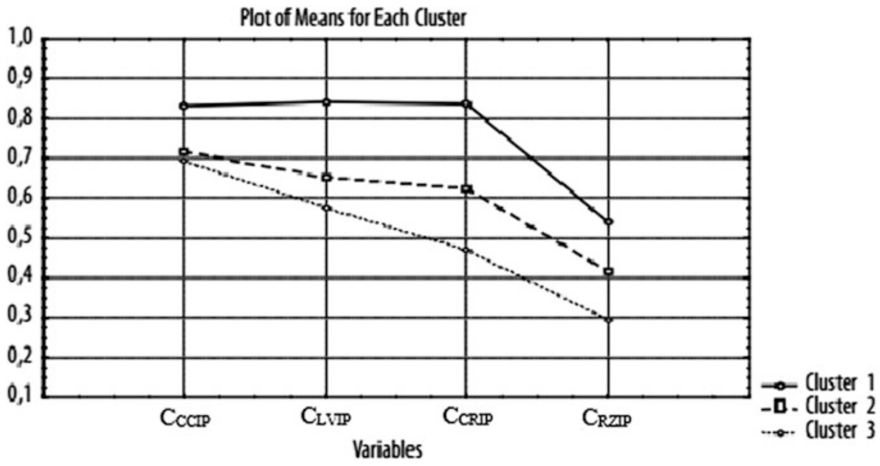


Fig. 7 Graph of characteristics the components integrated indicators average values of the surveyed countries in 2014

Table 3 Classification of countries by the level of components of the integral indicators of conditions the creation and the level and conditions and the implementation of the innovation potential

I cluster (B)	II cluster (C)	III cluster (H)
Austria	Hungary	Bulgaria
Belgium	Spain	Greece
Great Britain	Italy	Cyprus
Germany	Malta	Latvia
Denmark	Portugal	Lithuania
Ireland	Slovenia	Poland
Netherlands	Czech Republic	Romania
Finland	Estonia	Slovakia
France		Croatia
Sweden		Ukraine

(28.57%) belonged to the second cluster, and 10 countries (35.71%), including Ukraine, belonged to the third cluster [21].

The structural analysis of the components and the integral indicators of the creation conditions and the level and conditions and the implementation of the countries innovative potential is showed the expediency of further consideration of this economic phenomenon in the plane "creation conditions and level of innovation potential—conditions and implementation of innovation potential", which provides an opportunity for more objective analysis this phenomenon and development of strategic directions of development of countries in the innovation sphere.

In Table 4 shows the ratio of components values, elements and partial indicators of the integral index of the creation conditions and the level of innovation potential of Ukraine to the corresponding average values of the countries under study and to the average values of the cluster in which the country is located.

Table 4 Level of components, elements and partial indicators of the integral indicator of the creation conditions and the level of innovation potential of Ukraine

Indicator	Relation to the general average		Relation to the average cluster	
	Value	Rank	Value	Rank
Creation conditions and the level of innovation potential	0.895		1.015	
Creation conditions of innovative potential	0.902	1	0.973	2
Conditions for creating educational potential	0.914	2	1.013	2
Education costs	1.097	5	1.231	4
Expected duration of training	0.951	7	1.016	8
Access to ICTs	0.737	11	0.841	12
Creation conditions of institutional capacity	0.889	3	0.932	4
The ease of obtaining credit	1.209	2	1.093	6
The ease of investor protection	0.758	9	0.812	13
The development of clusters	0.699	12	0.846	11
Level of innovative capacity	0.887	2	1.067	1
Level of educational potential	1.032	1	1.207	1
Number of people with complete secondary education aged 20–24	1.147	4	1.085	7
Population aged 30–34 with complete higher education	1.083	6	1.126	5
Graduates in the field of science and technology	1.179	3	1.252	3
New graduates of the doctorate, the number of graduates 25–34 years old	1.438	1	1.732	2
Number of researchers	0.322	14	0.518	14
University Ratings	0.602	13	1.745	1
Level of institutional capacity	0.763	4	0.941	3
Quality of research institutions	0.756	10	0.928	10
The ability to innovations	0.771	8	0.955	9

As can be seen from Table 4, the positive impact on the level of the integral indicator of the creation conditions and the level of innovation potential in Ukraine in relation to the studied countries and the cluster of belonging created the following partial indicators: “New graduates of the doctorate”, “Graduates in science and technology”, “Education expenses”, “Number the population with complete secondary education aged 20–24” and “Population aged 30–34 with complete higher education”. To the partial indicators of this integral indicator, which least contributed to its growth include: “Number of researchers”, “Development of clusters”, “Access to ICT”, “Quality of research institutions” and “Ease of protection of investors” [21].

In Table 5 shows the ratio of the components values, elements and partial indicators of the integral indicator of conditions and the implementation of the innovative potential of Ukraine to the corresponding average values of the countries under study and to the average values of the cluster in which the country is located.

From Table 5 shows that the positive effects on the level of the integral indicator of conditions and the implementation of innovation potential in Ukraine in relation to the countries under study and the access cluster created the following partial indicators: "Internal applications for trade mark resolution", "Export of science-intensive services" and "Employment in science-intensive industries". To the partial indicators of this integral indicator, which least contributed to its growth, are: "Small and Medium Business with Innovative Goods", "Applications for patents", "Patents of the inhabitants of the country", "Scientific and technical products" and "High-tech exports".

In Table 6, on the results of the study, a comparative matrix of the creation conditions, the level and conditions of implementation of Ukraine's innovative potential was made.

The quadrants of the matrix are constructed according to the following principle: the quadrant "High (CCL)—High (CL)" is a partial indicator whose value exceeds the average values of the studied countries and the cluster to which Ukraine belongs; Quadrant "Low (CCL)—Low (CL)"—the value of partial indicators does not exceed the average values of the studied countries and the corresponding cluster; "Low (CL)—High (CCL)"—the value of partial indicators does not exceed

Table 5 The level of components, elements and partial indicators of the integral indicator of conditions and the implementation of Ukraine's innovation potential

Indicator	Relation to the general average		Relation to the average cluster	
	Value	Rank	Value	Rank
	0.574		0.796	
Conditions of the potential realization	0.586	1	0.802	1
Total costs of R&D	0.450	6	1.036	4
Using ICTs	0.321	9	0.419	10
Employment in science-intensive industries	0.884	3	1.011	5
Realization of innovative potential	0.557	2	0.786	2
Realization of R&D potential	0.600	1	0.913	1
Patent applications	0.058	13	0.419	12
Patents of the inhabitants of the country	0.145	10	0.636	9
Scientific and technical products	0.376	7	0.512	10
Internal applications for trademark permission	1.406	1	1.285	1
Revenues from patents and licenses abroad	0.101	12	1.220	2
Commercial realization of innovations	0.533	2	0.724	2
Exports of high-tech services	1.001	2	1133	3
High- and medium- high-tech production	0.584	5	0.861	7
High-tech export	0.366	8	0.717	8
Creating business- and ICT-models	0.753	4	0.865	6
Small and medium business with innovative goods	0.049	14	0.082	14
Exports of creative products	0.256	10	0.279	13

Table 6 Comparative matrix of the creation conditions and the level and conditions and implementation of innovative potential of Ukraine

		Creation conditions and the level of innovation potential	
		Low (CCL)	High (CCL)
Conditions and realization of innovative potential	High (CL)		*Prerequisites for creation* Education costs; the ease of obtaining credit *Level* New graduates of doctorate; graduates in science and technology; number of people with complete secondary education at the age of 20–24; population aged 30–34 with complete higher education *Realization* Internal applications for trademark permission; exports of knowledge-intensive services
	Low (CL)	*Level* The ability to innovate; quality of research institutions; access to ICTs; The ease investor protection; The development of clusters; number of researchers *Prerequisites for implementation* Using ICTs *Realization* Creation of business- and ICT-models; high- and medium-high-tech production; high-tech export; Scientific and technical products; Patents of the inhabitants of the country; exports of creative products; patent applications; small and medium business with innovative	*Prerequisites for creation* Expected duration of training *Level* University ratings *Prerequisites for implementation* Employment in science-intensive industries; total R&D expenses *Realization* Revenues from patents and licenses abroad

the average values of the studied countries and exceed the average values of the cluster to which Ukraine belongs [21].

As can be seen from Table 6 Ukraine has some relative advantages of innovative potential that characterize the financial provision of educational potential, the level and quality of education of the population, the domestic patent activity of the population and the export of knowledge-intensive services.

Even less potential relative advantages of the innovative potential of Ukraine are the characteristics of the duration of studies, the quality of university education, financing and employment in the field of R&D and revenues from inventions.

The greatest part of the characteristics of the innovative potential of Ukraine are relative disadvantages of the number of researchers and the quality of research institutions; access, use and practical implementation of ICT; the general patent activity of the population; the development of innovative clusters; the creation of scientific and technological productions and products; the export of high-tech and creative production; production of innovative products by small and medium enterprises; legal protection.

Thus, an objective comparative analysis of the characteristics of the innovative potential of Ukraine has shown a relatively low level of its creation and implementation, revealing its relative advantages and disadvantages. The results and conclusions obtained during the complex comparative analysis are an objective characteristic of the state of innovation potential of Ukraine and they will be used as input data for determining the trends of Green IT development in Ukraine.

5 Scientific and Methodological Support of Forecasting the Green IT Development in Ukraine

To ensure the proper level of efficiency of forecasting processes, the article describes the information technology of multilevel adaptive assessment of the current level and prospects for the development of Green IT.

The features of this technology are as follows:

– the application of the parametric identification method of Brown model based on the results of a retrospective analysis of the Green IT development in Ukraine, which, unlike the known ones, is based on the analysis of the configuration of retrospective equations roots, which allows analytically to determine the parameters of the model setting that ensure the accuracy of predictive estimates;
– implementation of the structural-parametric synthesis method, which takes into account the particularities of the type of source data used, which allows to ensure the adequacy of the predictive model of the Green IT development in Ukraine.

The above-described results of the innovative potential assessment at the state level provide an opportunity, based on forecast extrapolation, to construct a trendy long-term macro forecast [24] for the development of Green IT. At the same time, the extrapolation of trends does not allow for the possibility of taking into account random fluctuations, which makes it expedient to apply adaptive prognostic methods based on exponential smoothing. At the same time, it is possible to take into account the development of the trend existing at the moment, rather than the tendencies that have developed throughout the period under review [24].

It has been shown in [24] that for predicting of Green IT development, the method of two-parameter exponential smoothing (Hort's method) is the most suitable method [24]. A private case of the Hort method is the Brown method using the same name model.

For the implementation of a parametric analysis of Brown's predictive model, or exponential smoothing, is used the exponential average value of the stationary time series to evaluate its value at the next time point:

$$\hat{y}_t = \alpha y_{t-1} + \alpha(1-\alpha)y_{t-2} + \cdots + \alpha(1-\alpha)^{n-1}y_{t-n} = \sum_{i=1}^{n} \alpha(1-\alpha)^{i-1}y_{t-i},$$

where $y_{t-1}, y_{t-2}, \ldots, y_{t-n}$—the value of the series at the corresponding moments of time; n—sample length of the time series; α—smoothing parameter.

The quality of the retrospective forecast estimates obtained for the time series values in the past relative to the t points of time have tends to be preserved at a later point in time. For the a posteriori solution of the parametric synthesis problem, it is necessary to form retrospective equations of this kind:

$$\varepsilon_{t-1}(\alpha) = 100 \cdot \frac{\hat{y}_{t-1}(\alpha) - y_{t-1}}{y_{t-1}} = \frac{100}{y_{t-1}} \cdot \left(\sum_{i=1}^{n-1} \alpha(1-\alpha)^{i-1} y_{t-i-1} - y_{t-1} \right) = 0. \quad (2)$$

Process diagram of the method of Brown model parametric identification of Green IT development is shown in Fig. 8. This method can analyze the time series of two types: interval and numeric.

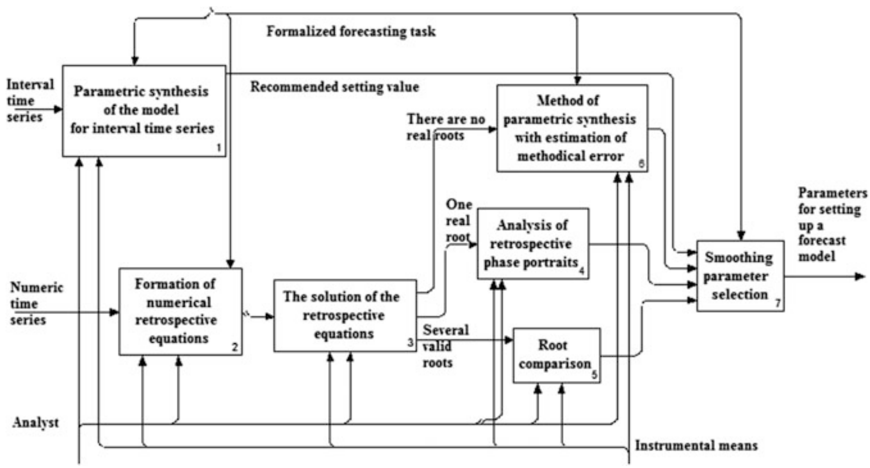


Fig. 8 Process diagram of the method of Brown's predictive model parametric identification for the Green IT development

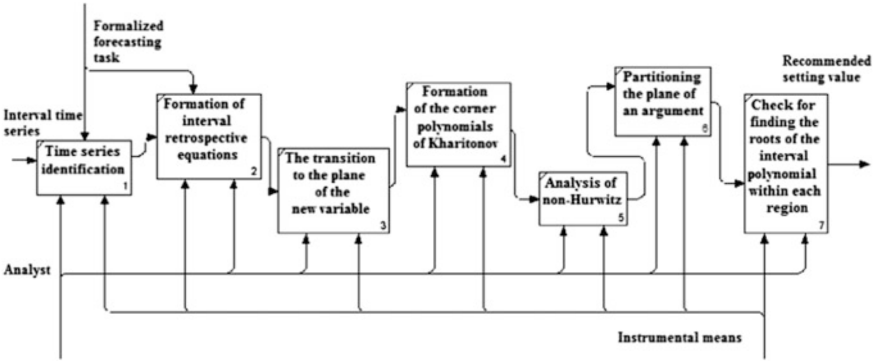


Fig. 9 Process diagram of the method of structural-parametric synthesis of the predictive Brown model for interval time series of innovation potential integral indicators

The method of structural-parametric synthesis (Fig. 9) contains the identification of the time series, which includes a preliminary analysis and processing of output data, namely the allocation of the trend and the formation a number of interval residues.

On the basis of identification, one or more interval retrospective equations should be formed. In order to establish the fact of finding the roots of the interval polynomial in the range $0 \leq \alpha \leq 2$, it is proposed to use a bilinear transform of the form

$$\alpha = \frac{2}{1 + w}, \tag{3}$$

which represents a single circle with center at the point (0; 1) in the plane of the parameter α in the right half-plane of the complex plane.

Thus, the article describes the method of parametric identification of the Brown model for creating a forecast for the development of Green IT technologies in Ukraine based on the results of a retrospective analysis, which, unlike the known, is based on an analysis of the configuration of the roots of the retrospective equations. [24].

Let us illustrate the application of the above method on the example of a retrospective analysis of the indices which are shown in Table 4 at that, the sample from 15 through 25 of the row will be considered.

Create the retrospective Eq. (4):

$$\begin{aligned} \varepsilon_{t-1}(\alpha) = {} & 101.3596193\alpha^{11} - 1117.063222\alpha^{10} \\ & + 5597.620666\alpha^{9} - 16837.11761\alpha^{8} + 33776.75051\alpha^{7} \\ & - 47445.75119\alpha^{6} + 47615.63562\alpha^{5} - 34148.1985\alpha^{4} \\ & + 17167.98097\alpha^{3} - 5777.702243\alpha^{2} + 1169.00068\alpha - 100 = 0. \end{aligned} \tag{4}$$

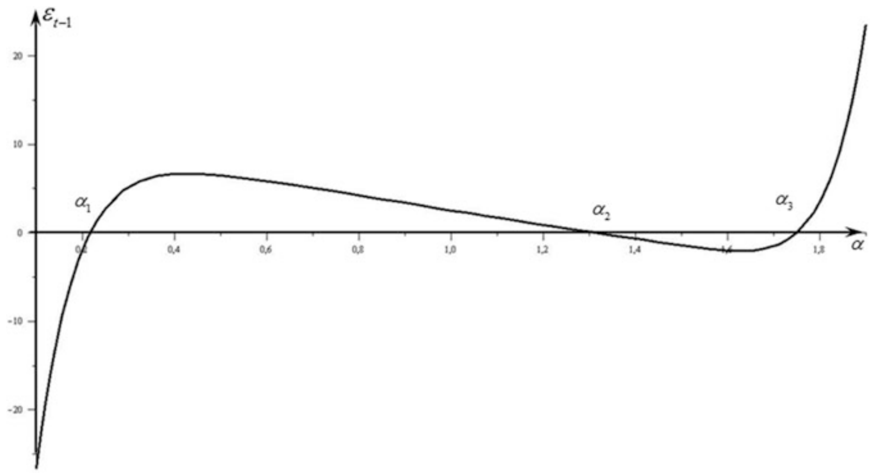


Fig. 10 Retrospective real roots of Eq. (4)

Real roots of Eq. (4) centered on the extended set of permissible parameter smoothing α (Fig. 10):

$$\alpha_1 = 0.2174;\ \alpha_2 = 1.3126;\ \alpha_3 = 1.7507. \qquad (5)$$

Specifies the values $\varepsilon^* = 0.5\%;\ \beta^* = 10\%$..

Determine the sensitivity of predictive estimates:

$$s_1 = \left| \left. \frac{d\varepsilon_{t-1}^{(2)}(\varepsilon_\alpha)}{d\varepsilon_\alpha} \right|_{\varepsilon_\alpha = 0} \right| = |tg\,\varphi_1| = 0.2245,\ s_2 = -0.0533,\ s_3 = 0.1144. \qquad (6)$$

Graphical analysis of sensitivity and robustness of predictive assessments is shown in Fig. 11.

The entire set of indicators and the relative errors of the current forecast estimates at the 27th time point when $\alpha = \alpha_j$ for a sliding sample of 11 elements $\varepsilon_j^{(11)}$ and for a cumulative sample of 12 elements $\varepsilon_j^{(12)}$ are given in Table 7.

As can be seen from Table 7, the choice of the smoothing parameter $\alpha = 1.3126$ provides not only the robustness of the retrospective forecast estimate to change α, but also the robustness of the current forecast estimate to the change in sample length n.

Using the method of parametric identification of the Brown model for creating a forecast for the Green IT development in Ukraine the following directions were assessed: green software engineering, software ecosystem, energy-saving green software and green telecommunications. It was found that the smallest prospects, in the future, has a direction of software ecosystem, and the largest—green software engineering.

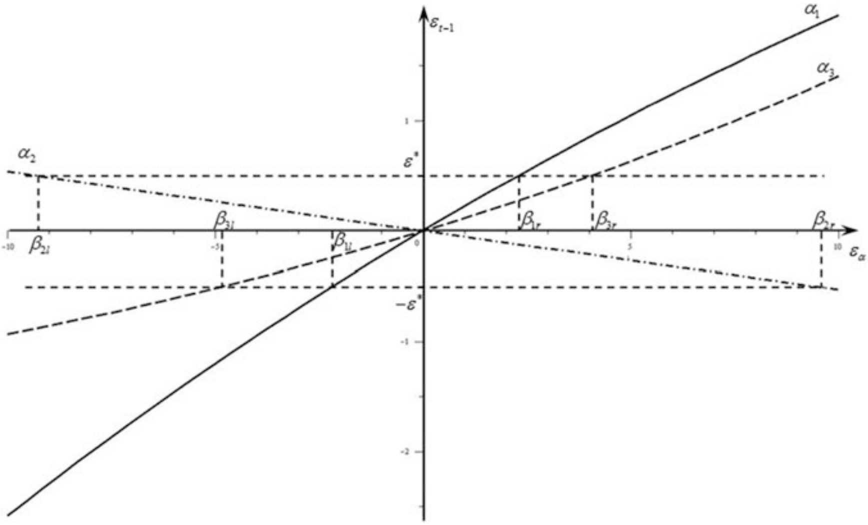


Fig. 11 Graphical analysis of sensitivity and robustness of predictive assessments

Table 7 The results of the time series retrospective analysis and analysis of the sensitivity and robustness of the forecast estimates

Elements series	15–26	15–26	15–26
j	1	2	3
α_j	0.2174	1.3126	1.7507
S_j	0.2245	−0.0533	0.1144
ε^* (%)	0.5		
β_{jl}, %	−2.1704	−9.3212	−4.7925
β_{jr} (%)	2.2886	9.4134	4.0485
$\Delta\beta_j$ (%)	4.459	18.7346	8.841
β^* (%)	10		
$r_j^{(\beta^*)}$	0.0443	0.1874	0.0865
$\varepsilon_j^{(11)}$ (%)	−2.3664	0.7907	7.2943
$\varepsilon_j^{(12)}$ (%)	−0.8478	0.7903	−0.4435

6 Methodical Approach to Conducting Foresight-Research on the Green IT Development

The results described above have provided the opportunity to generate initial data for the foresight research on Green IT development. The technology of carrying out the foresight-research consists of the following steps [25–27]:

1. Formation of expert panels, i.e., number and structure of expert members participating in forecasting based on the evaluation of their competence level.

2. Forming the initial list of the Green IT directions. It is necessary to analyze the status and prospects of Green IT development using methods of bibliometrics (calculating the number of publications), scientometrics (analysis of citing) and patent analysis (trace analysis of the dynamics of inventive activity). Then, for the obtained lists of trends of the Green IT development, it is calculated value of criteria for their assessment provided by technique of foresight-researches. Thus, all "leading" directions have quantitative assessments for each of the criteria, which will help to determine the number of priority.
3. Selection of priority directions for Green IT development. Initial data for the selecting of priorities is a list of directions of Green IT, as well as a set of values of criteria assessment for each direction. The procedure for the selection of priority directions of Green IT development is in ranking these directions criteria using the Pareto optimality principle and t-ordering method.
4. Coordination and approval of Green IT priority directions. In accordance with the current technique the strictly regulated procedure for coordination and approval of priorities is carried out.

The implementation this foresight procedure used the received estimates of the innovative potential level as quantitative criteria at the stage 2—forming the initial list of the Green IT directions. The results obtained by the method of parametric identification of the Brown model are used to support the decision making by experts when assessing the directions of Green IT development of at the stage 2—forming the initial list of the Green IT directions.

The authors previously conducted a foresight research on the Green IT development in Ukraine. It is described in [25]. The use of the obtained results of this research allowed to reduce the number of research directions on the 2nd stage, excluding software ecosystem from it, which reduced uncertainty in the decision making by experts, but did not affect its outcome. The most priority directions for the development of Green IT in Ukraine are still energy-saving green software and green software engineering.

7 Generalized Structure of an Interactive Decision Support System for Defining of Green IT Development

To automate the process of determining the Green IT perspective directions, a specialized tool environment was developed, that will formulate recommendations on perspective directions of the Green IT development for specialists who create state programs for the innovation policy of Ukraine. The system is implemented in the form of a two-tier structure consisting of the following components: database server (DB), which stores data and allows to scale the system in the event of increased load on it (may use the database management system (DBMS) Oracle, Microsoft SQL); an application server with an embedded Web server that implements data logic and also a flexible, scalable interface. To display the system interface, the client's user location must be equipped with a web browser (Internet Explorer, Firefox, Opera, etc.). The architecture of the system is shown in Fig. 12.

The input subsystem is intended for input of the necessary data, such as personal and professional data on experts, the values of indicators required at a particular stage of foresight research, etc. [26], as well as the mode of operation of the interactive decision support system (IDSS).

As operating modes are considered two modes: stand-alone mode and mode using the database. When working offline, the initial data can be entered manually or downloaded from the file.

Data import into IDSS is possible from three sources:

- external data sources (downloading data from a file);
- manual data entry (users: experts of all levels, decision makers, administrator);
- internal data sources (internal database).

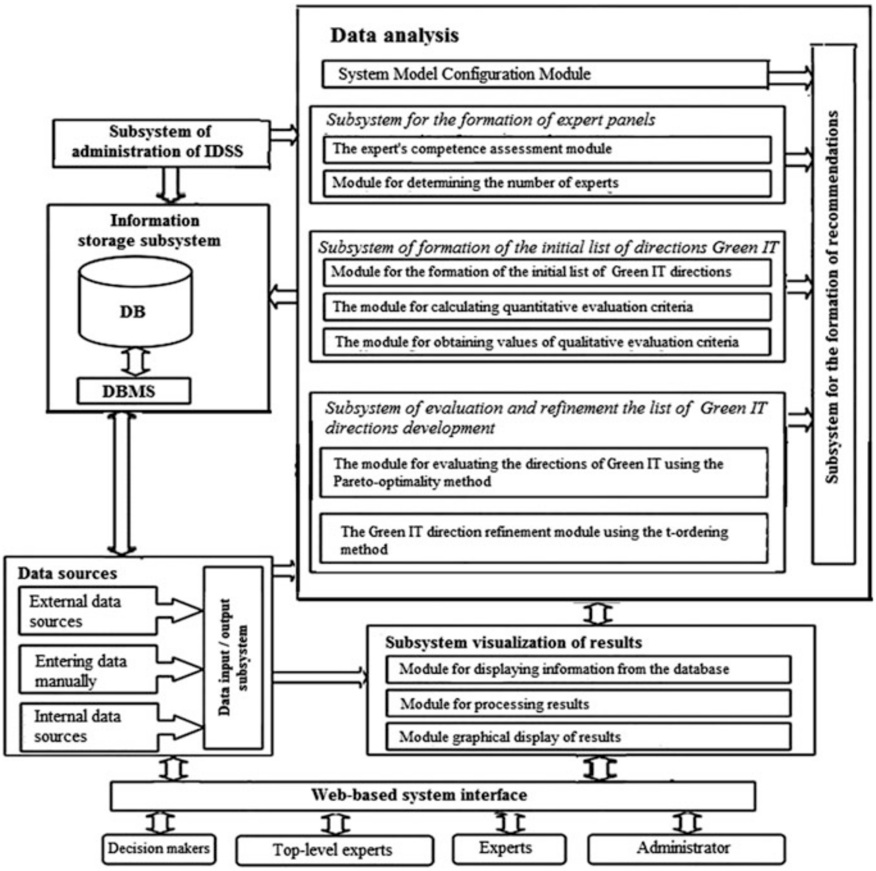


Fig. 12 The architecture of the interactive decision support system for defining of Green IT development strategies at the state level

Data can include:

– statistical information for updating the database;
– information that necessary for conducting an expert examination (lists of experts, criteria, Green IT directions, etc.);
– service information (for the administration of the IDSS) containing information about users, their access rights, etc.;
– other data necessary for the foresight research.

Output data can be presented in the form of lists, data samples, rankings.

The subsystem of information storage contains a relational DBMS, which implements access to DB data using the SQL query language. The proposed IDSS contains a DB of foresight research and has the ability to add other DB, if necessary.

Foresight research DB contains information of:

– experts (number of expertise, date of examination, lists of experts, assessment (advantages) of experts);
– directions of Green IT and criteria for their evaluation (lists of Green IT directions and criteria for their evaluation, the value of evaluation criteria for each Green IT directions, obtained during foresight research, ordinary information of the decision makers on the relative importance of the criteria);
– information necessary for conducting the foresight research (available expert lists, Green IT directions, criteria, statistical data required for calculations).

Module for setting up the system model is designed to select the configuration of foresight research by selecting the type and purpose of foresight research [26, 27]. On the basis of this choice it is given the sequence of modules execution and their set from the system "Data Analysis".

The subsystem for the formation of expert panels includes two modules. The expert's competence assessment module allows, with the help of calculation of the generalized indicator of the level of competence of each expert, to build a ranked list of participants in the examination. Based on the required number of experts for participation in foresight research, obtained from the module "Determining the number of experts", formed the final list of experts involved in the examination with the preservation of these data in the DB foresight research.

The subsystem of formation the initial list of Green IT directions contains the modules for forming the initial list of Green IT directions and obtaining the values of the criteria for their evaluation. In the implementation of these modules an initial list of Green IT directions in the form of vector functions is formed.

The module "Formation of the initial list of Green IT directions" contains functions which, based on the initial statistical data, using the methods of counting the number of publications, the analysis of the citation and the analysis of the curves of the dynamics of inventory activity, can form a preliminary list of Green IT directions.

Modules for obtaining values of qualitative and quantitative criteria for assessing Green IT directions contain a number of functions that allow, according to the received list of Green IT directions, to calculate the value of each criteria, using the necessary statistical and forecast data. In addition, the module "Obtaining values of qualitative criteria" implemented the functions of calculating the aggregated assessment of Green IT directions by experts on each of the criteria, as well as checking the consistency of expert opinions.

The subsystem of evaluation and refinement the list of Green IT directions is intended for the selection of priority directions of Green IT development by means of their estimation using the Pareto-optimality principle and the t-ordering method implemented in the corresponding modules of the subsystem.

In the module of evaluation of Green IT directions with using Pareto-optimality principle, is comparing of Green IT directions, in the form of received vector functions, results in the formation of a Pareto set of priority Green IT directions. The module of evaluation of Green IT directions with using the t-ordering method begins to work only if it is impossible to compare the obtained vector functions on the Pareto principle. It uses the basic information of the decision makers on the relative importance of the criteria for their evaluation. As a result, a list of priority directions of Green IT development is formed.

The subsystem for the formation of recommendations contains a mechanism for issuing the recommendation of the decision maker to select a better solution using the proposed model and data analysis methods, namely the list of Green IT development directions, which will become, in the near future, the most promising in Ukraine.

The subsystem visualization of results is designed to display the intermediate and resultant analysis data (in the form of graphs, tables or text arrays of data), as well as information contained in the database. The visualization subsystem includes a module for processing the results; module for displaying information (database information, results of performed calculations, analysis); module of graphic display of results (graphical presentation of the results obtained in the form of tables, graphs).

The administration subsystem is designed to differentiate the powers and access rights of users to information resources, control the program configuration of the system, execute the functions of DB administration.

It should be noted that the work of the subsystems occurs consistently, and the work of the next subsystem can not begin without successfully executing its functions by the previous subsystem.

8 Conclusion

It is proposed to consider the development of Green IT technologies at the state level of countries in the context of a neo-institutional paradigm. It is shown how Green IT can solve global problems facing humanity in the next 20 years and in the distant future. The concept of Green IT development as convergent technologies in Ukraine, is outlined for solving global problems in the conditions of the new industrial revolution.

An assessment of the innovative potential of Ukraine was conducted to substantiate the feasibility of introducing Green IT. As initial data used the indicators of annual rating “ōe Global Innovation Index” by the Cornell University, the French Business School and Research Institute (INSEAD), the World Intellectual Property Organization (WIPO) and the official EU statistics. To assess the innovation potential, an additive technique to convolve the normalized partial indicators on the basis of a simple average was used. As a result, three clusters were allocated: the first—10 countries (35.71%); the second—eight countries (28.57%); the third—10 countries (35.71%). Ukraine entered in the third cluster, that indicating of its rather low innovative potential.

To provide the appropriate level of efficiency of forecasting processes, information technology of the multilevel adaptive estimation of the current level and the prospects for the development of Green IT is described, which, unlike existing ones, is based on a set of developed methods of forecasting and complexation of forecast estimates, which allows ensuring the quality of forecasts in the conditions of uncertainty of data.

It is advisable to use the results of the structural and parametric synthesis of the Brown model for interval time series of integral indicators of the innovation potential as the initial data for the foresight research of the Green IT development. It was found that the smallest prospects, in the future, has a direction of software ecosystem, and the largest—green software engineering. The most perspective directions for the Green IT development in Ukraine are still energy-saving green software and green software engineering.

To automate the process of determining the Green IT perspective directions, a specialized IDSS was developed. The IDSS is intended for specialists who create state programs for the innovation policy of Ukraine namely of forming recommendations on perspective directions of the Green IT development. The generalized structure of the IDSS on the definition of the Green IT development is described. Structural parts are the subsystem of input and storage of data, the module of setting the system model, the subsystems of the formation of expert panels and the initial list of Green IT directions, the subsystem of evaluation and refinement the list of Green IT directions development, the subsystem of the formation of recommendations, the subsystem visualization of the results, as well as the subsystem of administration of IDSS.

The obtained results will serve as a basis for creating an integrated methodology for determining the opportunities of advanced technologies development in countries with different levels of technology development, which will ensure the effectiveness of innovation policy at the state level.

References

1. Yakovets, Y.V.: Global Economic Transformations of the 21st Century. Ekonomika. Moscow (2011) (in Russian)
2. Heyts, V.M., Mazaraki, A.A. (eds.): Priorities of National Economic Development in the Context Globalization Challenges, vol. 1. Kyyiv. nats. torh.-ekon. un-t, Kyiv (2008) (in Ukrainian)
3. Peres, C.: Technological Revolutions and Financial Capital. The Dynamics of Bubbles and Golden Ages. Edward Elgar Publisher, Cheltenham (2003)
4. Rifkin, J.: The Third Industrial Revolution: How Lateral Power is Transforming Energy, the Economy, and the World. St. Martin's Press, New York (2013)
5. Zhalilo, Y.A., Arkhiyeryeev, S.I., Bazilyuk, Y.B. et al.: Problems and Priorities of Formation the Innovation Model of Ukraine Economy Development. NISD, Kyiv (2006) (in Ukrainian)
6. Global Manufacturing Outlook: Preparing for battle: manufacturers get ready for transformation. https://www.kpmg.com/CN/en/IssuesAndInsights/ArticlesPublications/Documents/Global-Manufacturing-Outlook-O-201506.pdf
7. Akayev, A.A. et al. (eds.): Kondratieff Waves: Aspects and Perspectives. Uchitel', Volgograd (2012) (in Russian)
8. Bainbridge, W.S., Roco, M.C. (eds.): Handbook of Science and Technology Convergence. SpringerNature, Dordrecht (2016). http://www.springer.com/us/book/9783319070513
9. Kharchenko, V., Illiashenko, O.: Concepts of green IT engineering: taxonomy, principles and implementation. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Concepts, Models, Complex Systems Architectures, Studies in Systems, Decision and Control, vol. 74, pp. 3–20. Springer International Publishing, Berlin, Heidelberg (2017). https://doi.org/10.1007/978-3-319-44162-7_1
10. Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.): Green IT Engineering: Concepts, Models, Complex Systems Architectures, Studies in Systems, Decision and Control, vol. 74. Springer International Publishing, Berlin, Heidelberg (2017). https://doi.org/10.1007/978-3-319-44162-7
11. Matyushenko, I.Y.: Comprehensive modernization as a prerequisite economic recovery in Ukraine. In: Institutional Framework for the Functioning of the Economy in the Context of Transformation, pp. 23–28. Collection of scientific articles, Montreal (2015)
12. Pazowski, P.: Green computing: latest practices and technologies for ICT sustainability. http://www.toknowpress.net/ISBN/978-961-6914-13-0/papers/ML15-377.pdf (2015)
13. Calero, C., Piattini, M.: Introduction to green in software engineering. http://www.springer.com/978-3-319-08580-7 (2015)
14. Pavlenko, V., Shostak, I., Sobchak, A., Morozova, O., Danova, M.: The concept of virtual manufacturing enterprise operation as a green complex system. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Concepts, Models, Complex Systems Architectures, vol. 105, pp. 265–285. Springer International Publishing, Berlin, Heidelberg (2017). https://doi.org/10.1007/978-3-319-55595-9
15. Ardito, L., Procaccianti, G., Torchiano, M., Vetrò, A.: Understanding green software development: a conceptual framework. IT Prof. **17**, 44–50 (2017). https://doi.org/10.1109/MITP.2015.16
16. Kurfuss, Th.: Industry 4.0: Manufacturing in the United States. Bridges. http://ostaustria.org/bridges-magazine/item/8310-industry-4-0 (2014)
17. Foste, L.E.: Nanotechnology: Science, Innovation, and Opportunity. Prentice Hall PTR. New Jersey (2006)
18. Khaustova, V.Y.: Industrial Policy in Ukraine: Formation and Forecasting. INZHEK. Kharkiv (2015) (in Ukraine)
19. Geyman, O.A.: Nonlinearity of Economy and Uneven Development Regions. Liburkina L. M. INZHEK, Kharkov (2009)

20. Salam, M., Khan, S.U.: Identification of practices for green software development in multi-sourced project: a systematic literature review protocol. Int. J. Comput. Appl. **141**(8), 33–38 (2016)
21. Matyushenko, I.Y.: Development and Implementation of Converging Technologies in Ukraine Under Conditions of a New Industrial Revolution: Organization of State Support. Aleksandrova K. M., Kharkiv (2016) (in Ukrainian)
22. The Global Innovation Index 2012, 2013, 2014, 2015, 2016—Cornell University, INSEAD, WIPO, 2012–2016: Ithaca, Fontainebleau, and Geneva: Official Website. http://www.globalinnovationindex.org
23. Eurostat: Official Website. http://ec.europa.eu/eurostat/web/main
24. Romanenkov, Y.: Analysis of the predictive properties of Brown's model in the extended domain of the internal parameter. MOTROL. Comm. Mot. Energ. Agric. **17**(8), 27–34. Lublin (2015)
25. Shostak, I., Danova, M., Kuznetsova, Y.: Foresight-research for green IT engineering development. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Concepts, Models, Complex Systems Architectures, Studies in Systems, Decision and Control, vol. 74, pp. 3–20. Springer International Publishing, Berlin, Heidelberg (2017). https://doi.org/10.1007/978-3-319-44162-7_2
26. Kyzym, M.O., Matyushenko, I.Y.., Shostak, I.V., Danova, M.O.: Prospects for Foresight Forecasting of Priority Directions of Development of Nanotechnology and Nanomaterials in Countries of the World and Ukraine. INZHEK, Kharkiv (2015) (in Ukrainian)
27. Shostak, I., Danova, M.: Synthesis system models of choice the priorities at realization of national foresight-researches. ECONTECHMOD, **4**(3), 19–24. Lublin (2015)

The Application of Blockchain Technology in the Maritime Industry

Karen Czachorowski, Marina Solesvik and Yuriy Kondratenko

Abstract The maritime industry is one of the most polluting in the world. In this paper we present fresh insights related to the application of the novel blockchain technology in reducing pollutions. We analyse recent literature on blockchain technology and propose ways of the utilization of blockchain technology in the maritime industry. Special interest for maritime industry specialists represents the perspectives of the utilization of the blockchain technology in improving the environmental efficiency of the maritime industry. The technology has a broad range of applicability, allowing connecting the supply chain more efficiently, providing the exchange and visibility of time-stamped proofed data, decreasing the industry operational costs with intermediaries and increasing security. It also allows full visibility for all parties involved with proof of work, facilitating Class Societies inspections, Port State Control and audits compliance. The results of the study also show that cases on blockchain application in other fields increase the industry willingness to its application on the maritime industry. While having blockchain implementation specialized third parties would increase the implementation possibility and the industry willingness due to reduced costs and friction. The study will be interesting for scholars, policy-makers and practitioners from the maritime industry.

Keywords Blockchain technology · Maritime industry · Green supply chain Green IT

K. Czachorowski
University College of Southeast Norway, 3679 Notodden, Norway
e-mail: karenvcz@gmail.com

M. Solesvik (✉)
Nord University Business School, Nord University, 8049 Bodø, Norway
e-mail: mzs@hvl.no

Y. Kondratenko
Petro Mohyla Black Sea National University, Mykolaiv 54000, Ukraine
e-mail: y_kondrat2002@yahoo.com

© Springer Nature Switzerland AG 2019

V. Kharchenko et al. (eds.), *Green IT Engineering: Social, Business and Industrial Applications*, Studies in Systems, Decision and Control 171,
https://doi.org/10.1007/978-3-030-00253-4_24

1 Introduction

The steadily growing volumes of industrial production, growing population in the world and growing traffic exhaust lead to the contamination of environment. These issues concern policy-makers, scholars and practitioners around the globe. The IT sector cannot stay apart of the global trends and concerns and has much to offer in order to reduce the volume of pollution and thus save the environment [1–6]. One of such radical IT technologies that emerged ten years ago is the blockchain technology [7, 8]. The blockchain technology is the general-purpose technology [9] and can be applied in many industries. The blockchain technology is defined as “a distributed database of records or public ledger of all transactions or digital events that have been executed and shared among participating parties” [8, p. 7]. The system is secure—the records are verified by public ledgers and can never be erased in the future. The main benefits of the blockchain technology are transparency and cost-efficiency [10]. In the present, transactions between firms are made through intermediaries (i.e. banks) that charge significant amounts of money for their operations and currency conversion. The blockchain technology allows firms to make transactions directly between each other and is made in a very secure manner. The transaction costs that firms can save on bank operations can be used in the development of green technologies. Another advantage of the blockchain technology is the possibility to execute “smart contracts”. Smart contracts are computer programs that can automatically execute contracts’ conditions [8]. The amount of paper utilized for contracts and other paperwork would be significantly reduced.

The seaborne transport is one of the cheapest means of transportation in terms of the price per tonne of the cargo transported. However, the maritime transport is extremely polluting. For example, one ship produces as much CO_2 and NO_x exhaust as 70,000 cars. Different measures are approved by policy-makers in order to make the maritime transportation environment-friendly. For example, the new legislation specifies that exhaust-free vessels can come into European ports after 2020 (i.e. for example, the level of sulphur oxides, should not exceed 0.5% mm). Practitioners from shipbuilding and the shipping industry are concerned with the development of new vessel types, driven by electricity, hydrogen or hybrid [11]. Practitioners and policy-makers also search for other ways to reduce pollution in the industry. Without any doubts, digitalization opens new possibilities to save the environment and to provide sustainable development in the industry.

The research on the blockchain technology is quickly developing nowadays. However, two areas within the blockchain technology are barely researched yet. One of them is an application of blockchain in the development and support of green technologies. There are a few articles published in 2017 that shed light on how the blockchain technology assists in green innovation support [12, 13]. As mentioned above, the blockchain technology is a universal general-purpose instrument that can be applied in different areas. Thus, one of the most priority areas of sustainable development is the most important area of the blockchain application. This study intends to cover this gap and contribute to the emerging green blockchain stream of research.

The blockchain research is an interesting area for researchers and practitioners. Several special issues on the blockchain technology recently appeared in different ICT, technology and innovation journals. The topic is also popular in different conferences. The articles shed light on the application of the blockchain technology in different industries and areas, such as the energy sector [10], healthcare [14], finance [8, 15], government services [16] and others. The interest in the green IT technology in the maritime industry is steadily growing [17, 18]. However, there was still limited research on the application of the blockchain technology in the maritime sector [19]. Without any doubt, the highly international maritime business is one of the industries where the blockchain technology perspectives are quite interesting and thus more research attention is needed. Our paper intends to cover this gap in this knowledge and shed light on barriers on the dissemination of the blockchain technology in the maritime industry.

This research aims presenting the possibility of Blockchain technology applications within the offshore maritime industry. This chapter focuses on Scandinavian shipping companies. The Norwegian and Scandinavian maritime sector is highly innovative. It can be named as an innovation leader in the world maritime business. Norway and other Scandinavian countries are high-cost countries and practitioners constantly seek ways to reduce costs of operations in order to stay competitive in this highly volatile and cyclical market [20]. The research question that guides this study is: Why should firms in the maritime business implement the blockchain technology in order to preserve the environment? We have carried out interviews with the operators and suppliers of the Norwegian offshore shipping industry to answer the research question.

This paper is organized as follows. In the next section, we present the theoretical background of the study related to the blockchain technology. In Sect. 3, we explain the specifics of the maritime business, its need of the green IT technology and propose ways on how to utilize the blockchain technology to cover this need. The chapter terminates with the conclusions and implications for theory and practice.

2 The Blockchain Technology

In 2008, Satoshi Nakamoto (an alias for a person or group of people who has never had its identity publicly confirmed) developed Bitcoin, which is a peer-to-peer electronic cash system that allows two parties to perform payments directly, excluding the need for a trusted third party or intermediary, in financial transactions. Bitcoin is essentially a chain connecting several digital signatures, verified by a timestamp server [7]. Nakamoto developed a cryptocurrency to enhance trust among peers and to allow direct transactions, overcoming the need for intermediaries in financial transactions, thus reducing costs. He created a digital foundation, a technology to allow such transactions, which is now known as blockchain. Although blockchain was firstly created with Bitcoin together, the technology evolved to apply for several uses and businesses and the concept should not be confused with the concept of Bitcoin.

2.1 Development of Blockchain Technology

Since Bitcoin's start and popularization, several technology companies have begun working with blockchain, popularizing the technology within the "*fintech*" industry, as it is known. While most of these companies are still in their start-up phase, many of increasing interest over the technology. The biggest company to invest in blockchain so far has been IBM, which states that they expect blockchain to revolutionize transactions the same way the Internet has done for communications, allowing increased trust and efficiency for low or zero costs [21]. The company has started a big blockchain project, investing in academia and professional training, while tailoring solutions for other companies interested in applying the technology. According to IBM, blockchain is:

> …a shared, distributed ledger that facilitates the process of recording transactions and tracking assets in a business network. An asset can be tangible — a house, a car, cash, land — or intangible like intellectual property, such as patents, copyrights, or branding. It can also be used to help companies manage the flow of goods and related payments, or enable manufacturers to share production logs with original equipment manufacturers (OEMs) and regulators to reduce product recalls. Virtually anything of value can be tracked and traded on a blockchain network, reducing risk and cutting costs for all involved. [22, p. 5].

Blockchain is public, meaning that anyone can participate and contribute in the chain. It works based on Proof-of-Work system, in which each block contains a *hash*, a unique identifier or digital fingerprint of the data contained in the block (e.g. 000000rt5687ai85 in Block 456), the hash from the previous block (0000hj87yuki98 in Block 456), along with batches of time-stamped recent valid transactions, which also have their own hash (e.g. 0987hynjyf45lk87 in Block 456). The timestamp server publishes the previous blocks' hashes, proving that the data must have existed at the time (since it has a hash), also including the previous timestamp in the hash, thus forming a chain reinforced by each additional timestamp (n), shown on Fig. 1 in the sequence. The proof-of-work system allows the timestamp server's implementation on P2P basis by scanning for a value that when encrypted, the hash will begin with a number of zero bits. For this, it was added a nonce in the block, a number or key that needs to match exactly to the nonce created when the block/hash was created, therefore signed, until a value is found that gives the block's hash the required zero bits [7].

These replications mean that each node acts as both a publisher and a sub-them have already developed and have begun incorporating others, showing the scriber of the ledger; being allowed to send or receive transactions to and from other nodes; while the data is synchronized across the network as the transactions occurs [22].

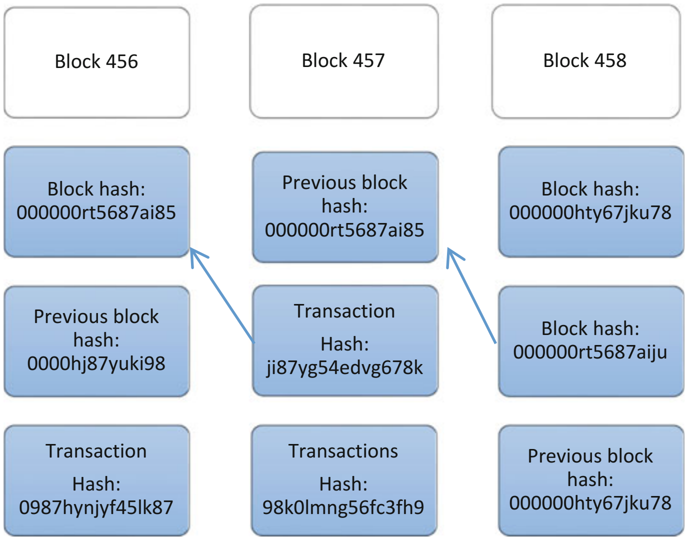


Fig. 1 Principle of the blockchain technology

The timestamp server, the ledger distribution, along with the proof-of-work concept constitutes the consensus model, which is blockchain's validation system [7], assuring that all transactions are authenticated, secure and verifiable [23]. Because of its characteristics, a blockchain network enforces trust within the chain for all the users [24].

2.2 *Types of Blockchain*

There are three types of blockchains, being Public (Bitcoin, Ethereum, Litecoin, etc.), Federated or Consortium (R3, B3I, EWF) and Private (company internal). The first one is the public and anonymous created technology, which is permission-less, meaning that there is no requirement for software, allowing anyone to participate, thus, completely decentralized. This public chain uses a Proof-of-Work consensus system to validate and maintain the nodes, while the other two are types are also decentralized within their users, or in other words, centralized to the permitted users to access the network, and require a solution provider to develop the chain. Their consensus system is similar to the one developed by Nakamoto [7], but they differ in accessibility. Table 1 shows their differences and particularities.

Permission-Less Blockchain
This Blockchain is public, meaning that anyone can participate and contribute in the chain. It works based on Proof-of-Work system, in which each block contains a *hash*, a unique identifier or digital fingerprint of the data contained in the block, the

Table 1 Types of blockchain and their characteristics

	Public	Consortium	Private
Participants	Without permission • Anonymous • Could be malicious	Permissioned • Identified • Trusted	Permissioned • Identified •Trusted
Consensus mechanisms	Proof of work, proof of stake, etc. • Large energy consumption • No finality • 51% attack	Voting or multi-party consensus algorithm • Lighter • Faster • Low energy consumption • Enable finality	Voting or multi-party consensus algorithm • Lighter • Faster • Low energy consumption • Enable finality
Transaction approval freq.	Long Bitcoin: 10 min or more	Short 100× ms	Short 100× ms

hash from the previous block, along with batches of time stamped recent valid transactions, which also have their own hash. The timestamp server publishes the previous blocks' hashes, proving that the data must have existed at the time (since it has a hash), also including the previous timestamp in the hash, thus forming a chain reinforced by each additional timestamp [7]. The proof-of-work system allows the timestamp server's implementation on P2P basis by scanning for a value that when encrypted, the hash will begin with a number of zero bits. For this, it was added a nonce in the block, a number or key that needs to match exactly to the nonce created when the block/hash was created, therefore signed, until a value is found that gives the block's hash the required zero bits [7].

The hash created shows the previous' block hash as well. If any alteration is attempted in the chain, the connection to the previous block will be broken, which will then cause the whole chain to break. So, the longer the chain, the stronger it is, requiring all previous blocks to be changed to match the new information. However, a process allows searching for the nonce number for the altered block, called "*mining*". Once any block in the chain is mined, it receives a new hash and nonce, while all the blocks in that ledger are now broken [7]. What makes this system work is that every node (user) in the chain has a copy of the chain, being the distributed characteristics of Blockchain (decentralization). Hence, if one block or more were mined to accept the alteration, the last block will have its hash altered and consequently, different from the other records from other users in the distributed chain. Therefore, the greater amount of identical hashes in the last blocks "wins" the distributed chain, keeping the immutable characteristics [7]. The mining concept also works as an incentive for keeping the Blockchain un-hackable. To discover the nonce number, several calculations are required, which demand very high processing capacities from a computer and/or server. Every time a "miner" can actually mine a nonce, it is paid a Bitcoin and the calculations become harder.

Permissioned Blockchains

In the permissioned blockchain, the users have a special permission to access the chain, working as a guarantee that only the allowed users can access the chain or specific parts of it, based on their assigned roles. It is also a distributed ledger; however, users may or may not be anonymous. Even though it also uses a consensus-base data validation, it does not apply the public blockchain Proof-of-Work, since the mining process explained before takes longer to process and requires advanced and high computing power, becoming expensive for private use. Thus, the permissioned types apply the concept of "Smart Contracts", which is:

> an agreement or set of rules that govern a business transaction; it's stored on the blockchain and is executed automatically as part of a transaction. Smart contracts may have many contractual clauses that could be made partially or fully self-executing, self-enforcing, or both. Their purpose is to provide security superior to traditional contract law while reducing the costs and delays associated with traditional contracts. [21, p. 17]

In this system, a private key distributed from a public key, which together form a "signature share", then creating a signature. If anyone can manage to discover the public key, this person will not be able to breach in, as there will be a private key to open the block [25].

All Blockchain types maintain the other characteristics mentioned before, with the permissioned ones returning to the same characteristics after the signature authentication. The main advantage from decentralization is the facility to identify and correct problems within the chain since it is easy to isolate a problematic node to investigate the matter further, then discovering if it is malfunctioning or showing signs of tampering, without losing the behaving nodes, which can continue to be available in the system [26]. Buterin [27] explains that there are three other main reasons for decentralization, being: fault tolerance, attack resistance and collusion resistance. The first explores Norton's [26] concept a little further, explaining that decentralized systems are less likely to fail by accident due to their reliance on separated and non-relational components. The system also becomes attack resistance due to its lack of sensitive central points, becoming more difficult to attack, destroy or manipulate; while becoming collusion resistant due to the difficulty of having most or all participants to conspire in favor of themselves over the other parts [27].

It is important to discover how to choose the correct type of blockchain to the required application. To do such, IBM suggests businesses to ask themselves a few questions to perform an analysis, such as:

- "Do you require a permissioned network?
- Do you need to know the identities in your business network? For example, to adhere to regulations such as anti-money laundering or know your customer?
- Do you have frequent exchanges with others that could be automated and pre-programmed, freeing up valuable time and resource?
- Would you benefit from transaction resolution in minutes rather than days or weeks?" [22].

Table 2 Blockchain implementation solutions

Approach	How it is done	Examples
IT services	Build on request	ConsenSys
Blockchain first	Develop using the tools provided by the blockchain	Ethereum, Bitcoin
Development platforms	Tools for IT professionals	ERIS, Tendermint, Hyperledger
Vertical solutions	Industry specific	Axoni, Chain, R3, itBit, clearmatics
Specials APIs and overlays	DIY building blocks	Blockstack, Factom, Open Assets, Tierion

Blockchain Storage and Implementation Solutions

The blockchain, disregarding which type, keeps the last block recorded in all the chain's nodes, thus saving disk space requirements. Yet, it is necessary to build the chain somewhere. When it is decided which type of blockchain will be implemented, it is necessary to choose the solution required to it. There are several providers in several possibilities as software and software as a service (SaaS); cloud based and Blockchain as a service (BaaS). Table 2 details some providers and solutions. The main players so far are Amazon, with AWS BaaS, Microsoft with Azure BaaS and IBM with their BlueMix BaaS [22]. However, the main Blockchain solution is the Hyperledger, an open-source Linux Foundation Project launched in 2015 alongside with 17 other companies to collaborate for the technology's development and advance into the cross-business use. There are other open-source platforms with the same purpose, such as Iroha, a C++ Blockchain platform, and Cello, a rapid cloud platform Blockchain deployment [21]. This way, developers have a single-click cloud-based environment for Blockchain deployment providing rapid smart contracts development [22].

3 The Blockchain Technology in Maritime Industry

As a result, the technology has several possible applications within other industries than finance, public sector, retail, manufacturing [28]. Even though not all companies might need the technology in their own operations, they might participate in other networks to facilitate and enhance businesses, comply with regulations and other requirements, among other examples, with a few applications on the maritime industry.

3.1 Industrial Specifics

The maritime industry has established itself as a key supply chain stakeholder over the years, either by its sub-industries or by supporting businesses and allowing their growth. Shipping itself has also become a differentiator among enterprises and an advantage for them in enlarging their market reach [29]. However, the industry faces some old known obstacles along the way, such as shipping cycles [30], along with crucial challenges and choices that can mean a company's survival in the industry. "The marine industry is undergoing a transformation. As well as managing today's rising operational costs and achieving cost-effective environmental compliance, ship operators are faced with tomorrow's "big decisions". Decisions about fuels, technology and whether it is possible to "future-proof" their fleet and assets" [31].

Maritime industry is one of the most affected industries by new stricter rules and legislations [32] mostly enforced by the International Maritime Organization (IMO) and the European Union (EU), as well as other global treaties. How to deal with such challenges call for diverse solutions and can be a good opportunity for innovative technologies. The industry has already conducted several studies, prototypes and other innovations, such as un-manned vessels prototypes, technologic control and engine rooms, artificial intelligence for learning and training, "green" fuels and batteries, etc. Environmental compliance fleet to allow local and global trade has been a great challenge for the marine industry as the legislation keeps updating constantly to new and higher standards [33].

Data is the focus of shipping digitalization, mainly driven by the offshore and containerized shipping, as poor information management can account for up to 20% of an operational budget [34]. A typical supply chain manages a data inflow of an average of 100 gigabytes per day [35], which is expected to reach 35 zeta bytes by 2020 [36]. Jan Wilhelmsson, shipping vice president of Eniram, a Wartsila systems developing company, lead a research for the company to develop their new fleet performance monitoring system to discover how the maritime industry is spread by digitalization. His findings showed that the cruise industry is on track to real digital transformation being visibly ahead of the others, while the cargo segment is mixed, being mostly advanced on shore. He divided the industry into digital evolution categories to understand the differences. The first category includes companies with management engaged to technology and innovation to support the business. The second, the ones that tried but are still struggling, with either methodology, data, results, etc.; while the third group comprises the ones that think their operations are too complex for any solution, therefore, faded to be out of the competition soon.

In an interview in the first quarter of 2017, Dr. Martin Stopford states that digitalizing shipping is the only solution for the industry. He argues that three methods would change the business model running through shipping cycles, allowing the digitalization that the industry requires: smart ships, smart fleet and smart global logistics. However, he adds that a great cultural change is needed in order to allow such changes, aligned with Wilhelmsson's findings. Stopford argues that "*having the technology is a first step, but the data needs to be used in order to show performance improvement*" [37]. One example is the concept of digital oil fields, in which information technology tools are applied to constantly collect data, providing big data for posterior decision-making and solving possible lack of qualified labour in the future [38]. The maritime transport and logistics has been applying data-driven technologies application for some time, with several examples, including the first blockchain application by Maersk. One more innovation comes from a start-up called Xeneta, which has begun collecting data from container-liners and tracking over 60,000 routes globally, allowing shippers to receive real pricing information for benchmarking, rather than past static data; thus, providing pricing transparency and supporting contracts negotiations [39]. Radio Frequency Identification (RFID) systems integration through Internet of Things (IoT) in order to create value for the chain based on data is another example driven by business needs [40, 41]. The use of RFID tags is estimated to increase to 209 billion units by 2021 [42], as the sensor-based technology could reduce operational costs by 10–25% [43]. IoT most current application in the industry is GPS tagging of shipping containers to assist managing their flow through transit nodes [44]. This also enables real time tracking of goods and vessels and delivers extra customer service as it provides real time information about one or more specific container. Another example is port operations. As shipping companies and other logistics modals compete fiercely for port space and resources, the collected data is not shared among the peers. This makes it extra difficult for a system to analyse and redistribute data from the port to its stakeholders, in order to optimize the use of available supply logistics; thus, turning ports in bottlenecks of everyday operations, as volumes keep increasing. The port of Hamburg estimates that the number of containers passing the port will increase from 9 million in 2013 to 25 million in 2025, and solved its data-sharing problem by requiring all parties to connect in a single data system [44]. The innovation scenario in the maritime industry seems to be evolving and growing rapidly. DNV-GL has already started providing cloud solutions and digital offerings, such as data analysis and value creation form it. Its main software is providing integration with machine learning and between historical data analysis and future forecast, with the goal of fully integrating to the IoT in the future. Such applications will allow data-smart operations and asset management, through another technology called "Digital Twin", which is basically an asset's digital model representing its deep profile characteristics, such as systems, software, behaviour, needs, demands and so on, providing analysis, perception and diagnostics while completely integrated with all stages/stakeholders [34]. With companies such as DNV GL, Kongsberg Maritime, etc. and several star-ups, Norway already has an established presence into the maritime and innovation development. In March 2017, it was announced a collaboration to test pilotless vehicles in the Fjord of Trondheim, among the Norwegian University of Science and Technology (NTNU) and industrial players as Kongsberg Seatex, Marintek, Maritime Robotics and Rolls-Royce Marine, becoming the world's first test location for self-directed vehicles such as unmanned ships and underwater drones.

Other countries also entered the digitalization race by fostering innovation. The Maritime and Port Authority of Singapore (MPA) launched the Smart Port Challenge 2017 to encourage start-up and organizations collaboration, pushing digital transformation into the industry, harnessing technologies to add value to the maritime logistics chain, also collaborating with the Port of Rotterdam in the same kind of endeavour. The start-ups are gaining space into the industry as entrepreneurs discovered its financial and size potentials, already claiming attention due to their innovative concepts. One start-up called Onboard, based in Rotterdam, is bringing the Internet of Things to the maritime industry by providing an open platform with full integration with other applications and customer's internal systems, providing full insight of vessels and operations [45]. Another start-up, Care4C's, is focusing on bringing the telemedicine into the vessels, monitoring and collecting data about cardio vascular and sleep patterns, providing predictive analysis and proactive risk management [46]. While Care4C's is digging into a needed and underdeveloped field, most of the start-ups are focusing on data analytics and artificial intelligence to reduce fuel consumption, CO_2 and other emissions, route optimization and integration with other technologies and trends, such as drones use and the afore mentioned unmanned vessels. However, data and analytics software, IoT applications, etc. require intermediaries in the chain to be accessible, which adds extra steps, resources and costs to it, as they require extra labour, training, systems integration, etc. Yet, all these examples show that there is willingness within the industry to pursue more innovations and innovative behaviour, developing the cultural aspects pointed by Stopford. Alongside, Blockchain technology could be a solution to support most of the ongoing and needed innovations, as it is a decentralized, application, building an encrypted and immutable ledger accessible and confirmed by all participants of the chain or only those who have access to it.

3.2 Blockchain Application in Maritime Industry

While still not completely diffused within the industry, the technology has been gaining space into discussions and possible applications. One good example is the Marine Transport International Limited (MTI) freight forwarder. In 2016, MTI announced that was using a public blockchain ledger called TrustMe™ to comply with the new verified gross mass (VGM) requirements of packed containers, implemented from July of 2016 by the Contracting Governments to the International Maritime Organization's SOLAS treaty (International Convention for the Safety of Life at Sea). The new regulation transferred to the shipper the responsibility of ensuring that the right VGM is provided to the terminal or carrier prior to being loaded to the vessel. The company then began using Blockchain through TrustMe™ in order to provide a permanent and visible record to port officials, shippers and cargo owners; thus eliminating the need for data intermediaries, private databases, logs and spreadsheets [19].

The technology has also been applied to enable single energy trade within commodities markets. A renewable energy trader, Volt Markets, implemented a public blockchain to provide assurance of trade and absolute tracking for renewable energy certificates, while in January 2017, another trading house, Mercuria, carried out the technology on a large oil trade within ING and Société Générale banks [47].

Several other uses are possible within the maritime industry, mainly in order to solve regulation compliance, documentation issues and origin assurance as well as to support communication and automation. Blockchain could solve the digitalization of the Bill of Lading (BoL), essential document within shipping, providing an immutable chain accessible by all parts required within 10 min of its creation. The BoL is a contract of carriage and an ownership/receipt document, and by the Maritime Law, the original version of the document is required to be produced by the consignee to allow cargo delivery. This document is frequently delayed due to banks and other issues, which leads to the cargo usually arriving at ports before the original BoL creation or to BoL fraud. Such includes the document being produced and signed before the actual loading of the vessel or being altered after cargo delivery; with forged signatures, bank guarantees and wrong description of cargo [48].

The blockchain application to support BoL was one of the first researched applications within maritime industry and is probably one of the most advanced ones, alongside with containerized shipping, already having adoption frameworks and models. An Israeli start-up called Wave is focusing on creating a paperless trade for shipping trade by applying Blockchain to create and track BoL's, Letters of Credit and the whole chain. Its main competition, Skuchain, located in California, USA; believes in the evolution of the "collaborative commerce", being the integration of all parts of all supply chains involved in commerce, exchanging information and collaborating towards commerce.

In order to find out whether Norwegian maritime companies are ready to adopt the blockchain technology, we carried out interviews with the representatives of shipping companies. There were seven interviews performed in total, being four with Offshore operators and three with suppliers, which were divided by these two categories and given numbers from 1 to 4, for Offshore operators players, and 1 to 3 for Offshore support players. Table 3 and Fig. 2 show the interviewees perception on how much shipping companies are willing to invest into blockchain innovation in general and divided by their own companies and in the offshore. The results from analyzing the interviews shows that there is need to innovate in the form of blockchain technology in the maritime industry and consensus regarding such need, however, with many limitations and uncertainties. There is also a consensus that costs are the main enabler and barrier to technological implementation, as well as secondary costs derived by it, such as reliance on consultants and third parties, need for training, need for specific personal and other possible requirements, such as software integrations. The willingness for innovation is directly connected to these factors, as well as what can be expected from the innovation. Again, the great measurement for success is cost, being cost reduction the biggest motivation for innovation.

Table 3 Perception to adopt blockchain innovation

	Company willingness		Industry willingness	
	Coding coverage (%)	Coding references	Coding coverage (%)	Coding references
Operator 1	5.79	6	6.23	6
Operator 2	8.18	6	0	0
Operator 3	2.63	3	1.66	1
Operator 4	5.28	3	8.42	6
Support player 1	7.66	6	1.45	1
Support player 2	5.10	7	4.02	6
Support player 3	7.56	9	4.07	4

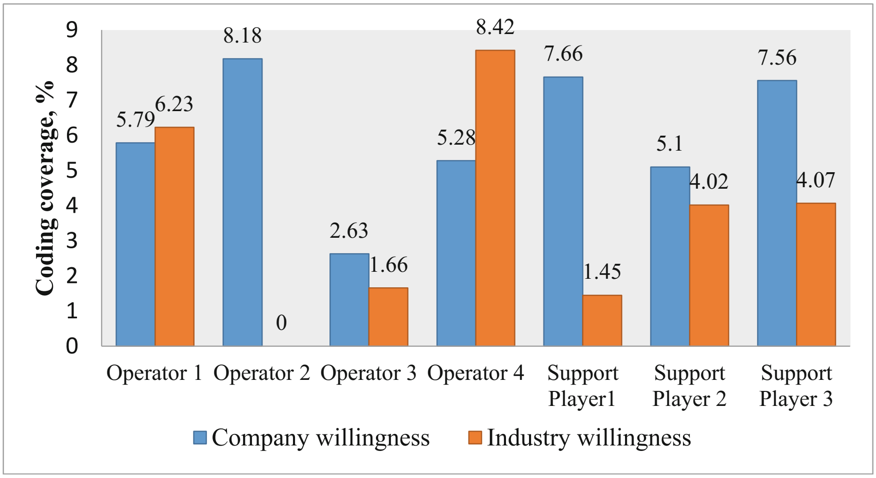


Fig. 2 Company and industry willingness to adopt blockchain technology: expert evaluations

4 Conclusions

Maritime industry is quite complex industry [49–51] that is characterized by high costs and the high level of pollutions. This study explored the possibility of blockchain utilization with the aim of pollution prevention within the maritime industry and other non-financial uses. We have identified the literature and its gaps related to the subject. The lack of academic literature about blockchain and its utilization in green innovations, especially in the maritime industry, was the key driver to this research. In a literature review of blockchain academic research, Yli-Huumo et al. [52] has not detected any study of its application within the maritime area.

Blockchain's applicability to architecture allows companies to keep an encrypted, immutable ledger of transactions, which can be shared trustfully within the selected network due to the peer-to-peer (P2P) proof of work concept, eliminating the third parties involved, thus saving money through reducing costs.

This study came from the perception that the industry lacks digital innovation in its operations and green supply chain [53], being one of the late adopters regarding new technology. The main research question driving this study attempted to foster an academic research of the topic. The answer to this question—Why should firms in maritime business implement the blockchain technology in order to preserve the environment—is quite simple. The industry needs to reduce costs to be able to remain lucrative and overcome the obstacles presented by global energy supply and demand, turning them into opportunities, while complying with the legislation requirements enforced by the responsible global regulation bodies. IT tools are excellent instruments that allow cost savings [54]. Maritime industry needs also an enhanced action against cybercrimes and piracy. Thus, the technology presents itself as a solution for such and opens many more opportunities within the industry.

Still, the industry needs a push towards the digital revolution it requires—external motivation. Based on the literature review and the results, it is clear that the industry is motivated by costs and by the need of implementing something based on legal requirements. Just as in the case of the Port of Rotterdam, if required by the Port Authority, Classification Societies, IMO or other similar bureaus, the industry rushes to adapt itself in an attempt to "future proof" its assets, which could be a main enabler for innovation and technological implementation. It is clear that the industry's lack of knowledge of blockchain's possibilities is the principal reason for its still rather slow adoption rate, which should change in the near future, based on the success of the pioneer's adoption.

The technology has a broad range of applicability, allowing connecting the supply chain more efficiently, providing the exchange and visibility of time-stamped proofed data, decreasing the industry operational costs with intermediaries and increasing security. It also allows full visibility for all parties involved with proof of work, facilitating Class Societies inspections, Port State Control and audits compliance. However, it is not enough to solve the operations need for innovation. It is important to differentiate the technology in order to engage with it, highlighting that it is not data analytics software, a data storage provider nor server. Yet, it does best on what it promises—creates and provides decentralized public or controlled access to a ledger, which is immutable, secure and time-stamped verified by proof of work hash-system that allows the elimination of third parties and the costs associated to it. However, the lack of knowledge about the technology and the industry's cultural barrier to innovation act against innovation and new technology application, even though the results showed a consensus on the industry's need for enhancements for both. Nevertheless, the results also showed that studies and cases on blockchain application in other fields increase the industry willingness to its application on the maritime industry, while having blockchain implementation specialized third parties would increase the implementation possibility and the industry willingness due to reduced costs and friction.

Costs present as the key driver to allow innovation, as both enabler and barrier, being the factor that can boost the industry's willingness of blockchain application, especially regarding operations, in addition to the industry's need for technological innovations.

References

1. Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.).: Green IT Engineering: Concepts, Models, Complex Systems Architectures. Studies in Systems, Decision and Control, vol. 74. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-44162-7
2. Kharchenko, V., Illiashenko, O.: Concepts of green IT engineering: taxonomy, principles and implementation. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Concepts, Models, Complex Systems Architectures. Studies in Systems, Decision and Control, vol. 74, pp. 3–19. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-44162-7_1
3. Kondratenko, Y., Korobko, O.V., Kozlov, O.V.: PLC-based systems for data acquisition and supervisory control of environment-friendly energy-saving technologies. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Concepts, Models, Complex Systems Architectures. Studies in Systems, Decision and Control, vol. 74, pp. 247–267. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-44162-7_13
4. Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.).: Green IT Engineering: Components, Networks and Systems Implementation. Studies in Systems, Decision and Control, vol. 105. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-55595-9
5. Kondratenko, Y., Korobko, V., Korobko, O., Kondratenko, G., Kozlov, O.: Green-IT approach to design and optimization of thermoacoustic waste heat utilization plant based on soft computing. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Components, Networks and Systems Implementation. Studies in Systems, Decision and Control, vol. 105, pp. 287–311. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-55595-9_14
6. Kuchuk, G., Kovalenko, A., Kharchenko, V., Shamraev, A.: Resource-oriented approaches to implementation of traffic control technologies in safety-critical I&C systems. In: Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.) Green IT Engineering: Components, Networks and Systems Implementation. Studies in Systems, Decision and Control, vol. 105, pp. 313–337. Springer, Cham (2017), https://doi.org/10.1007/978-3-319-55595-9_15
7. Nakamoto, S.: Bitcoin: A Peer-to-Peer Electronic Cash System (2008)
8. Crosby, M., Pattanayak, P., Verma, S., Kalyanaraman, V.: Blockchain technology: beyond bitcoin. Appl. Innov. **2**(6), 6–10 (2016)
9. McPhee, C., Ljutic, A.: Blockchain. Technol. Innov. Manag. Rev. **7**(10), 3–5 (2017)
10. Electricity journal homepage: blockchain technology: will it make a difference? Electr. J. **30** (3), 86–87 (2017
11. Solesvik, M.Z., Gulbrandsen, M.: Partner selection for open innovation. Technol. Innov. Manag. Rev. **3**(4), 11–16 (2013)
12. Imbault, F., Swiatek, M., De Beaufort, R., Plana, R.: The green blockchain: managing decentralized energy production and consumption. In: IEEE International Conference on Environment and Electrical Engineering, 2017 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe), pp. 1–5. IEEE (2017)
13. Castellanos, J.A.F., Coll-Mayor, D., Notholt, J.A.: Cryptocurrency as guarantees of origin: simulating a green certificate market with the Ethereum blockchain. In: IEEE International Conference on Smart Energy Grid Engineering, pp. 367–372. IEEE (2017)

14. Engelhardt, M.A.: Hitching healthcare to the chain: an introduction to blockchain technology in the healthcare sector. Technol. Innov. Manag. Rev. **7**(10), 22–34 (2017)
15. Zhu, H., Zhou, Z.Z.: Analysis and outlook of applications of blockchain technology to equity crowdfunding in China. Finan. Innov. **2**(1), 29–40 (2016)
16. Wolfond, G.: A blockchain ecosystem for digital identity: improving service delivery in Canada's public and private sectors. Technol. Innov. Manag. Rev. **7**(10), 35–40 (2017)
17. Solesvik, M.: Partner selection in green innovation projects. In: Berger-Vachon, C., Gil, L.A., Kacprzyk, J., Kondratenko, Y., Merigó, J., Morabito, C. (eds.) Complex Systems: Solutions and Challenges in Economics, Management and Engineering, vol. 125, pp. 471–480. Springer, Cham (2018)
18. Solesvik, M., Kondratenko, Y., Kondratenko, G., Sidenko, I., Kharchenko, V., Boyarchuk, A.: Fuzzy decision support systems in marine practice. In: 2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1–6. IEEE, 9 July 2017. https://doi.org/10.1109/FUZZ-IEEE.2017.8015471
19. Finextra: Marine Transport International Applies Blockchain to Shipping Supply Chain (2016). Retrieved from https://www.finextra.com/pressarticle/66223/marine-transport-international-applies-blockchain-to-shipping-supply-chain
20. Solesvik, M.Z.: Interfirm collaboration in the shipbuilding industry: the shipbuilding cycle perspective. Int. J. Bus. Syst. Res. **5**(4), 388–405 (2011)
21. Gupta, M.: Blockchain for Dummies, IBM Limited Edition. Wiley, Hoboken NJ (2017)
22. Blockchain Hub: Blockchains and Distributed Ledger Technologies. https://blockchainhub.net/blockchains-and-distributed-ledger-technologies-in-general/. Last accessed 15 Oct 2017
23. Watanabe, H., Fujimura, S., Nakadaira, A., Miyazaki, Y., Akutsu, A., Kishigami, J.J.: Blockchain contract: a complete consensus using blockchain. In: IEEE 4th Global Conference on Consumer Electronics (GCCE), pp. 577–578. IEEE (2015)
24. CB Insights: Banking is Only the Beginning: 30 Big Industries Blockchain Could Transform. https://www.cbinsights.com/research/industries-disrupted-blockchain/. Last accessed 14 Oct 2017
25. Stathakopoulous, C., Cachin, C.: Threshold Signatures for Blockchain Systems. Swiss Federal Institute of Technology—IBM Research, Zurich (2017)
26. Norton, J.: Blockchain: Easiest Ultimate Guide to Understanding Blockchain (2016)
27. Buterin, V.: The Meaning of Decentralization. https://medium.com/@VitalikButerin/the-meaning-of-decentralization-a0c92b76a274. Last accessed 09 Oct 2017
28. IBM: Ten Things Blockchain is not. https://www.ibm.com/blogs/insights-on-business/government/ten-things-blockchain-not/. Last accessed 21 Oct 2017
29. Hoffmann, J., Kumar, S.: International Seaborne Trade—Globalization, the Maritime Nexus. The Handbook of Maritime Economic and Business (2010)
30. Stopford, M.: Maritime Economics, 3rd edn. Routledge, London
31. Lloyds Register Marine. University College London. Maritime Fuel Trends. http://www.lr.org/en/_images/21334172_global_marine_fuel_trends_2030.pdf. Last accessed 15 Oct 2017
32. Stevens, L., Sys, C., Vanelslander, T., Van Hassel, E.: Is new emission legislation stimulating the implementation of sustainable and energy-efficient maritime technologies? Res. Transp. Bus. Manag. **17**(1), 14–25 (2015)
33. Den Norske Veritas—DNV GL: Alternative Fuels from Shipping. Strategic Research & Innovation Position—Paper 1 (2014)
34. Den Norske Veritas—DNV GL: Making your Asset Smarter with the Digital Twin. https://www.dnvgl.com/article/making-your-asset-smarter-with-the-digital-twin-63328. Last accessed 18 Oct 2017
35. The Economist. The data deluge, 25 Feb 2010. http://www.economist.com/node/15579717
36. Tien, J.M.: Internet of connected servgoods: considerations, consequences and concerns. J. Syst. Sci. Syst. Eng. **24**(2), 130–167 (2015)
37. Splash 24: Dr. Martin Stopford on the Future of Shipping. http://splash247.com/dr-martin-stopford-future-shipping/. Last accessed 18 Oct 2017

38. Elatab, M.: Five Trends in Oil & Gas Technology, and Why You Should Care. http://venturebeat.com/2012/03/28/5-trends-in-oil-gas-technologyand-why-you-should-care/. Last accessed 18 Oct 2017
39. Ball, B.: Reducing Global Logistics Cost with Benchmarking and Shipping Container Pricing Transparency. Aberdeen Group (2016)
40. Kondratenko, Y.P., Kozlov, O.V., Korobko, O.V., Topalov, A.M.: Internet of things approach for automation of the complex industrial systems. In: ICTERI-2017, CEUR Workshop Proceedings Open Access, vol. 1844, pp. 3–18 (2017). https://pdfs.semanticscholar.org/3ff6/4e4a07be1e8c2f0b16f4736397be1405218a.pdf
41. Kondratenko, Y.P., Kozlov, O., Gerasin, O., Topalov, A., Korobko, O.: Automation of control processes in specialized pyrolysis complexes based on web SCADA systems. In: Proceedings 2017 IEEE 9th International Conference on IDAACS, Bucharest, Romania, vol. 1, pp. 107–112, 21–23 Sept 2017. https://doi.org/10.1109/idaacs.2017.8095059
42. Tachizawa, E.M., Alvarez-Gil, M.J., Montes-Sancho, M.J.: How "smart cities" will change supply chain management. Supply Chain Manag. Int. J. **20**(3), 237–248 (2015)
43. Hahn, G.J., Packowski, J.: A perspective on applications of in-memory analytics in supply chain management. Decis. Support Syst. **76**(1), 45–52 (2015)
44. Lacey, M., Lisachuk, H., Giannopoulos, A., Ogura, A.: Shipping smarter: IoT opportunities in transport and logistics. The Internet of Things in Shipping. Dupress-Deloitte (2015)
45. Offshore Energy Today: Meet Onboard, the Maritime Internet of Things. http://www.offshoreenergytoday.com/oeec-meet-onboard-the-maritime-internet-of-things/?utm_source=emark&utm_medium=email&utm_campaign=daily-update-offshore-energy-today-2017-10-02&uid=207087. Last accessed 18 Oct 2017
46. Knect365 Maritime: Six Maritime Start-Ups That Are Changing the GAME. https://knect365.com/talentandtraining/article/1149354e-68d9-4e74-9f91-a900ac869526/6-maritime-startups-that-are-changing-the-game. Last accessed 19 Oct 2017
47. Dhanji, T.: Blockchain—Where Oil and Gas Traders Dare to Trade. Ernest Young Publications (2017). Retrieved from https://www.linkedin.com/pulse/blockchain-where-oil-gas-traders-dare-tread-talib-dhanji
48. UNICITRAL: Recognizing and Preventing Commercial Fraud. United Nations Commission on International Trade Law, NY (2013)
49. Solesvik, M.Z., Encheva, S.: Partner selection for interfirm collaboration in ship design. Ind. Manag. Data Syst. **110**(5), 701–717 (2010)
50. Encheva, S., Kondratenko, Y., Solesvik, M.Z., Tumin, S.: Decision support systems in logistics. In: AIP Conference Proceedings, vol. 1060, no. 1, pp. 254–256. AIP (2008). https://doi.org/10.1063/1.3037065
51. Werners, B., Kondratenko, Y.P.: Alternative fuzzy approaches for efficiently solving the capacitated vehicle routing problem in conditions of uncertain demands. In: Berger-Vachon, C., et al. (eds.) Complex Systems: Solutions and Challenges in Economics, Management and Engineering, Studies in Systems, Decision and Control, vol. 125, pp. 521–543. Springer International Publishing, Berlin, Heidelberg (2018). https://doi.org/10.1007/978-3-319-69989-9_31
52. Yli-Huumo, J., Ko, D., Choi, S., Park, S., Smolander, K.: Where is current research on blockchain technology? A systematic review. PLoS ONE **11**(10), 1–27 (2016)
53. Borch, O.J., Solesvik, M.Z.: Collaborative design of advanced vessel technology for offshore operations in Arctic waters. In: Luo, Y. (ed.) International Conference on Cooperative Design, Visualization and Engineering, pp. 157–160. Springer, Berlin, Heidelberg (2013)
54. Atamanyuk, I., Kondratenko, Y.P.: Computer's analysis method and reliability assessment of fault-tolerance operation of information systems. In: Batsakis, S., et al. (eds.) ICT in Education, Research and Industrial Applications: Integration, Harmonization and Knowledge Transfer, Proceedings of the 11th International Conference ICTERI-2015, CEUR-WS, vol. 1356, pp. 507–522, Lviv, Ukraine, 14–16 May 2015

Software Engineering Sustainability Education in Compliance with Industrial Standards and Green IT Concept

Igor Turkin and Yuliya Vykhodets

Abstract The chapter presents a scientific approach to building the course "Software Engineering Sustainability" to introduce the sustainable concept into educational programs for software engineers. The approach is based on the development and harmonization of the professional competence list with the application of formal mathematics and using graph theory along with open source visualization and exploration software for all kinds of graphs and networks. The proposed approach supports the flexible development of curricula, allowing each university to create programs, guided by their own specialties and demographics of their students. The authors demonstrate how they consider different factors to build a harmonized list of professional competencies for software engineers and to develop courses to support sustainability through the higher education system. This list of professional competences meets the requirements of employers, maintains the successful experience of national higher education and professional industry standards, and conforms to the European Union e-Competence Framework. The chapter contains review and analysis of the term "Software Engineering Sustainability" from different viewpoints. Authors provide description of elaborated and applied approach to building the course. They present the design of the course "Software Engineering Sustainability" with description of the course content and visible positioning Green IT. The course has been being delivered by the first author at National Aerospace University "Kharkiv Aviation Institute".

Keywords Course design · Curriculum · Sustainability · Green computing Software engineering · Descriptor · Competence · Standards · Education harmonization

I. Turkin · Y. Vykhodets (✉)
National Aerospace University «Kharkiv Aviation Institute»,
17 Chkalova Street, Kharkiv 61070, Ukraine
e-mail: yulialiya@gmail.com

I. Turkin
e-mail: i.turkin@khai.edu

© Springer Nature Switzerland AG 2019

V. Kharchenko et al. (eds.), *Green IT Engineering: Social, Business and Industrial Applications*, Studies in Systems, Decision and Control 171,
https://doi.org/10.1007/978-3-030-00253-4_25

1 Introduction

This chapter is devoted to the application of the sustainable concept in software engineering education. The ecosystem of the green computing industry and main principles of green computing for different applications has been well described by other authors in the previous volumes [1–3]. In this chapter the authors present the developed approach to building the course curriculum aimed to support Software Engineering (SE) sustainability through higher education. The developed course which has been delivered at our university is scientifically based: the attention was paid to building the list of professional competencies in the domain of SE sustainability to harmonize industrial requirements, educational standards, and societal needs to stimulate implementation and transition of green computing technologies through higher education.

In a recent article on this topic [4] the authors stated that the list of targeted professional competence must meet the requirements of employers; maintain a successful experience of national higher education and professional industry standards; conform to the European Union e-Competence Framework [5–8] and competence-based standard SWECOM [9].

In this chapter, we develop the approach and consider the State-of-the-Art in the domain of green computing and up-to-date standards of social responsibility for the business. Then we develop the above idea to build a course curriculum based on scientifically grounded harmonized lists of target competencies. Here we demonstrate the application of our approach for the case of introducing sustainable development into study programs for software engineers.

The main principles and concepts of the sustainable development in SE are considered further. It is shown that the concept of sustainable development must be considered from two different viewpoints. The first one is a view to the SE sustainability from the point of the software product itself and its ability to develop and evolve under the conditions of changing needs of stakeholders while changing is primarily driven by conversion in the business interests of these stakeholders. The second one is the sustainability of the software development processes, which, firstly, must ensure the creation of software products that correspond to the previous provision, and secondly, to guarantee the quality and efficiency of the software development processes properly. The analysis of inconsistencies between SE sustainable development concepts in professional and educational standards was performed with the following:

- European e-CF standard [5–8], common to the whole ICT industry in the European Union;
- Software Engineering Body of Knowledge (SWEBOK), which became the international standard ISO/IEC TR 19759: 2005 [10];
- Software Engineering Competency Model (SWECOM), which, unlike SWEBOK, describes the competences of software developers;
- Guidelines for teaching software engineering at universities (Curriculum Guidelines for Undergraduate (Graduate) Degree Programs in Software Engineering)—practical guidelines for organizing education in the domain of SE [11, 12].

2 Review of Professional Standards in the Domain of Software Engineering

General Standard for Information and Communications Technology (ICT) competences of the European Union is the e-Competence Framework (e-CF). European standard e-CF EN 16234-1 [5] was adopted in July 2016. It provides links to 40 ICT competencies demanded by employers [6–8]. There are five levels of competencies (skills, knowledge), relevant across Europe and overall ICT sector.

The overall structure of e-CF framework is four-dimensional. It is based on areas of competencies (descriptor 1) and competence (descriptor 2), and not on official duties. The competence-based approach is more flexible and personalized. Descriptor 3 is to describe the level to which qualifications may respond. The number of levels that may relate to the competency can vary depending on the content and range of the job activities. Descriptor 4 is a brief description of knowledge and skills, but it is not comprehensive. Definitions of competence, knowledge, skills, and attitudes have been worded in terms of organizations, not individuals. However, individual competencies can be included in descriptor 3, which ensures comparability and understanding of the connection between the concept of competence on the part of the organization and the individual. Determination of the descriptor 4 is consistent with the general definition of the knowledge, skills, and competencies in EQF (European Qualifications Framework, a meta-system intended to ensure the transparency and comparability in the recognition of qualifications, diplomas, certificates of education, to promote academic and labor mobility of citizens). There we see a generalized model, which describes information systems (IS) business—process, which includes the following phases of the IS life cycle: planning, implementation, commissioning, adjustment, and control. Stages of adaptation and management embrace all stages of the IS life cycle. This model has been applied for heuristic correlation of business processes with competences. To determine the appropriate level of competence the e-CF uses the following indicators EQF: “context complexity”, “independence”, and “behavior”. This made possible to agree on the e-CF educational levels with EQF.

Standard e-CF is common to the whole ICT sector in the European Union, but there are international models of knowledge, skills, and competencies specifically for professionals in the field of Software Engineering: Software Engineering Body of Knowledge (SWEBOK) and Software Engineering Competency Model (SWECOM). These documents are based on assumptions, that software development is a key skill for students in the chairs/departments that teach and graduate professionals in the field of SE. However, this skill is only the foundation for the entire field of SE because the term “software engineering” is interpreted as the use of technology for the design, development, implementation, testing, and maintenance of software in a systematic method [13].

Released SWEBOK, as a set of knowledge areas [14] is the international standard ISO/IEC TR 19759: 2005 [13]. In late 2013, SWEBOK V3 was approved for publication and published officially [14]. SWEBOK was created through collaboration between several professionals of the industry and supported by IEEE Computer Society. According to SWEBOK V3, the specialist in software engineering should have 15 major knowledge areas. SWECOM, unlike SWEBOK, describes capabilities of software developers. The knowledge and skills within the training field and employment process is an additional description of the required competencies. The model SWECOM is designed for managers, staff members and developers of programs [9].

SWEBOK and SWECOM both describe the requirements for ICT professionals who have practical experience as SWECOM is agreed with SWEBOK. The fundamental difference in setting up these requirements is: SWEBOK is knowledge-based but SWECOM is competency based. The second difference is that SWECOM subordinates to the model of IT competencies, which is implemented by the United States Department of Labor to determine the knowledge, skills, and abilities required for employees to work successfully in the SE.

The following guidelines for teaching SE at universities are the basis for the practical usage in SE (Curriculum Guidelines for Undergraduate [11, 12] and Graduate [15] Degree Programs in Software Engineering). They also have been applied to major existing standards and recommendations. Thorough elaboration of these recommendations varies with the level of education. In 2002, the specially created working group started developing the first version of recommendations for Bachelor of SE (level of higher education that meets the sixth level of qualification for the National Qualifications Framework Ukraine). The project was finished in 2004 and afterward, the second version of the recommendations was prepared. At the master's level, we have only one version of the recommendations dated 2009. Simultaneously, there is no guidance for the Ph.D. level. Recommendations for Bachelor of SE are thoughtful, logical and consistent guidance to practical educational decisions and actions. The author mentioned guidelines for creating and delivering master's programs which represents only one and not quite a successful attempt to establish effective standards.

GSwE2009 was developed as part of the Integrated Software and Systems Engineering (ISSEC) Curriculum Project, at Stevens Institute of Technology [16]. An underlying focus of GSwE2009 is how to advance the state of SE practice and to support better understanding and agreement about the nature of "professional software engineers". The key features of GSwE2009 include the following:

- A set of guiding principles for GSwE2009 development.
- A bottom line of student skills, knowledge, and experience is assumed by the core curriculum.
- A student's outcome is to guide the curriculum development.
- A description of the fundamental skills, knowledge, and experience are taught in the curriculum to achieve the outcomes, identified as the GSwE2009 Core Body of Knowledge (CBOK).

Below, on the example of Software Engineering, we'll describe an architectural framework, that supports flexible curriculum development by allowing each university to create programs guided by its own specialties and the demographics of its students.

3 Software Engineering Sustainability: Viewpoints and Interpretations

Sustainable living, i.e., living within the bounds of the available environmental, social, and economic resources, is the focus of many present-day social and scientific discussions [17]. Software engineering is one of the most demanding jobs in software industry nowadays and its demand keeps on increasing. A study of software systems effort distribution indicated that 55% of software system lifecycle effort involves sustainment activities [18]. The risks of inadequate sustainment can potentially undermine the stability, enhancement, and longevity of operational systems.

No authoritative definition of "software sustainment" exists. The Software Engineering Institute's working definition is, "The processes, procedures, people, material, and information required to support, maintain, and operate the software aspects of a system" [Lapham 2006].

The IEEE Standard Glossary of Software Engineering Terminology defies software maintenance as "The process of modifying a software system or component after delivery to correct faults, improve performance or other attributes, or adapt to a changed environment" [IEEE/ISO/IEC 24765-2010].

According to these definitions, software sustainment addresses issues of maintenance plus documentation, deployment, operation, security, configuration management, training (users and sustainment personnel), help desks, COTS product management, technology updates, and software retirement.

Three categories of software maintenance activities (corrective, adaptive, and perfective) involve modifying a software product/system after delivery. The fourth category of software maintenance activities focuses on preventive maintenance.

Successful software sustainment depends on the culture of the sustainment organization, the skills of the sustainment team (which is the focus of this skill area), the flexibility of the customer, and the operational domain of the product, in addition to skills needed to modify source code for corrective, adaptive, and perfective enhancements.

The industry claims that the SE graduates are not able to meet the software industry requirements. This phenomenon creates the gap between industry's requirements and graduates' readiness which indirectly reflects the gap between SE education and industry. Many studies and frameworks are available to ponder the issues on SE education and industry [19]. One of recent studies for the European Commission revealed that the main training/education barriers to the impact of newly emerging ICT trends that—in combination—drive economic transformation and growth include the lack of training/education in skills along with wrong skills and outdated education systems and educators.

3.1 Dimensions of the Sustainability

Sustainability is characterized by the three dimensions: economic, social, and environmental. This characterization is extended with the fourth dimension—human (or individual), portraying the individual development of every human over their lifetime. When analyzing the sustainability of IT systems, these four dimensions are applied, with an additional technical one, which supports the better structuring of concerns with respect to software systems. We are convinced that focus on environmental sustainability only makes sense when in balance with the other dimensions of sustainability.

More concisely, the dimensions of sustainability are characterized as follows [20]:

1. Individual sustainability refers to maintaining human fund (e.g., health, education, skills, knowledge, leadership, and access to services).
2. Social sustainability aims at preserving the societal communities in their solidarity and services.
3. Economic sustainability targets on maintaining major and added value.
4. Environmental sustainability refers to improving human welfare by protecting the natural resources: water, land, air, minerals and ecosystem services.
5. Technical sustainability relegates to the endurance of systems and infrastructure and their adequate evolution with changing surrounding conditions.

3.2 SE Sustainability

Referring to [21], "…in the context of sustainability, it is very important to explain and understand the different levels of effects software systems can have on sustainability. These can be classified as direct effects, enabling effects, and structural effects:

- Direct (or 'first order') effects are concerned with the immediate impacts resulting from the production, usage, and disposal of software systems. This can be measured using metrics based on performance requirements or network bandwidth for example. The direct environmental impact is quite often measured using the Life Cycle Assessment (LCA)…

- Enabling (or 'second order') effects are concerned with the benefits and impacts of the ongoing use in the software system. This might be for example how a web search engine reduces the cost of access to information.
- Structural (or 'third order ') effects are concerned with changes resulting from the use of software systems by a very large number of people over medium to long term. They are leading to substantial changes in societal structures such as new laws, politics, or social norms, or economic structures such as the networked economy."

Authors of [21] define Sustainable Software as "…software whose development, deployment, and usage results in minimal direct and indirect negative impacts or even positive impacts on the economy, society, human beings, and the environment" and propose the common criteria for sustainable software: "Sustainable Software arise from the well-known and standardized quality aspects for software, issued by the International Organization for Standardization. The proposed quality model takes such aspects into account as Efficiency, Reusability, Modifiability, and Usability. These quality aspects extend to the field of sustainable development… As for the quality aspects belonging to Efficiency are Runtime Efficiency, CPU Intensity, Memory Usage, Peripheral Intensity, Idleness, and Number of Methods. Runtime Efficiency considers the time needed to finish executing depending on its implementation… The aspects CPU-Intensity, Memory Usage, and Peripheral Intensity cover the resource utilization caused by software execution. In fact, the effects of the intensity of the resource consumption on the resulting energy consumption or even on system durability need to be analyzed. Idleness describes how often the system is idle. This aspect is relevant to certain types of software systems, such as virtual servers… The total Number of Methods reflects the size of applications…". Proposed in [21] Directly Related Criteria, "… In contrast to common efficiency criteria, which can be considered as indicators of energy efficiency …, the energy efficiency of a system itself can be ranked in a reference system… This means its goal is to optimize the energy consumption in relation to delivered service items. Thus, the quality criterion Energy Efficiency, belonging to Efficiency in general, is assigned to the directly related criteria… The Framework Entropy represents the grade of usage of external libraries and high-level constructs. The different types of applications with their different levels of energy efficiency are indicated by the criterion Functional Types…

Hardware Obsolescence … considers the amount of hardware that needs to be replaced before the sustainably worthwhile lifetime is reached, due to software with higher system requirements. This criterion goes together with the quality aspect Adaptability. Both aspects belong to the criterion Portability. Here, Adaptability describes the possibilities to adapt the software to the specific consumer needs.

The quality criterion Feasibility evaluates the effects of the product development, or rather the SE process, on sustainable development. In this context, Travel includes business trips that belong to a software project…The resulting Carbon

Footprint specifies the amount of CO_2 emissions caused by software development in the course of the software's life cycle… The aspect Energy Consumption indicates the consumed energy caused by software development. Waste counts the resources consumed during software development for activities that do not have any additional value for customers or end users… The aspect Infrastructure accounts for the conditions facilitating the usage of the software product.

From the social aspects perspective of sustainable development, there may be more qualities that require special measurements and assessment methods. As a first step, these aspects are called Accessibility, Usability and Organization Sustainability. Whereas Accessibility and Usability are well-known aspects in SE, the sustainability of the organization is usually not considered in common SE processes and software life cycle. Accessibility indicates whether the software product can be used by as many people as possible without any restrictions. Usability covers the comprehensibility, perspicuity, and operability of a software product, i.e. if the software usage is user-friendly and easily learned. The aspect Organization Sustainability covers the social situation in the software company, including working conditions, for example…".

Indirectly Related Criteria observed and discussed in [21] are "… indirectly related to the software product itself also need to be considered…Product Sustainability summarizes the effects of software on other products and services. It covers effects of usage, as well as systemic effects…The contribution of software within its application domain to reducing emissions, waste, and the like are indicated by the aspect Reduction. Beauty aims to value the software for its contribution to sustainable development… the quality Reflectivity refers to the quality aspect of efficiency. It specifies the manner in which the usage of software influences the resources that are indirectly needed by other persons or systems…".

Definition of non-functional requirements for "green" or sustainable software [22] includes the following:

- demand adaptivity;
- new versions do not make greater demands concerning memory, CPU performance, and bandwidth unless absolutely necessary for additional functions;
- basic functions can still be executed on older hardware in the long term;
- user-oriented configuration options for energy-saving modes;
- power awareness, optimum management of hardware concerning energy consumption; server software should also take the energy used by the client into account and should in no case hinder or discourage turning off end-user devices or local power management;
- "Power-down-friendliness": software should not incite people to leave hardware turned on all the time;
- support for data formats that are economical in terms of bandwidth and memory;
- support for open standards for data formats (no customer lock-in via formats);
- flexibility in terms of useable peripheral equipment (minimizing requirements to purchase new equipment);
- undesired advertising can be turned off.

3.3 Sustainability of SE Processes

A prerequisite to sustainable software is a persistent development process, which refers to the consideration of environmental and other impacts during the software life cycle and the pursuit of the goals of sustainable development [21]. Thus, authors of [21] define Sustainable SE "…as the art of developing sustainable software through a sustainable software engineering process… software products must be defined and developed in such a way that the negative and positive impacts on sustainable development that result or are expected to result from the software product over its whole life cycle are continuously assessed, documented, and used for further optimization of the software product".

For information: authors of [21] developed the reference model GREENSOFT for green software engineering (https://doi.org/10.1016/j.suscom.2011.06.004).

As it is noticed in [21], "Lami et al. point out that processes regarding sustainability are missing in ISO/IEC 12207. Therefore, they define a set of three complementary sustainability processes: Sustainability Management Process, Sustainability Engineering Process, and Sustainability Qualification Process. These additional reference processes aim at continuously improving the sustainability of both software products and software processes".

Citing [23], "…Green Software Development (GSD) is a methodical process which allows a systematic, disciplined and well-organized development of green software products".

3.4 Educational Aspect of Green IT Engineering

"Sustainability is not supported by traditional SE methods. Traditional SE has not fully supported sustainability as a relevant, first-class concern. Software engineers approach specific topics that have to do with sustainability in our discipline, for example, green IT, efficient algorithms, smart grids, agile practices and knowledge management, but we lack a common understanding of the concept of sustainability and if and how it can be encompassingly applied to software engineering" [24].

According to e-CF, A.8 competence impacts ICT solutions in terms of eco responsibilities including energy consumption.

Level 1: Advising business and ICT stakeholders on sustainable alternatives that are consistent with the business strategy.
Level 2: Applying an ICT purchasing and sales policy which fulfills eco-responsibilities.
Level 3: Promoting awareness, training, and commitment for the deployment of sustainable development and using the necessary tools for piloting this approach.

Level 4: Defining objective and strategy of sustainable IS development in accordance with the organization's sustainability policy.

Examples of knowledge: metrics and indicators related to sustainable development, corporate social responsibility (CSR) of stakeholders within the IT infrastructure.

Examples of skills: monitor and measure the ICT energy consumption, apply recommendations in projects to support the latest sustainable development strategies, master regulatory constraints and international standards related to ICT sustainability. [25] However, the most recent releases of the IEEE/ACM Curriculum Guidelines for Undergraduate Degree Programs in Software Engineering [26] or Computer Science [27] do not provide any indication to those topics. Nowadays there are not a lot of educational programs of BSc, MSc and Ph.D. level in green IT in Ukraine and other countries. There is a big need to fill in this gap [28]. A survey of 3860 software engineering practitioners working at IBM, Google, ABB, and Microsoft, shows that the current higher education curriculum does not prepare them to tackle sustainability, although these practitioners were willing to learn about sustainability [29, 30].

Sustainability within and through SE needs is supposed to be addressed better in near future. Nonetheless, sustainability and energy efficiency are regarded as key competencies for future software engineers and developers.

A few works exist that propose frameworks and guidelines on how to deliver content on sustainability in traditional education. For example, Mann et al. [30] provide a framework meant for educators to design modules/programs addressing sustainability. It classifies sustainability-focused education approaches in three types: centralized (i.e. concentrate sustainability topics in one or two focused courses), distributed (i.e. address sustainability topics across all the courses in the curriculum), and blended (i.e. a mix of the previous approaches). Distributed and blended approaches seem the most suitable for "infusing" sustainability principles in the computing community and ensuring that every aspect of computing contributes to the implementation of sustainable practices. Such approaches, however, require adapting many different courses and therefore necessitate the active contribution and participation of a wide range of educators within the institution. Centralized approaches, although likely to be less effective, may be easier to implement within an institution that is first moving to include sustainability in the computing curriculum and maybe the instrument for the establishment of a leadership that will guide the implementation of the distributed approach [31]. Referring to [26]: the quote by Dan Ariely regarding Big Data and teenage sex could aptly be applied to software sustainability: "Everyone is talking about it, nobody really knows how to do it, everyone thinks everyone else is doing it."

4 Proposed Solution

4.1 Research Methodology for Defining Target SE Competencies

In this section, we refer and cite the previously developed general approach [4] to drawing up an adapted list of professional competencies which should become the basis for elaborating courses, syllabi, and curricula. Sections 4.2 and 4.3 demonstrate its application for the case of sustainable development.

The approach to the selection and justification of targeted professional competence in the field of "Software Engineering" allows meeting the requirements of employers, using a successful experience of national higher education (expressed in professional industry standards) and EU compliant e-Competence Framework. The task of choosing and justification is solved mathematically using graph theory [4].

Input assumptions and limitations:

- target competencies for training specialists in SE should comply as fully e-CF, and be as close to the current industry standards for higher education in Ukraine;
- target competencies for training specialists in SE must provide training for profiles "developer" and "test specialist" because these profiles fully meet the qualifications of Degree "Specialist in the development and testing of software" by National classification of professions Ukraine. Since the profile "Digital media specialist" is almost identical for the necessary competencies to profiles "developer" and "specialist test", we assume that these three profiles will form the core for the specialty "Software engineering" [32];
- the core profiles are not enough to determine the competencies targeted training for the SE profession under the three-tier higher education system of Ukraine, as according to e-CF and EQF these profiles do not require the level of competence Ph.D.

The formal statement of the problem. The initial data is the interrelation of the "profile"—"competence" set in the e-CF in tabular form. In the mathematical representation, this table defines the mapping of the 23 profiles $P = \{P_1, P_2, \ldots P_{23}\}$ in the set of 40 competences $C = \{C_1, C_2, \ldots C_{40}\}$.

$$\phi : P \rightarrow C. \tag{1}$$

According to this table for e-CF display, each profile corresponds from three to five competencies:

$$\forall P_i \in P, 3 \leq \left\| \langle P_i, C_j \rangle \right|, C_j \in C \| \leq 5. \tag{2}$$

We need to define this subset of target profiles $P_{aim} \subseteq P$, for which the power of the set of target competencies $\|C_{aim}\|, C_{aim} \subseteq C$ will meet the guidelines for developing higher education standards approved by the Scientific and Methodological

Council of the Ministry of Education and Science of Ukraine [33]. "The estimated number of special competences usually does not exceed 10–20 competences considering the level of education." An additional limitation is a requirement for mandatory including profiles "developer", "test specialist" and "digital media specialist" to the target subset of profiles P_{aim}.

Given constraints, the task of identifying the target profile subset P_{aim} is optimization:

$$P^{*}_{aim} = \arg\max(\|P_{aim}\|), \tag{3}$$

with constraints:

$$P_{core} = \left\{ \begin{array}{c} \text{“Developer”}, \\ \text{“Test specialist”}, \\ \text{“Digital media specialist”} \end{array} \right\} \subseteq P_{aim}, \tag{4}$$

$$\|\phi(P_{aim})\| \leq 20. \tag{5}$$

Step 1. Interpretation of the problem in terms of graph theory. In the above formulation the problem (3–5), it is solvable, and the subset of target profiles exists. That means that a solution can be found in a limited time range, as this formulation provides a consistent analysis only of $2^{\|P\|-\|P_{core}\|} = 2^{20}$ variants by the method of exhaustive search. As a result, there will be received not only one solution, but a set of subsets of target profiles, which will satisfy both the objective function (3) and limitation (4, 5). In other words, the formal solution is not enough, we need to make meaningful analysis and justify one set of target profiles.

We take into consideration the weighted undirected graph profiles:

$$G = \langle P, V \rangle, \tag{6}$$

when each edge of the graph $v_{i,j} \in V$ has a corresponding value (weight of the edge) $w_{i,j}$, which is calculated as the number of shared competencies of i-th and j-th profiles.

Step 2. Selection and justification of the method of clustering profiles. By this, we mean the process of ordering structures into relatively homogeneous groups. A common feature of these groups is a structural proximity between profiles within the group (strong link inside) and weak links among groups. One of the most common metrics to calculate proximity, with the aim of clustering in graph theory is modularity. Modularity is calculated as the proportion of edges that are within the specified groups, excluding edges that would be used within specified groups in random distribution. For the proposed weighted graph (6), we calculate modularity considering the weight of edges:

$$Q = \sum_{i} (e_{ii} - a_i), \tag{7}$$

where e_{ii}—the relative sum of weights of all edges connecting profiles within the i-th cluster, calculated as the ratio of the sum of the weights of edges in the cluster to the sum weights of all edges vertices i-th cluster;

$a_i = \sum_{j,j \neq i} (e_{ij})$—the relative sum of weights of all edges connecting cluster c_i with other clusters.

Modularity metric performs calculation of the degree of “similarity” between any two clusters A and B according to the Jaccard similarity coefficient:

$$k = \frac{\|A \cap B\|}{\|A \cup B\|} = \frac{\|A \cap B\|}{\|A\| + \|B\| - \|A \cap B\|}. \tag{8}$$

Modularity value is within the range [−1, 1]. Modularity is positive if the sum of the weights of edges within groups exceeds than their sum a random distribution of vertices by groups. In other words, modularity reflects the concentration of edges’ weights within modules compared to the random distribution of the weights between all competences regardless of their competencies into clusters. It is known from the research of social networks communities that if the modularity value exceeds 0.3, it is a sign of existing community in the network.

The estimated value of modularity for the graph of profiles proximity by e-CF is 0.423, from which we can conclude about the validity of the hypothesis of the existence of groups of profiles with close requirements to the competencies.

Step 3. Clustering of the graph into relatively homogeneous groups with the help of Gephi [34]. It has a multi interface and a rich set of instruments for networks visualization. Gephi was written in the programming language Java and OpenGL project based on the modular NetBeans platform. This program is one of the most popular and allows importing data directly from a database MySQL. It supports extensions for different algorithms and clustering, gives statistics of proximity, development over time and in 3-dimensional space, allows users to update data via API (Application Program Interface) mode flow. System architecture and open source involve the addition of new features through the development of plug-ins.

To perform clustering a graph into relatively homogeneous groups Gephi uses the method of medoids when only one of the available vertices may act as the center of the cluster. Therefore, with using this system we may build a graph of profiles proximity according to the e-Competence Framework (e-CF) version 3.0 (edge thickness is proportional to the number of common competencies for top-profiles).

Experimental visualization (Fig. 1) has provided a clear argument that among e-CF profiles it is necessary to distinguish just 3 profiles groups which are closely related by the list of required competence (unlike to 6 families’ profiles correlated

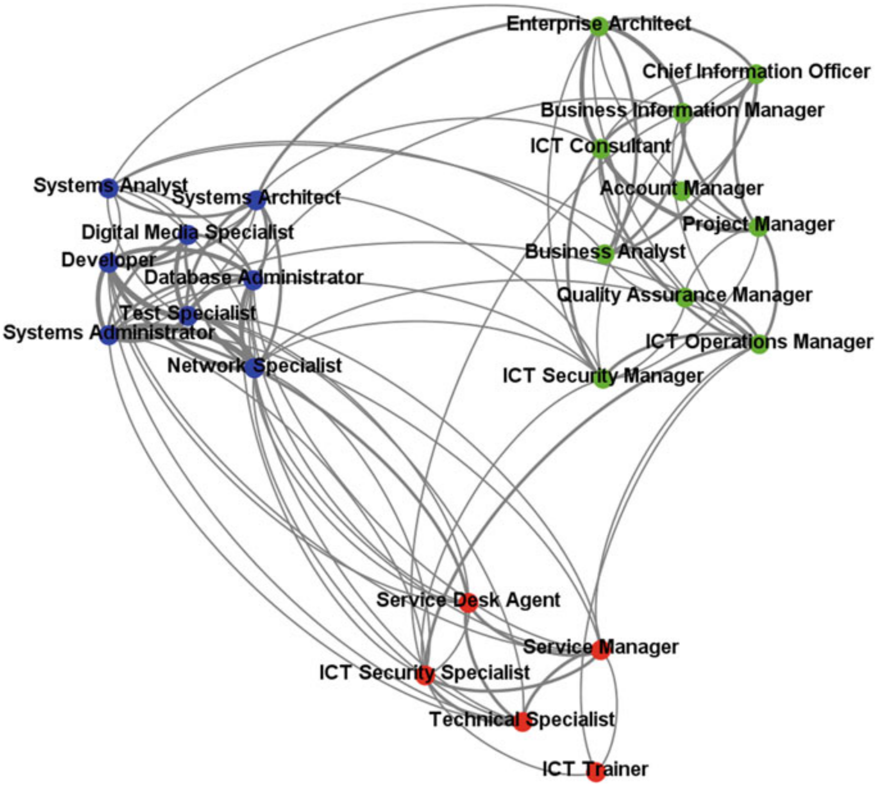


Fig. 1 Target competencies for SE

with business processes for e-CF). One of these three groups (blue) is fully consistent to the formulated optimization problem (3–5) to determine the target profiles cluster, which contains profiles “Developer”, “Test specialist” and “Digital Media Specialist”.

Table 1 shows the correspondences between European ICT Profiles titles, ICT Profiles Summary statements and alternative titles that may be used.

The proposed methodology for defining target SE profiles and competencies was approved at a meeting of the scientific and methodological commission of the Ministry of Education and Science of Ukraine.

The recommendations presented above, including a list of target training profiles and required competencies, were used in developing the educational standard of Ukraine for master’s degree programs in software engineering [35].

Table 1 Profile titles for SE with summary statements and examples of alternative titles used for similar profiles

European ICT profile title	ICT profile summary statement	Alternative titles that may be used
Database Administrator	Designs and implements; or monitors and maintains databases	Database Developer (Microsoft), Database Manager (Eucip), Network Administrator (ACS)
Developer	Builds/codes ICT solutions and specifies ICT products according to the customer needs	Component Developer (AITTS), Application Developer (ITA-J), Programmer (IBM)
Digital Media Specialist	Creates websites and multimedia applications combining the power of digital technology with effective use of graphics, audio, photographic and video images	Front-End Web Developer (IWA), User Experience Designer (IWA), Web and Multimedia Master (Eucip), Web Content Manager (UK-Gov, IWA), Web Developer (Bring-IT-On, Microsoft, UK-Gov), Web Editor (UK-Gov), Digital Media Developer (AITTS), Multimedia Designer (Bring-IT-On), Multimedia Developer (ACS)
Network Specialist	Ensures the alignment of the network, including telecommunication and/or computer infrastructure to meet the organization's communication needs	Network Engineer (Bring-IT-On, UK Gov), Network Manager (Eucip, UK Gov), Network Services Specialist (ITA-J), Network Support (ACS), Network Administrator (ACS)
Systems Administrator	Administers ICT System components to meet service requirements	Network Administrator (ACS), Server Administrator (Microsoft), System Administrator (SME), Database Administrator (Microsoft), Enterprise Administrator (Microsoft), Enterprise Messaging Administrator, (Microsoft), Web Server Administrator (IWA)
Systems Analyst	Analyses requirements and specifies software and systems	Information Scientist (UK-Gov), Information Systems Analyst (Eucip, ACS)
Systems Architect	Plans and is accountable for the implementation and integration of software and/or ICT systems	Telecommunications Architect (Eucip)
Test Specialist	Designs and performs testing plans	Computer Games Tester (UK-Gov), Software Tester (SME), Systems Integration and Testing Engineer (Eucip), Test Specialist (ITA-J), Tester (AITTS)

4.2 Defining Correspondences Between SE Profiles and E-CF Competence A.8 "Sustainable Development"

In the current version of the e-CF competence model [33, 36] the e-competence A.8 Sustainable Development is not assigned to any of the 23 ICT Profiles, because "E-competences A.8. could be included in a number of Profiles but not at this generic European level where the initial sense of Profiles specification is to provide basic characteristics and to make clear differences between them".

A.8 e-CF competence "Sustainable Development" associated with the EUCIP (The European Certification Model for ICT Professionals) Profiles "Data Centre and Configuration Manager", "IT Systems Architect" [33]. To find correspondences between A.8 e-CF competence and the target ones let's consider the Table 2.

The above Matrix provides a one-sight overview of which Deliverables have been assigned to each Profile in terms of Accountable (A), Responsible (R) or Contributor (C): To be Accountable (A) is to be the only "owner" of the work. The owner must sign off or approve when the task, objective or decision is complete. He/she must make sure that responsibilities are assigned for all related activities. There is only one owner accountable for each deliverable. Responsible (R)—The "Doers" of the work are responsible for the work. They must fulfill the task or objective or make the decision. Contributors (C) provide input before work can be completed and signed-off on. They are "in the loop" of active participants. Several people can be contributors to one deliverable. The resulting graph (Fig. 2) links profiles and competencies, including A.8. The thickness of the lines in the figure is proportional to the number of deliverables in Table 2.

4.3 Software Engineering Sustainability: The Overall Course Design and Positioning Green IT

The course "Software Development Sustainability" has been created for master's degree program "Software Engineering" being delivered at National Aerospace University "Kharkiv Aviation Institute". The subject of the course study is bifocal (Fig. 3):

1. Methods, models and tools that allow you to explore the software and its environment in the context of the concept of sustainable human development.
2. Influence of SE on the steady development of mankind.

The course requires knowledge of the courses "Formal software analysis", "Cloud computing", "Programming of mobile devices".

Course content modules (Fig. 3) are the following.

Module 1. Software Ecosystem Analysis Methods (related to e-CF competence A.8. Sustainable Development). Sustainability for Software Engineering.

Topic 1. Introduction. The main features, framework, and terminology.

Table 2 Correspondences between deliverables and profiles in terms of roles (Accountable (A), Responsible (R) or Contributor (C)) along with main sustainability areas assigned to each deliverable

N	Deliverables	ICT profiles								Sustainability areas	
		Database administrator	Developer	Digital media specialist	Network specialist	Systems administrator	Systems analyst	Systems architect	Test specialist	Software	Processes
1	Data model	A								+	
2	Development process							C			+
3	Hardware component		A								
4	ICT model						C			+	
5	ICT process definition						C				+
6	Integrated solution						R	A	C	+	
7	Multimedia component			A						+	
8	New technology integration proposal							R		+	+
9	Software component		A							+	
10	Solution documentation		R						C		
	Network				R						
11	Solution in operation			C		R					
	Network				R						
	Systems	R	C								
12	Solution specification	R					C	A			
	Network				R						
13	Solved incident	C			C	C					+
14	SW design description		C								
15	SW needs assessment						A			+	+

(continued)

Table 2 (continued)

N	Deliverables	ICT profiles								Sustainability areas	
		Database administrator	Developer	Digital media specialist	Network specialist	Systems administrator	Systems analyst	Systems architect	Test specialist	Software	Processes
16	Technical proposal						R			+	+
17	Test plan								R	+	
18	Test procedures	C	C						R	+	
19	Test result								R		
20	Validated solution								C		

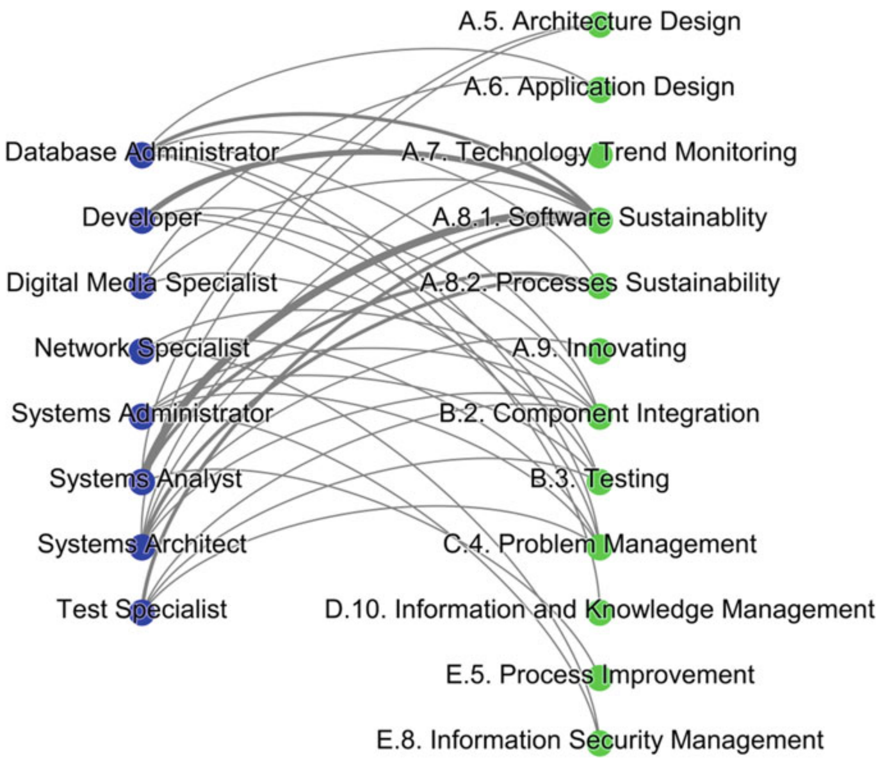


Fig. 2 Graph competencies, assigned to target profiles

Topic 2. Classification of software ecosystems.

Topic 3. Modeling and simulation of software ecosystems.

Topic 4. Forecasting the evolution of software ecosystems.

Module 2. Innovative IT Ecosystems (related to e-CF competences A7. Technology Trend Monitoring and A.9. Innovating). Software Engineering for Sustainability.

Topic 5. The ecosystem of Internet of Things.

Topic 6. Big data ecosystem.

Topic 7. The ecosystem of augmented reality.

Module 3. Methods and tools for building energy-efficient software (A.8. Sustainable software). Software Engineering for Sustainability.

Topic 8. Methods and means of the energy-saving use of computing resources.

Topic 9. Methods and means of energy profiling of software applications.

Topic 10. Innovations and prospects of development of energy-saving methods and means.

The course objectives and tasks. The learning objective for students is to acquire professional and personal competencies that will enable them to determine the goal and strategy of sustainable software development in accordance with the

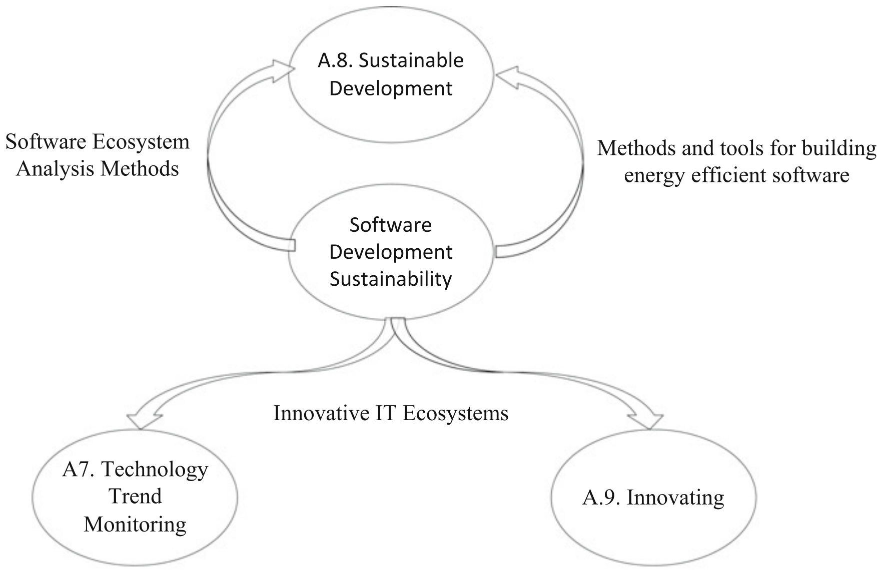


Fig. 3 Graph competencies, assigned to target profiles

organization's policy in the field of sustainable development, as well as to apply independent thinking and technological awareness to integrate disparate concepts into providing unique solutions.

The main tasks of the course study are the following: to learn to understand the principles of applying the scientific basis of ecology to solving the problems of SE, as well as to provide basic knowledge about modern innovations in the field of information and communication technologies and SE; to teach students to understand the principles of influence on the environment of processes for creating and maintaining software and the need for energy consumption and minimization of computational resources.

Target knowledge and skills. It is expected that as students complete the course successfully they will *know* the following:

- the nature of the ecological crisis; sustainable development as the only way out of the crisis;
- essence of "green" information technologies, "green" software, digital ecosystems, software ecosystems; constituents of the concept of sustainable development—economic, social, ecological and unity of components;
- the connection of ecological component of sustainable development with information systems, technologies, and software.

Students will *understand* the following:

- the role of software in solving the problems of sustainable development;

- value of the characteristics and software metrics as an object of research on the ecological approach.

Students ***will be able***:

- to receive and use knowledge about the characteristics of the software
- to apply an ecological approach to the research, creation and application of information systems, technologies and software;
- to model ecosystems of software in quantitative and qualitative aspects;
- to use means of solving problems of ensuring the environmental friendliness of the processes of creation and application of software (server-blades, zero clients, mini-frames, virtual machines, cloud services, green information systems, Grid-calculations, platforms for electronic document circulation);
- to perform energy profiling of algorithms and software.

The course content (4 ECTS credits).

Module 1. Methods of analysis of software ecosystems.

Topic 1. Introduction. Framework, terminology. Sustainable development as the only way out of the crisis. Environmental principles. An ecological approach to the research, creation and application of green information systems, technologies and software. Green Computing, Information Systems, Technology, Software.

Topic 2. Classification of software ecosystems. Principles of classification of software ecosystems. Metamodel of software ecosystems. Ways to present a model of software ecosystems. Models of software ecosystems in terms of business, software, innovation. Models of building business ecosystems based on platforms, innovations, commerce, things and interests.

Topic 3. Modeling of Software Ecosystems. Aspects of the study of the software ecosystem. The task of the study of the software ecosystem. Problems of modeling in software ecology. Tools for modeling software ecosystems. Tasks to explore ecosystems

Topic 4. Predicting the evolution of software ecosystems. Monitoring ecosystems software. Models, methods, and means of forecasting the evolution of software ecosystems.

Module 2. Innovative IT-ecosystems

Topic 5. The ecosystem of the Internet of things (IoT). IoT and its applications. Interaction of the Internet with existing networks and networks of the future. Wireless Touch Networks. Flying Networks. Tactile Internet. Medical networks. Nanonetworks. Software-defined networks (SDNs) and their interconnection with the IoT. Technology 5G and ultra-dense networks. Swarm intelligence. Modern network protocols and architectures. Cloud computing for IoT applications.

Topic 6. Ecosystem of Big Data. Technologies of big data storage and processing. Statistical methods for data analysis. Modern software tools for analyzing large volumes of information. Collection and storage of big data. Methods of big data processing and analysis. Visualization of input information and analytical data. Architecture and components of the Oracle NoSQL database. Comparison of NoSQL and RDBMS. Cloudera's Distribution Overview. Hadoop (distributed file system). Map Reduce model and data replication technology. OracleLoader for Hadoop (OLH).

Topic 7. Ecosystem of augmented reality. Methods of generating, examples of implementations (Google Glass), usage problems.

Module 3. Methods and tools for energy-efficient software engineering.

Topic 8. Methods and means of energy-saving use of computing resources. Methods of solving problems of ensuring the environmental friendliness of the processes of creating and applying software. Criteria for assessing "green" information systems, technologies and software. Methods, principles and tools for creating green systems, technologies and software.

Topic 9. Methods and tools for power profiling of software applications. Tools for power profiling applications: PowetTop, Joulemeter, Intel Power Gadget, Intel VTune Amplifier, XCode, Power Profiler.

Topic 10. Innovations and prospects for the development of energy saving computational methods and tools. Approximate and probabilistic calculations.

5 Conclusions

Green computing is a part of sustainable development values, activities and technologies being a comparatively new approach through which IT-industry supports sustainable development [37]. It is widely known that a framework of stakeholders' interests defines or affects how the new technology will develop. Despite clearly visible advantages of green computing technologies, there are some barriers to their implementation on different levels.

First, while business world seeks opportunities for economic growth and effective usage of existing production facilities, new green computing technologies may narrow some existing markets and question acceptability of widely used and well-promoted technologies. This creates a certain resistance of industrial sector for implementation and dissemination of GC technologies and we observe a conflict of interests between society and business. Thus, there is a need for strategies of solving this conflict. The certain potential in solving this conflict can be found in promoting green computing among consumers and transferring its values through the higher education system.

Second, knowledge of software engineers and consumers about green computing is under question, as well as their priorities for choice. Partly, the reason is that industrial requirements to professional competences in SE domain and educational standards are still not corresponded. Another reason is a low level of consumers' awareness about possible criteria for comparing different products and making choice.

The following barriers for implementation are considered: absence of educational standards which respond to industry requirements, best practices of national education, state-of-the-art in the domain of green computing technologies and best practices or standards for social responsibility of business. The authors demonstrate how all these factors may be considered to build a corresponded list of professional competencies for software engineers to support sustainable development through the higher education system.

The course "Software Sustainability" for the Master program "Software" is based on the qualification profile that clearly defines the goals and objectives of the course.

The qualification profile is based on professional competencies. In our opinion, this approach guarantees "compliance of objectives" when the selected objectives of the course are reasonable and consider the expectations of stakeholders (students, teachers, employers, as well as broader requirements of the Bologna process). The relevance of the course is confirmed during a broad consultation process, which involves leading scientists, representatives of higher education, and professional organizations. Learning outcomes follow Higher Education Standard of Ukraine (Second (master) level for Software engineering).

The second criterion of the course quality according to the TUNING project concept is a guaranteed "compliance with goals", i.e. the ability to achieve stated goals through the course. First, this possibility is determined by available resources, while the quality of resources directly affects the quality of the program. In this matter, our case can be described as follows. Department of Software Engineering is a part of Faculty of Software Engineering and Business of the National Aerospace University. The described course was elaborated there 3 years ago and introduced into master's degree program "Software Engineering". The first author is the lead lecturer of this course. Academic staff of the department includes 4 professors and 18 associate professors, among which 2 teachers are laureates of the State Prizes of Ukraine in the field of science and technology. Every year the course is delivered to 50–60 students, about 20% of them select thematic questions of the course as the topic for the future master's thesis.

During these three years, discipline evolved with expanding and adding new ideas. Today's presentation of course "Software Engineering Sustainability" is based on three intermediate milestones achieved during this time. These milestones include prepared and read courses "Energy profiling software," "Software Ecosystems", "Software Ecology". It can be argued that "Software Engineering Sustainability" is the fourth round of the development of an environmental approach to training software engineering masters at the National Aerospace University.

As for higher education in Ukraine as a whole, the pioneer of the environmental direction in Software Engineering is Professor M. Sidorov, who was the first in Ukraine to introduce the course "Software Ecology" in the master's program [38, 39] in the National Aviation on university. We are grateful to Prof. M. Sidorov for his ideas.

References

1. Kharchenko, V., Kondratenko, Y., Kacprzyk J. (eds.): Green IT Engineering: Concepts, Models, Complex Systems Architectures. Studies in Systems, Decision and Control, vol. 74. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-44162-7
2. Kharchenko, V., Illiashenko, O.: Concepts of green IT engineering: taxonomy, principles and implementation. In: Kharchenko, V., Kondratenko Y., Kacprzyk J. (eds.) Green IT Engineering: Concepts, Models, Complex Systems Architectures. Studies in Systems, Decision and Control, vol. 74, pp. 3–19. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-44162-7_1
3. Kharchenko, V., Kondratenko, Y., Kacprzyk, J. (eds.): Green IT Engineering: Components, Networks and Systems Implementation. Studies in Systems, Decision and Control, vol. 105. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-55595-9
4. Turkin, I., Narozhny, V., Vykhodets, Y.: The approach to drawing up a list of professional competences in the field of software engineering for harmonization of educational standards. In: The Proceedings of International Scientific Conference "Information Technologies and Computer Modelling", 15th–20th of May 2017, Ivano-Frankivsk, pp. 73–77. http://itcm.comp-sc.if.ua/2017/Turkin.pdf
5. EN 16234-1: e-Competence Framework (e-CF)—A common European Framework for ICT Professionals in All Industry Sectors—Part 1: Framework. Available: https://standards.cen.eu/dyn/www/f?p=204:110:0::::FSP_PROJECT,FSP_ORG_ID:41798, 1218399&cs = 17B0E0F8CABCDBDDB8066A46FA937510B
6. A common European framework for ICT. Available: http://www.ecompetences.eu/
7. European e-Competence Framework. Available: http://www.ecompetences.eu/e-cf-overview/
8. ICT profile European e-Competence Framework. Available: http://wp1087322.server-he.de/
9. Software Engineering Competency Model (SWECOM). Available: https://www.computer.org/web/peb/swecom
10. Software Engineering Body of Knowledge. Available: https://en.wikipedia.org/wiki/Software_Engineering_Body_of_Knowledge
11. SE 2004—Curriculum Guidelines for Undergraduate Degree Programs in Software Engineering.—2004. Available: http://sites.computer.org/ccse
12. SE 2014—Curriculum Guidelines for Undergraduate Degree Programs in Software Engineering. Available: http://www.acm.org/binaries/content/assets/education/se2014.pdf
13. ISO/IEC TR 19759: 2005 Software Engineering—Guide to the Software Engineering Body of Knowledge (SWEBOK). Available: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=33897
14. Board of Governors Approves New Version of Guide to the Software Engineering Body of Knowledge. Available: https://www.computer.org/web/pressroom/board-of-governors-approves-new-version-of-guide-to-the-software-engineering-body-of-knowledge
15. GSwE2009—Graduate Software Engineering 2009. Association for Computing Machinery and Computer Society of IEEE. Available: http://www.acm.org/binaries/content/assets/education/gsew2009.pdf
16. Hilburn, T.: Graduate Curricula in Software Engineering and Software Assurance: Need and Recommendations. In: Hilburn, T., Kornecki, A. (eds.) Electronics and Electrical Engineering. vol. 6, pp. 67–71 (2010)
17. Chitchyan, R.: What can software engineering do for sustainability: case of soft-ware product lines. In: Chitchyan, R., Noppen, J., Groher, I. (eds.) 15 Proceedings of the Fifth International Workshop on Product LinE Approaches in Software Engineering, pp. 11–14 (2015)
18. Software Engineering Competency Model Version 1.0 SWECOM
19. McLaughlin, S., Sherry, M., Doherty, E., et al.: e-SKILLS: The International Dimension and the Impact of Globalization Final Report, p. 684. European Commission (2014)
20. Penzenstadler, B.: Infusing Green: Requirements Engineering for Green in and Through Software Systems. Available: ceur-ws.org/Vol-1216/paper8.pdf

21. Naumann, S., Kern, E., Dick, M., Johann, T.: Sustainable software engineering: process and quality models, life cycle, and social aspects. In: Hilty, L.M., Aebischer, B. (eds.) ICT Innovations for Sustainability, Advances in Intelligent Systems and Computing, vol. 310, pp. 191–205
22. Hilty, L., Lohmann, W., Behrendt, S., Evers-Wölk, M., Fichter, K., Hintemann, R.: Green Software: Analysis of Potentials for Optimizing Software Development and Deployment for Resource Conservation: Establishing and Exploiting Potentials for Environmental Protection in Information and Communication Technology (Green IT) (2015). Available: https://doi.org/10.5167/uzh-110819
23. Abdullah, R., Abdullah, S., Din, J., Tee, M. (eds.) A systematic literature review of green software development in collaborative knowledge management environment a systematic literature review of green software development in collaborative knowledge management environment. Int. J. Adv. Comput. Technol. (IJACT) **4**(1), 63–80. http://ijact.org/volume4issue1/IJ0410057.pdf
24. Penzenstadler, B.: Towards a definition of sustainability in and for software engineering. In: Penzenstadler, B. (ed.) Proceeding SAC'13 Proceedings of the 28th Annual ACM Symposium on Applied Computing, pp. 1183–1185 (2013)
25. PERCCOM: A Master Program in Pervasive Computing and Communications for Sustainable Development
26. Sammalisto, K.: Integration of sustainability in higher education: a study with international perspectives. In: Sammalisto, K., Lindhqvist, T. (eds.) Innovative Higher Education, vol. 32/4. pp. 127–139 (2008)
27. Venters, C.C., Jay, C., Lau, L.M., et al.: Software sustainability: the modern tower of babel. In: Position Paper: RE4SuSy: Third International Workshop on Requirements Engineering for Sustainable Systems (2016). Available: http://ceur-ws.org/Vol-1216/paper2.pdf
28. Porras, J., Seffah, A., Rondeau, E., Andersson, K., Klimova, A.: 2016 IEEE 29th Conference on Software Engineering Education and Training, pp. 204–212 (2016)
29. Torre, D., Procaccianti, G., Fucci, D., et al.: On the presence of green and sustainable software engineering in higher education curricula. In: 1st International Workshop on Software Engineering Curricula for Millennials (SECM2017) (2017). Available: https://arxiv.org/pdf/1703.01078v1.pdf
30. Manotas, I., Bird, C., Zhang, R., et al.: An Empirical Study of Practitioners' Perspectives on Green Software Engineering (2016). Available: http://cabird.com/pubs/manotas2016green.pdf
31. Mann, S., Muller, L., Davis, J., et al.: Computing and Sustainability: Evaluating Resources for Educators. In: ITiCSE'09, pp. 144–156 (2009)
32. European ICT Professional Profiles, p. 5. ftp://ftp.cen.eu/CEN/Sectors/List/ICT/ CWAs/CWA %2016458.pdf. Accessed 22 Sep 2017
33. European e-Competence Framework 3.0. A Common European Framework for ICT Professionals in All Industry Sectors. CWA 16234:2014 Part 1 (2014). http://ecompetences.eu/wp-content/uploads/2014/02/European-e-Competence-Framework-3.0_CEN_CWA_16234-1_2014.pdf
34. The Open Graph Viz Platform. https://gephi.org/
35. Higher Education Standard of Ukraine. Second (Master) Level for Software Engineering. https://mon.gov.ua/storage/app/media/vyshcha/naukovo-metodychna_rada/proekty_standartiv_VO/121-inzheneriya-programnogo-zabezpechennya-magistr-31-05-17.doc. Accessed 6 Feb 2018
36. CEN Workshop ICT Skills CWA European ICT Professional Profiles. Updated by e-CF version 3.0 competences (2013). http://relaunch.ecompetences.eu/wp-con-tent/uploads/2013/12/EU_ICT_Professional_Profiles_CWA_updated_by_e_CF_3.0.pdf
37. Kharchenko, V., Illiashenko, O.: Concepts of green IT engineering: taxonomy, principles and implementation. In: Kharchenko V., Kondratenko Y., Kacprzyk J. (eds.) Green IT Engineering: Concepts, Models, Complex Systems Architectures. Studies in Systems, Decision and Control, vol. 74. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-44162-7_1

38. Sidorov, N.A.: Ekologiya programmnogo obespecheniya [Text]. Inzheneriya programmnogo obespecheniya, № 1, pp. 55–61 (2010) [Сидоров, Н. А. Экология программного обеспечения [Текст]/ Н. А. Сидоров // Інженерія програмного забезпечення, № 1, pp. 55–61 (2010)
39. Grinenko, O.O.: Software ecosystems. J. NTU **26**, 508–512 (2012)