

Quantum information technology on the Edge

Serhiy O. Semerikov^{a,b,c,d}, Tetiana A. Vakaliuk^{e,c}, Andrii M. Striuk^b and Andrii V. Morozov^e

^aKryvyi Rih State Pedagogical University, 54 Gagarin Ave., Kryvyi Rih, 50086, Ukraine

^bKryvyi Rih National University, 11 Vitalii Matusevych Str., Kryvyi Rih, 50027, Ukraine

^cInstitute of Information Technologies and Learning Tools of the NAES of Ukraine, 9 M. Berlynskoho Str., Kyiv, 04060, Ukraine

^dUniversity of Educational Management, 52-A Sichovykh Striltsiv Str., Kyiv, 04053, Ukraine

^eZhytomyr Polytechnic State University, 103 Chudnivsyka Str., Zhytomyr, 10005, Ukraine

Abstract

This is an introductory text to a collection of selected papers from the Joint Workshop on the Quantum Information Technologies and Edge Computing (QualnT & doors 2021) which were held in Zhytomyr, Ukraine, on the April 11, 2021. It consists of short summaries of selected papers and some observations about the events.

Keywords

Quantum Information Technologies, Edge Computing

1. Introduction

1.1. QualnT 2021: At a glance

Quantum Information Technologies (*QualnT*) is a peer-reviewed international workshop interdisciplinary between Computer Science, Physics and Mathematics, focusing on research advances and applications of quantum information science and technology.

QualnT topics of interest (inspired by [1, 2, 3, 4, 5, 6]):

- Annealing-based computing
- Cloud-based quantum computing
- Quantum information science education

QualnT 2021: Workshop on the Quantum Information Technologies, April 11, 2021, Zhytomyr, Ukraine
doors 2021: Edge Computing Workshop, April 11, 2021, Zhytomyr, Ukraine

✉ semerikov@gmail.com (Serhiy O. Semerikov); tetianavakaliuk@gmail.com (Tetiana A. Vakaliuk); andrey.n.striuk@gmail.com (Andrii M. Striuk); morozov@ztu.edu.ua (Andrii V. Morozov)

🌐 <https://kdpu.edu.ua/semerikov> (Serhiy O. Semerikov); <https://sites.google.com/view/neota/profile-vakaliuk-t> (Tetiana A. Vakaliuk); <http://mpz.knu.edu.ua/pro-kafedru/vikladachi/224-andrii-striuk> (Andrii M. Striuk); https://ztu.edu.ua/en/common/university_administration.php#tab3 (Andrii V. Morozov)

🆔 0000-0003-0789-0272 (Serhiy O. Semerikov); 0000-0001-6825-4697 (Tetiana A. Vakaliuk); 0000-0001-9240-1976 (Andrii M. Striuk); 0000-0001-6825-4697 (Andrii V. Morozov)

© 2021 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



CEUR Workshop Proceedings (CEUR-WS.org)

- Quantum machine learning
- Quantum algorithms
- Quantum coding theory
- Quantum communication
- Quantum complexity theory
- Quantum computing models
- Quantum cryptography
- Quantum finance
- Quantum image processing
- Quantum information processing
- Quantum intelligent systems
- Quantum optimization
- Quantum programming
- Quantum software engineering
- Quantum simulation
- Quantum and probability logic

The first part of this volume represents the proceedings of the Workshop on the Quantum Information Technologies (QuaInT 2021), held in Zhytomyr, Ukraine, on April 11, 2021 (figure 2). It comprises 1 contributed paper ([7]) that was carefully peer-reviewed and selected from 4 submissions (<https://notso.easyscience.education/quaint/2021/>). Each submission was reviewed by at least 3, and on the average 3.25, program committee members. The accepted paper present the state-of-the-art overview of successful cases and provides guidelines for future research.



Figure 1: QuaInT 2021 logo

1.2. QuaInT 2021 Program Committee

- *Gustavo Banegas*, INRIA, France
- *Wei Chen*, University of Science and Technology of China, China
- *Sebastian Feld*, Delft University of Technology, Netherlands



Figure 2: Joint workshop opening

- *Kamil Khadiev*, Kazan Federal University, Russia
- *Andrey Kupin*, Kryvyi Rih National University, Ukraine
- *Gyu Myoung Lee*, Liverpool John Moores University, United Kingdom
- *Claudia Linnhoff-Popien*, Ludwig-Maximilians-Universität, Germany
- *Orken Mamyrbayev*, Institute of Information and Computational Technologies, Kazakhstan
- *Bongkyo Moon*, Dongguk University, Korea
- *Serhiy Semerikov*, Kryvyi Rih State Pedagogical University, Ukraine

Additional reviewers:

- *Pawel Caban*, University of Lodz, Poland
- *Oleksandr Kolgatin*, Simon Kuznets Kharkiv National University of Economics, Ukraine
- *Xiongfeng Ma*, Tsinghua University, China
- *Natalia Moiseienko*, Kryvyi Rih State Pedagogical University, Ukraine
- *Oleg Pursky*, Kyiv National University of Trade and Economics, Ukraine
- *Andrii Striuk*, Kryvyi Rih National University, Ukraine

1.3. doors 2021: At a glance

Peter the Great hacked through a window to Europe. We use doors.

Edge Computing Workshop (*doors*) is a peer-reviewed international Computer Science workshop focusing on research advances and applications of edge computing, a process of building a distributed system in which some applications, as well as computation and storage services, are provided and managed by

- (i) central clouds and smart devices, the edge of networks in small proximity to mobile devices, sensors and end users, and
- (ii) others are provided and managed by the center cloud and a set of small in-between local clouds supporting IoT at the edge.

The goal of *doors* is to bring together researchers and practitioners from academia and industry working on edge computing to share their ideas, discuss research/work in progress, and identify new/emerging trends in this important emerging area. The emergence of the Internet of Things (IoT) and the demand for responsiveness, privacy, and situation-awareness are pushing computing to the edge of the Internet. There are many challenges in the design, implementation, and deployment of different aspects of edge computing: infrastructure, systems, networking, algorithms, applications, etc. *doors* would like to open discussions in these areas.

doors topics of interest are opened to:

- algorithms and techniques for machine learning and AI at the edge
- cellular infrastructure for edge computing
- distributed ledger technology and blockchain at the edge
- edge computing infrastructure and edge-enabled applications
- edge-based data storage and databases
- edge-optimized heterogeneous architectures
- fault-tolerance in edge computing
- fog computing models and applications
- geo-distributed analytics and indexing on edge nodes
- hardware architectures for edge computing and devices
- innovative applications at the edge
- interoperability and collaboration between edge and cloud computing



Figure 4: Pandemic times in Zhytomyr Polytechnic State University

- monitoring, management, and diagnosis in edge computing
- processing of IoT data at network edges
- programming models and toolkits for edge computing
- resource management and Quality of Service for edge computing
- security and privacy in edge computing

The second part of this volume represents the proceedings of the Edge Computing Workshop (doors 2021), held in Zhytomyr, Ukraine, on April 11, 2021 (figure 4). It comprises 7 contributed papers ([8, 9, 10, 11, 12, 13, 14]) that were carefully peer-reviewed and selected from 14 submissions (<https://notso.easyscience.education/doors/2021/>). Each submission was reviewed by at least 3, and on the average 3.2, program committee members. The accepted papers present the state-of-the-art overview of successful cases and provides guidelines for future research.

doors 2021

Figure 3: doors 2021 logo

1.4. doors 2021 Program Committees

Program Committee:

- *Mehdi Ammi*, University of Paris 8, France
- *Aleksandr Cariow*, West Pomeranian University of Technology, Poland

- *Olena Glazunova*, National University of Life and Environmental Sciences of Ukraine, Ukraine
- *Attila Kertesz*, University of Szeged, Hungary
- *Dmitry Korzun*, Petrozavodsk State University, Institute of Mathematics and Information Technology, Russia
- *Vyacheslav Kryzhanivskyy*, R&D Seco Tools AB, Sweden
- *Nagender Kumar*, University of Hyderabad, India
- *Gyu Myoung Lee*, Liverpool John Moores University, United Kingdom
- *Taras Maksymyuk*, Lviv Polytechnic National University, Ukraine
- *Franco Milano*, University of Florence, Italy
- *Bongkyo Moon*, Dongguk University, Korea
- *Leonardo Mostarda*, University of Camerino, Italy
- *Djamel Eddine Saidouni*, MISC Laboratory, University Constantine 2 – Abdelhamid Mehri, Algeria
- *Gwen Salaun*, University Grenoble Alpes, France
- *Serhiy Semerikov*, Kryvyi Rih State Pedagogical University, Ukraine
- *Pedro Valderas*, Universitat Politècnica de València, Spain
- *Xianzhi Wang*, University of Technology Sydney, Australia
- *Michael Wei*, VMware Research, USA
- *Eiko Yoneki*, University of Cambridge, United Kingdom
- *Pamela Zave*, Princeton University, USA

Additional reviewers:

- *Abhineet Anand*, Chitkara University, India
- *Dmitry Antoniuk*, Zhytomyr Polytechnic State University, Ukraine
- *Josef Cernohorsky*, Technical university of Liberec, Czech Republic
- *Lubomir Dimitrov*, Technical University-Sofia, Bulgaria
- *Mahmud Hossain*, The University of Alabama at Birmingham, US
- *Valerii Kontsedailo*, Inner Circle, Netherlands

- *Nadiia Lobanchykova*, Zhytomyr Polytechnic State University, Ukraine
- *Mykhailo Medvediev*, ADA University, Azerbaijan
- *Andrii Morozov*, Zhytomyr Polytechnic State University, Ukraine
- *Tetiana Nikitchuk*, Zhytomyr Polytechnic State University, Ukraine
- *Shadi Noghabi*, Microsoft Research, Redmond, WA, USA
- *Igor Puleko*, Zhytomyr Polytechnic State University, Ukraine
- *Etibar Seyidzade*, Baku Engineering University, Azerbaijan
- *Andrii Striuk*, Kryvyi Rih National University, Ukraine
- *Inna Suhoniak*, Zhytomyr Polytechnic State University, Ukraine
- *Tetiana Vakaliuk*, Zhytomyr Polytechnic State University, Ukraine
- *Tetiana Voloshyna*, National University of Life and Environmental Sciences of Ukraine, Ukraine
- *Volodymyr Voytenko*, Athabasca University, Canada
- *Valentyn Yanchuk*, Zhytomyr Polytechnic State University, Ukraine

Organizing Committee:

- *Nadiia Lobanchykova*, Zhytomyr Polytechnic State University, Ukraine
- *Andrii Morozov*, Zhytomyr Polytechnic State University, Ukraine
- *Serhiy Semerikov*, Kryvyi Rih State Pedagogical University, Ukraine
- *Andrii Striuk*, Kryvyi Rih National University, Ukraine
- *Tetiana Vakaliuk*, Zhytomyr Polytechnic State University, Ukraine

2. Articles overview

3. QualnT 2021 Article overview

Kamil Khadiev (figure 5) in the article [7] consider online algorithms as a request-answer game. An adversary that generates input requests, and an online algorithm answers. Author consider a generalized version of the game that has a buffer of limited size. The adversary loads data to the buffer, and the algorithm has random access to elements of the buffer. Author consider quantum and classical (deterministic or randomized) algorithms for the model. Kamil Khadiev provide a specific problem (The Most Frequent Keyword Problem) and a quantum algorithm that works better than any classical (deterministic or randomized) algorithm in terms of competitive ratio. At the same time, for the problem, classical online algorithms in the standard model are equivalent to the classical algorithms in the request-answer game with buffer model.

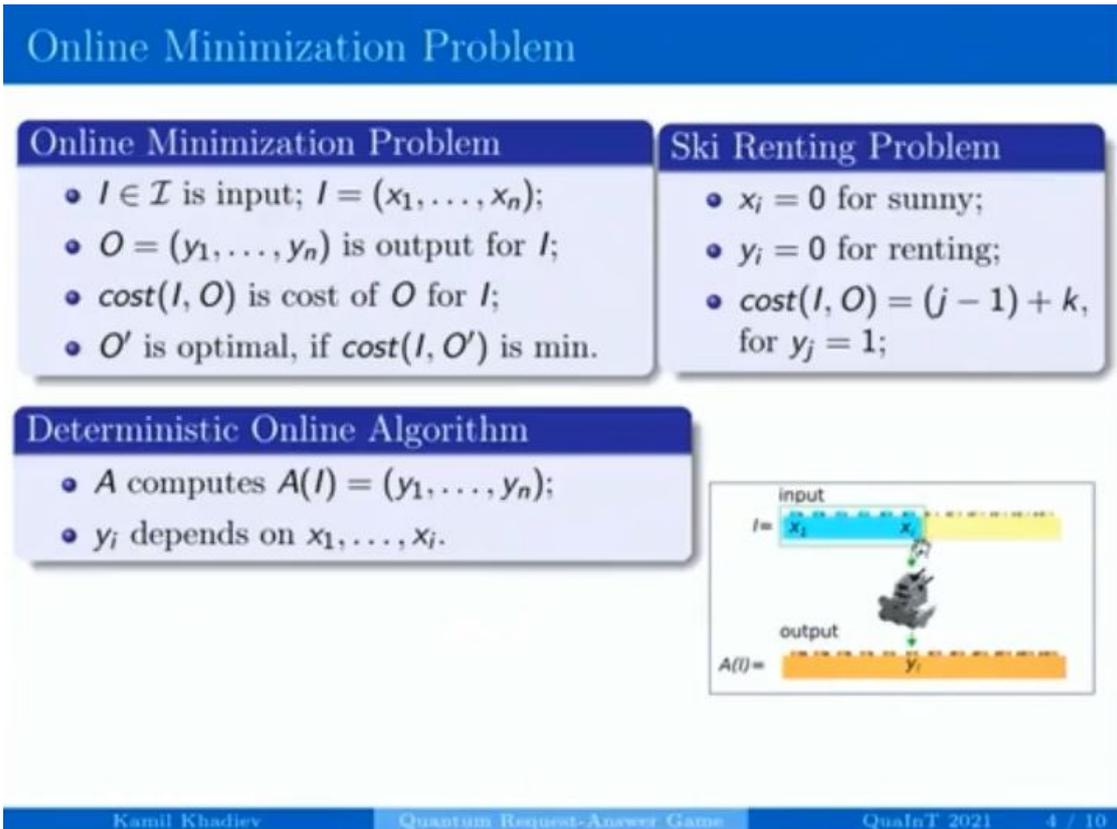


Figure 5: Presentation of paper [7]

4. doors 2021 Articles overview

The development and effective application of Fog Computing technologies require the most complex tasks related to the management and processing of huge data sets, including the tasks of rational construction of low-level networks that ensure the functioning of end devices within the IoT concept. The article “Graph model of Fog Computing system” [8] authored by Andriy V. Ryabko, Oksana V. Zaika, Roman P. Kukharchuk and Tetiana A. Vakaliuk (figure 6) describes the use of graph theory methods to solve such problems. The proposed graph model can provide the ability to determine the basic properties of systems, networks, and network devices within the concept of Fog Computing, the optimal characteristics, and ways to maintain them in working condition. This paper shows how to plot graphs, and then customize the display to add labels or highlighting to the graph nodes and edges of pseudo-random task graphs which can be used for evaluating Mobile Cloud, Fog and Edge computing systems. The graphs are described and visualized in Matlab code. Each task has an amount of computational work to perform, expressed in Megacycles per second. Each edge has an amount of data to transfer between tasks, expressed in kilobits or kilobytes of data. The set can be used by researchers to evaluate cloud/fog/edge computing systems and computational offloading algorithms. The

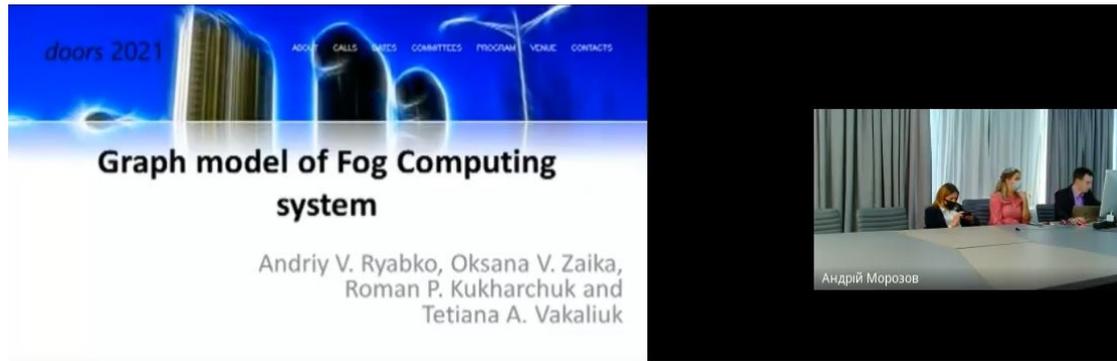


Figure 6: Presentation of paper [8]

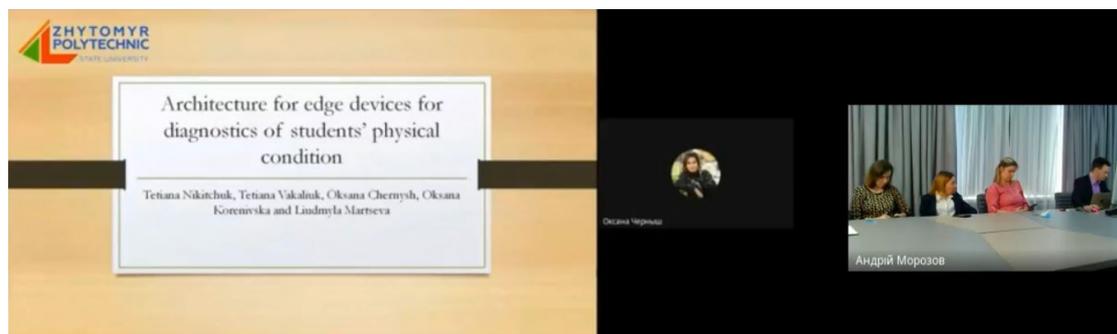


Figure 7: Presentation of paper [9]

task graphs can be used in single-user systems, where one mobile device accesses a remote server, or in multi user systems, where many users access a remote server through a wireless channel.

Tetiana M. Nikitchuk, Tetiana A. Vakaliuk, Oksana A. Chernysh (figure 7), Oksana L. Korenivska, Liudmyla A. Martseva and Viacheslav V. Osadchyi in the article “Architecture for edge devices for diagnostics of students’ physical condition” [9] investigates the possibility of technical realization of hardware complex. It presupposes the use of sensors of registration of a photoplethysmographic curve, which describes a pulse wave and defines the parameters of students’ cardiovascular system functional state. The method of photoplethysmography allows the use of non-contact sensors. Therefore, there is no artery compression, which eliminates circulatory disorders and allows the use of calculations to determine the saturation of oxygen by the pulse wave. It is recommended to use several optocouplers connected in series, parallel or parallel-series in a chain, with control of their mode of operation from the intensity of the received pulse wave signal depending on human body constitution. The edge device hardware is a part of the IoT system, which also includes another edge device, which instantly transmits data to the database on the edge server for the data further processing and storage.

The concept of the Internet of Things is increasingly defining the development of communication networks both now and in the future. The largest application of the IoT concept is

REACTIVE AND PROACTIVE MODES OF THE HWMP PROTOCOL

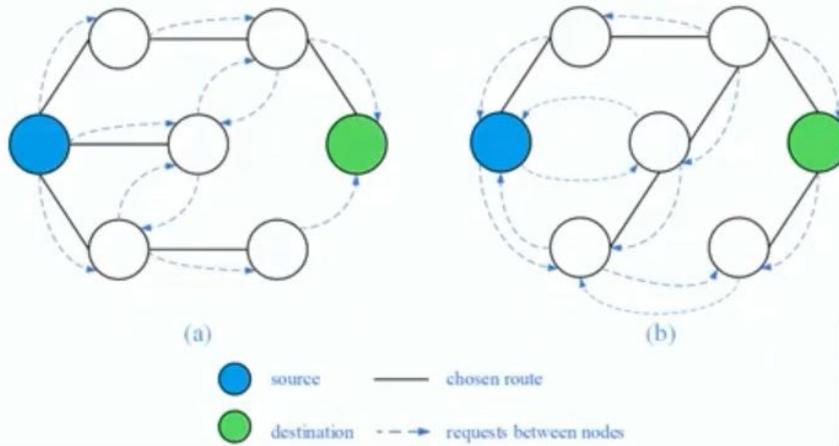


Figure 8: Presentation of paper [10]

wireless touch networks (WTN). Due to the potentially widespread use of WTN in all areas of human life, they are also called pervasive sensory networks. WTN belongs to the class of self-organizing networks, for which the construction principles, routing protocols, quality of service parameters, traffic models, and characteristics are significantly modified compared to traditional infrastructure networks, etc. The features of the application of dynamic routing protocols for the construction of a self-organizing network of autonomous IoT systems are considered. Anastasia D. Sverdlova (figure 8), Artur O. Zaporozhets, Ihor V. Bohachev, Oleksandr O. Popov, Anna V. Iatsyshyn, Andrii V. Iatsyshyn, Valeriia O. Kovach, Volodymyr O. Artemchuk and Nataliia M. Hrushchynska in the article “Self-organizing network topology for autonomous IoT systems” [10] provides an overview of the main methods for calculating the topology of self-organizing networks. A review of known dynamic routing protocols for mobile radio networks is given, the advantages and disadvantages of proactive and reactive approaches are shown.

Edge computing is an extension of cloud computing where physical servers are deployed closer to the users in order to reduce latency. Edge data centers face the challenge of serving a continuously increasing number of applications with a reduced capacity compared to traditional data center. Tania Lorigo-Botran (figure 9) and Muhammad Khurram Bhatti in the article “*ImpalaE*: Towards an optimal policy for efficient resource management at the edge” [11] introduces *ImpalaE*, an agent based on Deep Reinforcement Learning that aims at optimizing the resource usage in edge data centers. First, it proposes modeling the problem as a Markov Decision Process, with two optimization objectives: reducing the number of physical servers used and maximize number of applications placed in the data center. Second, it introduces an agent based on Proximal Policy Optimization, for finding the optimal consolidation policy,

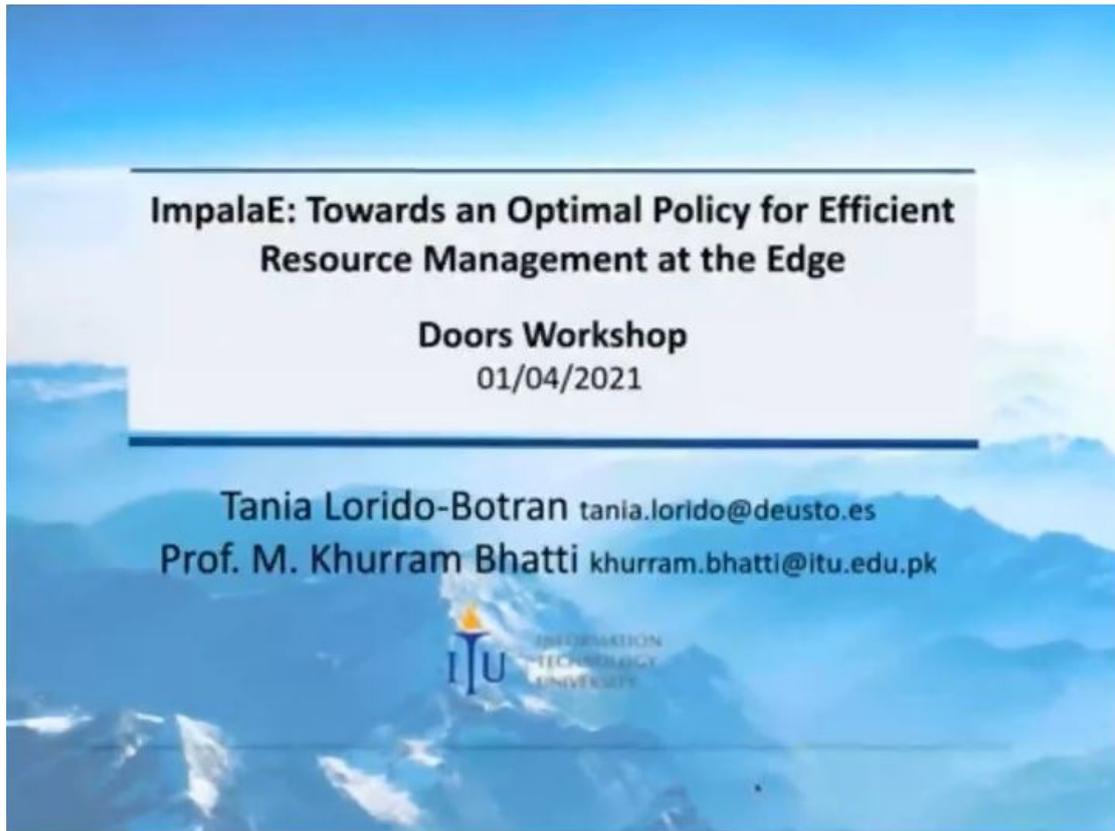


Figure 9: Presentation of paper [11]

and an asynchronous architecture with multiple workers-shared learner that enables for faster convergence, even with reduced amount of data. We show the potential in a simulated edge data center scenario with different VM sizes based on Microsoft Azure real traces, considering CPU, memory, disk and network requirements. Experiments show that *ImpalaE* effectively increases the number of VMs that can be placed per episode and that it quickly converges to an optimal policy.

Nadiia M. Lobanchykova (figure 10), Ihor A. Pilkevych and Oleksandr Korchenko in the article “Analysis of attacks on components of IoT systems and cybersecurity technologies” [12] presents the results of IoT analysis, methods and ways of their protection, prospects of using edge computing to minimize traffic transmission, decentralization of decision-making systems, and information protection. A detailed analysis of attacks on IoT system components was carried out and protection recommendations were developed.

Taras A. Uzdenov (figure 11) in the article “Task scheduling in Desktop GRID by FSA method: a practical example” [13] considers a new approach to solving the problem of dispatching task flows, the complexity of which is known, for GRID-systems with inalienable resources, the performance of which can be determined. A method based on this approach has been developed. The efficiency of the proposed method is compared with the well-known and widely used in

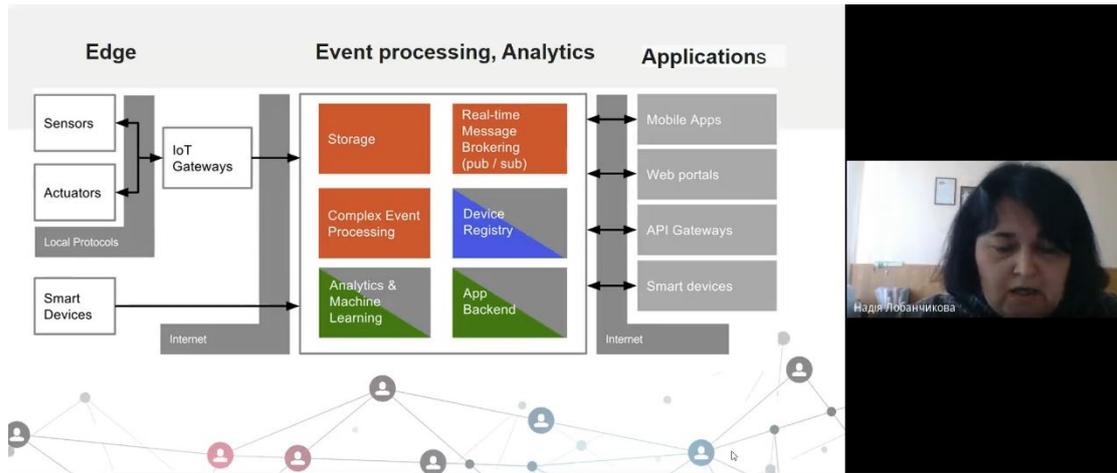


Figure 10: Presentation of paper [12]

Taras Uzdenov

**Task scheduling in Desktop GRID by FSA method:
a practical example**

Door - 2021

Figure 11: Presentation of paper [13]

various projects method FCFS. A feature of this method is the simplicity of implementation. An example of a simple practical problem that can be solved using the proposed method is described in this paper.

Volodymyr Kvasnikov (figure 12), Mariia Kataieva and Victor Kobylyansky in the article “Analysis of metrological support of nano-measurements” [14] analyzes the existing methods and means of measuring objects in the nanometer range and develops their classification based on the main principles of use. The main parameters on which each described method is based are considered and the conditions for their most effective application are determined. It is proved that the chemical and electrical sets of properties of the nanomaterial can change when the particle size decreases to the nanometer size, which requires the inclusion of additional chemical and electrical tests in existing methods. Based on the analysis, it was determined that

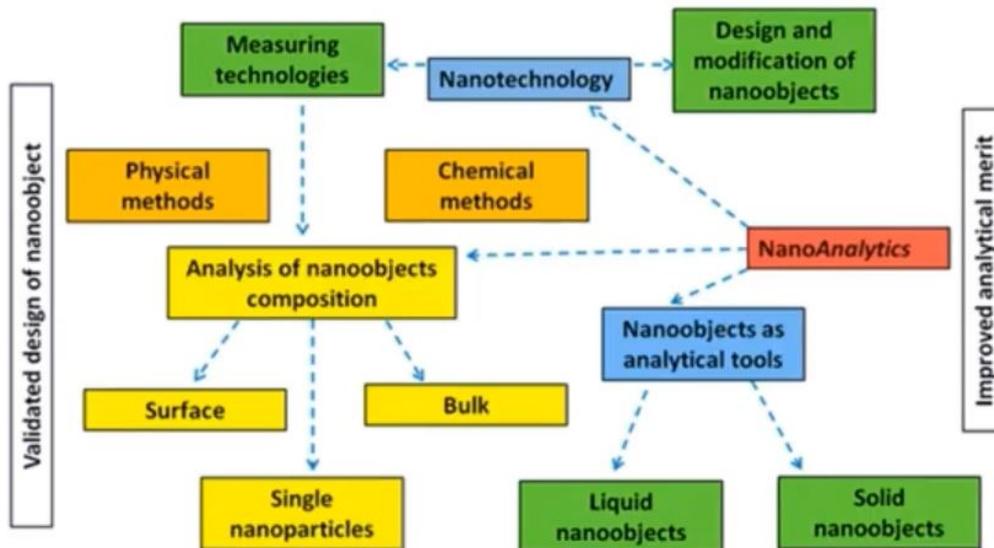


Figure 12: Presentation of paper [14]

the most functional and universal in solving a wide range of problems is the method of scanning probe microscopy. The classification of existing methods of scanning probe microscopy based on the nature of their applications is developed. The main information parameters on which each described method is based are considered, and the conditions of their most effective application are determined. To increase the accuracy of nanomeasurements, a methodology based on the principle of integration of information provided by different methods has been developed. The use of the differential-digital method is proposed, which includes the use of an additional information parameter in the mathematical model. An algorithm for including additional (a priori) information in the conditions for measuring the nanostructures has been developed, which leads the problem to the correct one according to the method of the control link, which characterizes the deviation of the parameters of measuring nanoobjects from their nominal values. It is proved that increasing the number of measurement methods used in the metrological analysis of nanoobjects will increase the reliability and accuracy of measurement results, and each method will provide additional information parameters to create a computerized method of calculating the control link. The main condition for correct comparison of the result is knowledge of the specific parameters on which each method is based.

5. Conclusion

The Joint Workshop on the Quantum Information Technologies and Edge Computing (QuaInT & doors 2021) was organized by Kryvyi Rih National University (with support of the rector Mykola I. Stupnik) and Zhytomyr Polytechnic State University (with support of the rector Viktor V. Ievdokymov) in collaboration with Kryvyi Rih State Pedagogical University (with support of the rector Yaroslav V. Shramko), Institute of Information Technologies and Learning

Tools of the NAES of Ukraine (with support of the director Valeriy Yu. Bykov) and University of Educational Management (with support of the vice-rector for research and digitalization Oleg M. Spirin).

We are thankful to all the authors who submitted papers and the delegates for their participation and their interest in QuaInT & doors as a platform to share their ideas and innovation. Also, we are also thankful to all the program committee members for providing continuous guidance and efforts taken by peer reviewers contributed to improve the quality of papers provided constructive critical comments, improvements and corrections to the authors are gratefully appreciated for their contribution to the success of the workshop.

References

- [1] V. Solovev, Atomic particle delocalisation effect in disordered media, *physica status solidi (a)* 83 (1984) 553–559. doi:10.1002/pssa.2210830216.
- [2] L. Lehka, S. Shokaliuk, Quantum programming is a promising direction of IT development, *CEUR Workshop Proceedings* 2292 (2018) 76–82.
- [3] V. Soloviev, A. Belinskiy, Complex systems theory and crashes of cryptocurrency market, *Communications in Computer and Information Science* 1007 (2019) 276–297. doi:10.1007/978-3-030-13929-2_14.
- [4] A. Bielinskiy, S. Semerikov, O. Serdyuk, V. Solovieva, V. Soloviev, L. Pichl, Econophysics of sustainability indices, *CEUR Workshop Proceedings* 2713 (2020) 372–392.
- [5] P. V. Zahorodko, Y. O. Modlo, O. O. Kalinichenko, T. V. Selivanova, S. O. Semerikov, Quantum enhanced machine learning: An overview, *CEUR Workshop Proceedings* 2832 (2020) 94–103. URL: <http://ceur-ws.org/Vol-2832/paper13.pdf>.
- [6] P. V. Zahorodko, S. O. Semerikov, V. N. Soloviev, A. M. Striuk, M. I. Striuk, H. M. Shalatska, Comparisons of performance between quantum-enhanced and classical machine learning algorithms on the IBM quantum experience, *Journal of Physics: Conference Series* 1840 (2021) 012021. URL: <https://doi.org/10.1088/1742-6596/1840/1/012021>. doi:10.1088/1742-6596/1840/1/012021.
- [7] K. Khadiev, Quantum request-answer game with buffer model for online algorithms. application for The Most Frequent Keyword Problem, *CEUR Workshop Proceedings* (2021, in press) 16–27.
- [8] A. V. Ryabko, O. V. Zaika, R. P. Kukharchuk, T. A. Vakaliuk, Graph model of Fog Computing system, *CEUR Workshop Proceedings* (2021, in press) 28–44.
- [9] T. M. Nikitchuk, T. A. Vakaliuk, O. A. Chernysh, O. L. Korenivska, L. A. Martseva, V. V. Osadchyi, Architecture for edge devices for diagnostics of students' physical condition, *CEUR Workshop Proceedings* (2021, in press) 45–56.
- [10] A. D. Sverdlova, A. O. Zaporozhets, I. V. Bohachev, O. O. Popov, A. V. Iatsyshyn, A. V. Iatsyshyn, V. O. Kovach, V. O. Artemchuk, N. M. Hrushchynska, Self-organizing network topology for autonomous IoT systems, *CEUR Workshop Proceedings* (2021, in press) 57–70.
- [11] T. Lorida-Botran, M. K. Bhatti, ImpalaE: Towards an optimal policy for efficient resource management at the edge, *CEUR Workshop Proceedings* (2021, in press) 71–82.

- [12] N. M. Lobanchykova, I. A. Pilkevych, O. Korchenko, Analysis of attacks on components of IoT systems and cybersecurity technologies, CEUR Workshop Proceedings (2021, in press) 83–96.
- [13] T. A. Uzdenov, Task scheduling in Desktop GRID by FSA method: a practical example, CEUR Workshop Proceedings (2021, in press) 97–109.
- [14] V. Kvasnikov, M. Kataieva, V. Kobylyansky, Analysis of metrological support of nano-measurements, CEUR Workshop Proceedings (2021, in press) 110–120.

Quantum request-answer game with buffer model for online algorithms. Application for The Most Frequent Keyword Problem

Kamil Khadiev

Kazan Federal University, 18 Kremlyovskaya str, Kazan, Tatarstan, 420008, Russia

Abstract

We consider online algorithms as a request-answer game. An adversary that generates input requests, and an online algorithm answers. We consider a generalized version of the game that has a buffer of limited size. The adversary loads data to the buffer, and the algorithm has random access to elements of the buffer. We consider quantum and classical (deterministic or randomized) algorithms for the model. In the paper, we provide a specific problem (The Most Frequent Keyword Problem) and a quantum algorithm that works better than any classical (deterministic or randomized) algorithm in terms of competitive ratio. At the same time, for the problem, classical online algorithms in the standard model are equivalent to the classical algorithms in the request-answer game with buffer model.

Keywords

quantum computation, online algorithm, request-answer game, online minimization problem, buffer, keywords search

1. Introduction

One of the applications for online algorithms is optimization problems [1]. The peculiarity is the following. An algorithm reads an input piece by piece and returns an answer piece by piece immediately, even if an answer can depend on future pieces of the input. The algorithm should return an answer for minimizing an objective function (the cost of an output). The most standard method to define the effectiveness is the competitive ratio [2, 3].

One of the possible point of view to online algorithms is a request-answer game [4]. Here we consider a game of an online algorithm and Adversary that holds input. Adversary requests and the algorithm returns answers. We suggest a reversed version of the game. The algorithm asks an input variable and Adversary returns an answer, but as a price for the answer, Adversary asks to return an output variable. The new version of the game is equivalent to the original one, but we can generalize it. We provide the new model for online algorithms that is called "Request-answer Game with Buffer". The model is a game of three players that are an online algorithm, Adversary and Buffer of limited size. The algorithm can do a request of one of two

QuaInT 2021: Workshop on the Quantum Information Technologies, April 11, 2021, Zhytomyr, Ukraine
doors 2021: Edge Computing Workshop, April 11, 2021, Zhytomyr, Ukraine

✉ kamilhadi@gmail.com (K. Khadiev)

🌐 https://kpfu.ru/Kamil.Hadiev?p_lang=2 (K. Khadiev)

🆔 0000-0002-5151-9908 (K. Khadiev)

© 2021 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
 CEUR Workshop Proceedings (CEUR-WS.org)

types:

- asking Adversary to load the next block of input variables to the Buffer;
- request Buffer for one of the holding variables.

For some integer parameter R , after each R requests Adversary asks an output variable. If the size of Buffer is 1 and $R = 1$, then the model is equivalent to the original one.

Motivation. Online algorithms have different applications. One of them is making a decision in current time with no knowledge about future data. Another one is processing a data stream and output a result data stream in online fashion, for example, streaming video on web sites and others. Many programming languages like Java, C++ [5, 6] and others use buffered data streams that store data in a fast buffer first, and then an algorithm reads data from the buffer. So, our model is like usage of buffered data streams. Additionally, we have asynchronous processing with online output. In other words, we focus on online behavior of the output stream, but when an algorithm reads an input stream, it can skip some data.

Quantum model. In the paper, we consider a quantum version of “Request-answer Game with Buffer” model. Quantum computing itself [7, 8, 9] is one of the hot topics in computer science. There are many problems where quantum algorithms outperform the best known classical algorithms [10, 11, 12, 13, 14]. Superior of quantum over classical was shown for different computational models like query model, streaming processing models, communication models and others [15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28].

Different versions of online quantum algorithms were considered in [21, 20] including quantum streaming algorithms as online algorithms [29, 30], quantum online algorithms with restricted memory [31, 32], quantum online algorithms with repeated test [33]. In these papers, authors show examples of problems that have quantum online algorithms with better competitive ratio comparing to classical online algorithms.

Our results. Here we provide a specific problem and a quantum online algorithm in “Request-answer Game with Buffer” model for it. We show that the quantum online algorithm has better competitive ratio than any classical (deterministic or randomized) counterpart. The problem is “The Most Frequent Keyword Problem”. Questions are strings of length k ; the problem is searching the most frequent keyword among words of a text and returning it after each word of the text immediately. The problem [34] is one of the most well-studied ones in the area of data streams [35, 36, 37]. Many applications in packet routing, telecommunication logging, and tracking keyword queries in search machines are critically based upon such routines. The similar problem in online fashion was considered in [38].

The paper is organized in the following way. Definitions are in Section 2. A description of the most frequent question problem and the quantum algorithm for the problem are described in Section 3. Section 4 contains lower bounds for classical algorithms.

2. Preliminaries

An online minimization problem consists of a set \mathcal{I} of inputs and a cost function. Each input $I = (x_1, \dots, x_n)$ is a sequence of requests, where n is a length of the input $|I| = n$. Furthermore, a set of feasible outputs (or solutions) $\mathcal{O}(I)$ is associated with each I ; an output is a sequence of answers $O = (y_1, \dots, y_n)$. The cost function assigns a positive real value $cost(I, O)$ to $I \in \mathcal{I}$ and $O \in \mathcal{O}(I)$. An optimal solution for $I \in \mathcal{I}$ is $O_{opt}(I) = \operatorname{argmin}_{O \in \mathcal{O}(I)} cost(I, O)$.

Let us define an online algorithm for this problem. **A deterministic online algorithm** A computes the output sequence $A(I) = (y_1, \dots, y_n)$ such that y_i is computed by x_1, \dots, x_i . We say that A is c -competitive if there exists a constant $\alpha \geq 0$ such that, for every n and for any input I of size n , we have: $cost(I, A(I)) \leq c \cdot cost(I, O_{opt}(I)) + \alpha$, where c is the minimal number that satisfies the inequality. Also we call c the **competitive ratio** of A . If $\alpha = 0, c = 1$, then A is optimal.

A randomized online algorithm R computes an output sequence $R^\psi(I) = (y_1, \dots, y_n)$ such that y_i is computed from ψ, x_1, \dots, x_i , where ψ is the content of the random tape, i. e., an infinite binary sequence, where every bit is chosen uniformly at random and independently of all the others. By $cost(I, R^\psi(I))$ we denote the random variable expressing the cost of the solution computed by R on I . R is c -competitive in expectation if there exists a constant $\alpha > 0$ such that, for every $I, \mathbb{E}[cost(I, R^\psi(I))] \leq c \cdot cost(I, O_{opt}(I)) + \alpha$. We can say that c is expected competitive ratio for the algorithm.

2.1. Request-answer game with buffer model

The standard model for online algorithms can be considered as a request-answer game [4]. Adversary holds an input, it sends request x_i to an algorithm, and the algorithm sends answer y_i . Here Adversary is an “active” player that rules the game and the algorithm is a “passive” player that answers on each response.

Let us change the point of view to this game. Both are “active” players in some sense.

Round 1. The algorithm asks an input variable x_1 . (The algorithm is active on this round).

Round 2. Adversary asks an output variable y_1 . (Adversary is active on this round).

...

Round $2i - 1$. The algorithm asks an input variable x_i . (The algorithm is active on this round).

Round $2i$. Adversary asks an output variable y_i . (Adversary is active on this round).

It is easy to see that the new game is equivalent to the original game and the standard model.

Let us consider the modification of the game that has a buffer. Assume that we have a buffer between the algorithm and Adversary. Let a positive integer K be a size of the buffer. Additionally, there is an integer parameter $R \leq K$. The algorithm will ask to load data to the buffer by blocks of K variables. Let i be a number of the loading block. The algorithm can do the following actions if it is active on some round:

- The algorithm asks to erase the buffer and load the next K input variables $x_{i \cdot K + 1}, \dots, x_{i \cdot K + K}$ to the buffer. After that, i is increased by 1. ($i \leftarrow i + 1$)
- The algorithm requests any variable from the buffer. We consider a query model (decision tree model) for the algorithm that queries variables from the buffer.

The game has the following scenario:

Round 0. We initialize $i \leftarrow 0$

Round 1. The algorithm is active and it does the possible actions that were described before.

Round 2. The algorithm is active and it does the possible actions that were described before.

...

Round R . The algorithm is active and it does the possible actions that were described before.

Round $R + 1$. Adversary is active. He asks output variables y_1, \dots, y_R .

...

Round $(R + 1) \cdot j + 1$. The algorithm is active and it does the possible actions that were described before.

Round $(R + 1) \cdot j + 2$. The algorithm is active and it does the possible actions that were described before.

...

Round $(R + 1) \cdot j + R$. The algorithm is active and it does the possible actions that were described before.

Round $(R + 1) \cdot j + R + 1$. Adversary is active. He asks output variables $y_{j \cdot R + 1}, \dots, y_{j \cdot R + R}$.

Comment. *In the case of $K = 1$ and $R = 1$, the new model is equivalent to the standard online algorithms model.*

In the randomized case, an algorithm that requests data from the buffer can be randomized, and we use a randomized query model in that case. We consider an expected competitive ratio for the model as for the standard model of randomized online algorithms. At the same time, the loading the next block to the buffer is deterministic action.

In the quantum case, an algorithm that requests data from the buffer can be quantum, and we use a quantum query model in that case. Because of the probabilistic behavior of quantum algorithms, we also consider an expected competitive ratio for the model. At the same time, the loading the next block to the buffer is deterministic action.

We skip details of the quantum model and quantum algorithms here because we use them as quantum subroutines and the rest part is classical. More details on quantum query model and quantum algorithms can be found in [7, 8, 9]

3. A quantum algorithm for The Most Frequent Keyword Problem

Let us present the problem formally.

Problem. For some positive integers m, d and k , the input is

$$I = (s^1, \dots, s^d, x^1, \dots, x^m).$$

Here (s^1, \dots, s^d) is a sequence of strings that are interesting keywords for us in the input, $s^j = (s_1^j, \dots, s_k^j) \in \{0, 1\}^k$, for $j \in \{1, \dots, d\}$. Strings x^1, \dots, x^m are words of a text, $x^j = (x_1^j, \dots, x_k^j) \in \{0, 1\}^k$, for $j \in \{1, \dots, m\}$. The input length is $n = (m + d) \cdot k$. A frequency of a string $t \in \{0, 1\}^k$ is $f(t) = \frac{\#(t)}{m}$, where $\#(t) = |\{i : t = x^i, i \in \{1, \dots, m\}\}|$ is a number of occurrence of t in (x^1, \dots, x^m) . The index i_0 of the most frequent string s^{i_0} is such that $f(s^{i_0}) = \max_{i \in \{1, \dots, d\}} f(s^i)$ and i_0 is minimal. We should return index i_0 after reading each string x^j . So, the right answer that returns offline algorithm is (z_1, \dots, z_n) where $z_{(j+d) \cdot k} = i_0$ for $j \in \{1, \dots, m\}$ and other output variables are not considered.

The cost of an output $O = (y_1, \dots, y_n)$ is

$$\text{cost}(I, O) = 1 + m - \sum_{j=1}^m \delta(y_{(j+d) \cdot k}, i_0)$$

Here $\delta(a, b) = 1$ if $a = b$ and $\delta(a, b) = 0$ if $a \neq b$

3.1. Quantum algorithm

Firstly, we discuss a quantum subroutine that compares two strings of length l for some integer $l > 0$.

3.1.1. The quantum algorithm for two strings comparing

Assume that the subroutine is `COMPARE_STRINGS(s, t)` and it compares s and t in lexicographical order. It returns:

- -1 if $s < t$;
- 0 if $s = t$;
- 1 if $s > t$.

As a base for our algorithm, we will use the algorithm of finding the minimal argument with 1-result of a Boolean-value function. Formally, we have:

Lemma 1. [39] *Suppose, we have a function $f : \{1, \dots, N\} \rightarrow \{0, 1\}$ for some integer N . There is a quantum algorithm for finding $j_0 = \min\{j \in \{1, \dots, N\} : f(j) = 1\}$. The algorithm finds j_0 with query complexity \sqrt{N} and error probability that is at most $\frac{1}{2}$.*

Let us choose the function $f(j) = (s_j \neq t_j)$. So, we search j_0 that is the index of the first unequal symbol of the strings. We search j_0 among indexes $1, \dots, \min(|s|, |t|)$, where $|s|$ is a length of s . Then, we can claim that s precedes t in lexicographical order iff s_{j_0} precedes t_{j_0} in the alphabet for strings. If there are no unequal symbols, then we have one of three options:

- if $|s| < |t|$, then $s < t$;
- if $|s| > |t|$, then $s > t$;
- if $|s| = |t|$, then $s = t$.

We use `THE_FIRST_ONE_SEARCH`(f, N) as a subroutine from Lemma 1, where $f(j) = (s_j \neq t_j)$. Assume that this subroutine returns $N + 1$ if it does not find any solution.

We apply the standard technique of boosting success probability that was used, for example, in [12]. So, we repeat the algorithm $3 \log_2 m$ times and return the minimal answer, where m is a number of strings in the sequence (x^1, \dots, x^m) . In that case, the error probability is $O\left(\frac{1}{2^{3 \log m}}\right) = O\left(\frac{1}{m^3}\right)$.

Let us present the algorithm.

Algorithm 1 `COMPARE_STRINGS`(s, t, k). The quantum algorithm for two strings comparing.

```

 $N \leftarrow \min(|s|, |t|)$ 
 $j_0 \leftarrow \text{THE\_FIRST\_ONE\_SEARCH}(f, N)$  ▷ The initial value
for  $i \in \{1, \dots, 3 \log_2 m\}$  do
   $j \leftarrow \text{THE\_FIRST\_ONE\_SEARCH}(f, N)$ 
  if  $j \leq k$  and  $s_j \neq t_j$  then
     $j_0 \leftarrow \min(j_0, j)$ 
  end if
end for
if  $j_0 = N + 1$  and  $|s| = |t|$  then
   $result \leftarrow 0$  ▷ The strings are equal.
end if
if  $((j_0 \neq N + 1)$  and  $(s_{j_0} < t_{j_0}))$  or  $((j_0 = N + 1)$  and  $(|s| < |t|))$  then
   $result \leftarrow -1$  ▷  $s$  precedes  $t$ .
end if
if  $((j_0 \neq N + 1)$  and  $(s_{j_0} > t_{j_0}))$  or  $((j_0 = N + 1)$  and  $(|s| > |t|))$  then
   $result \leftarrow 1$  ▷  $t$  succeeds  $s$ .
end if
return  $result$ 

```

Let us discuss the property of the algorithm:

Lemma 2. *Algorithm 1 compares two strings s and t in lexicographical order with query complexity $O(\sqrt{\min(|s|, |t|)} \log m)$ and error probability $O\left(\frac{1}{m^3}\right)$.*

Proof. The correctness of the algorithm follows from description and lexicographical order.

Let us discuss the error probability. The algorithm has error iff there are error in all $3 \log_2 m$ invocations of `THE_FIRST_ONE_SEARCH` algorithm. The probability of such event is at most $0.5^{3 \log_2 m} = O\left(\frac{1}{m^3}\right)$. \square

3.1.2. A quantum algorithm in request-answer game with buffer model

Firstly, we present an idea of the algorithm.

We use the well-known data structure a self-balancing binary search tree. As an implementation of the data structure, we can use the AVL tree [40, 41] or the Red-Black tree [42, 41]. Both data structures allow us to find and add elements in $O(\log N)$ running time, where N is a size of the tree.

The idea of the algorithm is the following. We store a triple (i, s, c) in a vertex of the tree, where i is the minimal index of a string from $\{s^1, \dots, s^d\}$ such that $s = s^i$ and c is a number of occurrences of the string s among $\{x^1, \dots, x^m\}$. We assume that a triple (i, s, c) is less than a pair (i', s', c') iff s precedes s' in the lexicographical order. So, we use `COMPARE_STRINGS`(s, s', k) subroutine as the comparator of the vertexes. The tree represents a set of unique strings from $\{s^1, \dots, s^d\}$ with a number of occurrences among (x^1, \dots, x^m) .

Firstly, we load all strings s^1, \dots, s^d one by one to Buffer and add a vertex $v = (j, s^j, 0)$ for each string s^j to the tree, here $j \in \{1, \dots, d\}$. We add only one node for each duplicate strings from s^1, \dots, s^d if they exist. The index j in v stores the index of s^j and if there is no a vertex that corresponds to s^j , then j is a minimal index from all possible indexes. 0 in v means that initially we assume that s^j does not occurs among (x^1, \dots, x^m) .

Secondly, we load questions (strings) from x^1 to x^m one by one to Buffer and search them in our tree. We increase the number of occurrences. If the string was not found in the tree, then it is not a keyword, i.e. it does not belong to s^1, \dots, s^d and we skip it. At the same time, we store

$$(i_{max}, s, c_{max}) = \operatorname{argmax}_{(i,t,c)} \text{ in the tree } c$$

and recalculate it in each step. When Adversary requests an output variable, then we return i_{max} .

Let us present the algorithm formally. Let BST be a self-balancing binary search tree such that:

- `FIND`(BST, x^i) finds a vertex (j, s, c) such that $s = x^i$, or *NULL* if x^i was not found. The standard algorithm for searching x^i in the tree is comparing with elements of vertexes and moving by the tree according to the result of the comparison. When we invoke the `COMPARE_STRINGS` subroutine, we request a variable from Buffer for checking a symbol of x^i and request to memory when we check a symbol of a string that is stored in a vertex.
- `ADD`(BST, j, s^j) adds a vertex $(j, s^j, 0)$ to the tree if a vertex with s^j does not exist; and does nothing otherwise.
- `INIT`(BST) initializes an empty tree.

Let us discuss the property of the algorithm.

Algorithm 2 A quantum algorithm for The Most Frequent Keyword Problem.

```

INIT(BST)                                ▷ The initialization of the tree.
 $c_{max} \leftarrow 1$                        ▷ The maximal number of occurrences.
 $i_{max} \leftarrow 1$                        ▷ The index of most frequent question.
 $step \leftarrow 0$ 
for  $j \in \{1, \dots, d\}$  do
    LOAD_TO_BUFFER                          ▷ Load  $s^j$  to Buffer
     $t \leftarrow \}$ ''                       ▷ Initially  $t$  is an empty string
    for  $q \in \{1, \dots, k\}$  do           ▷ Reading the string  $t$ 
         $t \leftarrow t + \text{REQUEST}(q)$     ▷ Requesting  $q$ -th variable from Buffer and appending the
        variable to  $t$ 
    end for
    ADD(BST,  $j, t$ )                        ▷ Adding the string  $t = s^j$  to the tree as a vertex ( $NULL, t, 0$ )
end for
for  $j \in \{1, \dots, m\}$  do
    LOAD_TO_BUFFER                          ▷ Load  $x^j$  to Buffer
     $v = (i, t, c) \leftarrow \text{FIND}(BST, x^j)$     ▷ Searching  $x^j$  in the tree.
    if  $v \neq NULL$  then                 ▷ If  $x^j$  belongs to  $(s^1, \dots, s^d)$ 
         $c \leftarrow c + 1$                 ▷ Updating the vertex by increasing the number of occurrences.
         $v \leftarrow (i, t, c)$            ▷ Updating the vertex by the new values
        if  $c > c_{max}$  then             ▷ Updating the maximal value.
             $c_{max} \leftarrow c$ 
             $i_{max} \leftarrow i$ 
        end if
    end if
end for
if Adversary request an output variable then return  $i_{max}$ 
end if

```

Theorem 3. The expected competitive ratio c for Algorithm 2 is at most C_Q where

$$C_Q = O\left(1 + \frac{(m \log m) \cdot (\log d)}{\sqrt{k}}\right).$$

Proof. The correctness of the algorithm follows from the description. Let us discuss the query complexity of $\text{FIND}(BST, x^j)$. The procedure requires $O(\log d)$ comparing operations $\text{COMPARE_STRINGS}(x^j, s^i, k)$. Due to Lemma 2, each comparing operation requires $O(\sqrt{k} \log m)$ queries. The total query complexity of the FIND procedure is $O(\sqrt{k}(\log m) \cdot (\log d))$. So, the algorithm checks all x^1, \dots, x^m in $O(m\sqrt{k}(\log m) \cdot (\log d))$ rounds and after that returns right answers for the requests of Adversary. Therefore, the first $O\left(\frac{m\sqrt{k}(\log m) \cdot (\log d)}{k}\right) = O\left(\frac{m(\log m) \cdot (\log d)}{\sqrt{k}}\right)$ “significant” output variables can be wrong and others are right. We call output variable $y_{(j+d) \cdot k}$, for $j \in \{1, \dots, m\}$, as “significant” because the cost depends on these variables. Hence, the cost is at most $1 + O\left(\frac{m(\log m) \cdot (\log d)}{\sqrt{k}}\right)$.

Let us discuss the error probability. Events of error in the algorithm are independent. So, all events should be correct. Due to Lemma 2, the probability of correctness of one event is $1 - \left(1 - \frac{1}{m^3}\right)$. Hence, the probability of correctness of all $O(m \log m)$ events is at least $1 - \left(1 - \frac{1}{m^3}\right)^{\gamma \cdot m \log m}$ for some constant γ .

Note that

$$\lim_{n \rightarrow \infty} \frac{\left(1 - \frac{1}{m^3}\right)^{\gamma \cdot m \log m}}{1/m} < 1;$$

Hence, the total error probability is at most $O\left(\frac{1}{m}\right)$.

In a case of an error, all “significant” output variables can be wrong.

Therefore, the expected competitive ratio of the algorithm is at most

$$C_Q = \frac{O\left(\frac{m-1}{m}\right) \cdot \left(1 + O\left(\frac{m(\log m) \cdot (\log d)}{\sqrt{k}}\right)\right) + O\left(m \cdot \frac{1}{m}\right)}{1} = O\left(1 + \frac{m(\log m) \cdot (\log d)}{\sqrt{k}}\right).$$

□

4. Lower bounds for classical algorithms for The Most Frequent Keyword Problem

There is an input I_B such that any classical (deterministic or randomized) algorithm returns output with the cost at least $O(m)$.

Theorem 4. *Any randomized algorithm for the problem has competitive ratio c at least $C_R = O(m) > C_Q$ in a case of $(\log_2 m) \cdot (\log_2 d) = o(\sqrt{k})$.*

Proof. Let us show that the problem is equivalent to unstructured search problem. Assume that $m = 2t$ for some integer t . Then, let $x^{t+1}, \dots, x^{2t} = 0^k$ where 0^k is a string of k zeros. We have two cases for other string:

- **case 1:** $x^1, \dots, x^t = 1^k$;
- **case 2:** there are $z \in \{1, \dots, t\}$ and $u \in \{1, \dots, k\}$ such that $x_u^z = 0$ and $x_{u'}^z = 1$ for all $u' \in \{1, \dots, u-1, u+1, \dots, k\}$, $x^{z'} = 1^k$ for $z' \in \{1, \dots, t\} \setminus \{z\}$.

Let $d = 2$, $s^1 = 0^k$ and $s^2 = 1^k$.

In the first case, the answer is 1^k . In the second case, the answer is 0^k . Therefore, the problem is equivalent to search 0 among the first $tk = mk/2$ variables.

Due to [43], the randomized query complexity of unstructured search among $mk/2$ is $\Omega(mk)$.

In a case of odd m , we assign $x^m = 1^{k/2}0^{k/2}$, and it is not used in the search. Then, we can consider only $m - 1$ strings. So, $m - 1$ is even.

Suppose, we have a randomized algorithm A for finding the most frequent question that uses $o(mk)$ queries to buffer when it reads x^1, \dots, x^m . Then, Adversary can construct the input I_B such that A obtains a wrong answer.

Therefore, all “significant” output variables will be wrong and $\text{cost}(I_B, A(I_B)) = 1 + m$. The competitive ratio in that case is $C_R = m + 1$.

If the algorithm do $O(mk)$ queries to Buffer for computing answer, then $O(m)$ “significant” output variables should be returned before getting a right answer. Therefore, $\text{cost}(I_B, A(I_B)) = O(m)$ and $C_R = O(m)$.

In the case of $(\log_2 m) \cdot (\log_2 d) = o(\sqrt{k})$ we have

$$C_Q = O\left(1 + \frac{m(\log_2 m) \cdot (\log_2 d)}{\sqrt{k}}\right) = o(m) < O(m) = C_R.$$

□

5. Conclusion

We consider a new setting or new model for online algorithms that is useful for real world problems. We show that in the case of $(\log_2 m) \cdot (\log_2 d) = o(\sqrt{k})$ the quantum algorithm shows a better competitive ratio than any classical (deterministic or randomized) algorithm. Note that this setting is reasonable.

Acknowledgments

The research was funded by the subsidy allocated to Kazan Federal University for the state assignment in the sphere of scientific activities, project No. 0671-2020-0065.

We thank Farid Ablayev and Aliya Khadieva from Kazan Federal University for helpful discussions and support.

References

- [1] D. Komm, An Introduction to Online Computation: Determinism, Randomization, Advice, Springer, 2016.
- [2] D. D. Sleator, R. E. Tarjan, Amortized efficiency of list update and paging rules, Communications of the ACM 28 (1985) 202–208.
- [3] A. R. Karlin, M. S. Manasse, L. Rudolph, D. D. Sleator, Competitive snoopy caching, in: FOCS, 1986., 27th Annual Symposium on, IEEE, 1986, pp. 244–254.
- [4] S. Albers, BRICS, Mini-Course on Competitive Online Algorithms, Aarhus University, 1996.
- [5] Oracle, Java Platform SE 8 documentation, 2021. URL: <https://docs.oracle.com/javase/8/docs/api/java/io/BufferedReader.html>.
- [6] S. B. Lippman, J. Lajoie, C++ Primer, third ed., Massachusetts: Addison-Wesley, 1998.
- [7] M. A. Nielsen, I. L. Chuang, Quantum computation and quantum information, Cambridge univ. press, 2010.
- [8] A. Ambainis, Understanding quantum algorithms via query complexity, in: Proc. Int. Conf. of Math. 2018, volume 4, 2018, pp. 3283–3304.

- [9] F. Ablayev, M. Ablayev, J. Z. Huang, K. Khadiev, N. Salikhova, D. Wu, On quantum methods for machine learning problems part i: Quantum tools, *Big Data Mining and Analytics* 3 (2019) 41–55.
- [10] R. de Wolf, Quantum computing and communication complexity, Ph.D. thesis, 2001.
- [11] S. Jordan, Quantum algorithms zoo, 2021. URL: <http://quantumalgorithmzoo.org/>.
- [12] K. Khadiev, L. Safina, Quantum algorithm for dynamic programming approach for dags. applications for zhegalkin polynomial evaluation and some problems on dags, in: *Proceedings of UCNC 2019*, volume 4362 of *LNCS*, 2019, pp. 150–163.
- [13] K. Khadiev, D. Kravchenko, D. Serov, On the quantum and classical complexity of solving subtraction games, in: *Proceedings of CSR 2019*, volume 11532 of *LNCS*, 2019, pp. 228–236.
- [14] K. Khadiev, I. Mannapov, L. Safina, The quantum version of classification decision tree constructing algorithm c5. 0, *CEUR Workshop Proceedings* 2500 (2019).
- [15] A. Ambainis, N. Nahimovs, Improved constructions of quantum automata, *Theoretical Computer Science* 410 (2009) 1916–1922.
- [16] F. Ablayev, A. Vasiliev, On quantum realisation of boolean functions by the fingerprinting technique, *Discrete Mathematics and Applications* 19 (2009) 555–572.
- [17] F. Ablayev, A. Gainutdinova, K. Khadiev, A. Yakaryılmaz, Very narrow quantum OBDDs and width hierarchies for classical OBDDs, in: *DCFS*, volume 8614 of *LNCS*, Springer, 2014, pp. 53–64.
- [18] F. Ablayev, A. Gainutdinova, K. Khadiev, A. Yakaryılmaz, Very narrow quantum OBDDs and width hierarchies for classical OBDDs, *Lobachevskii Journal of Mathematics* 37 (2016) 670–682.
- [19] F. Ablayev, A. Ambainis, K. Khadiev, A. Khadieva, Lower bounds and hierarchies for quantum memoryless communication protocols and quantum ordered binary decision diagrams with repeated test, In *SOFSEM 2018*, *LNCS* 10706 (2018) 197–211.
- [20] F. Ablayev, M. Ablayev, K. Khadiev, A. Vasiliev, Classical and quantum computations with restricted memory, *LNCS* 11011 (2018) 129–155.
- [21] K. Khadiev, A. Khadieva, I. Mannapov, Quantum online algorithms with respect to space and advice complexity, *Lobachevskii Journal of Mathematics* 39 (2018) 1210–1220.
- [22] K. Khadiev, A. Khadieva, Reordering method and hierarchies for quantum and classical ordered binary decision diagrams, in: *CSR 2017*, volume 10304 of *LNCS*, Springer, 2017, pp. 162–175.
- [23] R. Ibrahimov, K. Khadiev, K. Prūsis, A. Yakaryılmaz, Error-free affine, unitary, and probabilistic OBDDs, *Lecture Notes in Computer Science* 10952 *LNCS* (2018) 175–187.
- [24] F. Le Gall, Exponential separation of quantum and classical online space complexity, *Theory of Computing Systems* 45 (2009) 188–202.
- [25] K. Khadiev, A. Ilikaev, Quantum algorithms for the most frequently string search, intersection of two string sequences and sorting of strings problems, in: *International Conference on Theory and Practice of Natural Computing*, 2019, pp. 234–245.
- [26] D. Kravchenko, K. Khadiev, D. Serov, R. Kapralov, Quantum-over-classical advantage in solving multiplayer games, *Lecture Notes in Computer Science* 12448 (2020) 83–98.
- [27] A. Glos, N. Nahimovs, K. Balakirev, K. Khadiev, Upperbounds on the probability of finding marked connected components using quantum walks, *Quantum Information Processing* 20 (2021) 1–23.

- [28] A. Ambainis, K. Balodis, J. Iraids, K. Khadiev, V. Klevickis, K. Prūsis, Y. Shen, J. Smotrovs, J. Vihrovs, Quantum Lower and Upper Bounds for 2D-Grid and Dyck Language, in: 45th International Symposium on Mathematical Foundations of Computer Science (MFCS 2020), volume 170 of *Leibniz International Proceedings in Informatics (LIPIcs)*, 2020, pp. 8:1–8:14.
- [29] K. Khadiev, A. Khadieva, D. Kravchenko, A. Rivosh, R. Yamilov, I. Mannapov, Quantum versus classical online streaming algorithms with logarithmic size of memory, *Lobachevskii Journal of Mathematics* (2019). (in print). arXiv:1710.09595.
- [30] K. Khadiev, A. Khadieva, Quantum online streaming algorithms with logarithmic memory, *International Journal of Theoretical Physics* (2019). doi:10.1007/s10773-019-04209-1.
- [31] K. Khadiev, A. Khadieva, Two-way quantum and classical machines with small memory for online minimization problems, in: International Conference on Micro- and Nano-Electronics 2018, volume 11022 of *Proc. SPIE*, 2019, p. 110222T. doi:10.1117/12.2522462.
- [32] K. Khadiev, A. Khadieva, Two-way quantum and classical automata with advice for online minimization problems, in: Formal Methods. FM 2019 International Workshops, 2020, pp. 428–442.
- [33] Q. Yuan, Quantum online algorithms, Ph.D. thesis, University of California, Santa Barbara, 2009.
- [34] G. Cormode, M. Hadjieleftheriou, Finding frequent items in data streams, *Proceedings of the VLDB Endowment* 1 (2008) 1530–1541.
- [35] S. Muthukrishnan, Data streams: Algorithms and applications, *Foundations and Trends in Theoretical Computer Science* 1 (2005) 117–236.
- [36] C. C. Aggarwal, Data streams: models and algorithms, volume 31, Springer Science & Business Media, 2007.
- [37] L. Becchetti, I. Chatzigiannakis, Y. Giannakopoulos, Streaming techniques and data aggregation in networks of tiny artefacts, *Computer Science Review* 5 (2011) 27–46.
- [38] J. Boyar, K. S. Larsen, A. Maiti, The frequent items problem in online streaming under various performance measures, *International Journal of Foundations of Computer Science* 26 (2015) 413–439.
- [39] R. Kapralov, K. Khadiev, J. Mokut, Y. Shen, M. Yagafarov, Fast classical and quantum algorithms for online k -server problem on trees, arXiv preprint arXiv:2008.00270 (2020).
- [40] G. M. Adel'son-Vel'skii, E. M. Landis, An algorithm for organization of information, in: *Doklady Akademii Nauk*, volume 146, Russian Academy of Sciences, 1962, pp. 263–266.
- [41] T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein, *Introduction to Algorithms*, McGraw-Hill, 2001.
- [42] L. J. Guibas, R. Sedgwick, A dichromatic framework for balanced trees, in: *Proceedings of SFCS 1978*, IEEE, 1978, pp. 8–21.
- [43] C. H. Bennett, E. Bernstein, G. Brassard, U. Vazirani, Strengths and weaknesses of quantum computing, *SIAM journal on Computing* 26 (1997) 1510–1523.

Graph model of Fog Computing system

Andriy V. Ryabko^a, Oksana V. Zaika^a, Roman P. Kukharchuk^a and Tetiana A. Vakaliuk^{b,c}

^aOlexander Dovzhenko Glukhiv National Pedagogical University, 24 Kyievo-Moskovska Str., Glukhiv, 41400, Ukraine

^bZhytomyr Polytechnic State University, 103 Chudnivsyka Str., Zhytomyr, 10005, Ukraine

^cInstitute of Information Technologies and Learning Tools of the NAES of Ukraine, 9 M. Berlynskoho Str., Kyiv, 04060, Ukraine

Abstract

The development and effective application of Fog Computing technologies require the most complex tasks related to the management and processing of huge data sets, including the tasks of rational construction of low-level networks that ensure the functioning of end devices within the IoT concept. The article describes the use of graph theory methods to solve such problems. The proposed graph model can provide the ability to determine the basic properties of systems, networks, and network devices within the concept of Fog Computing, the optimal characteristics, and ways to maintain them in working condition. This paper shows how to plot graphs, and then customize the display to add labels or highlighting to the graph nodes and edges of pseudo-random task graphs which can be used for evaluating Mobile Cloud, Fog and Edge computing systems. The graphs are described and visualized in Matlab code. Each task has an amount of computational work to perform, expressed in Megacycles per second. Each edge has an amount of data to transfer between tasks, expressed in kilobits or kilobytes of data. The set can be used by researchers to evaluate cloud/fog/edge computing systems and computational offloading algorithms. The task graphs can be used in single-user systems, where one mobile device accesses a remote server, or in multi user systems, where many users access a remote server through a wireless channel.

Keywords

fogging computing, multi-level graph model, Internet of Things, Fog Computing, reference architecture OpenFog, graph theory

1. Introduction

Nowadays, at the same time as the rapid development of industrial and built electronics will lead to the fact that traditional equipment used in production processes and at home is more than more provided intellectual functions and objects in the network that requires virtually

QualnT 2021: Workshop on the Quantum Information Technologies, April 11, 2021, Zhytomyr, Ukraine

doors 2021: Edge Computing Workshop, April 11, 2021, Zhytomyr, Ukraine

✉ ryabko@meta.ua (A.V. Ryabko); ksuwazaika@gmail.com (O.V. Zaika); kyxap4yk1@ukr.net (R.P. Kukharchuk); tetianavakaliuk@gmail.com (Tetiana A. Vakaliuk)

🌐 <http://pfm.gnpu.edu.ua/index.php/struktura1/2015-04-01-14-50-26> (A.V. Ryabko);

<http://pfm.gnpu.edu.ua/index.php/struktura1/2015-04-01-14-50-26> (O.V. Zaika);

<http://pfm.gnpu.edu.ua/index.php/struktura1/2015-04-01-14-50-26> (R.P. Kukharchuk);

<https://sites.google.com/view/neota/profile-vakaliuk-t> (Tetiana A. Vakaliuk)

🆔 0000-0001-7728-6498 (A.V. Ryabko); 0000-0002-8479-9408 (O.V. Zaika); 0000-0002-7588-7406 (R.P. Kukharchuk);

0000-0001-6825-4697 (Tetiana A. Vakaliuk)

© 2021 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



CEUR Workshop Proceedings (CEUR-WS.org)

continuous data processing, growing and requiring computing power of this equipment. The need to control and manage individual mechanisms and machines, as well as the environment, transport flows, production, business, health and education, security, social processes has led to the creation of a large number of devices that interact with people, central data messages and between himself. For their effective functioning, it was necessary to create a global communication system, the quality of the natural state by visiting the Internet, the need to solve management and control problems to create a network concept focused on connecting devices and received from the Internet of Things, IoT.

The term Internet of Things was first used in 1999 by Kevin Ashton, head of the Massachusetts Institute of Technology's Auto-ID Research Center, who suggested that Procter & Gamble use radio frequency tags (RFID) on products to improve supply chain management [1]. Later, this name became commonplace, although its meaning has now expanded significantly.

It should be noted that within the framework of the IoT concept, the basis for the effective operation of both end devices with intelligent functions and the networks formed by them is a real-time operation, reliability, and security, which can not always be provided using a client-server interaction scheme, which is typical of the classic Internet and is based on cloud technology and computing (Cloud Computing) [2].

The concept of Fog Computing is designed primarily to bring data processing and storage closer to the devices that generate and use them. A modified concept of cloud technology applicable to IoT, called Fog Computing (FC), was proposed by Cisco researchers. According to the concept, the FC architecture is three-tiered. At the lowest level – the earth – are billions of things, at the top – many cloud data centers that provide resources for applications that require significant computing power and/or significant amounts of data. And, accordingly, between them is Fog – tens of thousands of geographically distributed smaller control centers, sufficient to solve local problems.

As mentioned above, the concept of Fog computing, in contrast to cloud computing, involves the data processing to the end devices of networks – computers, mobile devices, sensors, and more. It should be noted that now the term Edge Computing has become widespread, which in essence is quite close to Fog Computing, as it also involves the implementation of key operations to collect data processing outside the cloud. However, the key difference between Fog Computing and Edge Computing is considered to be the degree of convergence of data processing points to the edge devices of networks. If the concept of Fog involves sending data generated by end devices for processing and/or storage in local processing centers (Fog Nodes), then in the concept of Edge Computing the main tasks of data processing are solved directly on end devices.

Considering the differences between Cloud Computing and Fog Computing, it should be noted that from the authors' point of view, despite the differences, the contrast between the Fog/Edge Computing models and the Cloud Computing model is erroneous and should not be considered as alternatives but complementary. The basic idea of the Fog Computing concept is to ensure efficient, reliable, and secure interaction of a huge number of devices with each other, with local data centers, and with cloud data centers.

Fog Computing is characterized by the use by users of service functions of resources located on peripheral devices and in the distributed network. The data is located on the client nodes where they should be processed or nearby. The main method of data collection and transmis-

sion is wireless communication. The advantage of networks built according to the concept of Fog Computing over cloud systems is to reduce the latency of the response to the collected data by processing at the place of collection, and for real-time systems, this is one of the key factors. Also, in most cases, the security of Fog Computing systems is higher than for Cloud Computing systems. Another advantage of Fog Computing is the reduction in the amount of data transmitted to the cloud, which reduces network bandwidth requirements, increases data processing speed, and reduces decision delays. Thus, the use of Fog Computing allows you to completely or partially solve some of the most common problems, including:

- large network delays,
- difficulties associated with the mobility of endpoints,
- reliability of communication,
- the high cost of bandwidth;
- unpredictable network congestion,
- large geographical distribution of systems and customers.

OpenFog Consortium, founded in 2015, has proposed a specification of the OpenFog reference architecture, a universal technology model for projects primarily in the field of the IoT, mobile networks, and AI applications. The key members of the consortium are Cisco, Intel, ARM, Dell, Microsoft. According to the proposed model, OpenFog-infrastructure is a set of nodes (Fog Nodes) based on network smart devices that perform data processing, the specification also contains descriptions of options for hierarchical node construction, system deployment models, and examples of possible implementations.

The OpenFog reference architecture is based on the following eight technological principles (criteria):

- security,
- scalability,
- openness,
- autonomy,
- RAS (reliability, availability, suitability for service);
- adaptability;
- hierarchical principle of construction of input elements;
- programmability.

According to Cisco developers, the concept of Fog Computing is best suited for working with machine-to-machine (M2M) systems and devices that use a human-machine interface (HMI). They distinguish three main groups of such devices:

- Data acquisition devices are generated in series by different sensors with a frequency of a few milliseconds to fractions of a second. Examples are devices of security systems and control systems of industrial facilities. They are characterized by low latency requirements for data acquisition and high performance to calculate the required characteristics in real-time.
- Systems responsible for data processing, including operating. Here the input data comes with a frequency of a few seconds to a few minutes. Examples of such systems are devices for visualization of physical processes, technological industrial systems. The requirements for the latency of the collected data and their processing are not as high as for the devices of the first group, but all data are processed in real-time.
- Devices for the collection and processing of historical data, collected at a frequency of several minutes to several days. Example – visualization and reporting systems.

2. Theoretical background

Shanhe Yi, Zijiang Hao, Zhengrui Qin and Qun Li point to unresolved cloud computing issues such as unreliability, latency, lack of mobility support, and location awareness [3]. Fog computing can solve these problems by providing resources and services to end-users at the network boundary, while cloud computing is more about providing resources distributed on the core network. Fog provides IoT data processing and storage locally on IoT devices, instead of sending it to the cloud [4].

Case studies Weisong Shi and Schahram Dustdar [5], Weisong Shi, Jie Cao, Quan Zhang, Youhuizi Li and Lanyu Xu [5], Nasir Abbas, Yan Zhang, Amir Taherkordi and Tor Skeie [6], Mahadev Satyanarayanan [7] provide a detailed description, definition, and capabilities of Edge-Computing, from cloud technology unloading to smart home and city, mobile networks.

Luis Miguel Vaquero and Luis Rodero-Merino [8] define Fog, considering a variety of technologies such as cloud, sensor networks, peer-to-peer networks, network virtualization functions, or configuration management techniques [8]. Redowan Mahmud, Ramamohanarao Kotagiri and Rajkumar Buyya note that Fog computing is located closer to IoT devices/sensors and expands the capabilities of cloud-based computing, storage, and networking technologies [9]. Ivan Stojmenovic, Sheng Wen, Xinyi Huang and Hao Luan point to Smart Grid technologies, smart traffic lights, software-defined networks [10].

Flavio G. Bonomi, Rodolfo A. Milito, Jiang Zhu and Sateesh K. Addepalli indicate the main characteristics of Fog: a) low latency and location awareness; b) wide geographical distribution; c) mobility; d) a very large number of nodes; e) the predominant role of wireless access; f) the strong presence of streaming and real-time programs; g) heterogeneity [11]. These features make Fog an appropriate platform for several critical IoT services and applications, namely, automotive, smart grids, smart cities, and wireless sensors and network devices (WSANs) in general.

The results of studies by Subhadeep Sarkar and Sudip Misra show that for the scenario where 25% of IoT applications will run with low real-time latency, the average energy consumption for Fog calculations will be 40.48% lower, than in the usual model of cloud computing [12].

Expanding the concepts of Fog Computing is the driving force behind the introduction of Industry 4.0. The development of algorithms and optimization methods is complicated by the complexity of such systems and the lack of real data on Fog systems, which leads to the use of algorithms that are not adapted to real scenarios. Graph-based system parameters allow you to scale and design more realistic test scenarios for future optimization attempts, as well as to determine the features of Fog systems compared to other types of networks [13].

Graph theory is used to construct a load balancing algorithm for Fog network computations [14] based on a dynamic graph distribution [15] that the Fog Cloud Atomization computing system can flexibly build a system network, and the dynamic load balancing mechanism can efficiently configure system resources as well as reduce the consumption of node migration caused by system changes.

Some authors propose a Fog network caching scheme based on the Steiner tree, in which Fog servers, by caching resources, first create a Steiner tree in graphs to minimize the total path weight (or cost) so that the cost of resource caching with this tree can be minimized [3, 16].

Xu Chen and Junshan Zhang propose a hybrid HyFog system for unloading tasks in Fog Computing based on a three-level graph for efficient distribution of tasks between devices. The problem of minimizing the total cost of the task is reformulated as the problem of the minimum weight ratio in the constructed three-level graph, which can be effectively solved using the Blossom Edmonds algorithm [17].

Dmitry Korzun, Aleksey Varfolomeyev, Anton Shabaev and Vladimir Kuznetsov consider two emerging IoT-enabled paradigms: Edge-centric Computing and Fog Computing. They are elaborating their potential for development of smart applications with focus on the dependability and using a mobile assistant for e-tourism as a reference application. They analyze possible concept elements for smart application development and provide recommendations in respect to the dependability [18].

Isaac Lera, Carlos Guerrero and Carlos Juiz propose a fog computing simulator for analyzing the design and deployment of applications through customized and dynamical strategies. They model the relationships among deployed applications, network connections, and infrastructure characteristics through complex network theory, enabling the integration of topological measures in dynamic and customizable strategies, such as the placement of application modules, workload location, and path routing and scheduling of services [19].

Ted H. Szymanski presents describes 300 task graphs which can be used for evaluating mobile cloud, fog and edge computing systems. The task graphs are organized as 3 sets of 100 graphs. Each graph in the first set has the same topology, with $N = 9$ tasks and 6 offloadable tasks. Each graph in the second set has the same topology, with $N = 29$ tasks and 20 offloadable tasks. Each graph in the third set has the same topology, with $N = 23$ tasks and 19 offloadable tasks. Users can also change the number of offloadable components per task graph, in which case the total number of task graphs specified in this paper exceeds 5,000, providing a good basis for the evaluating cloud computing systems [20].

3. Research methods

When modeling the operation of both individual elements and the system, designed and/or built on the concept of fog computing as a whole, for example, to assess efficiency, performance, bandwidth, performance, equipment reliability, the delay time for certain types of service, data, software, the system should be presented in the form of several interconnected levels:

- equipment,
- interfaces,
- transport system (network),
- operating systems (OS),
- data,
- services,
- applications.

Such a representation of the object of modeling allows you to use the apparatus of graph theory.

Figure 1 in the form of an oriented graph $G = (Z, L)$ presents a multilevel graph model of the system built according to the concept of Fog Computing. The set of vertices of the graph $Z_{Equip}, Z_{ConDev}, \dots, Z_{ComDev}, \dots; Z_{Interf}, Z_{HMI}, \dots, Z_{M2M}, \dots; Z_{TS}, Z_{NetDev}, \dots, Z_L, \dots; Z_{os}, Z_{os1}, \dots, Z_{osn}, Z_{Data}, Z_{OI}, \dots, Z_{HD}, \dots, Z_{BD}, \dots; Z_S, Z_{S1}, \dots, Z_{Sm}; Z_{App}, Z_{App1}, \dots, Z_{Appk}$ are tasks that are solved at each specific level.

The set of arcs contains: X_{pr} – the set of parameters needed to solve these problems and the set of information links – H . The set of parameters X_{pr} consists of H_{pr}^{in} – input and H_{pr}^{out} – output parameters.

$$X_{pr} = \{H_{pr}^{in}, H_{pr}^{out}\}$$

with $L = X_{pr} \cup H$ and $X_{pr} \cap H = \emptyset$.

It should be noted that the output parameters of some tasks (vertices) may be input for others. Information connections are converted into a Boolean matrix $n \cdot n$:

$$H = \|h_{gl}\|_{n \times n}.$$

The matrix element h_{gl} characterizes the presence of information connections of problems g and l :

$$h_{gl} = \begin{cases} 1 & \text{if the tasks are related by parameters,} \\ 0 & \text{otherwise.} \end{cases}$$

The solution of any problem can be represented as

$$Z_i : \{X_{pr}^{in}\} \Rightarrow Z_{pr}^{out}.$$

It should be noted that there are many mathematical models for solving the problems that correspond to the vertices of the graph $G = (Z, L)$, in the future, they are used to estimate the parameters and characteristics of Fog Computing elements. Each vertex of the graph corresponds to one or more nodes, and the output parameters of some nodes can be input to others. Michaela Iorga, Larry Feldman, Robert Barton, Michael J. Martin, Nedim Goren and Charif Mahmoudi note that Fog nodes are either physical components (e.g., gateways, switches, routers, servers, etc.) or virtual components (e.g., virtual switches, virtual machines, clouds, etc.) that are closely related to intelligent endpoints, devices or access networks, and provide computing resources for these devices [21].

Each parameter for solving problems is characterized by a vector of characteristics of parameters, which includes: units of measurement of the parameter; the vertex in which this parameter is the source; the vertex in which the parameter is input, the level number, and other individual characteristics. The units of measurement of parameters are determined by the International System of Units (SI) and by the semantics of specific modeling tasks. In this case, all parameters should be divided into two groups: 1) parameters corresponding to the SI system; 2) all other parameters.

For the same parameters, the units of measurement may be different; therefore, it is necessary to perform parameter matching. As mentioned above, the OpenFog reference architecture is based on the following eight criteria:

$$K = \{K_1^u, \dots, K_8^u\},$$

where K is a compound criterion; u is the level number ($u = 1, \dots, 7$); K_i^u – criterion of a specific level, for example, security at the level of services.

This method can be used in this study, because using the graph $G = (Z, L)$, it is possible to integrate models for calculating and evaluating various parameters depending on specific tasks, to carry out multivariate calculations, and to effectively evaluate almost all parameters and characteristics of Fog Computing elements. The graph model clearly shows the mutual influence and interrelation of tasks and parameters, and modeling can be started with any task that corresponds to the top of the graph (one or more).

4. Results

Let us dwell in more detail on the results of modeling the graph of a system designed and/or built on the concept of Fog Computing (figure 1).

1st level – equipment. The vertices of the graph Z_{Equip} , Z_{ConDev} , ..., Z_{ComDev} , ... correspond to this level. The vertex of Z_{Equip} defines all problems which are solved at this level; vertices of Z_{ConDev} , ... – (connecting devices) tasks that are solved at the level of a wide range of devices connected to the network; vertices Z_{ComDev} , ... – tasks for devices that allow you to perform the necessary calculations. An example of the tasks that are solved at this level is the calculation of the reliability and performance of devices.

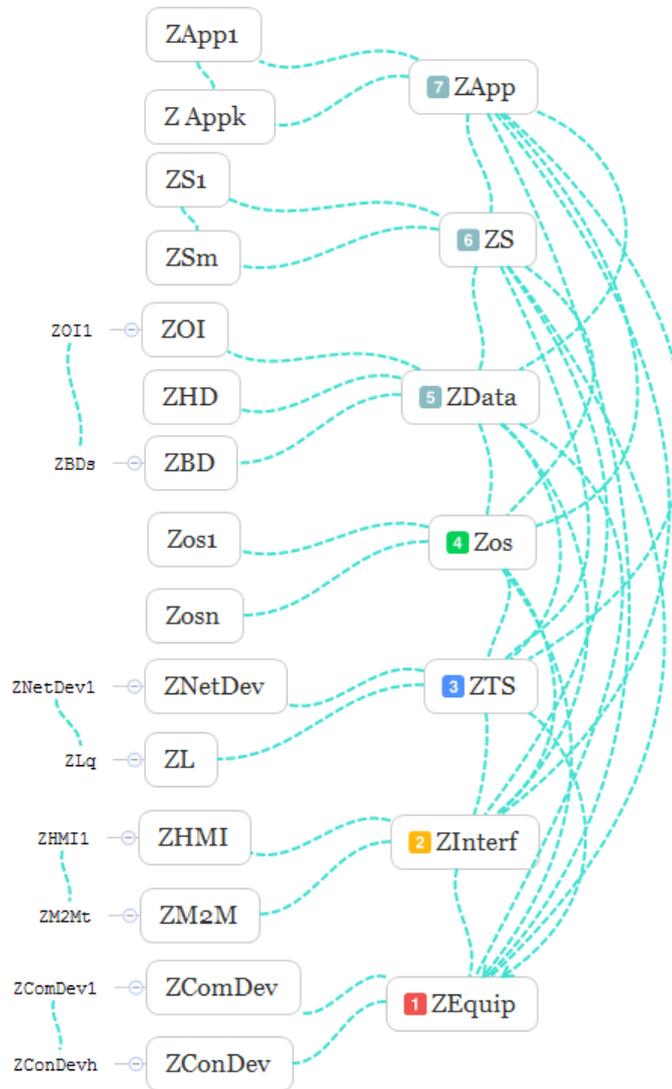


Figure 1: Graph model of the system, which is implemented according to the concept of Fog Computing

2nd level – interfaces. As mentioned above, Fog calculations are best suited for working with inter-machine interaction systems – M2M, and devices that use a human-machine interface – HMI. The interface level corresponds to the vertices of the graph Z_{Interf} , Z_{HMI} , ..., Z_{M2M} , The vertex of the Z_{Interf} describes the tasks specific to this level; vertices Z_{HMI} ,

..., Z_{M2M} , ... – define tasks for systems with inter-machine and human-machine interface, respectively. M2M – machine-machine interaction (Machine-to-Machine, Mobile-to-Machine, Machine-to-Mobile) – the name of the technology (sum of technologies), which allows data transfer between different devices, and it can be groups of devices, such as public transport. Work on M2M is coordinated by the following organizations: the Eclipse Foundation, the Focus Groupon Machine-to-Machine, a member of the International Telecommunication Union, and the TR-50 M2M Intelligent Devices Engineering Committee.

Types of M2M: stationary M2M, such as process control, payment terminals, meters, etc., and mobile M2M, for example, for fleet management, where M2M is used as an on-board device for monitoring, diagnostics, navigation, security, and mobile communications. M2M applications include access systems, premises security systems, security systems, remote control and management of equipment, transport and monitoring of moving objects, vending machines, payment terminals, healthcare, etc. HMI – human-machine interface – a concept that encompasses engineering solutions that provide human interaction with controlled objects (machines, systems, devices).

The HMI can be a computer, standard software, a simple remote control with a set of LED indicators, and so on. Modern computers are focused on streaming architecture, the implementation of intelligent human-machine interface, which provides not only a systematic solution but also the ability of the machine to logical thinking and self-learning, to associative information processing and drawing logical conclusions. The requirements that different users impose on the HMI vary widely. The implementation of an intelligent human-machine interface is associated with the ability to solve problems of recognition and understanding of natural language, for this, there are recognition systems (language, handwritten texts, images). Creating a user-friendly and efficient human-machine interface is an urgent task. Also among the tasks of the interface level can be distinguished, for example, the choice of standard interface buses; reliability assessment at the interface level, and others.

3rd level – transport system (network). The transport system is used to transmit information and contains nodes of the Fog infrastructure – switches, routers, etc. The vertex Z_{TS} of the graph $G = (Z, L)$ defines the general tasks characteristic of the 3rd level, and the vertices Z_{NetDev} , ... – describe the tasks that are solved at the level of network devices; vertices Z_L , ..., Z_{Lq} , ... – tasks that are solved at the level of communication channels. As the tasks are solved at this level it is possible to result in the following – a choice of a communication channel; channel bandwidth estimation; calculation of the delay factor of network equipment; estimation of message delay and many others.

4th level – operating systems (OS). Here you should consider the presence of different types of operating systems (UNIX-like OS, Windows, macOS, etc.). The vertex Z_{os} of the graph describes the general tasks characteristic of the OS level.

Vertices Z_{os1} , ..., Z_{osn} are tasks that are solved for each specific operating system, for example:

- calculation of the coefficient of relative losses of OS performance for a multiprocessor system,
- determining the average processing time of the OS request,
- estimation of the average time spent on access to external memory and analysis of the

- intensity of OS requests to external memory devices,
- assessment of the reaction time of the OS in solving specific problems,
- an estimate of the average time required to transmit the OS request,
- estimate the time of access to RAM,
- optimization of the core structure of open OS by the criterion of information security,
- estimation of time of detection of errors in processes,
- calculation of the probability of skipping the controlled signal (quantitative characteristics for the tasks of monitoring the integrity of OS files) and many others.

5th level – data. The vertices $Z_{Data}, Z_{OI}, \dots, Z_{HD}, \dots, Z_{BD}, \dots$ of the graph $G = (Z, L)$ correspond to this level. The Z_{Data} vertex describes general tasks, $Z_{OI}, \dots, Z_{HD}, \dots, Z_{BD}, \dots$ vertices – tasks specific to operational information (real-time analysis), historical data (transaction analysis), and long-term storage (BigData analysis).

Examples of tasks to be solved at this level:

- prognostic calculation of the speed of new data generation,
- optimization of file placement and processing of requests to the database,
- estimation of data volume,
- data compression,
- distributed calculations when planning requests to the database,
- assessment of the integrity of information at the level of links and other tasks.

6th level – services. These can be various services, such as online services (like Uber), streaming services (like Netflix, Amazon Prime, Hulu, and Crunchyroll), etc.

The level of services corresponds to the vertices $Z_S, Z_{S1}, \dots, Z_{Sm}$ graph model. Vertex Z_S defines general tasks for service level, vertices Z_{S1}, \dots, Z_{Sm} – tasks for different types of service.

Examples of tasks:

- calculation of productivity for the 6th level,
- assessment of service quality in virtual VPN channels,
- optimization of system services by network resources,
- assessment of the security of transmission of confidential information in broadcast communication channels,
- maximum support for different types of 6th level traffic, etc.

7th level – applications. Applications are research software, computer-aided design systems, games, applications for artists, geographically distributed applications for pipeline monitoring, smart devices in the car, SmartGrid, traffic light control systems, etc.

The vertices Z_{App} , Z_{App1} , ..., Z_{Appk} of the graph $G = (Z, L)$ correspond to this level. Vertex Z_{App} describes the general tasks of the application level, the vertices of Z_{App1} , ..., Z_{Appk} – tasks for different types of applications. These can be applications installed on computers, tablets, smartphones of users.

Some examples of tasks:

- calculation of maximum productivity for the 7th level,
- distribution of application tasks between users according to the criterion of weighted average route length,
- prognostic estimate of the conditional average service time of the application required to perform the task lasting in nt period,
- estimation of the average time of the decision of applied problems,
- calculation of exchange time with external memory in the process of solving applied corporate tasks,
- prognostic calculation of the time required for data processing in the application system,
- calculation of the average service time of the application for algorithms of non-priority service disciplines,
- scalability for the 7th level.

The main properties of the multilevel graph model of Fog Computing system are:

1. integration – the ability to solve individual (partial) problems depending on the specific situation,
2. universality,
3. adequacy,
4. accuracy,
5. efficiency,
6. property of development – the model is created and functions taking into account additions, improvements, and updates.

The requirements of a high degree of universality, accuracy, adequacy on the one hand, and its high efficiency, on the other hand, are contradictory.

In the process of analyzing the graph model of the system, which is implemented by the concept of Fog Computing, the graph $G = (Z, L)$ is decomposed into typical subgraphs (the necessary subgraphs are allocated to save time and computing resources), the parameters of the corresponding vertices are determined by set of connections between vertices (tasks). Then there is a selection of appropriate models (formulas) to calculate the necessary parameters for solving specific problems. Next, you need to analyze the entire graph model. Using graph

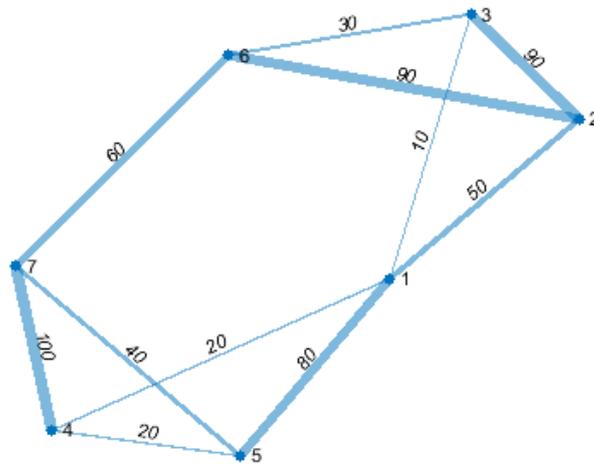


Figure 2: Graph with the width of the edges, which is proportional to their weight

$G = (Z, L)$ it is possible to integrate models for calculation and estimation of various parameters depending on concrete tasks, to carry out various calculations, to estimate practically all parameters and characteristics of Fog Computing system elements. The graph model clearly shows the mutual influence and relationship of tasks and parameters, and modeling can begin with any problem that corresponds to the vertex of the graph (one or more).

Fog Computing network simulations were performed in Matlab. Since the 2015 release, Matlab has been able to work with graphs using the $G = graph()$ function. After you create a graph that simulates a network, you can get information about the graph by using object functions to perform object queries. For example, you can add or remove nodes or edges, define the shortest path between two nodes, or find a specific graph node:

```
G = graph([1 1], [2 3]);
e = G.Edges
G = addedge(G, 2, 3)
G = addnode(G, 4)
plot(G)
```

The next step in describing the relationships between objects using graphs is to give the edges certain symbolic values, qualitative characteristics, called weights. In the simplest cases, this may be the ordinal numbering of the edges, which is checked by the order of their consideration (priority or hierarchy). Rib weight can mean length (message paths), bandwidth (communication lines), load. Weight can be attributed not only to the ribs but also to the vertices. For example, the vertices that correspond to the Fog nodes of the network can characterize the number, bandwidth, and so on. Next, a graph is constructed, indicating the weight of the edges and making the width of the edges proportional to their weight (figure 2).

We used a set of pseudo-random task graphs which can be used for evaluating Mobile Cloud, Fog and Edge Computing systems. The pseudo-random task graphs are based upon graphs that

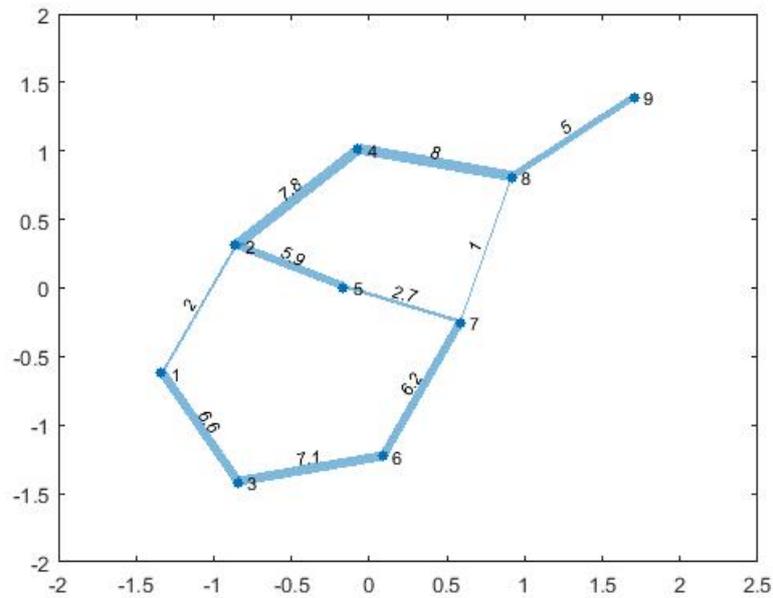


Figure 3: Three task graphs each with $N = 9$ nodes and 6 offloadable tasks. Tasks 1,6 and 9 execute locally. Edge weights are in Kilobits

have previously developed by Ted H. Szymanski [20]. Researchers can download these 300 task graphs and analyse their own configurations of mobile cloud, fog and edge computing systems with different mobile device parameters and cloud/fog/edge server parameters, but the code does not provide the ability to visualize graphs and properties of nodes and edges.

For example, the first set of graphs is based upon a task graph presented in [20]. Reference [20] provided the 3 task graphs shown in figure 3, and it provided the optimal energy for each graph (under certain assumptions). In figure 3, the task graphs consists of 9 tasks and 10 edges. Tasks 1, 6 and 9 must execute locally, and the remaining tasks can execute locally or remotely. The task complexities are expressed in Mega-cycles per task, and the edges are annotated with the data size expressed in kilobits. The task graph in Fig. 3 contains only in one variable, where the complexity of task 8 changes from 3 to 20 Mega-cycles.

Graph plots are the primary way to visualize graphs and networks created using the graph and digraph functions. After you create a GraphPlot object, you can modify aspects of the plot by changing its property values. This is particularly useful for modifying the display of the graph nodes or edges. GraphPlot properties control the appearance and behavior of plotted graphs. By changing property values, you can modify aspects of the graph display. Use dot notation to refer to a particular object and property. Let's plot a graph by marking the edges with their weights and making the widths of the edges proportional to their weights. Use a scaled version of the edge weights to define the width of each edge so that the widest line has a width of 5. This example shows how to plot graphs:

```
EDGE_src(1, 1:10)=[1, 1, 2, 2, 3, 4, 5, 6, 7, 8, ];
```

```
EDGE_dst(1,1:10)=[2,3,4,5,6,8,7,7,8,9,];
EDGE_bits(1,1:10)=[200,660,780,590,710,80,270,620,100,500,];
G=graph(EDGE_src,EDGE_dst,EDGE_bits)
```

```
G =
graph with properties:
Edges: [10x2 table]
Nodes: [9x0 table]
>> LWidths = 5*G.Edges.Weight/max(G.Edges.Weight);
plot(G, 'EdgeLabel',G.Edges.Weight, 'LineWidth',LWidths)
```

The second task graph is an extended version of the previous task graph, and the seed graph is shown in figure 3. Three instances of the task graph from set 1 (in figure 2) are placed in parallel, between the entry node 1 and exit node 29. The seed graph has 29 nodes and 36 edges. Nodes 1, 29 and 7 other randomly selected nodes execute locally. Each task graph in set 2 thus has 20 offloadable tasks. This example shows how to plot graphs:

```
>>EDGE_src(1,13:24)=[1,11,11,12,12,13,14,15,16,17,18,19,];
EDGE_dst(1,13:24)=[11,12,13,14,15,16,18,17,17,18,19,29,];
EDGE_bits(1,25:36)=[200,200,660,780,590,710,80,270,620,100,500,300,];
G=graph(EDGE_src,EDGE_dst,EDGE_bits)
```

```
G=
Graph with properties:
Edges: [36x2table]
Nodes: [29x0table]
>>LWidths=5*G.Edges.Weight/max(G.Edges.Weight);
plot(G, 'EdgeLabel',G.Edges.Weight, 'LineWidth',LWidths)
```

The first task graph in set 2 is identical to the one shown in figure 4.

The practical application of the model can, for example, help mobile devices overcome resource constraints by unloading computing tasks on cloud servers. The task of the cloud is to minimize the time of data transfer and execution of tasks to the user, whose location changes due to the mobility and power consumption of the mobile device.

Ensuring satisfactory computational performance is particularly difficult in Fog Computing. The graph model of calculations will allow bringing effectively computing power Fog to the mobile user. The graph model consists of remote cloud nodes and local cloud nodes that are connected to the wireless access infrastructure. Evaluating the effectiveness of our method using experimental modeling in Matlab shows good results, which show that this method allows you to calculate the ability to reduce task execution time and power consumption of mobile devices.

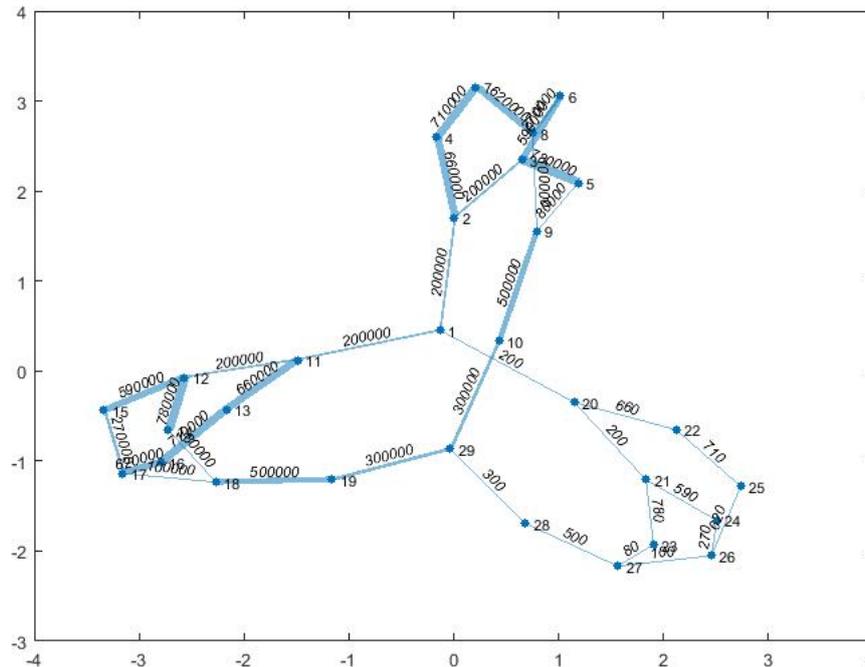


Figure 4: Task graph with $N = 29$ nodes and 20 offloadable tasks based upon previous figure

5. Conclusions

The developed graph model can be a very versatile tool for optimizing existing networks and for implementing new systems and networks created by the concept of Fog Computing, and, above all, focused on the IoT. The IoT and Fog computing is just entering everyday life, and the model provides the ability to calculate, perform a prognostic assessment, optimize the necessary characteristics of the entire system and individual elements, identify bottlenecks and redundancy at the design stage taking into account OpenFog reference architecture, criteria (technological principles) and types of data processing systems. The proposed graph model can provide the ability to determine the basic properties of systems, networks, and network devices within the concept of Fog Computing, the optimal characteristics, and ways to maintain them in working condition. A promising task is to create an algorithm for calculating the shortest path between nodes when dynamically changing the weight of the edges of the graph (for example, when changing the location of a mobile device) to solve practical problems of unloading mobile resources.

This paper shows how to plot graphs, and then customize the display to add labels or highlighting to the graph nodes and edges of pseudo-random task graphs which can be used for evaluating Mobile Cloud, Fog and Edge computing systems. The graphs are described and visualized in Matlab code.

We plan to develop a set of use-case scenarios that we analyze to determine the graph based parameters of the system that allows us to scale and generate a more realistic testing scenario for future optimization attempts as well as determine the nature of fog systems in comparison to other networks types.

References

- [1] T. Kramp, R. van Kranenburg, S. Lange, *Introduction to the Internet of Things*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 1–10. doi:10.1007/978-3-642-40403-0_1.
- [2] O. Markova, S. Semerikov, A. Striuk, H. Shalatska, P. Nechypurenko, V. Tron, Implementation of cloud service models in training of future information technology specialists, *CEUR Workshop Proceedings* 2433 (2019) 499–515.
- [3] S. Yi, Z. Hao, Z. Qin, Q. Li, Fog computing: Platform and applications, in: 2015 Third IEEE workshop on hot topics in web systems and technologies, HotWeb, 2015, pp. 73–78. doi:10.1109/HotWeb.2015.22.
- [4] H. Atlam, R. Walters, G. Wills, Fog computing and the internet of things: A review, *Big data and cognitive computing* 2 (2018) 10. doi:10.3390/bdcc2020010.
- [5] W. Shi, J. Cao, Q. Zhang, Y. Li, L. Xu, Edge computing: Vision and challenges, *IEEE Internet of things journal* 3 (2016) 637–646. doi:10.1109/JIOT.2016.2579198.
- [6] N. Abbas, Y. Zhang, A. Taherkordi, T. Skeie, Mobile edge computing: A survey, *IEEE Internet of Things Journal* 5 (2017) 450–465. doi:10.1109/JIOT.2017.2750180.
- [7] M. Satyanarayanan, The emergence of edge computing, *Computer* 50 (2017) 30–39. doi:10.1109/MC.2017.9.
- [8] L. Vaquero, L. Rodero-Merino, Finding your way in the fog: Towards a comprehensive definition of fog computing, *ACM SIGCOMM Computer Communication Review* 44 (2014) 27–32. doi:10.1145/2677046.2677052.
- [9] R. Mahmud, R. Kotagiri, R. Buyya, Fog computing: A taxonomy, survey and future directions, *Internet of everything* (2018) 103–130. doi:10.1007/978-981-10-5861-5_5.
- [10] I. Stojmenovic, S. Wen, X. Huang, H. Luan, An overview of fog computing and its security issues, *Concurrency and Computation: Practice and Experience* 28 (2016) 2991–3005. doi:10.1002/cpe.3485.
- [11] F. Bonomi, R. Milito, J. Zhu, S. Addepalli, Fog computing and its role in the internet of things, in: *Proceedings of the first edition of the MCC workshop on Mobile cloud computing*, 2012, pp. 13–16. doi:10.1145/2342509.2342513.
- [12] S. Sarkar, S. Misra, Theoretical modelling of fog computing: a green computing paradigm to support iot applications, *IET Networks* 5 (2016) 23–29. doi:10.1049/iet-net.2015.0034.
- [13] N. Verba, K. Chao, A. James, J. Lewandowski, X. Fei, C. Tsai, Graph analysis of fog computing systems for industry 4.0, in: 2017 IEEE 14th International Conference on e-Business Engineering (ICEBE), 2017, pp. 46–53. doi:10.1109/ICEBE.2017.17.
- [14] I. Lera, C. Guerrero, C. Juiz, Availability-aware service placement policy in fog computing

- based on graph partitions, *IEEE Internet of Things Journal* 6 (2018) 3641–3651. doi:10.1109/JIOT.2018.2889511.
- [15] S. Ningning, G. Chao, A. Xingshuo, Z. Qiang, Fog computing dynamic load balancing mechanism based on graph repartitioning, *China Communications* 13 (2016) 156–164. doi:10.1109/CC.2016.7445510.
- [16] S. Yi, C. Li, Q. Li, A survey of fog computing: concepts, applications and issues, in: *Proceedings of the 2015 workshop on mobile big data*, 2015, pp. 37–42. doi:10.1145/2757384.2757397.
- [17] X. Chen, J. Zhang, When d2d meets cloud: Hybrid mobile task offloading in fog computing, in: *2017 IEEE international conference on communications (ICC)*, 2017, pp. 1–6. doi:10.1109/ICC.2017.7996590.
- [18] D. Korzun, A. Varfolomeyev, A. Shabaev, V. Kuznetsov, On dependability of smart applications within edge-centric and fog computing paradigms, in: *2018 IEEE 9th International Conference on Dependable Systems, Services and Technologies (DESSERT)*, 2018, pp. 502–507. doi:10.1109/DESSERT.2018.8409185.
- [19] I. Lera, C. Guerrero, C. Juiz, Yafs: A simulator for iot scenarios in fog computing, *IEEE Access* 7 (2019) 91745–91758. doi:10.1109/ACCESS.2019.2927895.
- [20] T. H. Szymanski, 300 pseudo-random task graphs for evaluating mobile cloud, fog and edge computing systems, 2018. doi:10.21227/kak5-8n96.
- [21] M. Iorga, L. Feldman, R. Barton, M. Martin, N. Goren, C. Mahmoudi, Fog computing conceptual model, *Natl. Inst. Stand. Technol. Spec.* (2018) 1–15. doi:10.6028/NIST.SP.500-325.

Architecture for edge devices for diagnostics of students' physical condition

Tetiana M. Nikitchuk^a, Tetiana A. Vakaliuk^{a,b}, Oksana A. Chernysh^a,
Oksana L. Korenivska^a, Liudmyla A. Martseva^a and Viacheslav V. Osadchyi^c

^aZhytomyr Polytechnic State University, 103 Chudnivsyka Str., Zhytomyr, 10005, Ukraine

^bInstitute of Information Technologies and Learning Tools of the NAES of Ukraine, 9 M. Berlynskoho Str., Kyiv, 04060, Ukraine

^cBogdan Khmelnytsky Melitopol State Pedagogical University, 20 Hetmanska Str., Melitopol, 72300, Ukraine

Abstract

The article investigates the possibility of technical realization of hardware complex. It presupposes the use of sensors of registration of a photoplethysmographic curve, which describes a pulse wave and defines the parameters of students' cardiovascular system functional state. The method of photoplethysmography allows the use of non-contact sensors. Therefore, there is no artery compression, which eliminates circulatory disorders and allows the use of calculations to determine the saturation of oxygen by the pulse wave. It is recommended to use several optocouplers connected in series, parallel or parallel-series in a chain, with control of their mode of operation from the intensity of the received pulse wave signal depending on human body constitution. The edge device hardware is a part of the IoT system, which also includes another edge device, which instantly transmits data to the database on the edge server for the data further processing and storage.

Keywords

pulse wave, saturation, edge device, sensor, biotechnical system, photoplethysmography, photoplethysmograph

1. Introduction

2020 is the year of the COVID-19 pandemic [1], which forced people to change their attitude to health. In the period of morbidity, when the number of the infected is constantly increasing

QualnT 2021: Workshop on the Quantum Information Technologies, April 11, 2021, Zhytomyr, Ukraine
doors 2021: Edge Computing Workshop, April 11, 2021, Zhytomyr, Ukraine

✉ tnikitchuk@ukr.net (T.M. Nikitchuk); tetianavakaliuk@gmail.com (T.A. Vakaliuk); chernyshoxana@gmail.com (O.A. Chernysh); o.l.korenivska@gmail.com (Oksana L. Korenivska); l.a.martseva@gmail.com (L.A. Martseva); poliform55@gmail.com (V.V. Osadchyi)

🌐 <https://ztu.edu.ua/ua/structure/faculties/fikt/krt.php> (T.M. Nikitchuk); <https://sites.google.com/view/neota/profile-vakaliuk-t> (T.A. Vakaliuk); <https://ztu.edu.ua/ua/structure/pv/> (O.A. Chernysh); https://ztu.edu.ua/ua/structure/faculties/fikt/teachers_krt.php (Oksana L. Korenivska); https://ztu.edu.ua/ua/structure/faculties/fikt/teachers_kikm.php (L.A. Martseva); <http://osadchyi.mdpu.org.ua> (V.V. Osadchyi)

🆔 0000-0002-9068-931X (T.M. Nikitchuk); 0000-0001-6825-4697 (T.A. Vakaliuk); 0000-0002-2010-200X (O.A. Chernysh); 0000-0002-3735-7690 (Oksana L. Korenivska); 0000-0001-5037-6565 (L.A. Martseva); 0000-0001-5659-4774 (V.V. Osadchyi)

© 2021 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
 CEUR Workshop Proceedings (CEUR-WS.org)



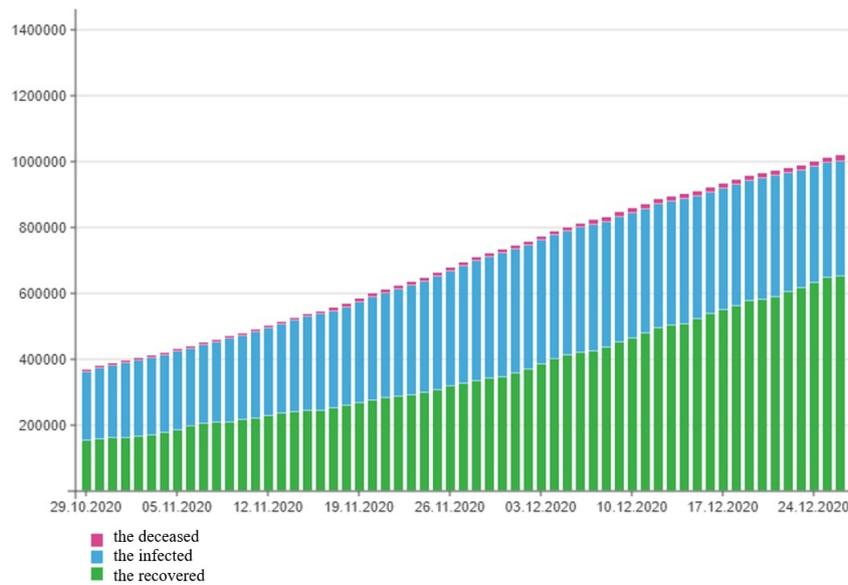


Figure 1: The number of people infected with COVID-19 at the end of 2020

exponentially (figure 1), early detection of certain abnormalities in health is a precautionary measure.

One of the requirements to participate in any event as well as attend classes is a satisfactory health condition. Therefore, in an educational institution, the problem of determining students' state of health arises.

Due to the COVID-19 outbreak, it has become impossible to monitor the health condition of a student body. Therefore, it is proposed to develop edge devices, the components of which will be partially located in classrooms.

The system screens and monitors the functional parameters of students' cardiovascular system and other organs for coronavirus symptoms, pre-existing and health condition. It displays the result on the edge device screen, or transfers it server, or a mobile device.

In the last decades of the last century – at the beginning of the current for the functional diagnosis of the cardiovascular and respiratory systems and, to some extent, the functional features of the circulatory system, pulse oximeters have become widespread. These are devices designed to determine the relative saturation of blood hemoglobin with oxygen in the natural conditions of human life and the dynamics of its performance of various functional tests [2, 3, 4].

In the middle – the last quarter of the 20th century photooxyhemographs were widely used to solve the problem mentioned above. Modern integrated pulse oximeters, in contrast, allow obtaining high-quality curves of peripheral arterial pulse due to periodic heart activity – photoplethysmograms (PPG) [2, 5].

The contour and form of photoplethysmograms greatly resembles the peripheral pulse curve (sphygmogram) obtained with mechanical pressure sensors that convert the oscillations of the wall of the pulsating artery into an electrical signal [2, 4, 5]. Therefore, pulse oximetry can

be used as a source of primary biological information about heart rate and natural heart rate variability [5].

Edge devices are viewed as a complex automated system [6]. It combines memory hardware (considered in the paper), data transmission and visualization unit, and the database. The database contains students' medical records, medical check-up data, and the data of patients' health condition monitoring.

1.1. Theoretical background

Prior researches prove that the introduction of ICT in the field of health care greatly contributes to health promotion and maintenance [7, 8, 9, 10, 11, 12, 13, 14]. Moreover, it improves the demographic situation, upgrades the quality and efficiency of health care. Furthermore, it ensures the human rights to health care [7]. V. Avramenko, V. Kachmar and A. Khvyshchun [7, 15, 16] have made a significant scientific and practical contribution to the introduction of modern information technologies in the educational process as well as in health care institutions work in Ukraine.

A closer look to the literature on medical field informatization, however, reveals a number of gaps and shortcomings. Although there are many studies, the research in the assessment of medical information systems effectiveness remains limited. Nonetheless, there exists a considerable body of literature on organizational and economic efficiency of industrial, corporate, accounting and other types of information systems introduction in large enterprises, government agencies, and in the tourism industry [7, 17, 18, 19, 20, 21].

The study addresses several further questions on edge computing, which is a comparatively new area of research [22]. However, it has been successfully established and described by Najmul Hassan, Saira Gillani, Ejaz Ahmed, Ibrar Yaqoob and Muhammad Imran. The scientists bring some information about the role of edge computing in the internet of things [23]. They propose a layered model for the delivery of IoT services based on CloudEdge, as well as the taxonomy of the Edge Environment based on IoT (see figure 2). Moreover, the researchers provide a clear illustration of cloud computing complementary role in the IoT environment (see figure 3) [23].

A more comprehensive description can be found in the works of Inés Sittón-Candanedo and Juan Manuel Corchado. The scientists consider the concept of Edge Computing, and the possibility of Edge Computing integration. They suggest that it significantly contributes to optimizing the processes that are usually performed in a cloud computing environment [24]. In addition, the scientists successfully establish the relation between Edge and Cloud Computing (figure 4) [24].

Jun-Ho Huh and Yeong-Seok Seo present the framework, preconditions and discuss the advantages and disadvantages of edge calculations. The researchers describe how they function and provide their structure hierarchically with the concepts of artificial intelligence [25]. Moreover, the scientists draw a comparison of the cloud and edge computing paradigm; propose a three-tier edge computing architecture, and develop the design of an edge computing environment with AI support (see figure 5).

The aim of this research is to develop the hardware of edge devices of pulse rapid diagnostics of human body functional state. Its parameters enable to identify the early symptoms of

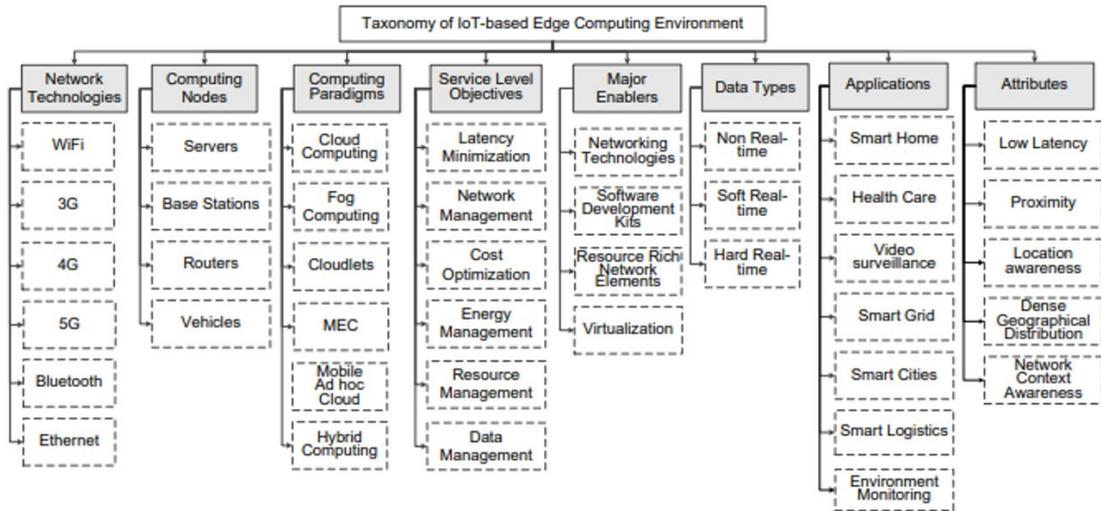


Figure 2: Taxonomy of IoT-based Edge Environment [23]

COVID-19 and determine the functional state of cardiovascular system. The hardware serves as a means of determining the parameters of human body functional state and can be installed in the places of student' study.

2. Results

The hardware complex is located in the classrooms and consists of 2 units:

1. the unit for determining students' health condition according to 4 parameters:
 - body temperature
 - saturation (oxygen saturation) of blood
 - heart rate (HR)
 - rapid diagnostics of cardiovascular system functional state
2. indoor air quality monitoring unit.

The article reviews the possibility of technical realization of hardware complex. It presupposes the use of sensors of registration of a photoplethysmographic curve, which describes a pulse wave and defines the parameters of students' cardiovascular system functional state.

The method of photoplethysmography is designed to study the cardiovascular system of biological objects in which the measurement of characteristics and parameters of blood circulation (pulse curve, blood pressure, arterial oxygen saturation level, etc.), vascular reactions and metabolic processes are performed by recording the fluxes intensity of electromagnetic radiation in the optical range (from visible – 0.4 μm to near-infrared – 1.5 μm) after their interaction with the tissues of a living organism [26].

There are two types of photoplethysmographic methods: transmitted-light photoplethysmography and side-scattered photoplethysmography (figure 6).

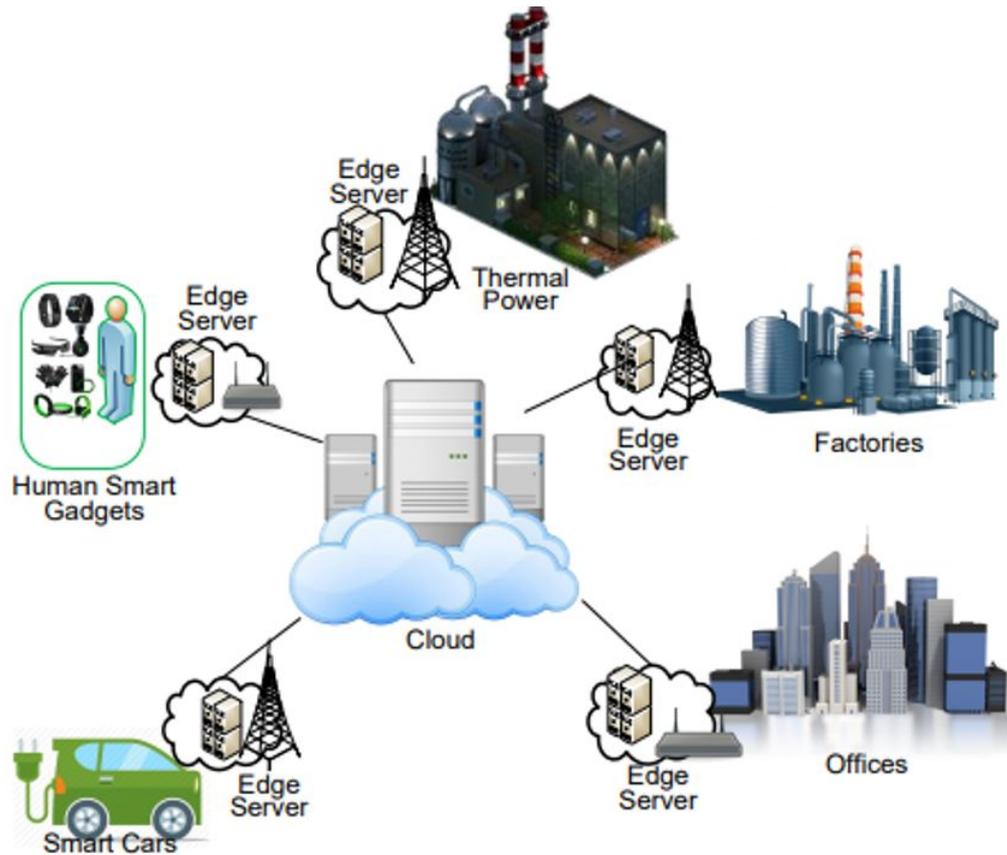


Figure 3: Illustration of Edge Cloud Computing Complimentary Role in IoT Environment [23]

The on the lumen method allows to install the sensor on a finger or an earlobe, as the radiating unit should fully X-ray the area. Moreover, the receiver, which is located perpendicularly, captures the light quantity that has passed through the finger. The on the reflection method presupposes that the light quantity from the radiating unit falls on a certain part of the body. In such a case, some of the light is absorbed, and some is reflected and enters the code receiver. This method is more universal, as it is possible to place the sensor on any part of the body, if full contact with him is provided [26, 27, 28].

The basis of photopulse oximetry method lies in the measurement of light absorption of a certain wavelength by blood hemoglobin. Hemoglobin serves as a filter, what is more, the “color” and “thickness” of this natural filter can vary [27, 28]. The “color” of the filter depends on the percentage of oxyhemoglobin. That is how pulse oximetry determines the level of blood oxygenation.

Changes in the “thickness” of the filter are affected by the pulsation of the arterioles: each pulse wave increases the amount of blood in the arteries and arterioles. The doctor defines this as a pulse rate, and the pulse oximeter considers that as a “thickening” of the filter. In such a way, the pulse rate and amplitude of the pulse wave are measured.

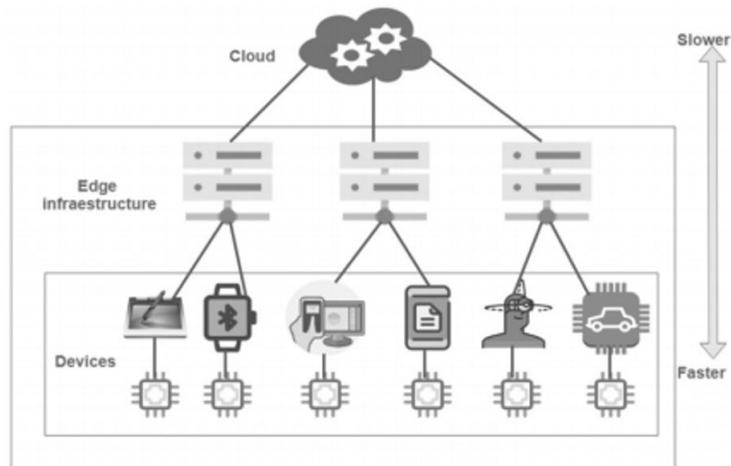


Figure 4: Edge and Cloud Computing [24]

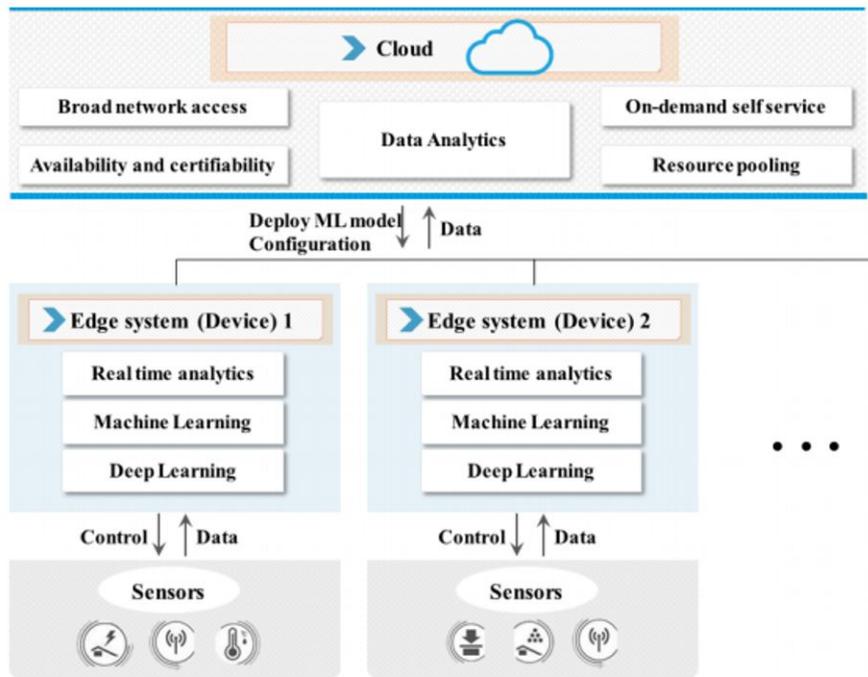


Figure 5: A design of Edge Computing Environment Supported by AI [25]

Therefore, the use of one measurement principle allows determining three diagnostic parameters: the levels of saturation of hemoglobin with oxygen, the pulse rate and its “volume” amplitude. In addition, it enables further processing and analysis of pulse waves to determine the functional state of cardiovascular system.

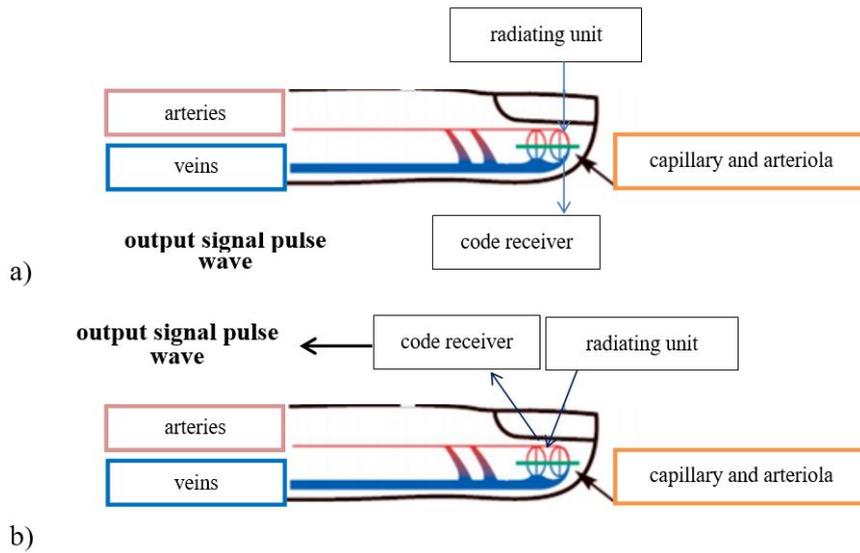


Figure 6: Ways of Sensors Location for Blood Circulation Registration. a) on the lumen, b) on the reflection

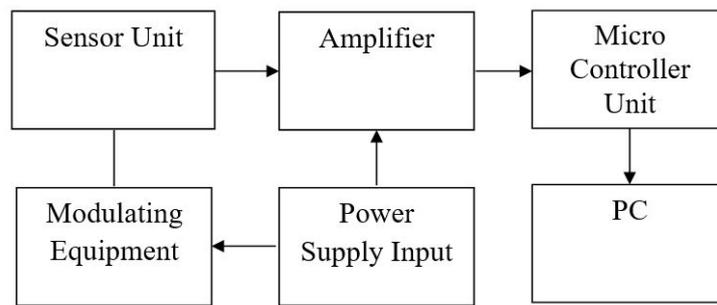


Figure 7: The Survey Plan of Information System Hardware

The registration of photoplethysmographic signals [5] is performed using the scheme shown in figure 7.

The analog part of the hardware consists of an optical sensor unit, an amplifier, and modulating equipment. The other units belong to the digital part. Amplified signals coming from the sensor unit, via USB-input are transmitted to the PC and the program window displays a pre-processed photoplethysmographic signal.

To transfer data from the microcontroller to the PC, USB port is used.

The method of photoplethysmography allows the use of non-contact sensors. Therefore, there is no artery compression, which eliminates circulatory disorders and allows the use of calculations to determine the saturation of oxygen by the pulse wave. Taking into consideration that the hardware of the signal recording system is required for further transmission, processing and analysis of pulse waves, the method of finger photoplethysmography is insuf-

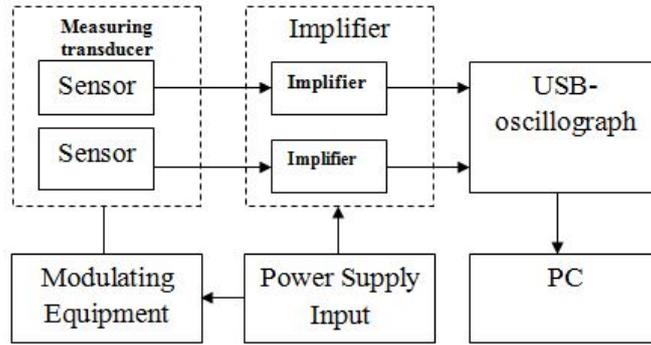


Figure 8: The Survey Plan of BTS Pulsogram Hardware Based on the Parametres of the Photopulse Oxygenation on the basis of PC

ficient in its practical use.

On the one hand, the method is sufficient if the end phalanx of the finger or foot is X-rayed on one side by ordinary incoherent light, which after side-scattering enters the photodetector; however, on the other hand, this method is not appropriate to obtain sufficiently intense signals from radial artery. However, it should be mentioned that the signal from radial artery is the most informative for cardiovascular system diagnosis. Nonetheless, this signal also depends on the human body constitution, its anthropometric parameters in particular. To consider this and to make the study of pulse waves and cardiovascular system more reliable we recommend controlling the intensity of infrared light depending on the human body constitution. The use of one optocoupler is not enough for this due to the low power of the light quantity and the depth of its penetration. It is recommended to use several optocouplers connected in series, parallel or parallel-series in a chain, with control of their mode of operation from the intensity of the received pulse wave signal depending on human body constitution (figure 8). Small optocouplers design allows doing it on a small plane, which the sensor itself has.

In the case of photopulse oxygenation, we are interested in the absorption of light quantity by blood running through veins, arterial blood in particular. Thus, the aim of pulse oximetry is to measure the level of saturation of hemoglobin in arterial blood with oxygen.

Hemoglobin is the common name for blood proteins found in red blood cells. Oxyhemoglobin is fully oxygenated hemoglobin, each molecule of which contains four oxygen molecules. Deoxyhemoglobin is hemoglobin that does not contain any oxygen.

The tissues through which both light quantity pass are a non-selective filter and evenly attenuate the radiation of both LEDs. The degree of attenuation depends on tissues thickness, skin pigment and other obstacles in the way of light. Hemoglobin, in contrast to tissues, is a color filter, and the color of this filter is affected by the level of oxygen saturation of hemoglobin. Deoxyhemoglobin has a dark cherry color. It intensively absorbs red light and weakly delays infrared. Therefore, if to put blood that does not contain any oxygen under the red and infrared light, the first one will be almost completely held, and the second one will be only slightly weakened. Conversely, oxyhemoglobin scatters red light (therefore, it has a red color), but intensively absorbs infrared radiation.

Thus, the ratio of two light quantities under the photodetector depends on blood oxygen sat-

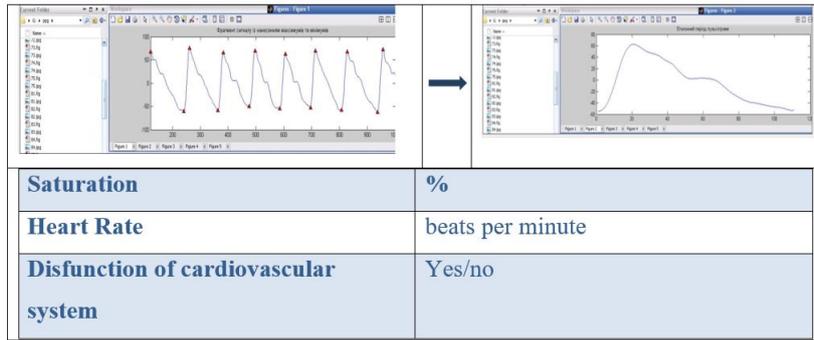


Figure 9: The Program Window



Figure 10: The Hardware and Software Complex Overview and the window with pulsegrams

uration level. According to these data, using a certain algorithm, the microprocessor calculates the percentage of oxyhemoglobin in the blood.

Therefore, using the unit of photoplethysmography and implying the methods of photoplethysmographic signal digital processing, we obtain the result as shown in figure 9.

In order to read the pulse signal, it is possible to connect the sensor directly to the laptop, previously pre-amplifying the signal. Moreover, it is also possible to implement a small model in the MATLAB package [29, 30] for further analysis of pulsegrams. It should be noted that the display of the pulse graph is in real time. What is more, data can be stored in the database. It is rather convenient for keeping the records and dealing with statistics.

Non-invasive methods of registration, analysis and evaluation of amplitude-time parameters of pulse signals [31, 32, 33, 34, 35, 36] are viewed as a set of modern technical means and mathematical methods of processing biosignals. Nowadays, they define the current trends

in cardiovascular system as well as other systems diagnostics. Furthermore, determination of additional values of saturation and body temperature is an important issue which is not restricted to rapid students' diagnostics only.

3. Conclusions

The paper proposes a device for recording pulse signals, which can record not only the heart rate but also measure saturation, which significantly minimizes the design of the device.

The edge device hardware is a part of the IoT system, which also includes another edge device, which instantly transmits data to the database on the edge server for the data further processing and storage. In addition, further detailed study of edge device data as a part of the IoT System is needed. Furthermore, the development of a mobile application to display the data is planned. This will allow you to monitor changes in the physiological parameters of the student in real-time around the clock and/or record on the server and, if necessary, view them.

References

- [1] S. Semerikov, S. Chukharev, S. Sakhno, A. Striuk, V. Osadchyi, V. Solovieva, T. Vakaliuk, P. Nechypurenko, O. Bondarenko, H. Danylchuk, Our sustainable coronavirus future, volume 166, EDP Sciences, 2020. doi:10.1051/e3sconf/202016600001.
- [2] I. B. Isupov, R. S. Zatrudina, Electronic module for photoplethysmography and pulse oximetry, *Natural Systems and Resources* 8 (2018).
- [3] D. Rogatkin, Physical basis of optical oximetry, *Medical Physic* 2 (2012) 97–114.
- [4] S. Munir, A. Guilcher, T. Kamalesh, B. Clapp, S. Redwood, M. Marber, P. Chowienczyk, Peripheral augmentation index defines the relationship between central and peripheral pulse pressure, *Hypertension* 51 (2008) 112–118. doi:10.1161/HYPERTENSIONAHA.107.096016.
- [5] T. N. Nikitchuk, V. F. Manoilov, P. P. Martynchuk, Recording technique of photoplethysmographic signals for study in the phase plane, in: 2012 22nd International Crimean Conference "Microwave Telecommunication Technology", 2012, pp. 981–982.
- [6] A. V. Ryabko, O. V. Zaika, R. P. Kukharchuk, T. A. Vakaliuk, Graph model of Fog Computing system, *CEUR Workshop Proceedings* (2021).
- [7] K. Kopnyak, Estimation of efficiency of medical information systems introduction, *Economy and organization of management* 2 (2017) 109–119.
- [8] O. Gorna, T. Stanishevskya, T. Kopulova, O. Yusupova, D. Horban, Research of the somatic health of student youth using information and communication technologies, *E3S Web of Conferences* 166 (2020). doi:10.1051/e3sconf/202016610034.
- [9] O. Klochko, V. Fedorets, O. Maliar, V. Hnatyuk, The use of digital models of hemodynamics for the development of the 21st century skills as a components of healthcare competence of the physical education teacher, *E3S Web of Conferences* 166 (2020). doi:10.1051/e3sconf/202016610033.
- [10] O. Klochko, V. Fedorets, A. Uchitel, V. Hnatyuk, Methodological aspects of using aug-

- mented reality for improvement of the health preserving competence of a physical education teacher, CEUR Workshop Proceedings 2731 (2020) 108–128.
- [11] H. Meshko, O. Meshko, N. Drobyk, O. Mikheienko, Psycho-pedagogical training as a mean of forming the occupational stress resistance of future teachers, E3S Web of Conferences 166 (2020). doi:10.1051/e3sconf/202016610023.
- [12] Y. Nosenko, A. Sukhikh, The method for forming the health-saving component of basic school students' digital competence, CEUR Workshop Proceedings 2393 (2019) 178–190.
- [13] Y. Nosenko, A. Sukhikh, O. Dmytriienko, Organizational and pedagogical conditions of ict health-saving usage at school: Guidelines for teachers, CEUR Workshop Proceedings 2732 (2020) 1069–1081.
- [14] D. Shiyani, I. Ostapchuk, O. Lakomova, Geographical analysis of ecology-dependent diseases of kryvyi rih population in order to provide a sustainable development of the industrial regions, E3S Web of Conferences 166 (2020). doi:10.1051/e3sconf/202016601012.
- [15] V. Kachmar, V. Avramenko, Directions of development of information technologies in medicine, Medicine of transport of Ukraine 3 (2011) 96–103.
- [16] V. Kachmar, A. Khvyshchun, Electronic medical record of the patient. mutual compatibility and standardization, Ukrainian Journal of Telemedicine and Medical Telematics 6 (2008) 76–79.
- [17] O. Hanchuk, O. Bondarenko, I. Varfolomyeyeva, O. Pakhomova, T. Lohvynenko, Couch-surfing as a virtual hospitality network and a type of sustainable youth tourism, E3S Web of Conferences 166 (2020). doi:10.1051/e3sconf/202016609005.
- [18] A. Kiv, V. Soloviev, S. Semerikov, H. Danylchuk, L. Kibalnyk, A. Matviychuk, Experimental economics and machine learning for prediction of emergent economy dynamics, CEUR Workshop Proceedings 2422 (2019) 1–4.
- [19] I. Lutsenko, E. Vihrova, E. Fomovskaya, O. Serdiuk, Development of the method for testing of efficiency criterion of models of simple target operations, Eastern-European Journal of Enterprise Technologies 2 (2016) 42–50. doi:10.15587/1729-4061.2016.66307.
- [20] V. Soloviev, A. Belinskiy, Complex systems theory and crashes of cryptocurrency market, Communications in Computer and Information Science 1007 (2019) 276–297. doi:10.1007/978-3-030-13929-2_14.
- [21] S. Zelinska, A. Azaryan, V. Azaryan, Investigation of opportunities of the practical application of the augmented reality technologies in the information and educative environment for mining engineers training in the higher education establishment, CEUR Workshop Proceedings 2257 (2018) 204–214. 1st International Workshop on Augmented Reality in Education, AREdu 2018 ; Conference Date: 2 October 2018.
- [22] T. Lorido-Botran, M. K. Bhatti, ImpalaE: Towards an optimal policy for efficient resource management at the edge, CEUR Workshop Proceedings (2021).
- [23] N. Hassan, S. Gillani, E. Ahmed, I. Yaqoob, M. Imran, The role of edge computing in internet of things, IEEE Communications Magazine 56 (2018) 110–115. doi:10.1109/MCOM.2018.1700906.
- [24] I. Candanedo, J. Corchado, An edge computing tutorial, Orient. J. Comp. Sci. and Technol. 12 (2019). doi:10.13005/ojcst12.02.02.
- [25] J. Huh, Y. Seo, Understanding edge computing: Engineering evolution with artificial in-

- telligence, *IEEE Access* 7 (2019) 164229–164245. doi:10.1109/ACCESS.2019.2945338.
- [26] V. Alekseev, A. Perminov, S. Yuran, Mutual arrangement of the source and receiver of the sensor radiation for photoplethysmography, *Instruments and methods of measurements* 1 (2011).
- [27] S. Pavlov, V. Kozhemyako, V. Petruk, P. Kolisnyk, Photoplethysmographic technologies of control of the cardiovascular system, UNIVERSUM-Vinnytsia, Vinnytsia, 2007.
- [28] I. Shurygin, Respiration monitoring: pulse oximetry, capnography, oximetry, BINOM, Moscow, 2000.
- [29] T. Nikitchuk, Devising an information system for the analysis of pulse signals, *Eastern-European Journal of Enterprise Technologies* 5 (2015) 19–23. URL: <http://journals.uran.ua/eejet/article/view/51219>. doi:10.15587/1729-4061.2015.51219.
- [30] A. Leonenkov, Fuzzy modeling in MATLAB and fussy TECH, BHV-Petersburg, 2005.
- [31] T. H. Nikitchuk, V. F. Manoylov, The technique of pulse signals processing in a phase plane, in: 2011 21st International Crimean Conference "Microwave Telecommunication Technology", 2011, pp. 1040–1041.
- [32] L. Ayusheeva, V. Boronoev, I. Lebedintseva, I. Ledneva, Time parameters of the pulse wave in the diagnostics of human diseases according to the tradition of tibetan medicine, *Biomedical Radioelectronics* 3 (2009) 17–23.
- [33] V. Mosiychuk, Multisignal digital registration and processing of pulse wave parameters, Ph.D. thesis, Nat. tech. University of Ukraine 'Kyiv Polytechnic Institute', Kyiv, 2011.
- [34] T. Kozlovskaya, Optical-electronic device for diagnosing the state of peripheral circulation with high reliability, Ph.D. thesis, Vinnytsia National Technical University, Vinnytsia, 2012.
- [35] O. Voloshin, V. Oleinik, S. Kulish, A. O. Sami, Ekg method for diagnostics of human functional state on the basis of fractal analysis and wavelet transform, *Radioelectronic and Computersystems* 4 (2010) 29–34.
- [36] L. Fainzilberg, New opportunities of phasegraphy in medical practice, *Science and Innovation* 13 (2017) 37–50.

Self-organizing network topology for autonomous IoT systems

Anastasia D. Sverdlova^a, Artur O. Zaporozhets^{a,b}, Ihor V. Bohachev^a,
Oleksandr O. Popov^{b,c,d}, Anna V. Iatsyshyn^b, Andrii V. Iatsyshyn^{b,c},
Valeriia O. Kovach^{b,e}, Volodymyr O. Artemchuk^{b,c} and Nataliia M. Hrushchynska^e

^a*Institute of Engineering Thermophysics of NAS of Ukraine, 2a Marii Kapnist Str., Kyiv, 03057, Ukraine*

^b*State Institution "The Institute of Environmental Geochemistry of National Academy of Sciences of Ukraine", 34a Palladin Ave., Kyiv, 03680, Ukraine*

^c*G.E. Pukhov Institute for Modelling in Energy Engineering of NAS of Ukraine, 15 General Naumova Str., Kyiv, 03164, Ukraine*

^d*Interregional Academy of Personnel Management, 2 Frometivska Str., Kyiv, 03039, Ukraine*

^e*National Aviation University, 1 Liubomyra Huzara Ave., Kyiv, 03058, Ukraine*

Abstract

The concept of the Internet of Things is increasingly defining the development of communication networks both now and in the future. The largest application of the IoT concept is wireless touch networks (WTN). Due to the potentially widespread use of WTN in all areas of human life, they are also called pervasive sensory networks. WTN belongs to the class of self-organizing networks, for which the construction principles, routing protocols, quality of service parameters, traffic models, and characteristics are significantly modified compared to traditional infrastructure networks, etc. The features of the application of dynamic routing protocols for the construction of a self-organizing network of autonomous IoT systems are considered. This article provides an overview of the main methods for calculating the topology of self-organizing networks. A review of known dynamic routing protocols for mobile radio networks is given, the advantages and disadvantages of proactive and reactive approaches are shown.

Keywords

IoT, networks, protocols, data transmission methods, error correction method

1. Introduction

In recent years, mobile devices have become widespread: cell phones, laptops, smartphones, and tablets. This has opened up new opportunities for the developers of network solutions [1]. One of the areas of development of network technologies for mobile devices is the Internet of Things.

QualnT 2021: Workshop on the Quantum Information Technologies, April 11, 2021, Zhytomyr, Ukraine
doors 2021: Edge Computing Workshop, April 11, 2021, Zhytomyr, Ukraine

✉ a.o.zaporozhets@nas.gov.ua (Artur O. Zaporozhets)

🆔 0000-0001-8222-1357 (Anastasia D. Sverdlova); 0000-0002-0704-4116 (Artur O. Zaporozhets);

0000-0001-7781-5767 (Ihor V. Bohachev); 0000-0002-5065-3822 (Oleksandr O. Popov); 0000-0001-8011-5956

(Anna V. Iatsyshyn); 0000-0001-5508-7017 (Andrii V. Iatsyshyn); 0000-0002-1014-8979 (Valeriia O. Kovach);

0000-0001-8819-4564 (Volodymyr O. Artemchuk); 0000-0002-5606-4666 (Nataliia M. Hrushchynska)

© 2021 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

The current direction of development of communication networks is the concept of the Internet of Things. The main task of the Internet of Things is to create a single network that includes objects of the information (virtual) and physical (real) worlds and will ensure the interaction of objects with each other.

The technological base of the first stage of development of the Internet of Things is all-pervasive (wireless) sensor networks, which are widely used in the modern world in almost all spheres of life, due to their low cost, rapid deployment and efficiency.

The implementation of the concept of the Internet of Things is expressed in the penetration of telecommunication technologies into all spheres of human activity. Currently, this is reflected in the expansion of the field of application of wireless sensor networks.

The concept of the Internet of Things includes data exchange between devices (M2M), sensor networks, and self-organizing networks of mobile devices (MANET). This article provides an overview of the main methods for calculating the topology of self-organizing networks.

The task of improving the quality of data transmission in self-organizing networks of mobile devices can be solved by different methods: the method of retransmission request (ARQ) [2], the method of redundant coding (FEC) [3, 4, 5, 6, 7, 8, 9], the method of Network Coding [10]. One of the approaches to data transmission in self-organized networks is the use of the method of superimposed networks (P2P) [11, 12]. P2P protocol sets the rules for streaming data between nodes [13, 14, 15, 16, 17]. Streaming data is transmitted between the nodes of the superimposed network along the routes selected by the underlying protocols [18, 19]. Controlling the data transmission process will avoid congested areas in the network, increase the throughput and improve the reliability of the network as a whole [20, 21, 22, 23, 24, 25]. Overlay networks rely on tree and multi-link structures [26, 27, 28]. To improve network reliability, some researchers [29] use different types of multipath redundancy, such as “routing braids”, which demonstrate improved reliability and stability in self-organizing networks. The environmental sector demonstrates particular interest in IoT, where modern air quality monitoring systems can be built using sensor networks [30, 31, 32, 33, 34, 35]. Also as a part of complex diagnostic systems of energy facilities [36, 37, 38, 39, 40] which build on hierarchical structures, IoT can be used in energy sector.

2. Analysis of routing protocols in IoT systems

Self-organizing networks are an alternative to infrastructure networks. In such a network, each node in the network can act as a router. The possibility for each node to leave the network or connect to it will lead to the fact that an important issue in the organization of the self-organizing network is the choice of a routing protocol. The routing protocols developed are classified according to the approach to update the network topology information into reactive, proactive, and hybrid [41]. Figure 1 shows the classification of routing protocols.

The reactive approach to routing involves constructing routes as they are needed. When a connection to a network node is attempted, a complete enumeration of all options is performed and the best route to it is found according to the routing metric. This route is used as long as there is a connection to the destination.

With a proactive approach, the network topology must be monitored and updated at specific

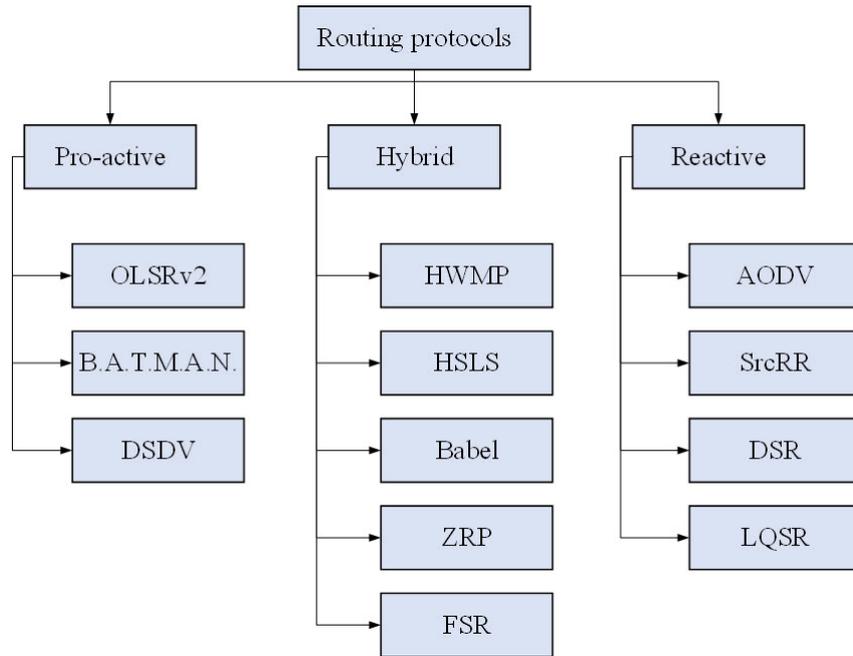


Figure 1: Routing protocols in self-organizing networks

intervals. Proactive protocols update the network topology with periodic queries. Protocols belonging to this group may use different numbers of databases with information about the network topology and different ways of keeping this information up to date. The proactive approach relies on keeping track of the network topology, so nodes are constantly exchanging messages, which can lead to higher power consumption compared to the reactive approach. On the other hand, a node in a network using the reactive approach has to wait for the route variants to be enumerated, which can affect the transmission speed in networks with changing topologies. The hybrid approach involves combining reactive and proactive approaches within the same network. The best route between network nodes is selected based on metrics: number of routing steps, ETX, ETT, Air Time Link, etc. Metrics can take into account information from the physical, data link, and network layers of the OSI model.

2.1. Proactive protocols

The OLSR (Optimized Link State Routing) protocol is proactive and oriented for use in large networks with a high density of nodes. Each node uses HELLO broadcast messages that are transmitted at regular intervals to nodes within one routing step. After receiving the HELLO, the destination node tries to establish a two-way connection with the sender node. The number of control messages in OLSR is reduced due to the MPR (Multipoint Relays) approach [42]. In OLSRv2, the exchange of control messages in the network has become more efficient, and the message form itself has been standardized and simplified. OLSR interacts with the network layer by managing routing tables and using IP addresses for packet transmission. The

B.A.T.M.A.N. protocol (Better Approach To Mobile Ad hoc Networks) also uses a proactive approach [43], in which all nodes produce an Originator Message (OGM) broadcast. An OGM contains an originator address, a recipient address, and a unique sequential number. Each neighboring node changes the recipient address to its own and sends the message back to the originator. OGM messages do not include any additional information such as QoS metrics and routing tables. The B.A.T.M.A.N. protocol. Has lower non-productive costs in networks with more nodes than the OLSR protocol. One of the first proactive protocols was DSDV (Destination Sequenced Distance Vector), developed in 1994 [44]. Its main feature was the addition of an ordinal number field in control messages because this bypassed the problem of loops between nodes in the network (Loop free) since each node now knew whether its information about the network topology was obsolete. DSDV proved ineffective in large networks with rapidly changing topologies but influenced the development of other protocols, such as AODV.

2.2. Reactive protocols

Reactive DSR (Dynamic Source Routing) protocol uses a special DSR Options Header Format that can be added to any packet and contains the route from source to destination node [45]. A node can perform a route discovery process to the desired node (Route Discovery) using broadcast messages. The Route Maintenance process is to monitor the link-layer notifications. If a link-layer notification is accepted or node requests are left unanswered, the discovery process is repeated. Disadvantages and advantages of DSR include its reactivity, which reduces the cost of sending control messages but makes it necessary to buffer packets for the duration of route discovery. Besides, the special header format can lead to a large header for small payloads, reducing the efficiency of the network. Further evolution of the reactive approach was the AODV protocol [46]. Instead of relying on the transmission of voluminous headers, AODV reintroduced routing tables that accumulated all the information about the network topology as messages were received from other nodes. To avoid looping, two sequence numbers were introduced, one for the source and one for the destination, allowing you to track the novelty of topology information as you use the route from the destination to the source. The use of AODV is recommended for networks of 10 to 1000 mobile nodes. The main goal of its development was to reduce the cost of sending control messages and to improve the scalability and performance of the network. Another protocol based on DSR was reactive SrcRR [47]. Its main difference from DSR was the use of an ETX metric, which was measured by periodic broadcasts to neighboring nodes, and the total ETX of its parts was used for the entire route. Also, SrcRR was independent of the network layer and could use MAC addresses to find the path. Microsoft developed and patented the LQSR (Link Quality Source Routing) protocol, which is also based on DSR [48]. It is implemented between the link layer and the network layer using a virtual network adapter, allowing it to handle multiple physical connections at once. The LQSR header is located between the Ethernet header and the frame payload. Each node, as in SrcRR, measures the QoS metric to neighboring nodes, propagates this information through the network, and it is taken into account in selecting the best path to the destination. Guided by the rule that the shortest path does not mean the best path, LQSR allows the use of three QoS metrics: ETX, RTT, and PktPair. VNF sees network latency as a critical attribute for reliability, availability, and QoS requirements by most researchers. By automating and elastically

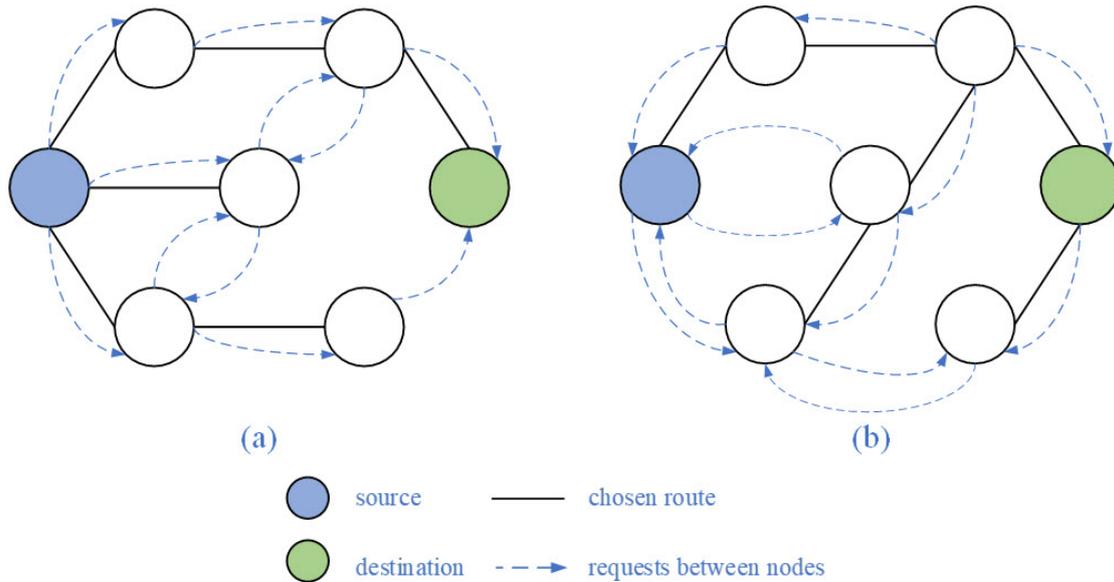


Figure 2: Reactive (a) and proactive (b) modes of the HWMP protocol

allocating resources, these enhanced service offerings are implemented [49].

2.3. Hybrid protocols

The hybrid approach enables the use of reactive and proactive approaches within a single network. It is used in 802.11s to provide WMN support at the link layer [50, 51]. In previous 802.11 family standards, there was no way to obtain link-layer QoS metrics. For the QoS metric to be more accurate, it should be obtained at a lower layer of the OSI network model. As a default protocol in the standard 802.11s recommended hybrid HWMP (Hybrid Wireless Mesh Protocol), and the optional protocol can act as OLSR. The reactive approach is implemented based on AODV (Ad hoc On-demand Distance Vector). In this case, the node looks for the best route as needed, taking into account QoS metrics. Using a proactive approach, a root node (Root) is assigned to a WMN that polls nodes at intervals, thus updating the network map. The connected node, can contact the root node and get information about the routes to all nodes in the network. Both approaches can be used separately or simultaneously in the same network (figure 2).

The hybrid approach has been used not only in HWMP but also in earlier protocols, such as the HSL (Hazy Sighted Link State) Routing Protocol. Intending to reduce non-productive costs, HSL controls the interval at which network topology information is updated to reduce the number of control messages [52]. If the route is obsolete, HSL begins to operate in reactive mode. The lack of up-to-date network topology information is a major drawback of this protocol. Another hybrid protocol for WMN is Babel. Based on the ideas of DSDV, AODV, and the Cisco EIGRP (Enhanced Interior Gateway Protocol), Babel takes a proactive approach and is aimed at working in networks with mobile nodes [53]. It allows the implementation of

different QoS metrics, although by default it uses ETX. Reactive mode is used in Babel if no route from a node is suitable for reliable packet transmission. In a hybrid protocol, ZRP (Zone Routing Protocol) node applies proactive route lookup within a certain section of the network and reactive outside of it [54]. The FSR (Fish-eye State Routing Protocol) protocol is characterized by the fact that the accuracy of the network topology information decreases with distance from the node [55].

3. Analysis of data transmission methods and correction of transmission errors

3.1. Retransmission request method

Many different methods are used to recover lost and corrupted data. In self-organizing networks, the network topology and transmission environment change rapidly, it is very difficult to ensure reliable communication, to overcome the high mobility of nodes and external interference. Therefore, many packets are received with errors, which means that error correction methods play a significant role in data transmission processes [56].

In the transmission error correction method with a repeated request (ARQ), data reception acknowledgment messages (ACK) are used for reliable data transmission. For example, if the source has not received an acknowledgment from the recipient within a certain time interval (timeout), it will retransmit until it receives an ACK message. The ARQ method relies on sum and sequence number check fields in each packet header to detect corrupted and retransmitted data. The retransmission request is used in the Stop-and-Wait ARQ, Go-Back-N ARQ, and Selective-Repeat ARQ methods.

The Stop-and-Wait ARQ and Selective-Repeat ARQ methods are used in the 802.11 families of standards at the data link layer of the OSI model, and all three approaches are used in various transport layer protocols. The methods differ in the size of the transmit window and the receive window. The Stop-and-Wait ARQ method starts a timer for each packet that is sent, and the source waits until the ACK message arrives. If the ACK message has not been received and the timeout has already expired, the source will repeat the packet. Thus, the interaction between source and destination occurs from packet to packet.

The Go-Back-N ARQ method is more efficient than Stop-and-Wait ARQ. With this approach, the source transmits several packets at a time and stores them in a buffer until it receives a group ACK message. After the timeout expires, the source repeats all packets for which no ACK message arrived. The Stop-and-Wait ARQ and Go-Back-N ARQ methods are very similar, but they use different transmit window sizes.

In the Selective-Repeat ARQ method, the source transmits several packets at once but waits for an individual acknowledgment for each packet. The receive and transmit windows sizes are the same, and the destination can receive and store packets received in any order. The source repeats those packets for which the timeout has expired.

The ARQ method can improve connection reliability but is not suitable for use in video broadcasting because of the large and unstable delay.

3.2. Direct error correction method

The FEC method adds redundancy to the data being sent, which allows the addressee to detect and correct errors without a second request from the source, and the maximum number of bits recovered depends on the code used. The FEC method is usually implemented at the physical layer and is responsible for correcting errors caused by interference in the communication channel. The application layer FEC method uses Reed Solomon codes or BCH codes. By introducing redundancy, this code can detect and correct bit errors in transmission. But the introduction of redundancy reduces the efficiency of communication channels if the transmission is error-free. Therefore, an adaptive noise coding method has been developed [57, 58, 59, 60]. This method allows controlling redundancy at byte or packet-level [61, 62], using video characteristics or quality of service metrics, such as information fragment delivery ratio.

3.3. Network coding method

One method that is very similar to the FEC method is network coding (NC) [63]. In the NC method, data are encoded by intermediate nodes. The self-organizing network provides new opportunities for the implementation of the NC method. Today, the NC method for reliable video data transmission is most often implemented based on random linear network coding (RLCN). The source node groups the data into generations encodes each generation with SLSC and writes the encoding coefficients in the header of each packet. In the NC method, redundancy can be controlled: k line-independent packets of a given generation are required to decode all packets of the same generation. Thus, more encoded packets can be transmitted in unreliable transmission channels. SLSC has advantages over other coding methods (e.g., fountain code, block code) in that it makes the handling of encoded packets more flexible, reduces the network delay for video transmission, eliminates the transmission of identical packets, and uses the bandwidth more efficiently. The SLSC method can be used in conjunction with the ARQ method to prevent the loss of the entire packet generation. Packets needed for decoding the current generation can be requested from neighboring nodes to obtain k line-independent packets of each generation. The SLSC method has much in common with FEC coding, but FEC is implemented only on the destination node, while the SK method is also implemented on the intermediate nodes. FEC and SK methods can work simultaneously without additional modifications, but more complex hybrid solutions are possible. The SK method can improve network reliability and resilience when used in conjunction with the multipath redundancy method.

3.4. Multipath redundancy method

The multipath redundancy method is to transfer data from the source to the destination via multiple routes. This method of routing has different goals: to distribute the load on the network routes or to increase the bandwidth and reliability of the network as a whole [64, 65]. Applying the multipath redundancy method, it is possible to get rid of congested sections in the network or simultaneously deliver streaming data via multiple routes. However, a self-organizing network is usually congested at the section between the source and the destination, and its network topology is inconstant and therefore requires recalculation of independent routes. It is because of this that multipath redundancy is more often used to improve network

reliability. Multipath redundancy can be provided by a superimposed network [66, 67]. The simplest structure of a superimposed network is a single-layer tree. The root of such a tree is the source node. The short existence time of a connection between nodes imposes restrictions on the application of this structure in a self-organizing network. A multi-layer structure can also be used in superimposed networks. This structure is more resilient to user outages, the load is evenly distributed throughout the network, and does not require centralized coordination, both during normal network operation and during emergencies. The structure is adapted both for single-source transmissions and to provide transmissions from multiple sources. The disadvantage of this structure is that the networks built on their basis are more complex than their counterparts. This entails that to maintain this structure, a large volume of control messages are transferred between the nodes of the network. This can significantly limit the applicability of such a structure to nodes that vary considerably in self-organizing networks. The multi-layer tree structure seeks to eliminate two major drawbacks of an overlapping network with a single-tree structure [68]. First, in single-tree structures, the few closest nodes are loaded much more heavily than the rest of the network, since the “leaves” of the tree have not been involved in content transmission. Secondly, the disconnection of these highly stressed nodes leads to mass switching of network users looking for a new data source and a new connection point. Node outages could lead to the degradation of streaming data quality. In a multilayer tree, each node must stream data in multiple trees with a common root, distributing complementary content. Such a structure ensures that all content is not lost if one of the trees loses connectivity and better utilizes the available resources of each node in the network. Source S distributes streaming data to all nodes at once, but some video (e.g., every third fragment) can be transmitted along the intended paths between the nodes themselves. Thus, instead of a single tree, we consider three single-layer trees at once in the case of a multilayer tree with a multilayer coefficient equal to three [69]. To improve network reliability, some algorithms use different types of multipath redundancy, such as multipath “braid” routing. This move allows the use of multiple routes instead of a single route to apply the SC method, which improves the reliability and robustness of self-organizing networks [29]. The multilayer tree structure helps to overcome packet loss in multicast in case of multiple node outage [70]. But the use of multiple paths can increase the unproductive cost of forwarding data, so more research on this method is needed. The multilayer tree structure can be combined with SC, FEC, and ARQ methods to improve data transmission efficiency [71].

4. Conclusions

Known routing protocols used in self-organizing networks are considered and analyzed. The main disadvantage of existing routing protocols is that they cannot respond to abrupt changes in the network topology, which entails their inefficiency in self-organizing networks with highly mobile nodes. When a source is broadcast to one destination, the quality of the transmission depends entirely on the performance of the routing protocol. In the event of loss of one route due to movement of one or more nodes of the network or due to a change in the state of communication channels between nodes, routing protocols cannot effectively use existing routes in the network to transfer streaming data from the source to destination. An overview of

transmission error correction techniques has been made and a process for transmitting streaming data using a selective retransmission request method has been described.

References

- [1] F. Zambonelli, H. V. D. Parunak, Signs of a revolution in computer science and software engineering, in: P. Petta, R. Tolksdorf, F. Zambonelli (Eds.), *Engineering Societies in the Agents World III. ESAW 2002. Lecture Notes in Computer Science*, volume 2577, Springer, Berlin, Heidelberg, 2019, pp. 13–28. doi:/10.1007/3-540-39173-8_2.
- [2] D. Claypool, K. McNeill, Automatic repeat request (arq) over tdma-based mesh network, in: *Proceedings of the Military Communications Conference (MILCOM 2008)*, IEEE, 2016, pp. 1–7.
- [3] L. Baldantoni, H. Lundqvist, G. Karlsson, Adaptive end-to-end FEC for improving TCP performance over wireless links, in: *Proceedings of IEEE International Conference on Communications*, IEEE, 1999, pp. 4023–4027.
- [4] J.-C. Bolot, S. Fosse-Parisis, D. Towsley, Adaptive FEC-based error control for internet telephony, in: *Proceedings of Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. The Future is Now (IEEE INFOCOM '99)*, IEEE, 1999, pp. 1453–1460.
- [5] S.-H. Chan, X. Zheng, Q. Zhang, W.-W. Zhu, Y.-Q. Zhang, Video loss recovery with FEC and stream replication, *IEEE Transactions on Multimedia* 8 (2006) 75–380. doi:10.1109/TMM.2005.864340.
- [6] K. French, M. Claypool, Repair of streaming multimedia with adaptive forward error correction, in: *Proc. SPIE 4518, Multimedia Systems and Applications IV*, 2001. doi:10.1117/12.448220.
- [7] A. Nafaa, T. Ahmed, A. Mehaoua, Unequal and interleaved FEC protocol for robust MPEG-4 multicasting over wireless LANs, in: *Proceedings of IEEE International Conference on Communications (IEEE Cat. No.04CH37577)*, IEEE, 2004, pp. 1431–1435.
- [8] H. Wu, M. Claypool, R. Kinicki, Adjusting forward error correction with quality scaling for streaming MPEG, in: *Proceedings of the 15th ACM International Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV)*, 2005, pp. 111–116.
- [9] S. Pleisch, Efficient flooding in mobile ad-hoc networks, in: *Proceedings of the 7th ACM Int'l Symp. on Mobile Ad Hoc Networking and Computing*, 2006, pp. 1–12.
- [10] C. Chen, S. Oh, J. Park, M. Gerla, M. Sanadidi, Combocoding: Combined intra-/inter-flow network coding for TCP over disruptive MANETs., *J. Adv. Res.* 2 (2011) 241–252. doi:10.1016/j.jare.2011.05.002.
- [11] J. Kuo, C. Shih, C. Ho, Y. Chen, A cross-layer approach for real-time multimedia streaming on wireless peer-to-peer ad hoc network, *Ad Hoc Netw.* 11 (2013) 339–354. doi:10.1016/j.adhoc.2012.06.008.
- [12] N. Qadri, M. Fleury, M. Altaf, M. Ghanbari, P2P layered video streaming over wireless ad hoc networks, in: *Proceedings of the 5th International ICST Mobile Multimedia Communications Conference (Mobimedia '09)*, ICST, 2006, pp. 1–7.

- [13] S. Banerjee, B. Bhattacharjee, C. Kommareddy, Scalable application layer multicast, in: Proceedings of the 2002 conference on Applications, technologies, architectures, and protocols for computer communications (SIGCOMM '02), Association for Computing Machinery, 2002, pp. 205–217.
- [14] J. Jannotti, D. K. Gifford, K. L. Johnson, M. F. Kaashoek, J. W. O'Toole, Overcast: reliable multicasting with on overlay network (2000) Article 14.
- [15] A. Nicolosi, S. Annapureddy, P2PCast: A peer-to-peer multicast scheme for streaming data, 2003. URL: <https://www.cs.stevens.edu/~nicolosi/papers/isw03.pdf>.
- [16] V. Padmanabhan, H. Wang, P. Chou, Resilient peer-to-peer streaming, in: Proceedings of 11th IEEE International Conference on Network Protocols, IEEE, 2003, pp. 16–27.
- [17] M. Castro, P. Druschel, A.-M. Kermarrec, A. Nandi, A. Rowstron, A. Singh, Splitstream: high-bandwidth multicast in cooperative environments, in: Proceedings of the nineteenth ACM symposium on Operating systems principles (SOSP '03), Association for Computing Machinery, 2003, pp. 298–313.
- [18] H. Deshpande, M. Bawa, H. Garcia-Molina, Streaming live media over a peer-to-peer network, 2002. URL: <http://ilpubs.stanford.edu:8090/501/1/2001-30.pdf>.
- [19] H. Deshpande, M. Bawa, H. Garcia-Molina, Streaming live media over peers, 2002. URL: <http://ilpubs.stanford.edu:8090/863/1/2002-21.pdf>.
- [20] T. Yun, S. Lifeng, L. Jianguang, Y. Shiqiang, Z. Yuzhuo, How scalable is cache-and-relay scheme in p2p on-demand streaming?, IEICE Trans. Commun. E90-B (2007) 987–989.
- [21] S. Oh, M. Gerla, Resilient peer-to-peer streaming, in: Proceedings of Communication Systems and Networks and (COMSNETS 2009), IEEE, 2009, pp. 1–10.
- [22] C. Vassilakis, N. Laoutaris, I. Stavrakaki, The impact of playout policy on the performance of p2p live streaming: or how not to kill your p2p advantage, in: Proc. SPIE 6818, Multimedia Computing and Networking, 2008. doi:10.1117/12.775147.
- [23] V. Babak, S. Babak, M. Myslovych, A. Zaporozhets, V. Zvaritch, Methods and models for information data analysis, in: Diagnostic Systems For Energy Equipments. Studies in Systems, Decision and Control, volume 281, Springer, Cham, 2020, pp. 23–70. doi:10.1007/978-3-030-44443-3_2.
- [24] V. Babak, V. Eremenko, A. Zaporozhets, Research of diagnostic parameters of composite materials using johnson distribution, Int. J. Comput. Appl. 18 (2019) 483–494. URL: <https://www.computingonline.net/computing/article/view/1618>.
- [25] V. Eremenko, A. Zaporozhets, V. Isaenko, K. Babikova, Application of wavelet transform for determining diagnostic signs, in: V. Ermolayev, F. Mallet, V. Yakovyna, H. Mayr, A. Spivakovsky (Eds.), Proceedings of the 15th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer. Volume I: Main Conference (ICTERI 2019), CEUR Workshop Proceedings, volume 2387, 2019, pp. 202–214. URL: <http://ceur-ws.org/Vol-2387/20190202.pdf>.
- [26] J. Chakareski, P. Frossard, Utility-based packet scheduling in p2p mesh-based multicast (2009) 72571S. doi:10.1117/12.806783.
- [27] Y. hua Chu, S. Rao, S. Seshan, H. Zhang, A case for end system multicast, IEEE J. Sel. Areas Commun. 20 (2002) 1456–1471. doi:10.1109/JSAC.2002.803066.
- [28] D. Zappala, A. Fabbri, V. Lo, An evaluation of shared multicast trees with multiple active cores, Telecommunication Systems 19 (2002) 461–479. doi:10.1023/A:1013854808626.

- [29] S. Oh, B.-H. Shen, M. Gerla, Automatic repeat request (arq) over tdma-based mesh network, in: Proceedings of IEEE Military Communications Conference (MILCOM 2012), IEEE, 2012, pp. 1–6.
- [30] O. Popov, A. Iatsyshyn, V. Kovach, V. Artemchuk, I. Kameneva, D. Taraduda, V. Sobyna, D. Sokolov, M. Dement, T. Yatsyshyn, Risk assessment for the population of Kyiv, Ukraine as a result of atmospheric air pollution, *J. Health Pollut.* 10 (2020) 200303. doi:10.5696/2156-9614-10.25.200303.
- [31] A. Iatsyshyn, A. Iatsyshyn, V. Artemchuk, I. Kameneva, V. Kovach, O. Popov, Software tools for tasks of sustainable development of environmental problems: peculiarities of programming and implementation in the specialists' preparation, *E3S Web Conf.* 166 (2020) 01001. doi:10.1051/e3sconf/202016601001.
- [32] A. Iatsyshyn, A. Iatsyshyn, V. Kovach, I. Zinovieva, V. Artemchuk, O. Popov, O. Cholyskhina, O. Radchenko, O. Radchenko, A. Turevych, Application of open and specialized geoinformation systems for computer modelling studying by students and PhD students, in: O. Sokolov, G. Zholtkevych, V. Yakovyna, Y. Tarasich, V. Kharchenko, V. Kobets, O. Burov, S. Semerikov, H. Kravtsov (Eds.), Proceedings of the 16th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer. Volume II: Workshops (ICTERI 2020), CEUR Workshop Proceedings, volume 2732, 2020, pp. 893–908. URL: <http://ceur-ws.org/Vol-2732/20200893.pdf>.
- [33] A. Zaporozhets, V. Babak, V. Isaienko, K. Babikova, Analysis of the air pollution monitoring system in ukraine, in: V. Babak, V. Isaienko, A. Zaporozhets (Eds.), Systems, Decision and Control in Energy I. Studies in Systems, Decision and Control, volume 298, Springer, Cham, 2020, pp. 85–110. doi:10.1007/978-3-030-48583-2_6.
- [34] A. Zaporozhets, V. Khaidurov, Mathematical models of inverse problems for finding the main characteristics of air pollution sources, *Water, Air, & Soil Pollution* 231 (2020) 563. doi:10.1007/s11270-020-04933-z.
- [35] A. Zaporozhets, O. Redko, V. Babak, V. Eremenko, V. Mokiychuk, Method of indirect measurement of oxygen concentration in the air, *Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu* 5 (2018) 105–114. doi:10.29202/nvngu/2018-5/14.
- [36] A. Zaporozhets, Experimental research of a computer system for the control of the fuel combustion process, in: Control of Fuel Combustion in Boilers. Studies in Systems, Decision and Control, volume 287, Springer, Cham, 2020, pp. 89–123. doi:10.1007/978-3-030-46299-4_4.
- [37] A. Zaporozhets, Hardware and software implementation of modules of the system of the fuel combustion control process, in: Control of Fuel Combustion in Boilers. Studies in Systems, Decision and Control, volume 287, Springer, Cham, 2020, pp. 61–87. doi:10.1007/978-3-030-46299-4_3.
- [38] V. Eremenko, A. Zaporozhets, V. Babak, V. Isaienko, K. Babikova, Using hilbert transform in diagnostic of composite materials by impedance method, *Period. Polytech. Electr. Eng. Comput. Sci.* 64 (2020) 334–342. doi:10.3311/PPee.15066.
- [39] V. Babak, S. Babak, M. Myslovych, A. Zaporozhets, V. Zvaritch, Principles of construction of systems for diagnosing the energy equipment, in: Diagnostic Systems For Energy Equipments. Studies in Systems, Decision and Control, volume 281, Springer, Cham, 2020, pp. 1–22. doi:10.1007/978-3-030-44443-3_1.

- [40] A. Zaporozhets, Application of wavelet transform for determining diagnostic signs, in: V. Ermolayev, F. Mallet, V. Yakovyna, H. Mayr, A. Spivakovsky (Eds.), Proceedings of the 15th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer. Volume I: Main Conference (ICTERI 2019), CEUR Workshop Proceedings, volume 2387, 2019, pp. 223–230. URL: <http://ceur-ws.org/Vol-2387/20190223.pdf>.
- [41] R. Baumann, S. Heimlicher, M. Strasser, A. Weibel, A Survey on Routing Metrics, TIK Report 262, Computer Engineering and Networks Laboratory ETH-Zentrum, Switzerland, 2007. URL: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.304.7863&rep=rep1&type=pdf>.
- [42] T. Clausen, C. Dearlove, P. Jacquet, U. Herberg, The optimized link state routing protocol version 2, 2014. URL: <https://tools.ietf.org/html/rfc7181>.
- [43] A. Neumann, C. Aichele, M. Lindner, S. Wunderlich, Better Approach To Mobile Ad-hoc Networking (B.A.T.M.A.N.) draft-openmesh-b-a-t-m-a-n-00, Technical Report, 2008. URL: <https://tools.ietf.org/html/draft-openmesh-b-a-t-m-a-n-00>.
- [44] C. E. Perkins, P. Bhagwat, Highly dynamic Destination-Sequenced Distance-Vector routing (DSDV) for mobile computers, in: Proc. SPIE 6818, Multimedia Computing and Networking, Association for Computing Machinery, 2008, pp. 234–244. doi:10.1145/190314.190336.
- [45] D. Johnson, Y. Hu, D. Maltz, The Dynamic Source Routing Protocol (DSR) for Mobile Ad Hoc Networks for IPv4, 2007. URL: <http://www.ietf.org/rfc/rfc4728.txt>.
- [46] C. Perkins, E. Belding-Royer, S. Das, Ad hoc On-Demand Distance Vector (AODV) Routing, 2003. URL: <http://www.ietf.org/rfc/rfc3561.txt>.
- [47] D. Aguayo, J. Bicket, R. Morris, Srcrr: A high throughput routing protocol for 802.11 mesh networks (draft), 2011. URL: <https://pdos.csail.mit.edu/archive/rtm/srcrr-draft.pdf>.
- [48] R. P. Draves, B. D. Zill, J. D. Padhye, System and method for link quality source routing, 2011. Patent No. US 7,978,672 B2, Filed Jan. 19, 2010, Issued Jul. 12, 2011.
- [49] R. Mijumbi, J. Serrat, J.-L. Gorricho, Autonomic resource management in virtual networks, 2015.
- [50] IEEE Standard for Information technology—Telecommunications and information exchange between systems Local and metropolitan area networks—Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, IEEE, 2016.
- [51] A. Zaporozhets, V. Babak, A. Sverdlova, V. Isaienko, K. Babikova, Development of a system for diagnosing heat power equipment based on iee 802.11s, in: A. Zaporozhets, V. Artemchuk (Eds.), Systems, Decision and Control in Energy II, Springer International Publishing, Cham, 2021, pp. 141–151. doi:10.1007/978-3-030-69189-9_8.
- [52] C. A. Santivanez, R. Ramanathan, Hazy Sighted Link State (HSLs) Routing: A Scalable Link State Algorithm, Internetwork Research Department, BBN Technologies, 2001.
- [53] J. Chroboczek, RFC 6126. The Babel Routing Protocol, 2011. URL: <https://datatracker.ietf.org/doc/rfc6126/>.
- [54] Z. J. Haas, M. R. Pearlman, P. Samar, The Zone Routing Protocol (ZRP) for Ad Hoc Networks, 2002. URL: <https://tools.ietf.org/html/draft-ietf-manet-zone-zrp-04>.
- [55] G. Pei, M. Gerla, T.-W. Chen, Fisheye state routing: a routing scheme for ad hoc wireless

- networks, in: Proceedings of IEEE International Conference on Communications. ICC 2000. Global Convergence Through Communications. vol.1, IEEE, 2000, pp. 70–74.
- [56] D. Vasiliev, A survey on routing protocols and error correction methods for data delivery in MANETs, in: Proceedings of the 4th Forum of Young Researchers in the Framework of International Forum “Education Quality, 2014, pp. 375–384.
- [57] S. ryong Kang, D. Loguinov, Impact of FEC overhead on scalable video streaming, in: Proceedings of the international workshop on Network and operating systems support for digital audio and video (NOSSDAV ’05), Association for Computing Machinery, New York, NY, USA, 2005, pp. 123–128.
- [58] K. Park, W. Wang, AFEC: An adaptive forward error correction protocol for end-to-end transport of real-time traffic, in: Proceedings of 7th International Conference on Computer Communications and Networks (Cat. No.98EX226), IEEE, 1998, pp. 196–205.
- [59] I. Sanchez, On adaptive forward error correction for real-time traffic, Master’s Degree Project, KTH, Stockholm, Sweden, 2004.
- [60] Y. Sohn, J. Hwang, S.-S. Kang, Adaptive packet-level fec algorithm for improving the video quality over ieee 802.11 networks, *Int. J. Softw. Eng. its Appl.* 6 (2012) 27–34.
- [61] S. Karande, H. Radha, Rate-Constrained Adaptive FEC for Video over Erasure Channels with Memory, in: Proceedings of IEEE International Conference on Image Processing (ICIP), Vol.4, IEEE, 2004, pp. 2539–2542.
- [62] Y. Liu, M. Claypool, Using redundancy to repair video damaged by network data loss, in: Proc. SPIE 3969, Multimedia Computing and Networking, 2000. doi:10.1117/12.373536.
- [63] C. Fragouli, E. Soljanin, Network coding applications, *Found. Trends Netw.* 2 (2007) 135–269. doi:10.1561/13000000013.
- [64] Y. Oh, M. Gerla, A. Tiwari, Robust MANET routing using adaptive path redundancy and coding, in: Proceedings of First International Communication Systems and Networks and Workshops, IEEE, 2009, pp. 1–10.
- [65] P. Koohyun, S. Yong-Sik, L. Hyun-Chan, Multicast routing by multiple tree routes, in: M. M. abd A. Bianco (Ed.), *Lecture Notes in Computer Science*, volume 1989, Springer, Berlin, Heidelberg, 2001, pp. 285–298. doi:10.1007/3-540-44554-4_19.
- [66] A. Biernacki, U. Krieger, Session level analysis of p2p television traces, in: S. Zeadally, E. Cerqueira, M. Curad, M.Leszczuk (Eds.), *Lecture Notes in Computer Science*, volume 6157, Springer, Berlin, Heidelberg, 2010, pp. 285–298. doi:10.1007/978-3-642-13789-1_15.
- [67] D. Bradler, J. Kangasharju, M. Muhlhauser, Optimally Efficient Multicast in Structured Peer-to-Peer Networks, in: Proceedings of 6th IEEE Consumer Communications and Networking Conference, IEEE, 2009, pp. 1–5.
- [68] V. Emelyanov, A. Abilov, Robustness of multiple-tree-based P2P streaming networks, in: Proceedings of the 2nd Forum of Young Researchers in the Framework of International Forum ’Education Quality–2010’, 2010, pp. 331–339.
- [69] F. Pianese, P2P Live Media Streaming: Delivering Data Streams to Massive Audiences within Strict Timing Constraints, Master Thesis, Institut Eurecom, Sophia-Antipolis, 2004.
- [70] B. R. Tamma, A. Badam, C. S. R. Murthy, R. R. Rao, K-tree: A multiple tree video multicast protocol for ad hoc wireless networks, *Comput. Netw.* 54 (2010) 1864–1884. doi:10.1016/

j.comnet.2010.02.013.

- [71] S. Mao, S. Lin, S. Panwar, Y. Wang, Reliable transmission of video over ad-hoc networks using automatic repeat request and multipath transport, in: Proceedings of IEEE 54th Vehicular Technology Conference. VTC Fall 2001 (Cat. No.01CH37211), IEEE, 2001, pp. 615-619.

ImpalaE: Towards an optimal policy for efficient resource management at the edge

Tania Lorido-Botran^a, Muhammad Khurram Bhatti^b

^aBilbao, Spain

^bInformation Technology University, Arfa Software Technology Park, Ferozepur Road, Lahore, Pakistan

Abstract

Edge computing is an extension of cloud computing where physical servers are deployed closer to the users in order to reduce latency. Edge data centers face the challenge of serving a continuously increasing number of applications with a reduced capacity compared to traditional data center. This paper introduces *ImpalaE*, an agent based on Deep Reinforcement Learning that aims at optimizing the resource usage in edge data centers. First, it proposes modeling the problem as a Markov Decision Process, with two optimization objectives: reducing the number of physical servers used and maximize number of applications placed in the data center. Second, it introduces an agent based on Proximal Policy Optimization, for finding the optimal consolidation policy, and an asynchronous architecture with multiple workers-shared learner that enables for faster convergence, even with reduced amount of data. We show the potential in a simulated edge data center scenario with different VM sizes based on Microsoft Azure real traces, considering CPU, memory, disk and network requirements. Experiments show that *ImpalaE* effectively increases the number of VMs that can be placed per episode and that it quickly converges to an optimal policy.

Keywords

Edge Computing, Policy Gradient, Reinforcement Learning, Efficient Resource Management

1. Introduction

Cloud Computing providers have popularized and quickly replaced private data centers. Many businesses, government organizations and research centers rely on external clouds to run their workloads. However, Cloud data centers are usually located far away from the end-user and the perceived latency might not be up to the standard. In recent years, the Edge Computing paradigm has augmented Cloud capabilities by placing computing facilities and services close to end users. Thus, Edge data centers are able to provide low latency and mobility to delay-sensitive applications. According to a Markov Growth study [1], Edge Computing was valued at USD 1.93 Billion in 2018 and is projected to reach USD 10.96 Billion by 2026. With this high growth in revenue, it is clear the increased interest in this services.

The Edge computing platform is expected to deliver consistent performance despite the rapid increase of application demand, specially coming from Internet-of-Things applications, such us

QuaInT 2021: Workshop on the Quantum Information Technologies, April 11, 2021, Zhytomyr, Ukraine
doors 2021: Edge Computing Workshop, April 11, 2021, Zhytomyr, Ukraine

✉ tania.lorido@deusto.es (T. Lorido-Botran); khurram.bhatti@itu.edu.pk (M.K. Bhatti)

🌐 <https://itu.edu.pk/faculty-itu/dr-khurram-bhatti/> (M.K. Bhatti)

🆔 0000-0002-9132-4435 (T. Lorido-Botran); 0000-0002-1974-8268 (M.K. Bhatti)

© 2021 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



CEUR Workshop Proceedings (CEUR-WS.org)

self-sufficient vehicles producing data from their various cameras, radar or accelerometers. The new challenge for edge service providers is to perform efficient resource management of their edge data centers with reduced computation and storage capabilities [2]. In particular, providers will look for automated solutions that can adapt to the varying demand and diverse workloads.

Reinforcement Learning (RL) is a family of self-adaptive algorithms that has been successfully applied to multiple domains. From the popular AlphaGo [3] for playing the game of Go, to autonomous driving, drug discovery, personalized recommendations and optimizing chemical reactions. RL has also been applied to for cloud resource optimization, both horizontal and vertical scalability [4] [5] [6]. Similarly, RL has the potential to provide an efficient and automated solution to the management of resource at the Edge.

2. Related work

Edge computing has received increasing attention in recent years. A common use case scenario is the off-loading of certain requests to different Edge data centers. Liu et al. [7] focus on the task scheduling problem and proposed an RL-based scheduling solution and successfully offload certain tasks to other data centers. Some authors have proposed DRL-based solutions for the offloading of VMs [8]. However, computation offloading might lead to unbalancing issues, as some edge data centers in the region could be overloaded while some others are in idle state [9].

Unbalanced data centers lead to performance degradation and wasted resources. One approach would be to spread the load equally among the different edge data centers. Puthal et al. [10] take this approach and propose a solution based on Bread-First-Search to keep the application load equally distributed. However, edge data centers are characterized from scarce resources compared to traditional servers and a load balancing approach will not maximize the number of applications that can be served.

There are clashing objectives between the end-user and the service provider. The end-user expects guaranteed application performance, while the provider wants to maximize its revenue by increasing the number serviced applications. In order to meet both end-user and provider's expectation, it seems reasonable to define the overall objective as a consolidation problem: placing as many requests as possible using the minimal capacity, always subject to resource constraints. With this goal in mind, some authors have focused on the execution of tasks on edge data centers [7, 11]. Zhu et al. [11] successfully introduce two approximation scheduling algorithms focused on minimizing energy consumption and reducing the overall task execution delay.

As stated by Khan et al. [2], edge data centers can benefit from the use of Virtual Machines to co-allocate multiple applications in the same physical server. Tao et al. [12] gather a list of proposed solutions that handle the VM placement on edge data centers. Proposed optimization methods range from Mixed-Linear Non-Linear Programming [13, 14] to Particle Swarm Optimization [15]. However, there seems to be a lack of solutions exploring the potential of RL for optimal VM placement in edge data centers, aiming at minimizing resource wastage.

To the best of our knowledge, this is the first attempt to explore the application of policy-

gradient RL methods to achieve efficient resource management in edge data centers. This paper introduces an agent (named *ImpalaE* that uses policy-gradient method to find the optimal placement policy and a distributed architecture that enables fast training. The resource management problem is formulated with a bi-objective function that tries to (1) reduce the number of physical servers utilized and (2) maximize the number of applications that can be placed in the edge data center.

3. Background: Policy-Based Reinforcement Learning

The basic elements in an RL problem are the *agent* and the *environment*. The agent continuously interacts with the environment, observes the current *state* and decides the best *action* to take. After some time, the agent will observe the *reward* obtained after applying that action. The goal is to learn an optimal *policy* $\pi_\theta(s|a)$ that maps each state with its optimal action.

3.1. Vanilla Policy Gradient (PG)

There are different approaches to learn the optimal policy. As the name suggests, *Policy-based* algorithms directly learn the policy without an intermediary function. The policy $\pi_\theta(s|a)$ is approximated with deep neural network that has a vector of policy parameters θ . The goal is to adjust the values of these parameters, such that the policy maximizes the reward obtained from the environment.

Policy gradient methods rely on applying stochastic gradient descent as an iterative process. At each step, the algorithm estimates the gradient of some estimated scalar performance objective $J(\theta_k)$ and updates the policy parameters θ :

$$\theta_{k+1} = \theta_k + \alpha \nabla_\theta J(\theta_k) \quad (1)$$

The gradient of $J(\pi_\theta)$ for the *Vanilla Policy Gradient* can be calculated as follows:

$$\nabla_\theta J(\pi_\theta) = \mathbb{E}_{\tau \sim \pi_\theta} \sum_{t=0}^T \nabla_\theta \log \pi_\theta(a_t|s_t) A^{\pi_\theta}(s_t, a_t), \quad (2)$$

where τ is an episode, that is a sequence of states and actions, e.g. a pre-defined sequence of requests and their corresponding placements in the edge data center; and \mathbb{E} denotes calculating average over a batch of samples.

The main drawback in Vanilla PG is the high gradient variance, that will hinder the convergence to an optimal policy. The advantage function A^{π_θ} included in the gradient function helps in reducing such variance. Without going deep into the details, the advantage function evaluates how good an action is compared to the average action for a specific state.

3.2. Proximal Policy Optimization (PPO)

PPO [16] aims to optimize the gradient update taken at each step, ensuring that it minimizes the objective function, while ensuring that the difference to the previous policy is relatively

small. Too big of an update might cause a divergence from the optimal policy. PPO imposes a constraint to the policy gradient updates as follows:

$$J(\theta) = L^{CLIP}(\theta) = \mathbb{E}_t [\min(r_t(\theta)A_t, \text{clip}(r_t(\theta), 1 - \epsilon, 1 + \epsilon)A_t)] \quad (3)$$

There are two main modifications with respect to the vanilla PG method. The first one is $r_t = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)}$, which computes a ratio between the current policy (after update) and the older policy (just before the update). Additionally, PPO relies on a clipping function $\text{clip}(r_t(\theta), 1 - \epsilon, 1 + \epsilon)$ that keep the value of r_t between certain range defined by $1 - \epsilon$ and $1 + \epsilon$.

PPO with Clipping is used as the core agent for *ImpalaE*. The full logic is depicted in Algorithm 1:

Algorithm 1: PPO with clipping

Input: initial policy parameters θ_0 , clipping threshold ϵ

for $0, 1, 2, \dots$ **do do**

 Collect set of partial trajectories (episodes) τ on policy $\pi = \pi(\theta)$

 Estimate advantages A_t using any advantage estimation algorithm

 Update the policy by maximizing the policy the PPO-Clip objective:

$\theta_{k+1} = \arg \max_{\theta} L^{CLIP}(\theta_k)$, typically, by taking K steps of minibatch stochastic gradient descent with Adam optimization

end

3.3. Importance Weighted Actor-Learner Architectures (IMPALA)

IMPALA [17] is a state-of-the-art algorithm produced by DeepMind. It uses the vanilla Policy Gradient at its core, but also introduces two significant improvements: a distributed architecture, and a correction algorithm V-trace. First, it introduces a highly-scalable architecture that relies on a single (or multiple) learner and multiple workers (see figure 1. In traditional RL approaches [18], each worker updates its local model parameters before each episode and communicates gradients to the main learner. IMPALA proposes a loosely coupled architecture where each worker focuses on collecting trajectories of experience (states, action, rewards). Then, the learner asynchronously samples batches of experiences from the workers, computes the policy gradients and updates the current model. This architecture enables the learner to be accelerated by a GPU and to distribute the workers across different nodes and collect experience on different domains (e.g. independent edge data centers).

The high scalability of the IMPALA architecture comes at a cost. Each worker interacts with its environment based on a policy that is slightly older than the main learner’s policy, since the learner broadcasts the updated weights in a period and asynchronous manner. In order to address this divergence, Espeholt et al. [17] introduce a correction algorithm called *V-trace* that readjusts the value function $V(s)$ for each state and account for the lag in each action decision.

4. *ImpalaE*: efficient resource management at the Edge

This paper introduces *ImpalaE*, an agent designed to address the specific resource management needs from Edge Computing paradigm. The agent specializes in edge data centers that use Virtual Machines as an abstraction layer to place applications. It relies on the use of Policy Gradient Reinforcement Learning to learn and adapt to different VM request arrival patterns and dynamic resource usage. By leveraging a combination of PPO with an asynchronous architecture, it quickly finds the optimal placement policy that squeezes the maximum performance out of the reduced capacity of an edge data center. As a first step, the Edge computing environment is formulated to be suitable for an RL-based agent.

4.1. Environment modeling

The scenario is one or more edge data centers composed of n physical servers. Each physical server has a given capacity for a set of resources, m . The agent has to learn the optimal policy π that matches each incoming request, expressed as a VM type with specific resource requirements, with the best physical server available. The overall goal is to maximize the number of requests that can be served given the current capacity. With this goal in mind, the resource management problem on edge data centers can be formulated as a Markov Decision Process (MDP) as follows:

State space: The state s at time t is defined as the current resource usage in the data center, together with the request received at time t . The resource usage of each physical server is expressed as a normalized variable, ranged $[0, 1]$, for each of the resources considered m . Additionally, each physical server has a binary variable associated s , which indicates if it is active (it has any load assigned to it) or not. Overall, the resource usage of the data center is a multi-dimensional vector $[n, m + 1]$. Each request v corresponds to a VM type, defined a set of m resource requirements that need to be satisfied. For the current case, we will consider $m = 4$ resources, namely CPU, memory, disk and network capacity.

Action space: The action space A is the set of n physical servers available in the data center. At time t , A_t is defined as the subset of servers where the current request v could be placed, that is, never exceeding the capacity of the machine:

$$A_t = \{a \in A \mid \sum_{i=1}^m u_{a,i} + v_i \leq 1\} \quad (4)$$

where $u_{a,i}$ is the current utilization value for physical server a and resource i and v_i is the capacity requested for resource i .

Reward definition: The primary goal in the edge data center is to maximize the number of requests that can be served with the available capacity. The reward function R is defined with this goal in mind and it is composed of two objectives. The first objective R_1 accounts for the amount of unused resources in the data center, normalized by the total capacity, $n * m$:

$$R_1 = - \frac{\sum_{i=1}^n s_i * f_i}{n * m} \quad (5)$$

where f_i is the total amount of free capacity across m resources for physical server i . The reward only accounts for free resources in active physical servers, defined with $s_i = 1$.

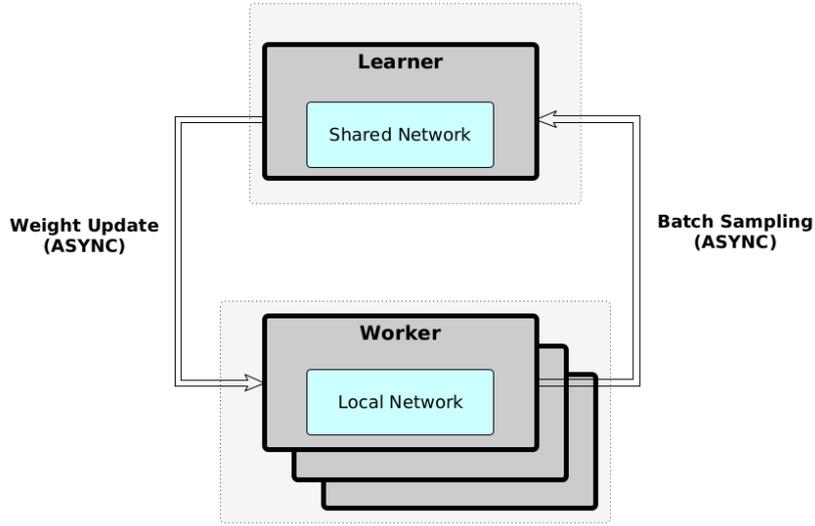


Figure 1: Architecture for *ImpalaE*

The second part of the reward function directly accounts for the number of requests remaining to be placed in the current episode:

$$R_2 = -\frac{P - V}{V} \quad (6)$$

The final reward function is simply the linear combination of R_1 and R_2 with equal weights.

4.2. Agent architecture

The proposed agent is based on the asynchronous architecture introduced by [17], from which it takes its name, *ImpalaE*. It consists of a main learner and one or more workers (see figure 1). Each worker interacts with the environment using their local copy of the network (only performing inference) and store (state, action, reward) samples. The main learner asynchronously samples batches from each of the workers and uses them to update the central network. After that, the learner broadcasts the network updated new weights to each of the learners in an asynchronous manner. This architecture enables for faster, parallel collection of environment info, which in turn leads for a quick convergence toward the optimal policy.

The learner is based on PPO algorithm with clipping (see Algorithm 1) for finding the optimal policy, that is, the best placement of each incoming VM request to the edge data center. The network model uses a shared architecture for the policy and the value function. It consists of feed-forward neural network with TanH activation function. In order to speed up the convergence, the learner makes use of a *buffer replay*. This buffer stores all the instances composed of (*state, action, reward, next_state*). Periodically, the learner samples *batch_size* instances sampled from the buffer to perform a gradient update in the policy network. Finally, the learner leverages *V-trace*[17], a correction algorithm that fixes discrepancies in the instances as a result

Table 1
Parameter configuration for *ImpalaE*

Type	Parameter	Symbol	Value
Scenario	Number of physical servers	n	500
	Number of resources	m	4
	Number of actions	$ A $	n
<i>ImpalaE</i>	Learning rate	α	0.005
	Train Batch size		500
	Optimization algorithm		Adam
	Clipping parameter		0.4
	Number of workers		2
Network Model	Input layer		$(n + 1) * (m + 1)$
	Hidden layer 1		1024
	Hidden layer 2		1024
	Output layer		n

of the asynchronous architecture. Table 1 contains a summary of the configuration used in the experimental evaluation:

5. Experimental evaluation

The following set of experiments are defined to evaluate the general performance of *ImpalaE*, compared against other policy-gradient methods from the state-of-art, and also the convergence and scalability of the agent architecture.

Testing environment: A simulated environment of an edge data center with certain number of homogeneous physical servers (same capacity). Each physical server and VM request is defined in terms of their CPU, memory, network and disk requirements. The resource specification is normalized between 0 and 1 (required by the model input). The experiments are based on real-world traces collected from Microsoft Azure data center [19, 20] (in particular, 15 VM types assigned to a machine identified with id 0). All algorithms are implemented in Python v3.8 and models are implemented using Tensorflow v2.5.0, and trained on a GPU. The hardware for the experiments is a machine with Intel Cor i7-10510U, 16GB of RAM, NVIDIA GeForce MX330.

Baseline methods: *ImpalaE* is compared against one heuristic method, Round Robin, and two other state-of-the-art RL algorithms: (vanilla) Policy Gradient (PG) and Proximal Policy Optimization (PPO).

5.1. Convergence and performance evaluation

The main goal of *ImpalaE* is to quickly converge to the optimal placement policy, the one that optimizes resource usage and maximises the number of requests that can be accommodated in the edge data center. In the first scenario, the data center is composed of 500 physical servers and has enough capacity to serve an episode consisting of 1000 VM requests. Requests are randomly drawn from a set of 14 VM types extracted from Azure data center traces (machineID 0).

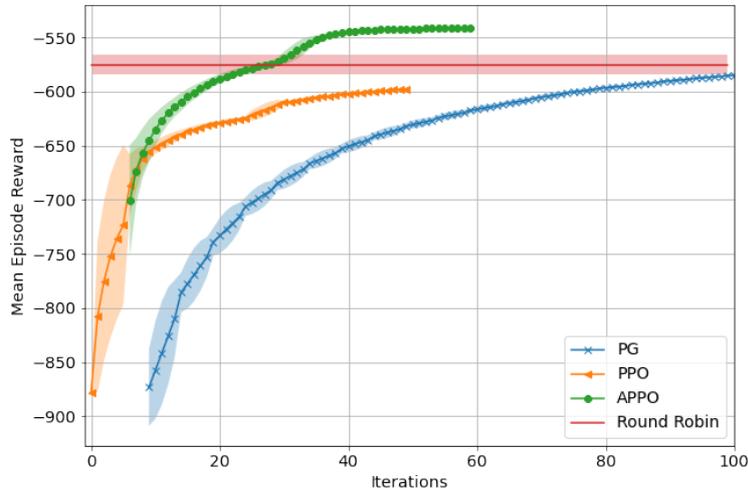


Figure 2: Training results for *ImpalaE*, PPO and PG

For fairness of results, the same network architecture is used for *ImpalaE*, PPO and PG. The network contains 2 hidden layers, with 1024 units each. When the agent architecture allows, two workers are used in the training process.

Figure 2 shows the convergence results for *ImpalaE*, PPO, PG and Round Robin. In less than 30 iterations, *ImpalaE* quickly converges to the optimal policy. In contrast, both PG and PPO achieve a sub-optimal policy (lower than the heuristic-based agent, Round Robin), with lower mean reward per episode. PG takes a high number of iterations to converge.

The second scenario is designed to stress the agent ability to make optimal placement decision in cases of high occupancy. The data center consists again of 500 physical servers, but in this case, 2000 VM requests have to be placed in each episode. The data center does not have enough capacity to serve all of them. Figure 3 shows the percentage of placed requests, calculated as the mean of the last 5 iterations. The heuristic-based agent (Round Robin) only manages to accommodate 25% of the requests. This is inherent to the nature of Round Robin algorithm, that tries to spread out the load across different nodes. This naturally leads to resource fragmentation and limits the amount of resources that can be placed in a data center. In contrast, RL-based agents quickly learn a policy that tries to maximize the resource utilization. Both state-of-the-art baseline methods, PPO and PG, achieve a higher rate of successful placements in contrast to the heuristic agent, 89% and 91% respectively. Thanks to its parallel architecture, *ImpalaE* agent is able to explore more scenarios in a shorter amount of time and thus, further train the policy to score the highest placement rate, 94% of the 2000 VM requests within the same edge data center.

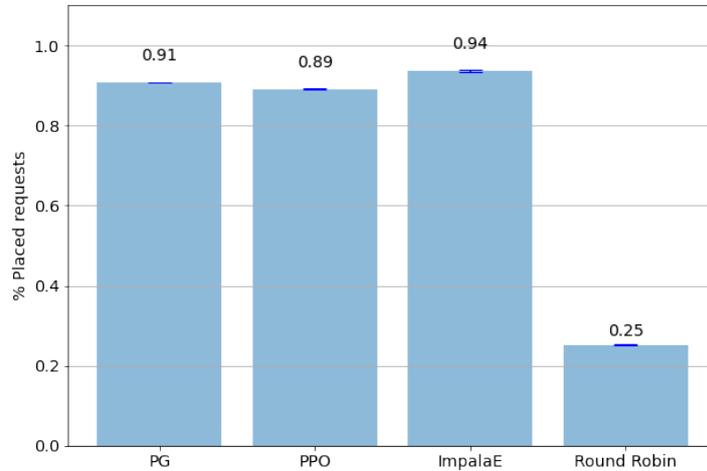


Figure 3: Mean percentage of placed requests per episode

5.2. Agent scalability

The single learner-multiple worker architecture makes the proposed agent highly scalable, which in turns allows for faster convergence. The next experiment explores the impact of the number of workers in the training process. The scenario uses 500 physical servers and 1000 VM requests per episode, and compares the performance of PPO and *ImpalaE* (see figure 4). As expected, PPO shows the slowest convergence rate, easily surpassed by *ImpalaE* with a single worker. At its core, *ImpalaE* relies on several workers interacting with the environment and gathering as much information as possible, that is, they explore different data center scenarios and placement decisions and record the outcome of such decision (did it improved the request acceptance?). For this reason, increasing the number of works naturally improves the placement policy (higher reward) and leads to an earlier convergence. In this particular case, *ImpalaE* achieves the best results with 4 works. However, it is interesting to note that an additional worker (up to 5) actually achieves a slightly worse policy, which might be due to high variance in the sampling. We leave for future work the deeper analysis of the algorithm stability during training.

A well-known drawback of RL-based agents is their extremely long times (hours) needed to converge to an optimal policy, which makes it unfeasible to deploy such agent in a production environment. This experiment analyses the overall training time of the agent for a data center composed of 500 physical servers. As figure 5 shows, the baseline method, PPO, requires around 37 minutes of total training time. In contrast, the parallel architecture of *ImpalaE* allows it to further reduce the training time to only 4.4 minutes with 4 workers. This is especially appealing feature for highly dynamic environments, where the workload request patterns and resource usage change abruptly.

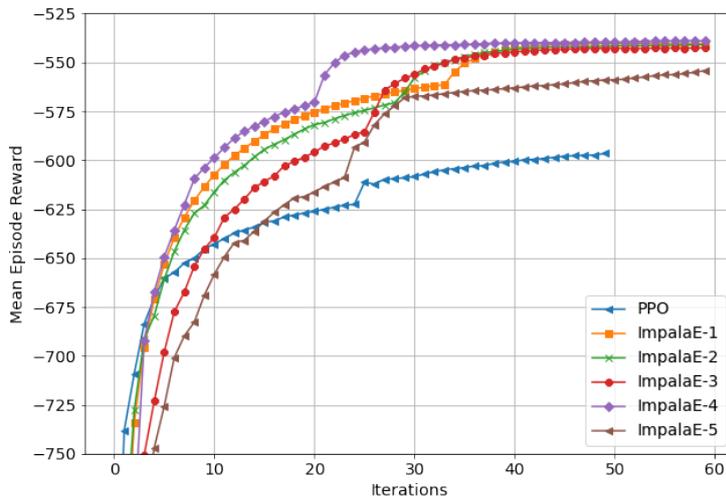


Figure 4: Mean reward per episode during training time

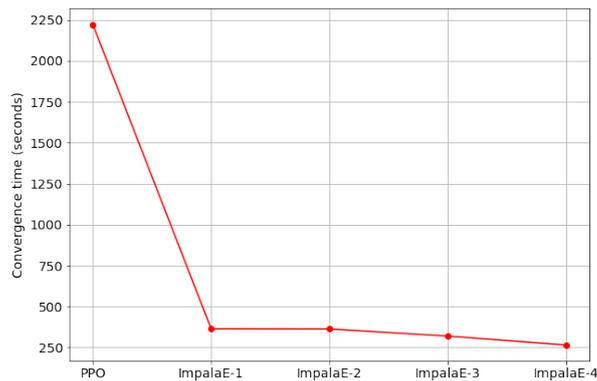


Figure 5: Convergence time

6. Conclusions and future work

Edge computing was born as an extension of widely used Cloud computing, with the differences that computing resources are located closer to the end-user and this is imperative for latency-critical applications. Edge computing providers face an additional challenge when making an optimal resource management of their data centers with reduced capacity, while trying to meet the client demand. This paper introduces *ImpalaE*, an agent based on Deep Reinforcement Learning, specially designed to optimize resource usage at the edge. It leverages Proximal Pol-

icy Optimization for finding the best placement policy for applications in edge data centers. It is also based on the IMPALA architecture, an asynchronous paradigm composed of one learner and multiple parallel workers that speed up the convergence, even with reduced amount of data. The paper also introduces modeling of the edge computing environment as a Markov Decision Process with a bi-objective reward function specially designed to squeeze maximum performance. The validity of *ImpalaE* is assessed in a simulated environment considering VM requests based on real Microsoft Azure traces and considering CPU, memory, disk and network requirements.

The full potential of IMPALA architecture is yet to be explored. It has demonstrated higher performance with less data and ability to transfer information among tasks [17]. One natural extension would be to expand *ImpalaE* to multiple data centers, that learn an optimal policy per data center, but also benefit from asynchronously exchanging information among different agents. However, there is also a need for deeper experimentation about the training stability for larger number of workers.

The current environment model takes into account the network bandwidth needs of each application. However, it could be further extended to consider the communication pattern among different nodes or VMs within the application. The reward function could be augmented with other objectives, such as application latency experienced by end-user or the data center energy utilization.

References

- [1] M. I. Reports, Global Edge Computing Market Size, Status And Forecast 2020-2026, 2021-02.
- [2] W. Z. Khan, E. Ahmed, S. Hakak, I. Yaqoob, A. Ahmed, Edge computing: A survey, *Future Generation Computer Systems* 97 (2019) 219–235.
- [3] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, Others, Mastering the game of Go with deep neural networks and tree search, *nature* 529 (2016) 484–489.
- [4] Z. Wang, C. Gwon, T. Oates, A. Iezzi, Automated cloud provisioning on aws using deep reinforcement learning, *arXiv preprint arXiv:1709.04305* (2017).
- [5] B. Du, C. Wu, Z. Huang, Learning resource allocation and pricing for cloud profit maximization, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 2019, pp. 7570–7577.
- [6] S. Zhang, T. Wu, M. Pan, C. Zhang, Y. Yu, A-SARSA: A Predictive Container Auto-Scaling Algorithm Based on Reinforcement Learning, in: *2020 IEEE International Conference on Web Services (ICWS)*, IEEE, 2020, pp. 489–497.
- [7] J. Liu, Y. Mao, J. Zhang, K. B. Letaief, Delay-optimal computation task scheduling for mobile-edge computing systems, in: *2016 IEEE International Symposium on Information Theory (ISIT)*, IEEE, 2016, pp. 1451–1455.
- [8] X. Qiu, L. Liu, W. Chen, Z. Hong, Z. Zheng, Online deep reinforcement learning for computation offloading in blockchain-empowered mobile edge computing, *IEEE Transactions on Vehicular Technology* 68 (2019) 8050–8062.

- [9] K.-K. R. Choo, R. Lu, L. Chen, X. Yi, A foggy research future: Advances and future opportunities in fog computing research, 2018.
- [10] D. Puthal, M. S. Obaidat, P. Nanda, M. Prasad, S. P. Mohanty, A. Y. Zomaya, Secure and sustainable load balancing of edge data centers in fog computing, *IEEE Communications Magazine* 56 (2018) 60–65.
- [11] T. Zhu, T. Shi, J. Li, Z. Cai, X. Zhou, Task scheduling in deadline-aware mobile edge computing systems, *IEEE Internet of Things Journal* 6 (2018) 4854–4866.
- [12] Z. Tao, Q. Xia, Z. Hao, C. Li, L. Ma, S. Yi, Q. Li, A survey of virtual machine management in edge computing, *Proceedings of the IEEE* 107 (2019) 1482–1499.
- [13] Q. Fan, N. Ansari, Cost aware cloudlet placement for big data processing at the edge, in: 2017 IEEE International Conference on Communications (ICC), IEEE, 2017, pp. 1–6.
- [14] S. Mondal, G. Das, E. Wong, CCOMPASSION: A hybrid cloudlet placement framework over passive optical access networks, in: IEEE INFOCOM 2018-IEEE Conference on Computer Communications, IEEE, 2018, pp. 216–224.
- [15] Y. Li, S. Wang, An energy-aware edge server placement algorithm in mobile edge computing, in: 2018 IEEE International Conference on Edge Computing (EDGE), IEEE, 2018, pp. 66–73.
- [16] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O. Klimov, Proximal policy optimization algorithms, arXiv preprint arXiv:1707.06347 (2017).
- [17] L. Espeholt, H. Soyer, R. Munos, K. Simonyan, V. Mnih, T. Ward, Y. Doron, V. Firoiu, T. Harley, I. Dunning, Others, Impala: Scalable distributed deep-rl with importance weighted actor-learner architectures, in: International Conference on Machine Learning, PMLR, 2018, pp. 1407–1416.
- [18] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, K. Kavukcuoglu, Asynchronous methods for deep reinforcement learning, in: International conference on machine learning, PMLR, 2016, pp. 1928–1937.
- [19] O. Hadary, L. Marshall, I. Menache, A. Pan, E. E. Greeff, D. Dion, S. Dorminey, S. Joshi, Y. Chen, M. Russinovich, Others, Protean: VM Allocation Service at Scale, in: 14th USENIX Symposium on Operating Systems Design and Implementation (OSDI 20), 2020, pp. 845–861.
- [20] Trace, Azure Public Dataset, 2021-02. URL: <https://github.com/Azure/AzurePublicDataset>.

Analysis of attacks on components of IoT systems and cybersecurity technologies

Nadiia M. Lobanchykova^a, Ihor A. Pilkevych^b and Oleksandr Korchenko^c

^aZhytomyr Polytechnic State University, 103 Chudnivsyka Str., Zhytomyr, 10005, Ukraine

^bS.P. Korolev Zhytomyr Military Institute, 22 Mira Ave., Zhytomyr, 10004, Ukraine

^cUniversity of Bielsko-Biala, Willowa 2, 43-300 Bielsko-Biala, Poland

Abstract

This article presents the results of IoT analysis, methods and ways of their protection, prospects of using edge computing to minimize traffic transmission, decentralization of decision-making systems, and information protection. A detailed analysis of attacks on IoT system components was carried out and protection recommendations were developed.

Keywords

Internet of Things, IoT, edge computing, cybersecurity, attack, protection

1. Introduction

In the last few years, the Internet of Things (IoT) systems has been widely developed and implemented. The Internet of Things market research notes a steady and rapid increase in the number of such devices every year. If analysts currently estimate the number of active IoT devices at 21 billion, in a few years their number will exceed 50 billion [1, 2]. Due to the development and widespread introduction of IoT technologies, information security experts are concerned about their level of protection [3, 4, 5, 6, 7]. According to them, the huge number of poorly protected Internet devices gives new opportunities to cybercriminals. Yes, there are already known cases of breakage of several IoT systems. This task is especially relevant when using these tools at critical infrastructure.

New technologies and new tools are creating new types of cyber threats. Many companies today have introduced their protection models, which are constantly trying to standardize, correlate and implement.

The development of information technology makes its adjustments in the field of information security. Therefore, the advent and edge of computing technologies allow solving several

QualnT 2021: Workshop on the Quantum Information Technologies, April 11, 2021, Zhytomyr, Ukraine

doors 2021: Edge Computing Workshop, April 11, 2021, Zhytomyr, Ukraine

✉ lobanchikovanadia@gmail.com (N.M. Lobanchykova); igor.pilkevich@meta.ua (I.A. Pilkevych);

agkorchenko@gmail.com (O. Korchenko)

🌐 <https://ztu.edu.ua/ua/structure/faculties/fikt/> (N.M. Lobanchykova);

<https://scholar.google.com.ua/citations?user=UK3h53oAAAAJ> (I.A. Pilkevych);

<http://www.kinf.ath.bielsko.pl/pl/oleksandr-korchenko> (O. Korchenko)

🆔 0000-0003-4010-0308 (N.M. Lobanchykova); 0000-0001-5064-3272 (I.A. Pilkevych); 0000-0003-3376-0631 (O.

Korchenko)



© 2021 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

cybersecurity problems. The main trend of edge computing is remote monitoring and data processing directly on IoT devices. The main advantage of this approach is the minimization of processing time and decision-making due to the absence of the need to transfer all data to a data center (data center) or cloud. The combination of IoT and edge computing is a promising area and can be used in industry, hospitals, climate control systems, and “smart” buildings, in the management of the infrastructure of the city or region, in trade and logistics networks [8]. Of particular interest is the use of edge computing for network security monitoring and access control systems. This technology is quite effective in preventing certain types of attacks and the spread of malicious software. Also, performing calculations immediately after receiving a signal allows you to decide whether to generate an alarm, move the “object” to quarantine, isolate, if necessary, several IoT devices to prevent network compromise or system failure. The widespread introduction of IoT devices creates large amounts of information that are increasingly difficult to transfer to a data center or cloud, process and store them, so the use of edge computing is a necessity for many areas of the digital society. The study of traffic minimization technologies, data storage, resources, and security in IoT using edge computing is a crucial task today for the development of digital society and the entry of humanity into the fourth industrial revolution (Industry 4.0) [9].

2. Theoretical background

The analysis of works [1, 2, 10] confirms the relevance of research in the field of IoT, which is associated with the benefits of these devices and technologies, as well as the transition of mankind to the use of Industry 4.0. In [1, 2, 3], the authors note the incredibly rapid pace of IoT implementation in various areas of the digital society. Immerman [1] testifies that at the beginning of the implementation of IoT, sensors sent data to the cloud, where they were processed, analyzed, and stored, and making management decisions. As the number of devices increased exponentially, the load on both the data channels and the storage cloud (trillions of gigabytes) increased, so the use of edge computing became a necessity, not a whim. The author notes that the use of edge computing and cloud technologies together is possible, and in some cases necessary, especially in industry. Edge computing is the most important component of IoT, which helps reduce latency and increase the reliability of deployed systems [1]. In [2] the models of IoT architecture are presented, the need for IoT protection is determined, the results of research on the construction of information protection systems for IoT devices, including shared and centralized, conducted simulation load depending on the number of devices.

Security issues are quite relevant and aimed at the comprehensive protection of information. Thus, Blyler [3] focuses on the complexity of IoT protection and presents eight key security technologies: network security, authentication, encryption, security attack, security analytics, and threat forecasting, interface protection, delivery mechanisms.

Prospects for implementation and threats facing IoT systems are presented in [4, 5, 6, 10, 11, 12]. The analysis of these works confirm the relevance of security issues, areas of protection, and the main conceptual approaches to security. Loud cyberattacks have occurred more than once and the number of hacker attacks is growing [7, 13, 14, 15]. The urgency of the problem is underscored by incidents, the loss of capital from which is measured in billions of dollars.

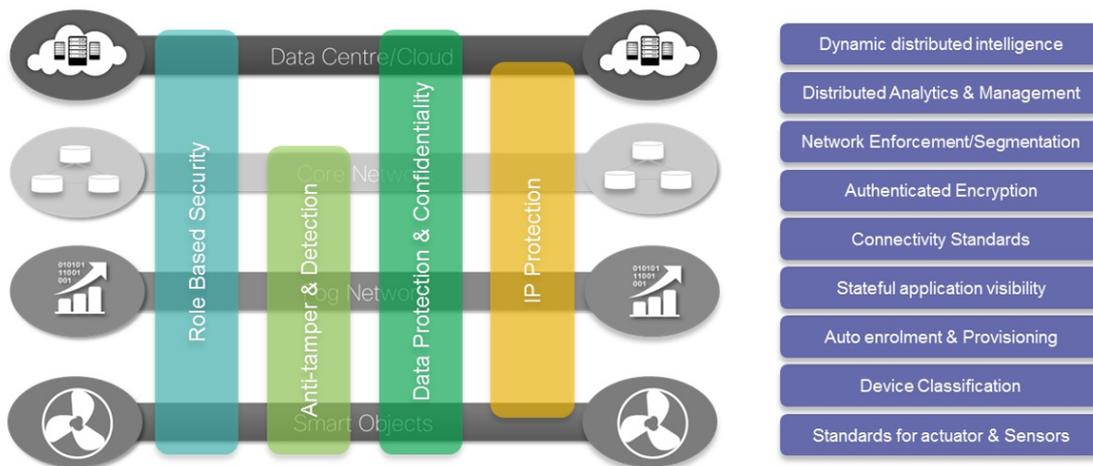


Figure 1: IoT security environment

In total, HP experts have identified about 25 different vulnerabilities in each of the studied devices and their mobile and cloud components [13]. The conclusion of HP experts is disappointing: a secure IoT system does not exist today. The particular danger to the Internet of Things is hidden in the context of the spread of targeted attacks. It is only necessary for intruders to show interest in anyone, and our helpers from the world of IoT turn into traitors, openly open access to the world of their owners.

Because the issue is extremely acute, companies that develop equipment, communications, network devices, software, and cybersecurity companies are looking for means to protect IoT devices [15]. One of the leading companies in the development of security in IoT is Cisco Systems, which played a leading role in the development of the IoT model at the World IoT Forum, developed the IoT security framework, which became a useful addition to the reference model [13]. Figure 1 shows the security environment associated with the logical structure of the IoT.

The Cisco IoT model is a simplified version of the World IoT Forum model. Figure 1 shows specific functional areas of security on top of the four levels of the IoT model. The Cisco document also proposes an IoT security concept that defines the components of the IoT security feature, covering all levels: authentication, authorization, network policy and security analytics.

Humanity's entry into the Industry 4.0 [10] creates new challenges and opportunities for Ukraine. The new cyber threats are associated with the widespread use of Industry 4.0 technology, which can have catastrophic consequences when attacking regime facilities. This task is especially relevant in the organization of temporary protection of the perimeter of the regime object when there are limited resources, adverse weather conditions, and unforeseen terrain.

Portable devices account for the largest number of attacks, and the use of wireless communication technologies between system elements creates the preconditions for a cyber-attack on the system. According to [4, 5, 6, 10, 11, 12], unauthorized access is most often carried out

by hackers through entry points (access) to the corporate network or used to launch a DDoS attack. Given the large number of sensors connected to the system, the use of wireless networks, cloud services, etc. does not provide a reliable perimeter of cybersecurity of the object. Another area is the theft of confidential user data (companies). The powerful potential of cyber threats has the technology of machine learning and the use of artificial intelligence systems through dual-purpose (the algorithms used can both counteract cyber-attacks and create them). New technologies create new cyber threats, which can be resisted only with the use of new information technologies.

Global statistics compiled by Cisco in 2017 [10] show:

- vulnerabilities (“holes”) in modern security systems allow up to 65% of cyber incidents,
- human factor – critical (if we scale it to the number and complexity of cyber threats) reduction of the level of literacy of users – up to 48% of incidents,
- 55% of organizations are unable to establish the cause of the incident,
- the average time to establish such a cause in the modern informational security and cybersecurity industry is 100 days.

Leading companies and specialists implement multi-level comprehensive protection systems based on the use of the latest technical tools, qualified personnel, control procedures, administrative regulations with strict compliance with them. In such systems, the emphasis is on setting up early warning systems that monitor the operation of IT equipment in real-time, notify administrators in the event of any abnormal activity, allow timely detection of attacks, as well as analyze potential threats. The criteria for the stability of such a defense system are the ability to respond to attacks in a timely and adequate manner and to restore the operation of the object with minimal losses [10].

3. Results

Our research system is a wireless IoT system, the hardware of which can be divided into the following elements [3, 4, 11, 6, 12]:

1. communication subsystem (wireless communication in the sensor network, includes a radio receiver),
2. computing subsystem (data processing, node functionality),
3. sensor subsystem (network connection with the “outside world”),
4. power subsystem.

Tasks facing the system to the hardware:

- low electricity consumption,
- the ability to work with a large number of nodes at relatively short distances,
- relatively low cost,

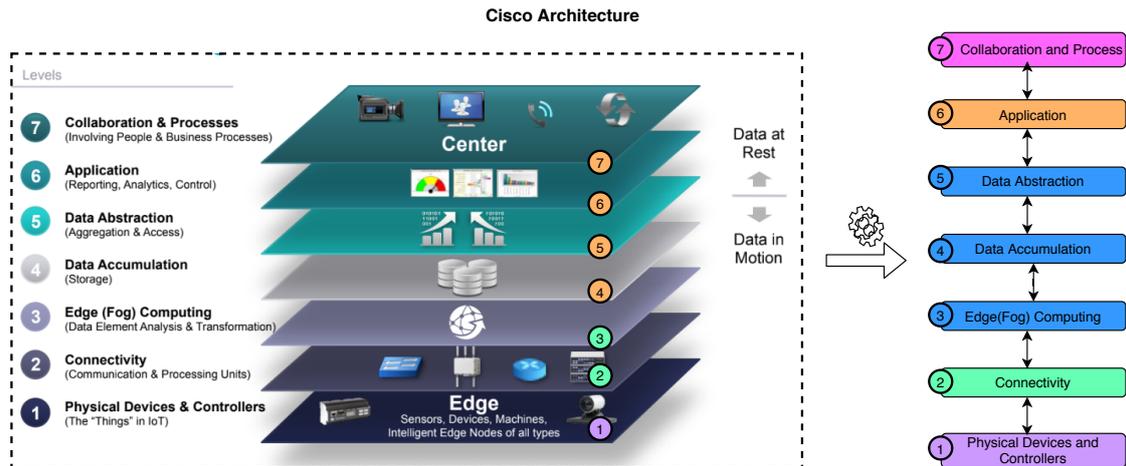


Figure 2: Cisco IoT Architecture

- work autonomously and without maintenance,
- have a camouflage effect,
- be resistant to the environment.

We choose the 7-level architecture of IoT systems, proposed by Cisco (figure 2).

Given the fact that sensor networks are vulnerable to many attacks, the issue of cybersecurity is especially relevant in the implementation of IoT systems to protect the perimeter of the regime object.

We assume that it is necessary to carry out temporary protection of the perimeter during the transportation of cargo/person/reconnaissance operation. Created using Cisco Packet Tracer simulation of one protection zone of the IoT perimeter security system is presented in figure 3. This scheme contains a set of devices used to create a zone of the temporary perimeter security system.

Also performed modeling of a typical fire alarm system of a separate room on the example of a garage (figure 4). The set of devices is typical.

The constructed computer models, figures 3 and 4 allow us to research to identify potential cyber threats and develop recommendations for the protection of IoT components. The results of modeling and countering cyberattacks will be presented in future studies.

Modeling of systems allowed to determine that the main areas that need attention from cybersecurity are:

- communication security,
- protection of the devices themselves,
- control over the operation of devices,
- control of network interaction.

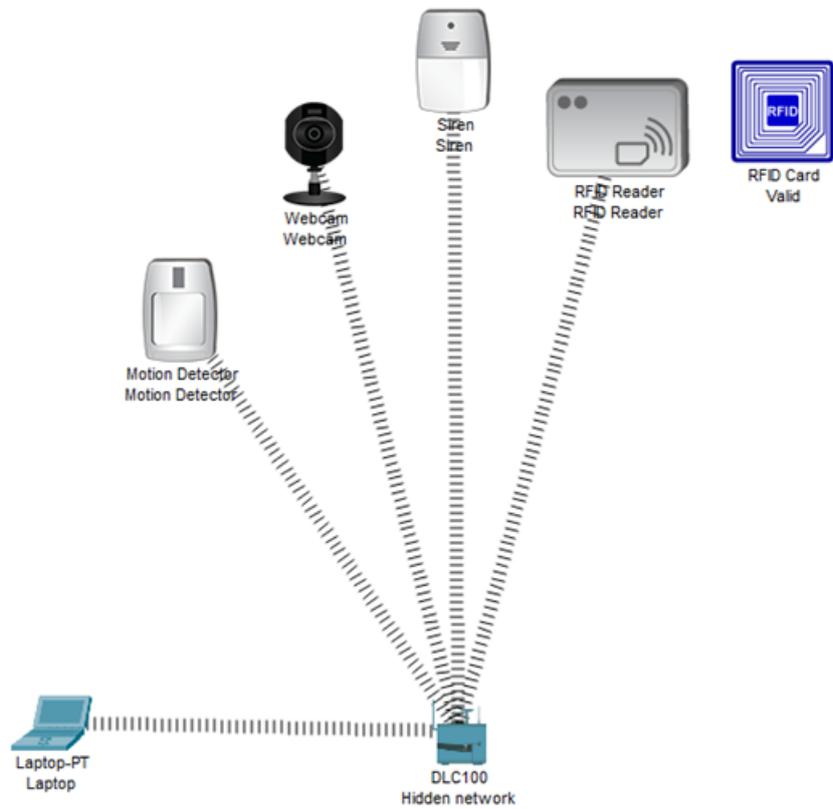


Figure 3: Cluster protection zone

As a result of research and analysis of the most likely attacks on simulated systems, the following classification of attacks is proposed (figure 5):

- Denial-of-Service (DoS) (D):
 - physical level (H):
 - * obstacle attack (H_1)
 - * attack of interference in the IoT system (H_2)
 - channel level (C):
 - * collision attack (C_1)
- attacks on routing protocols (R):
 - “Black Hole” attack (R_1)
 - selective forwarding attack (R_2)
 - “Rapid onslaught” attack (R_3)
 - “Funnel” attack (R_4)

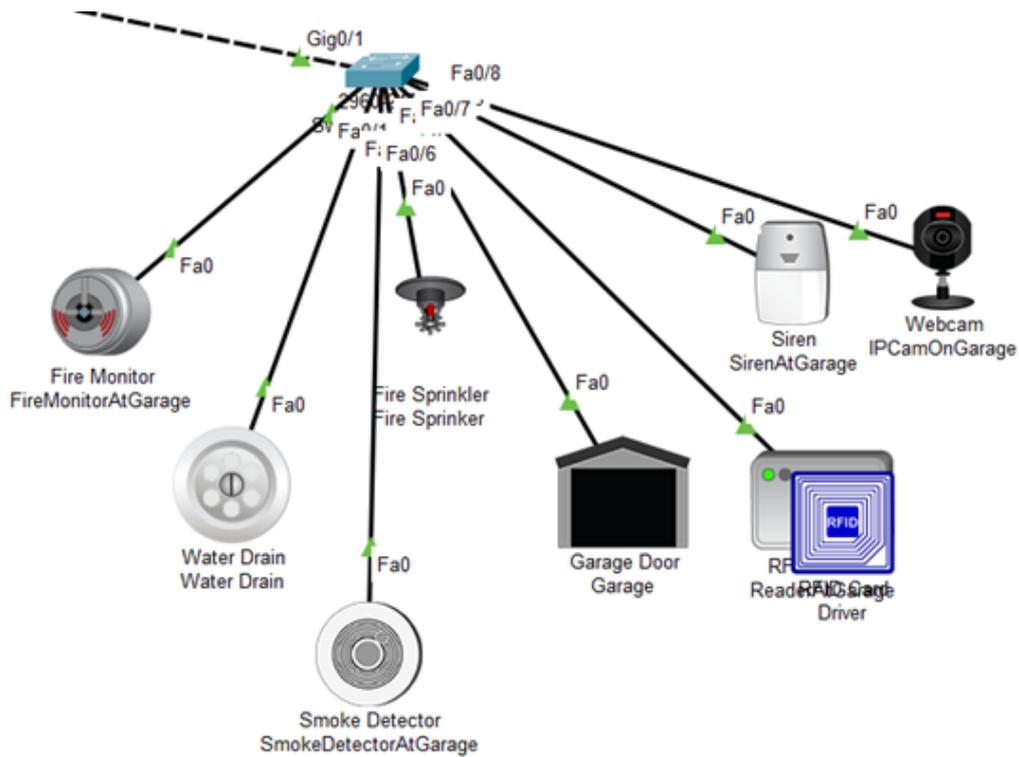


Figure 4: Scheme of fire alarm system of a separate room on the example of a garage

- Sybil attack (R_5)
- “wormholes” attack (R_6)
- flood attack (R_7)
- attacks at the transport level (T):
 - avalanche attack (T_1)
 - desynchronization attack (T_2)
- attacks on data aggregation (G);
- privacy attacks (P).

Attacks can be represented in the form of open classification groups.

$D = H \cup C$ – a set of attacks that lead to denials of service, involves combining sets of attacks at the physical and channel level.

Many attacks that lead to denials of service at the physical level:

$$H = \bigcup_{i=1}^n H_i$$

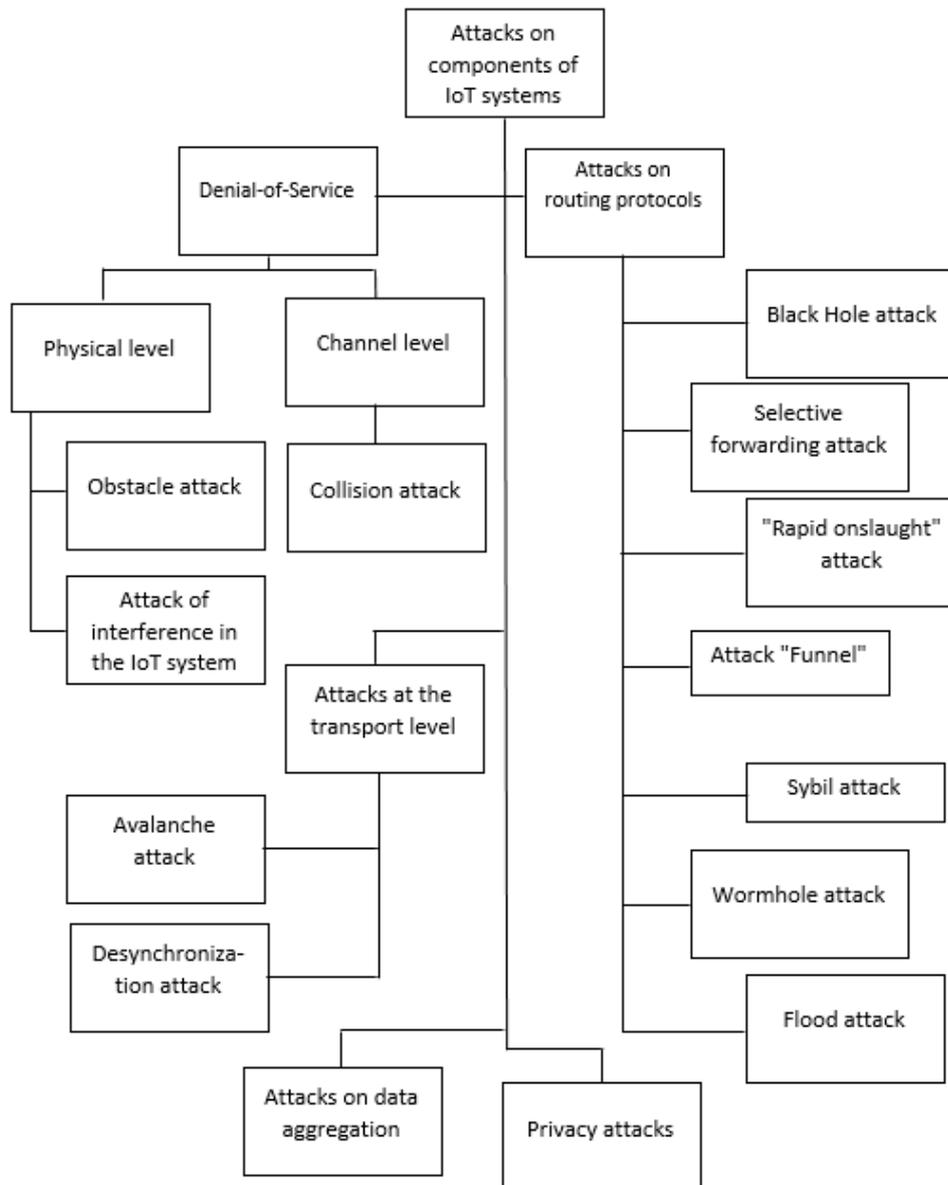


Figure 5: Attacks on IoT system components

The set of attacks that lead to denial of service link-level:

$$C = \bigcup_{k=1}^z C_k$$

The set of attacks on routing protocols:

$$R = \bigcup_{v=1}^s R_v$$

The open classification grouping of transport layer attacks is presented in the form of a set:

$$T = \bigcup_{\alpha=1}^l T_\alpha$$

The set of attacks on data aggregation is represented as follows:

$$G = \bigcup_{j=1}^m G_j$$

The set of attacks on privacy:

$$P = \bigcup_{\gamma=1}^{\delta} P_\gamma$$

In general, attacks can be represented as a union of all classification groups:

$$A = D \cup R \cup T \cup G \cup P$$

Let's analyze each attack that is part of the classification group.

DoS attack on the physical level. A DoS attack is characterized by an attempt by an enemy to stop a network or destroy a network security service. In an IoT system, a DoS attack can occur at different levels of the protocol stack, can affect several levels simultaneously, and use the interaction between them. DoS attack at the physical level can be carried out by interfering with the radio frequencies on which the system operates. In such an attack, one attacking node may disconnect all or part of the network (for example, blocking data transmission).

An attack on the IoT system's detection of a sensor (in our case, a sensor/camera around the perimeter of a security object) and an attempt to physically access it is critical to our system. In this case, an attacker can destroy the device, try to replace the data, access sensitive information (including cryptographic keys), use the device to log on to the network.

DoS channel level attack. DoS collision attack at the channel level is usually aimed at depleting the resources of nodes. This attack affects the packet transmission process, causing exponential delay and packet retransmission procedures in some MAC protocols. Thus, when a large number of bits are damaged in a packet, the node will try to use error correction codes to recover the damaged bits, thus wasting limited energy resources. Another example of such an attack is a "collision" at the end of the frame, which leads to the retransmission of the entire packet. Another embodiment of the attacks inherent in the IEEE 802.11 protocols may be the generation of an RTS message to a base station or neighboring node, which will lead to the processing of this message and generate a CTS message, followed by waiting for signal reception, and all other nodes stop transmitting data to receiving node for the time specified in the RTS message. Handshake methods can also be implemented.

Let us analyze attacks on routing protocols. The known Black Hole attack aims to use a routing protocol to redirect packets from or to the target node through a specific node. This

attack can be used to drop packets or a “middle man” (a method of compromising a communication channel in which an attacker, by joining a channel between counterparties, interferes with the transmission protocol by deleting or modifying information). Another type of attack is a selective forwarding attack, which is similar to a Black Hole attack, but in this attack will be rejected packets that meet certain criteria, not all.

When implementing the “Rapid Pressure” attack, the procedure of opening the route at the request of routing protocols is used. The malicious node generates and transmits a route request to its neighbors, and as a result, the node is more likely to be part of the selected route between the source and destination.

The “Funnel” attack is characterized by the fact that the attacker tries to either compromise the node, or place its own in the path of as many networks flows as possible, and the latter then begins to act on the type of funnel – collecting all the traffic of the sensor network. In protocols that use broadcast, the attacker, listening to the channel, informs neighbors that he “knows” the shortest route to the base station. Once it has managed to stand between the transmitting sensor node and the base station, it can perform any action with the data packets coming to it.

Sybil attack is characterized by the fact that the attacker tries to compromise the existing node, or connect your own with several pseudo-identifiers and thus pretending to be several nodes at once. Thus, neighboring nodes may perceive it as “their own”. Such attacks are used to disrupt the mechanism of distributed storage, routing mechanisms, data aggregation mechanisms, voting mechanisms in the network.

A wormhole attack poses a serious threat to the security of sensor networks because it does not require compromising the sensor node. For example, an attacker listens to a channel, receives a broadcast to request a route from the base station, and forwards it to the nearest neighbor. The node that received this message will consider it the parent, that is, the one closest to it, although this is not the case. The attack is based on creating a special path between two or more network nodes to transmit intercepted packets, and the nodes will think that they transmit packets by the shortest path.

One type of attack is a flood attack (HELLO flood attack). The peculiarity of this attack is the attempt to transmit to the network many optional messages that will deprive the network of various resources (computing power, channel capacity, energy resources). Having a high-frequency radio transmitter with sufficient computing power, the attacker sends Hello packets of many nodes of the sensor network. Upon receipt of this message, the nodes perceive the compromised node as a neighbor and include the received address of the sender in the mailing list. In this way, the attacker gains access to data sent from the nodes.

Transport layer functions include the delivery of packets (TCP) and datagrams (UDP) from sender to recipient. Attacks at the transport level are aimed at analyzing the regularity of traffic and sending parallel duplicates of messages in other ways used at this level. Given the fact that most transport protocols support sensitive information and are therefore vulnerable to memory depletion, an avalanche attack attacker makes new connection requests each time increasing the amount of confidential information in the attacking node, gradually leading to the node becomes faulty (failure of the node from further connections) due to resource depletion) and uses this shortcoming.

Another typical attack of this level is the desynchronization attack, because of which an attacker tries to break the connection between two working nodes in the network, repeatedly

forging messages to them. In particular, transport layer protocols can use sequence numbers to track successfully received packets, identify packet loss, and detect copies. Attacker-generated packets can use these sequence numbers to reassure the node that packets have been lost and to provoke retransmission, which can have the effect of depleting the resource and filling the data channel when valid information does not arrive at the database or arrives with a delay.

Attacks on data aggregation are aimed at changing the behavior of the network. Data aggregation and merging procedures are used in networks where the location of typical sensors is close to each other. Such procedures are used to combine multiple data to eliminate redundant information. To save resources, this is positive, but it is dangerous from the point of view of cybersecurity. Thus, the calculation of simple mathematical functions (minimum, maximum, average, sum) used in aggregation in the presence of a single malicious node or the replacement of real data from sensors can change the behavior of the network in part or completely.

Privacy attacks are aimed at capturing information collected by sensors and can be implemented by listening to the network, analyzing traffic, and/or capturing the node. This is especially true for those networks that do not use data encryption.

3.1. Recommendations for counteracting attacks on components of the IoT system

Resist DoS attacks at the physical level. IEEE 802.11 (Broadband) standards use frequency hopping. In this case, the interference transmitter must “know” the sequence of hopping or create interference with a larger frequency band. It is proposed to use spectrum expansion technology to protect against such attacks. The transmission of such a signal will be similar to noise, which will reduce the risk of intentional interference with the information signal.

Besides, when the signal disappears from any part of the network or node, network element, DSS should generate an alarm on the unit. Nodes that have detected an interference attack must send a short message to their “neighbors” and the base station about the attack on the network. In this case, if the message “does not reach” the base station from the attacked node, it is likely to receive an alarm message from the node that was not attacked.

To counteract IoT intrusion attacks, each sensor used in the system must be equipped with a tamper (a miniature button on the board of the device that is squeezed when opening the case or disconnecting it from the mounting location). When the tamper is triggered, the hub sends push-messages and SMS to all users of the security system (if there are such messages in the devices to be used), as well as the transmission of the message to the base station.

Besides, it is desirable to provide software that when the tamper is triggered during “arming”, all data stored on the device was destroyed automatically. To avoid detecting sensors, they should be placed in hidden places, but suitable for their installation, use materials that are resistant to external influences. Sensors and cameras have their range, so when placing such devices should take into account this figure and install them with an overlap to avoid insensitivity. If installed correctly, the sensor will detect the danger and send an alarm to the base station until the attacker approaches it.

The proposed system uses an RFID tag to identify a person. The decision support system provides a situation where the RFID tag and motion sensor is activated, but we do not receive a signal from the camcorder. This situation may indicate that the tag was “removed” or “replaced”

and the motion sensor detected movement, but attackers could disable the camera to avoid being identified as violators. This set of parameters will generate an alarm on the unit.

To counter a DoS attack at the channel layer, there is authentication to verify that the node generating the message is authorized on the network in combination with encryption. In our case, we use the WPA2-PSK authentication standard with an AES encryption type. Given the energy limit, the use of asymmetric encryption becomes impossible in such systems. The main disadvantage of using symmetric encryption is the problem of key distribution. When using a symmetric cryptographic scheme, it is necessary to ensure the reliable and secure installation of shared cryptographic keys between two nodes before they can exchange data. Key installation and management techniques should be suitable for use with hundreds and thousands of nodes.

Another way to improve security is to install an RFID tag on all devices on the network and conduct a combined (two-factor) node authentication procedure.

It is proposed to use blockchain technology to protect against interference with the program code and substitution of sensors. This technology is a distributed database that is potentially available to everyone. Thanks to the use of blockchain technology, it is possible to counteract fraud, manage identification, transactions, verify the status of elements of various systems, and ensure data integrity. Combining blockchain and Internet of Things technologies can solve several security issues, namely: tracking sensor data measurements and preventing duplication of any other malicious data; authentication and secure data transmission.

Cryptography is proposed to protect against eavesdropping, injection, and packet modification.

To counter aggregation attacks, it is proposed to use aggregation delay and authentication methods. To prevent routing attacks, we use channel-level encryption and authentication using a global public key. Sybil attacks can be prevented by verifying the identity of the sensor nodes (using a shared symmetric key from a trusted base station) and limiting the number of neighbors that the node may have. In this way, the compromised node will only be able to contact trusted neighbors. You can counter a funnel attack using a geo-routing protocol, in which traffic “naturally” directed to the physical location of the base station is difficult to redirect to create a funnel.

The proposed system uses static sensors that require one-time authentication in the network.

Edge computing in information security systems can be used to counter several considered attacks and is the subject of further research. The use of clusters of security systems, IoT clusters in combination with edge computing creates new approaches to technologies for building secure IoT with decentralized data processing.

The list of attacks is an open classification group that can be supplemented and expanded.

The implementation of IoT clusters in combination with edge computing requires further research.

They need to develop a cluster model and mathematical software for IoT systems in combination with edge computing to minimize information processing and decision-making time.

4. Conclusions

The analysis allowed us to generalize cyber threats to the components of IoT systems. As a result, it is determined that the largest number of attacks occur on network nodes, and the use of wireless communication technologies between the elements of the system creates the preconditions for a cyber-attack on the system.

It is determined that today multi-stage complex protection systems are being implemented, based on the use of the latest technical means, qualified personnel, control procedures, administrative regulations with their strict observance.

The analysis of attacks allowed determining their list and exploring the features of implementation. As a result of the analysis and generalization, recommendations for counteracting attacks on the components of the IoT system have been developed.

References

- [1] G. Immerman, The importance of edge computing for the iot, 2020. URL: <https://www.machinemetrics.com/blog/edge-computing-iot>.
- [2] S. Khomich, A. Fedosiuk, M. Kulikovskiy, Research of system of iot devices information security, *Digital technologies* 18 (2015) 166–171.
- [3] J. Blyler, 8 critical iot security technologies, 2020. URL: <https://www.electronicdesign.com/industrial-automation/article/21805420/8-critical-iot-security-technologies>.
- [4] Q. Jing, A. V. Vasilakos, J. Wan, J. Lu, D. Qiu, Security of the internet of things: perspectives and challenges, *Wireless Networks* 20 (2014) 2481–2501. doi:10.1007/s11276-014-0761-7.
- [5] D. Kuznetsov, L. Ryabchina, Information security of the internet of things systems, *Bulletin of Kryvyi Rih National University* 49 (2019) 80–83.
- [6] O. Turanska, Development of methods of information protection in wireless sensor networks: master's thesis, Master's thesis, NTU of Ukraine "KPI named after Igor Sikorsky", 2018.
- [7] C. Systems, The internet of things reference model, 2014. URL: http://cdn.iotwf.com/resources/71/IoT_Reference_Model_White_Paper_June_4_2014.pdf.
- [8] A. Herts, I. Tsidylo, N. Herts, L. Barna, S.-I. Mazur, Photosynq - cloud platform powered by iot devices, *E3S Web of Conferences* 166 (2020). doi:10.1051/e3sconf/202016605001.
- [9] S. Shokaliuk, Y. Bohunenko, I. Lovianova, M. Shyshkina, Technologies of distance learning for programming basics on the principles of integrated development of key competences, *CEUR Workshop Proceedings* 2643 (2020) 548–562.
- [10] S. Gnatyuk, Cybersecurity in the context of the fourth industrial revolution (industry 4.0): challenges and opportunities for ukraine, 2019. URL: <https://niss.gov.ua/doslidzhennya/informaciyni-strategii/kiberbezpeka-v-umovakh-rozgartannya-chetvertoi-promislovoi>.
- [11] A. Vovk, Methods of information security IoT, Master's thesis, NTU of Ukraine "KPI named after Igor Sikorsky", 2018.
- [12] O. Korchenko, M. Alexander, R. Odarchenko, A. Nadzhi, O. Petrenko, Analysis of threats

and mechanisms for information security in sensor networks, *Information protection* 1 (2016) 48–56.

- [13] H. Packard, Hp study reveals 70 percent of internet of things devices vulnerable to attack, 2020. URL: <https://www8.hp.com/us/en/hp-news/press-release.html?id=1744676>.
- [14] J. Frahim, C. Pignataro, J. Apar, M. Morrow, Securing the Internet of Things: A Proposed Framework, 2015. URL: http://web.archive.org/web/20210323170935/https://tools.cisco.com/security/center/resources/secure_iiot_proposed_framework.
- [15] M. G. dos Santos, D. Ameyed, F. Petrillo, F. Jaafar, M. Cheriet, Internet of things architectures: A comparative study, 2020. URL: <https://arxiv.org/pdf/2004.12936.pdf>.

Task scheduling in Desktop GRID by FSA method: a practical example

Taras A. Uzdenov^{a,b}

^aZhytomyr Polytechnic State University, 103 Chudnivska Str., Zhytomyr, 10005, Ukraine

^bG.E. Pukhov Institute for Modelling in Energy Engineering of NAS of Ukraine, 15 General Naumova Str., Kyiv, 03164, Ukraine

Abstract

The paper considers a new approach to solving the problem of dispatching task flows, the complexity of which is known, for GRID-systems with inalienable resources, the performance of which can be determined. A method based on this approach has been developed. The efficiency of the proposed method is compared with the well-known and widely used in various projects method FCFS. A feature of this method is the simplicity of implementation. An example of a simple practical problem that can be solved using the proposed method is described in this paper.

Keywords

GRID-systems, Desktop GRID, task scheduling, non-alienable resources, FCFS

1. Introduction

Every modern organization has a number of computers on which its staff works, their use is not the most efficient, as most of them tasks performed on them do not take up 10–20% of the maximum performance of the PC. Therefore, it makes sense to use free resources to solve other problems.

Therefore, the idea arose to create on the basis of such resources computer systems that would allow other tasks to be performed in parallel with the current ones for each of the nodes. Such systems are called GRID systems with non-alienable resources. Such systems are also known as Desktop GRID.

Thus, Desktop GRID is a GRID system that uses non-specialized computing resources as computing nodes, but disparate computing nodes (computers, laptops, smartphones, etc.) using local and global networks and special software.

Back in 1999, the first large-scale project of distributed voluntary computing SETI @ home was launched. Today, Desktop Grid is part of the high-performance computing industry along with clusters and GRID.

One of the main tasks in creating a GRID-system with inalienable resources, as well as for GRID-systems, is the task of task scheduling. Therefore, in GRID-systems, a planning mecha-

QualnT 2021: Workshop on the Quantum Information Technologies, April 11, 2021, Zhytomyr, Ukraine

doors 2021: Edge Computing Workshop, April 11, 2021, Zhytomyr, Ukraine

✉ uzdenov.taras@gmail.com (T.A. Uzdenov)

ORCID 0000-0002-0731-7620 (T.A. Uzdenov)

© 2021 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

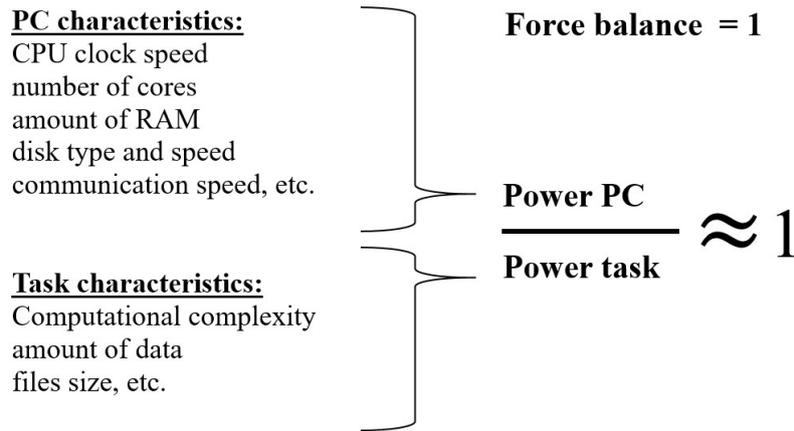


Figure 1: New approach

nism must be implemented. It is necessary for the distribution of tasks for execution between the nodes of the system, in order to minimize the execution time and balance the load of the system.

One of the problems that needs to be solved when developing software for a GRID-system with inalienable resources is the task of task scheduling. Task scheduling is quite complex and now there is no clear and unambiguous solution [1, 2]. Analysis of scheduling methods used in real systems shows that most systems use mainly the FCFS method [3]. In addition, were analyzed a number of publications in recent years on this topic , in which developers offer new methods and various modifications of known ones. They are compared with FCFS and SJF (Shortest Job First) [4, 5, 6, 7, 8, 9, 10, 11, 12]. But since in real systems the FCFS method is most often used, it was decided to make a comparison with it in this study as well.

This is due to the fact that this method is very simple and reliable in both development and operation. The use of other methods significantly complicates the system, making it less reliable. Since such systems are quite unstable, it is clear why developers abandon complex methods and prefer FCFS. This leads to the conclusion that it is necessary to develop new methods, the main characteristics of which should be simplicity and better performance compared to FCFS.

2. New approach

This article further studies the methods developed on the basis of the new approach outlined in [13]. In particular, a simple practical task that could be performed on a GRID-system with non-alienable resources is considered, and the FSA method is used to distribute tasks between nodes.

In figure 1 schematically shows the main essence of the proposed in [13] approach. The proposed methods are developed on the basis of a new approach, which proposes to consider tasks as one force, and the nodes on which they should be performed as another force, like

Newton's third law. And distribute the tasks in such a way as to maintain a balance of forces. Given that such concepts as the strength of the task and the strength of the node are quite abstract concepts, the author proposes to use the concepts of task power and node power. And to carry out distribution already according to balance of capacities.

This choice is not accidental, and can be explained as follows. It is known from physics that power is equal to the ratio of work to time. And the work, in turn, is equal to the force multiplied by the path to be traversed. Given that when performing a task on a computing node, the time to determine the power of the task will be equivalent to the time to determine the power of the node, and the path that the task must pass is equal to the path that the node must pass, such a replacement is quite logical and acceptable. In the developed methods, these concepts are quantities relative and therefore can be calculated in different ways, depending on different characteristics of tasks (volume, computational complexity, algorithmic complexity, etc.) and different characteristics of nodes (CPU clock speed, RAM volume, communication channel speed and others). The power of the task means the totality of all the characteristics of the task, and the power of the PC means the totality of all the characteristics of the PC, compiled in some way. Moreover, if for a PC it is still possible to use some general formula for calculation, then for the power of the problem such a formula will change each time, depending on the type of problem that needs to be solved.

The approach described above is quite simple and effective to use, but it actually divides the scheduling task into three subtasks:

1. Calculation powers of tasks
2. Calculation powers of nodes
3. Distribution (FSA method)

Both the first and the second subtasks are quite complex and today there are no unambiguous and universal solution for them. The fact is that any task has a number of characteristics, which have already been written above, and to compare them and somehow reduce to one value is quite difficult.

This requires the development of additional methods that would provide such an opportunity. On the other hand, the task of calculating the powers of nodes is no less complex and also requires a separate study and solution. But at the same time there are a number of tasks for which the calculation of capacity will not cause much difficulty. This is well illustrated by the example of a simple practical problem described in the last section.

2.1. Formulation of the problem

Suppose there is a GRID system with N task and M nodes. By nodes we mean a computational element. We introduce the concept of task power P_n and node power R_m . Therefore, we have the set of power of tasks $P = \{P_1, P_2, P_3, \dots, P_n\}$ and the set of power of nodes $R = \{R_1, R_2, R_3, \dots, R_m\}$. We need to optimally distribute tasks across nodes.

Schematically, a given task is shown in figure 2.

2.2. Flow Scheduling Algorithm (FSA)

The method of flow scheduling has the following form:

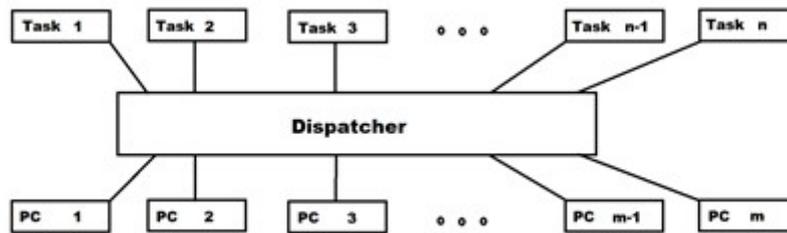


Figure 2: Scheduling task

1. calculate the power of tasks and the power of nodes
2. choose the i -th task
3. find the pair $i - j$, for which the ratio of the power of the task P_i to the power of the node R_j will be as close as possible to unity
4. send the i -th task to the j -th node
5. recalculate the power without P_i and R_j
6. if there are unsent tasks, then go to point 2
7. completion

2.3. Flow Scheduling Algorithm Parallel (FSA_P)

The described approach is universal and allows to develop not only methods for distribution of consecutive tasks but also for tasks which can be parallelized. Below is the following method:

1. calculate the power of the nodes, and the total power P_{sum}
2. select the i -th task from the task queue
3. distribute it proportionally, according to the power of the nodes
4. send for execution
5. if there are unallocated tasks in the queue, go to point 2
6. completion

3. Software package

To create software that allows you to test and investigate the effectiveness of the proposed methods, developed its client-server architectural model (figure 3).

This model is based on the WCF service [14] for software that requires distributed computing in computer networks and the Internet, as well as to create a Desktop GRID.

Based on this model, a software package for simulating the operation of a GRID-system with inalienable resources and the study of scheduling algorithms was developed, which was named SgridAR-1 [15].

A number of tests were performed using the developed simulator. To determine their effectiveness, the FCFS scheduling method was chosen because it is most often used for GRID

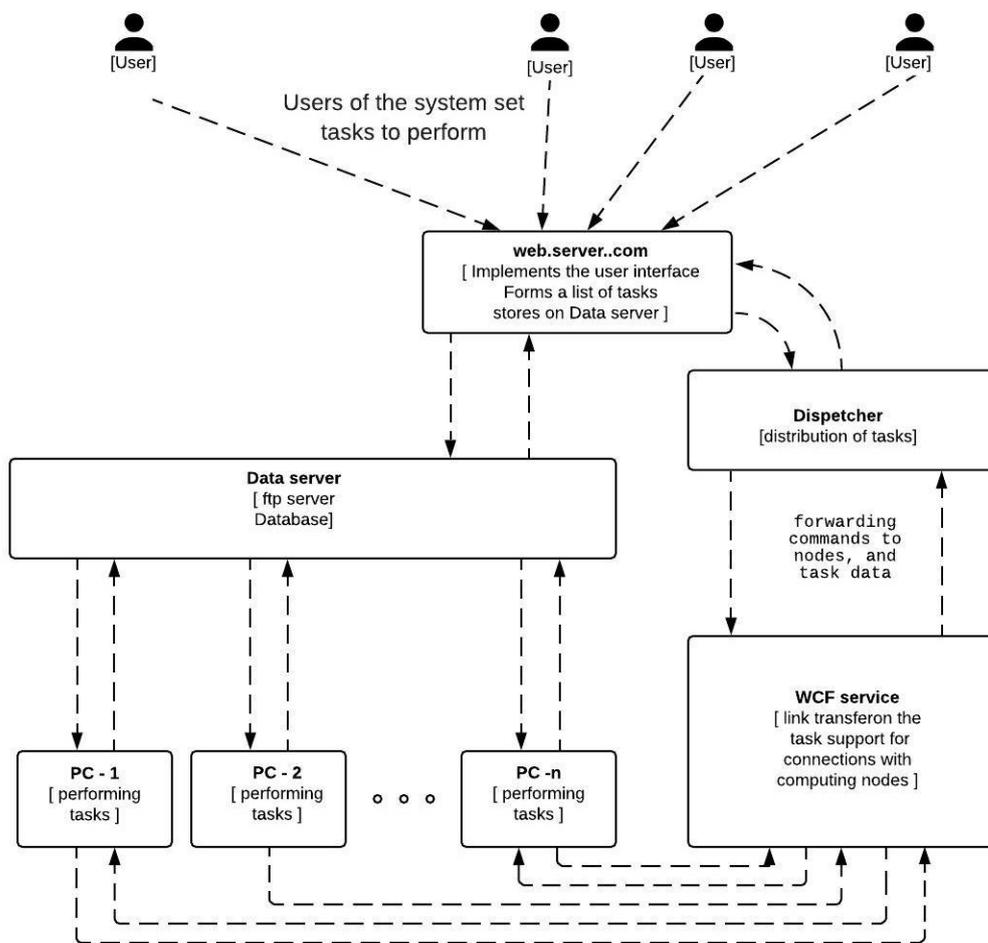


Figure 3: Architectural model of Desktop GRID

systems with inalienable resources. In addition, other methods cannot be used because in this case they do not involve a change in the power of the nodes, and other characteristics, such as time quantum or priority, it was decided not to enter.

Figure 4 highlights nine areas in the Server window, which during testing displays information about the results of the simulation:

1. the total volume of all tasks to be solved
2. total execution time of all tasks for each algorithm
3. the area in which messages about the beginning and end of calculations are displayed, as well as information about which algorithm is currently working
4. in this area the conditions for testing are set: the number of tasks, their minimum and maximum value, the choice of algorithm

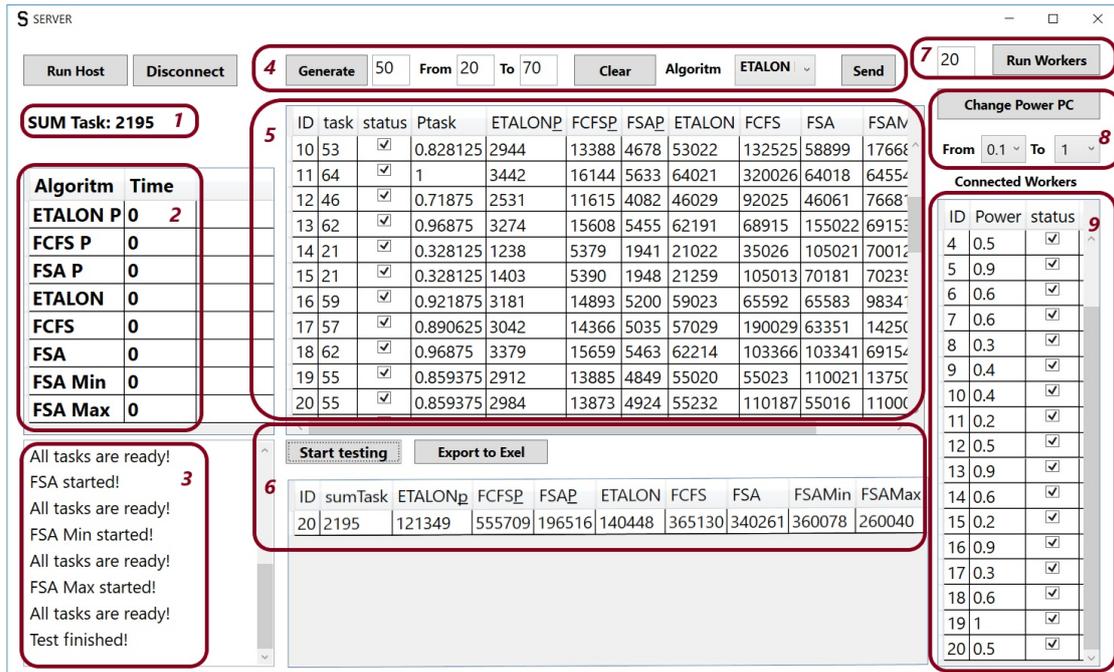


Figure 4: SGridAR-1 server window

5. list of all generated tasks, their scope and power, status (performed or not), execution time for each algorithm
6. after completing the test, this area receives summary information about the operating time of each algorithm, after which in area 2 the information is updated to perform a new test, and the button “Export to Excel” becomes active
7. the number of Clients to run
8. the maximum and minimum value of the power of the system nodes, after pressing the “Change Power PC” button, the system will automatically change the power in any way for all nodes, in the range from minimum to maximum values
9. information about the nodes connected to the system, their capacity and status (free or busy)

SGridAR-1 allows to show work of the offered method and to compare results of its work with other, well-known, methods of dispatching. With the help of this software you can conduct experiments and explore the work of algorithms, changing the number and size of tasks, the number and power of the PC.

In this program the mechanism of generation of tasks was implemented. The task of which is to create a tasks to be executed in an arbitrary way, specifying the time required for the task to be performed. Power of tasks are calculated in proportion to the given time. The power of the same nodes is also generated arbitrarily. But depending on their size slow down the timer.

The implemented visual interface clearly shows how the GRID system works. This set of programs can be used not only for research but also for the educational process.

4. Test results

The SGridAR-1 system implements a testing mechanism that provides for the execution of all tasks generated by the system, using different algorithms for the distribution of tasks between nodes. That is, first the tasks are distributed according to algorithm 1, then according to algorithm 2 and so on until the last task.

The following scheduling methods are introduced into the system: ETALON_P; FCFS_P; FSA_P; ETALON; FSA, FSA_Min, FSA_Max.

Parallel methods:

- ETALON_P – the method is taken as a reference. This is a FCFS method that does not take into account the power of the personal computer (PC) and tasks. A system is modeled in which all nodes are the same in power. Parallel tasks are possible
- FCFS_P – FCFS method, which takes into account the power of the PC and the ability to distribute one task to several PCs
- FSA_P is a proposed method that takes into account the power of PCs and the ability to distribute one task to multiple PCs

The following methods:

- ETALON – the method is taken as a reference. This is a FCFS method that does not take into account the power of the PC and the task cannot be distributed
- FCFS – a method that takes into account the power of the PC and the task can not be distributed between nodes
- FSA – the proposed methods, which take into account the power of the PC and the task can not be distributed
- FSA_Min – modified FSA method combined with Min-Min method [16]
- FSA_Max – modified FSA method combined with Max-Min method [17]

For comparative experiments, 100 tasks with a runtime of 1 to 10 seconds were generated and 10 nodes were run on which they should be executed. Testing was as follows: first, the calculation of the total execution time of all tasks for 1 method, entered into the system, at an average power of the system 10%, then programmatically increased the system power by 5% and again performed calculations for 1 method. Then again increased the power of the system by 5% and performed the calculation by 1 method. And so on until the power of the system has grown to 100%. Thus, testing was performed for each of the methods. Each time performing the same tasks, changing only the average power of the system.

A total of 19 average power tests were performed for each of the planning methods introduced into the system. A total of 152 tests were conducted.

Such studies were conducted to show how the average power of the system affects the efficiency of the methods.

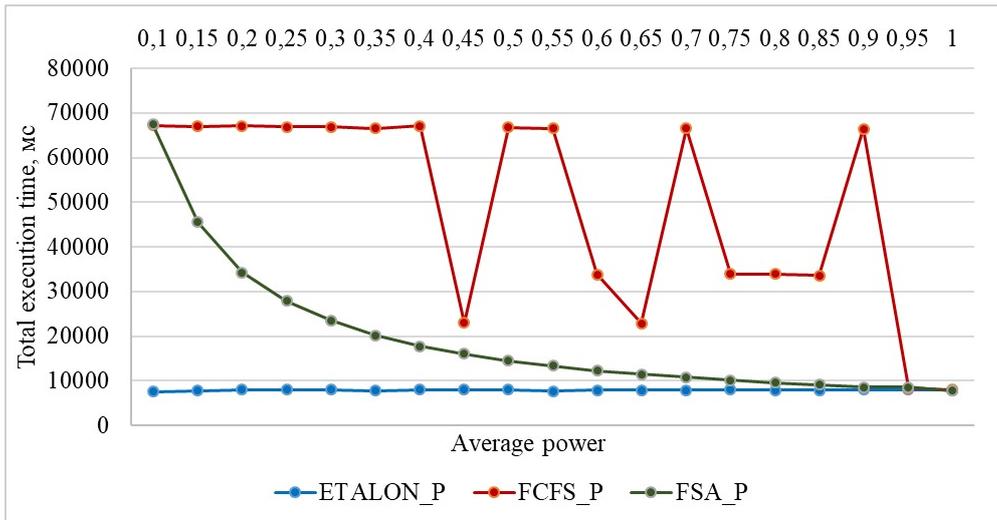


Figure 5: Test results for parallel methods

In figures 5-7 are diagrams based on test results. Based on the results obtained, it can be concluded that the FSA and FSA_P methods give better results than the FCFS and FCFS_P methods.

As can be seen in figure 5, the curve of the FSA_P method has a smooth shape, and the curve of the FCFS_P method is broken. This allows us to conclude that the FSA_P method is well predictable and, knowing the power of the system and the amount of tasks, it is possible to predict the completion of calculations. However, solving problems by the FCFS_P method, it is quite difficult to predict the completion of calculations, because a lot depends on which node, which task will be performed.

In figure 6, it is seen that the sequential FSA method has a smoother curve than the FCFS method, which also indicates a better predictability.

In figure 7, which shows additional methods FSA Min and FSA Max in comparison with the FSA method, it is seen that the differences in the results are insignificant, but this is only for the situation when the number of tasks exceeds the number of nodes.

As can be seen from the graphs, the system also has ETALON and ETALON_P methods, and at first glance it may seem that their efficiency is much better than the FSA and FSA_P methods. But this is not the case. The fact is that the methods ETALON and ETALON_P are shown not for comparison with others, but to demonstrate the reference state of the system, when all the nodes in it have a capacity of 100%. As can be seen from the graphs, when the average power increases to 1, the execution time by different methods approaches the maximum possible reference value.

Figures 5, 6 show that the behavior of the FCFS method is very unstable, large jumps in the results. This is because the distribution by the FCFS method very much depends on a random factor, and which node will have what task. This is quite logical, because if, for example, a node with a capacity of 0.1 receives a task with a capacity of 1, then the execution time in this case will increase significantly, because there may be a situation that other nodes will work

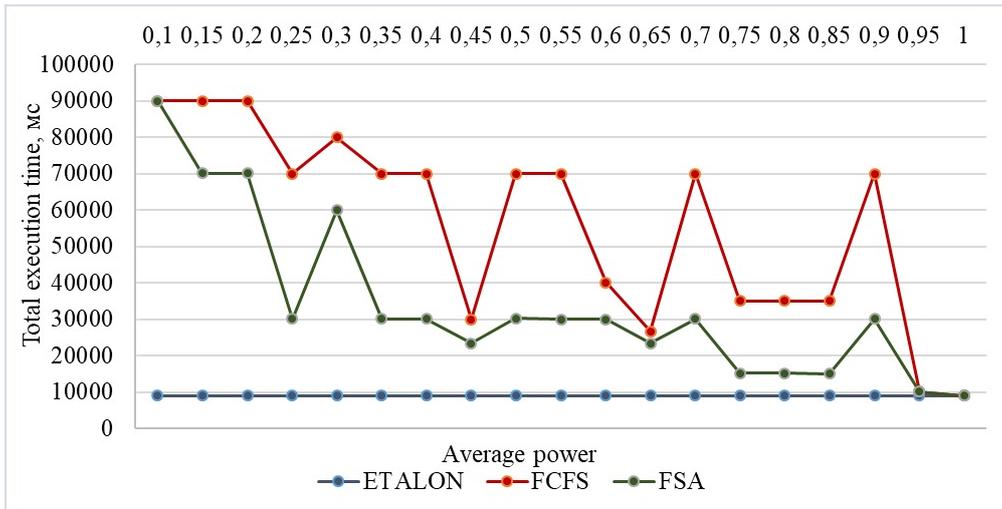


Figure 6: Test results for following methods

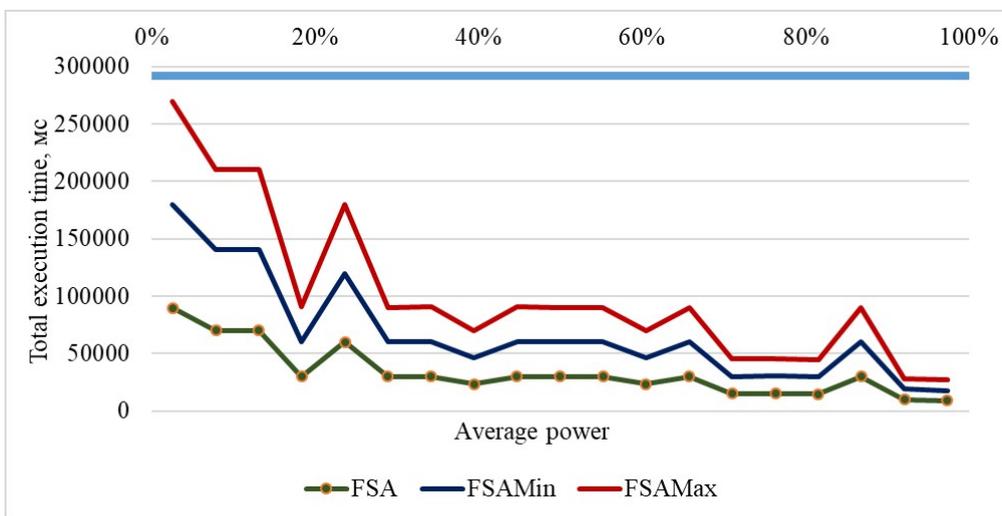
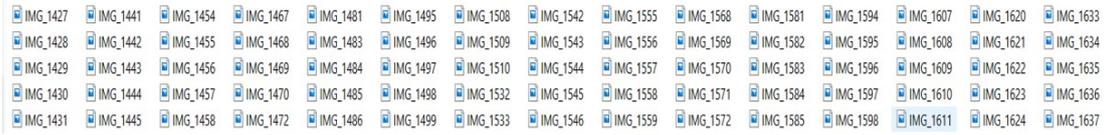


Figure 7: Test results for methods FSA, FSA_Min, FSA_Max

and will wait too long to complete this task.

5. Practical task

It is proposed to consider such a task. If, for example, we have a web resource with a large number of images (for example, a portfolio, an online store, etc.), and we want to promote it in search engines in order to attract more users and increase revenue, then we need will perform page optimization according to search engine rules (for example, using Google PageSpeed [18]), which includes image optimization.



There are 1000 images ranging in size from 100 KB to 10 MB

Task
 Resize image with an optimizer built into each node client

There are 15 nodes with power from 0.1 to 1



Figure 8: Practical task

In figure 8 schematically shows the problem described above. Let's say we have 1000 images that need to be optimized according to Google Page Speed. Total files size is 5 GB. If you perform this task on one PC, it will take a long time (maybe even a full day). It all depends on the speed of the PC on which to do it.

Therefore, to speed up this task, it would be advisable to divide it into different PCs that we have. To do this, we will need to determine which method will be used for distribution.

In fact, as mentioned above, there is a new method of FSA, based on the described approach.

The main difficulty of this method is to calculate the power of tasks and nodes. But for this practical example, this problem is quite simple to solve. Given that all images differ from each other only in size, the largest image (S_{max}) is assigned power ($P_{max} = 1$), and the remaining ($n - 1$) images are calculated by the power of the tasks by the formula:

$$P_i = \frac{S_i \cdot P_{max}}{S_{max}}$$

Given that ($P_{max} = 1$) we can shorten the expression:

$$P_i = \frac{S_i}{S_{max}} \tag{1}$$

On the other hand, we have m PCs (computing nodes), which differ from each other, for example, only the clock speed of the processor. Then, similarly to the calculation of the power of the problem, we can find the power of the nodes, according to the proportion.

The largest by the clock frequency of the node (F_{max}) is assigned power ($R_{max} = 1$), and the remaining ($m - 1$) power of the nodes is calculated by the formula:

$$R_j = \frac{F_j \cdot R_{max}}{F_{max}}$$

Given that ($R_{max} = 1$) we can shorten the expression:

$$R_j = \frac{F_j}{F_{max}} \quad (2)$$

Then the algorithm of the method of scheduling FSA tasks can be reduced to the following form:

1. calculate $P = P_1, P_2, P_3, \dots, P_n$ by formula (1)
2. calculate $R = R_1, R_2, R_3, \dots, R_m$ by formula (2)
3. choose the i -th task
4. find the pair $i - j$, for which the condition

$$\min \left\{ \left| \frac{P_i}{R_j} - 1 \right| \right\}$$

is fulfilled.

5. send the i -th task to the j -th node
6. recalculate the powers without P_i and R_j
7. if there are unsent tasks, then go to point 2
8. completion

Thus, we showed that calculating the power of tasks and the power of nodes is not such a difficult task when you need to distribute the image between nodes in order to optimize the size. If you use different nodes, but different in speed of connection to the network, then the power of the nodes must be calculated in some other way, because not always the node with the processor with the highest clock speed will have the highest power. And if this point is not taken into account, then the distribution will not be as effective as in the first case.

As mentioned above, this method is universal and the distribution of tasks according to it makes it possible to significantly speed up the execution of the task queue and thus increase the efficiency of GRID-systems with inalienable resources compared to the FCFS method.

6. Conclusions

The results of the study of the effectiveness of the proposed methods showed that their use for task allocation in GRID-systems with non-alienable resources provides a significant reduction in task queue time compared to the FCFS scheduling method, provided that the number of tasks exceeds the number of nodes.

All the proposed methods are quite stable and well-predicted, which means that their use in GRID-systems will give advantages not only in time and performance, but also allow more efficient planning of the system work. The FCFS method works well, but for GRID systems that

have different power resources, its performance depends heavily on a random fact that can be considered a significant drawback.

The main difficulty in using these methods is the need to somehow reduce all the characteristics of tasks and nodes to one relative value. But there are a number of tasks for which this can be done quite easily. As shown in the example of the problem of image optimization, it is quite easy to calculate both the power of tasks and the power of nodes.

References

- [1] V. Kropyvnytska, B. Klim, A. Romanchuk, M. Slabinoga, Investigation of scheduling algorithms in computer systems, *Rozvidka ta rozrobka naftovykh i hazovykh rodovyshch* 2 (2011) 93–105.
- [2] S. Node, Desktop grids, connecting everyone to science, 2021. URL: <https://sciencenode.org/feature/desktop-grids-connecting-everyonescience.php>.
- [3] S. Choi, H. Kim, E. Byun, C. Hwang, A Taxonomy of Desktop Grid Systems Focusing on Scheduling, Technical Report, KU-CSE-2006-1120-02, 2006.
- [4] S. Sahana, Evolutionary based hybrid GA for solving multi-objective grid scheduling problem, *Microsystem Technologies* (2019). URL: <https://doi.org/10.1007/s00542-019-04673-z>. doi:10.1007/s00542-019-04673-z.
- [5] D. Carastan-Santos, R. Y. De Camargo, D. Trystram, One can only gain by replacing EASY Backfilling, a simple scheduling policies case study, in: 19th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID), IEEE, 2019, pp. 1–10. doi:10.1109/CCGRID.2019.00010.
- [6] K. Dheenadayalan, V. N. Muralidhara, G. Srinivasaraghavan, Storage load control through meta-scheduler using predictive analytics, in: N. Bjørner, S. Prasad, L. Parida (Eds.), *Distributed Computing and Internet Technology*, Springer International Publishing, Cham, 2016, pp. 75–86.
- [7] A. A. Haruna, L. T. Jung, N. Zakaria, Design and Development of Hybrid Integrated Thermal Aware Job Scheduling on Computational Grid Environment, in: 2015 International Symposium on Mathematical Sciences and Computing Research, 2015, pp. 13–17. doi:10.1109/ismsc.2015.7594020.
- [8] Y. Thet, H. Hlaing, T. T. Yee, Static Independent Task Scheduling on Virtualized Servers in Cloud Computing Environment, in: 2019 International Conference on Advanced Information Technologies, IEEE, 2019, pp. 55–59. doi:10.1109/aitc.2019.8920865.
- [9] P. S. Kumar, L. Parthiban, V. Jegatheeswari, Privacy and security issues in cloud computing using idyllic approach Latha Parthiban, *Networking and Virtual Organisations* 21 (2019) 30–42. doi:10.1504/IJNVO.2019.101146.
- [10] M. Kaur, Multi-objective Evolution-Based Scheduling of Computational Intensive Applications in Grid Environment, in: *Proceedings of the International Conference on Data Engineering and Communication Technology*, 2017, pp. 457–467. doi:10.1007/978-981-10-1678-3_44.
- [11] P. Naithani, Genetic Algorithm Based Scheduling To Reduce Energy Consumption In

- Cloud, in: 2018 Fifth International Conference on Parallel, Distributed and Grid Computing, IEEE, 2018, pp. 616–620. doi:10.1109/pdgc.2018.8745801.
- [12] A. Pujiyanta, L. E. Nugroho, Planning and Scheduling Jobs on Grid Computing, in: 2018 International Symposium on Advanced Intelligent Informatics, IEEE, 2018, pp. 162–166. doi:10.1109/icic47613.2019.8985978.
- [13] T. Uzdenov, A New Task Scheduling Algorithm for GRID Systems with Non-alienable Resources, Springer International Publishing, Cham, 2021, pp. 207–220. doi:10.1007/978-3-030-69189-9_12.
- [14] M. Docs, What Is Windows Communication Foundation – WCF, 2021. URL: <https://docs.microsoft.com/en-us/dotnet/framework/wcf/whats-wcf>.
- [15] T. Uzdenov, Simulator of Task Sheduling in Geographicaly Distributed Computer systemswith Non-Alienable Resources, Electronic Modeling 42 (2021).
- [16] T. Kokilavani, D. I. G. Amalarethinam, Load Balanced Min-Min Algorithm for Static Meta-Task Scheduling in Grid Computing, International Journal of Computer Applications 20 (2011) 43–49.
- [17] D. Ramyachitra, P. P. Kumar, Frog leap algorithm for homology modelling in grid environment, Journal of Emerging Technologies and Innovative Research 7 (2016).
- [18] G. Docs, Pagespeed insights, 2021. URL: <https://developers.google.com/speed/docs/insights/v5/about>.

Analysis of metrological support of nano-measurements

Volodymyr Kvasnikov, Mariia Kataieva and Victor Kobylyansky

National Aviation University, 1 Lubomyr Husar Ave., Kyiv, 03058, Ukraine

Abstract

The article analyzes the existing methods and means of measuring objects in the nanometer range and develops their classification based on the main principles of use. The main parameters on which each described method is based are considered and the conditions for their most effective application are determined. It is proved that the chemical and electrical sets of properties of the nanomaterial can change when the particle size decreases to the nanometer size, which requires the inclusion of additional chemical and electrical tests in existing methods. Based on the analysis, it was determined that the most functional and universal in solving a wide range of problems is the method of scanning probe microscopy. The classification of existing methods of scanning probe microscopy based on the nature of their applications is developed. The main information parameters on which each described method is based are considered, and the conditions of their most effective application are determined. To increase the accuracy of nanomeasurements, a methodology based on the principle of integration of information provided by different methods has been developed. The use of the differential-digital method is proposed, which includes the use of an additional information parameter in the mathematical model. An algorithm for including additional (a priori) information in the conditions for measuring the nanostructures has been developed, which leads the problem to the correct one according to the method of the control link, which characterizes the deviation of the parameters of measuring nanoobjects from their nominal values. It is proved that increasing the number of measurement methods used in the metrological analysis of nanoobjects will increase the reliability and accuracy of measurement results, and each method will provide additional information parameters to create a computerized method of calculating the control link. The main condition for correct comparison of the result is knowledge of the specific parameters on which each method is based.

Keywords

nanotechnology, nanomeasurement, metrology, methods and means of measurement, nanomaterials

1. Introduction

In recent years, the study of submicron, nano-, and cluster materials has developed rapidly in many fields of science and technology [1, 2, 3]. For several decades, almost all developed countries in various sectors of the economy have seen rapid progress in the field of nanotechnology.

Improving production efficiency and product quality, development of electronic equipment and biomedical devices, as well as the creation of new nanostructured materials with special properties are largely determined by the accuracy and precision of metrological support. It

QualnT 2021: Workshop on the Quantum Information Technologies, April 11, 2021, Zhytomyr, Ukraine
doors 2021: Edge Computing Workshop, April 11, 2021, Zhytomyr, Ukraine

✉ kvp@nau.edu.ua (V. Kvasnikov); kataeva.mariia@gmail.com (M. Kataieva)



© 2021 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

should be noted that research aimed at improving the instrumentation of nanotechnology has reached a new progressive level. Of course, large-scale development of the nanoindustry is impossible without appropriate metrological (MS) and software support.

Most of the theoretical and applied studies of new measurement methods with a minimum error have no practical application. The existing methods [4, 5] are unproductive, have low accuracy, noise immunity, reliability and cannot be used as part of flexible computerized measuring systems. These methods do not provide the required measurement accuracy for nanoobjects with complex, pronounced topography and do not meet modern requirements for the accuracy and speed of measurements.

In this regard, the problem of improving existing and developing new automated methods for measuring and evaluating the physical and mechanical properties of nanoobjects is relevant.

Therefore, the aim of the work is a thorough analysis of existing methods and measuring instruments in the nanometer range, their classification, and determination of the main characteristics. The analysis is carried out to create a new unified computer program for automatic correction of nano-measurements errors online.

2. Results

The essence of nano-measurements is to work at the molecular level, to study structures with fundamentally new properties. One of the main problems when working with nano-objects and nanostructures is associated with ultra-low signal levels. Another problem [6] is the wide range of behavior that objects and components can exhibit when measured, associated with the influence of destabilizing factors.

The essence of nano-measurements is to work at the molecular level, to study structures with fundamentally new properties. One of the main problems when working with nanoobjects and nanostructures is related to ultra-low signal levels. Another problem [7] is the wide range of behaviors that measuring objects and components can exhibit.

For example [8], measuring objects made of polymeric materials may have a resistance of more than one GM. However, being drawn into fibers with a diameter of less than 100 nm and doped with various nanoparticles, the polymer can be transformed from an excellent insulator into a high-conductivity wire. The result is an extremely wide range of test signals. High-sensitivity, high-resolution instruments are required to detect weak electrical signals at the bottom of the range.

Also, a prominent aspect is that DC measurements may require the characterization of some devices and structures by radio frequency signals. This requires a strict design of measuring instruments with reliable RF connections with low losses of the measuring head and a special electronic circuit for each signal path [9]. Otherwise, it will not be possible to achieve the resolution required for precision re-measurements. Therefore, measurement methods and tools should minimize noise and other sources of error that may interfere with the signal. No less important is the fact that metrological means of nano measurement should be easy to use and economical.

Nowadays, raster electron microscopy (SEM), transmission electron microscopy (TEM), atomic force microscopy (AFM), scanning tunneling microscopy (STM), microscopy, and focusing on

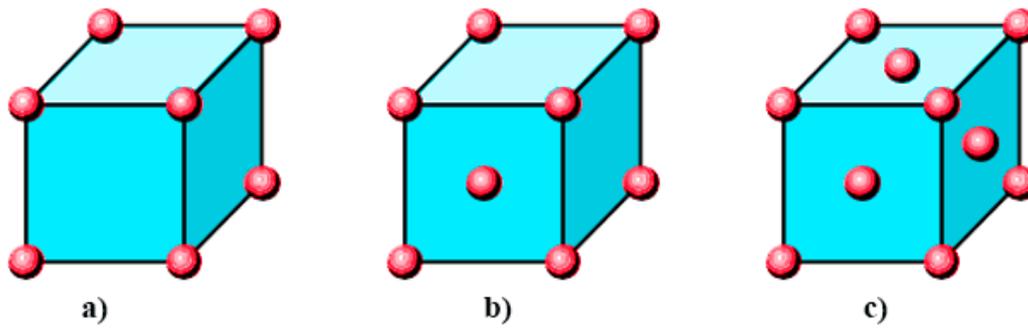


Figure 1: General crystalline atomic structures: (a) simple cubic structure, (b) volume-centered cubic structure, (c) bounded cubic structure

microscopy have become the most widespread. mass spectrometry, Auger spectroscopy, etc. From the point of view of research of the relief and physical properties of structures with high lateral resolution (less than 10 nm), one of the most perspectives is scanning probe microscopy (SPM).

The size and structure of the nanoobject have a great influence on the type of technique and the characteristics of the measurement methods used [10]. It is proved that optical microscopes are the most suitable for the study of macroscopic materials, for nanoscopic materials with a particle size less than 200 nanometers it is better to use STM, AFM, or combined methods of scanning probe microscopy.

SPM, STM, AFM methods are especially useful for the structures of crystalline nanoparticles. Figure 1 shows the atomic structures for some typical crystals of well-known shapes, such as a simple cube, a volume-centered cube, and a border-centered cube. Knowing the location of atoms in these structures helps to predict the properties of particles.

However [11, 12], on a nanoscopic scale, it is the particle size that radically changes the physics of its behavior and dictates the need to use other measurement methods.

Importantly, the chemical and electrical sets of properties of the nanomaterial can change as the particle size decreases to nanometer size. Therefore, in practice, additional chemical and electrical tests are required to determine the characteristics of most of these materials, which also affects the choice of the measurement method. Depending on the means of measurement, methods of analysis of nanomaterials can be divided into two main groups: discrete and ensemble methods of measurement.

Discrete methods of measuring nanoobjects typically use powerful microscopes, such as scanning electron microscopy (SEM), transmission electron microscopy (TEM), electron microscopy (SEM), atomic force microscopy (AFM), and scanning probe microscopy (SPM). EM and SPM are fairly simple methods and reflect particle shapes, but are insufficient in terms of statistics [13, 14].

These methods have a resolution much higher than optical microscopy and can detect and measure discrete particles by scattering a high-energy electron beam (SEM and TEM), or by detecting through a probe attached to a slightly sprung console (AFM). In figure 2 shows a comparison of discrete methods of nano measurement concerning the indicators of the speed

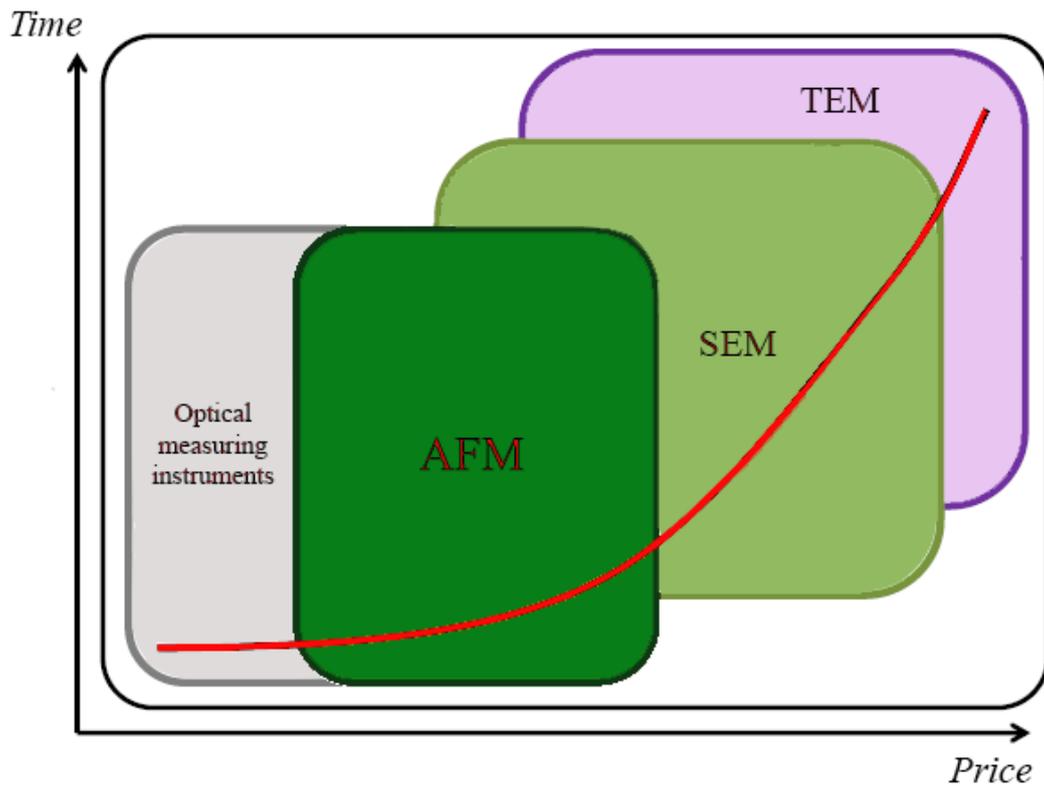


Figure 2: Comparison of discrete methods of nano measurements

of obtaining results and price.

Ensemble methods are usually indirect methods when information from many nanoparticles is obtained simultaneously. The methods of this group make it possible to establish the average size of nanoparticles as low-angle X-ray scattering (SAXS) and to detect photon interference.

One of the most frequently used and user-friendly methods for determining the size of nanoobjects in this group is the method of dynamic light scattering (DRS), which is based on the principles of coherence of light waves, and metrological characteristics are obtained by determining the phase difference of these waves after interaction with nanoparticles.

One method that potentially combines the advantages of a group of discrete methods and an ensemble approach is the nanoparticle tracking method, which involves determining the position of particles suspended in a liquid by detecting the light they scatter when irradiated by a laser source and viewing the suspension by using a camera with a charged connected device.

In figure 3 presents the ratio of the main methods of research of nanomaterials with different relief characteristics of nanoobjects.

Based on the analysis, it was determined that the most functional and universal in solving a wide range of problems is the method of scanning probe microscopy. SPM, in turn, covers

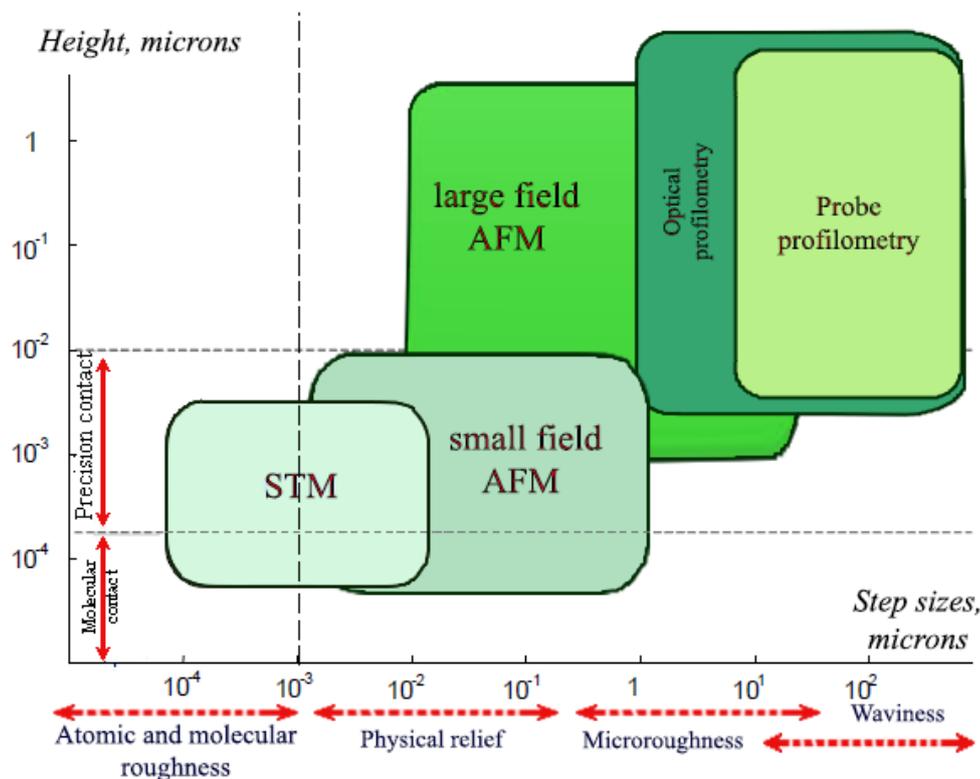


Figure 3: The ratio of the main methods of research of nanomaterials with different indicators of the relief of nanoobjects

several different experimental methods for studying the structure and properties of the surface, both at the micro-level and the level of individual molecules and atoms [4, 15].

The indisputable advantage of this method is the fact that with the help of SPM you can get information directly from a relatively large area of the surface, which allows you to use this method on-line. Therefore, it is not surprising that they are now widely used for the research, diagnosis, and modification of surfaces.

Common to all methods is the presence of a pointed probe as a tool for working with the surface of the samples. There are contact, semi-contact, and non-contact modes of operation, as well as various modes of operation, including tunnel mode, atomic force mode, spectroscopy mode, Kelvin probe method, electric power, magnetic force, near field, optical, confocal microscopy, etc.

With these methods, you can measure not only the topology of the structure but also many special properties, such as modulus of elasticity, distribution of various substances on the surface, the degree of surface roughness, static charge distribution, the orientation of magnetic domains, etc. [16].

Based on the nature of the applications of existing methods in SPM, they can be classified as

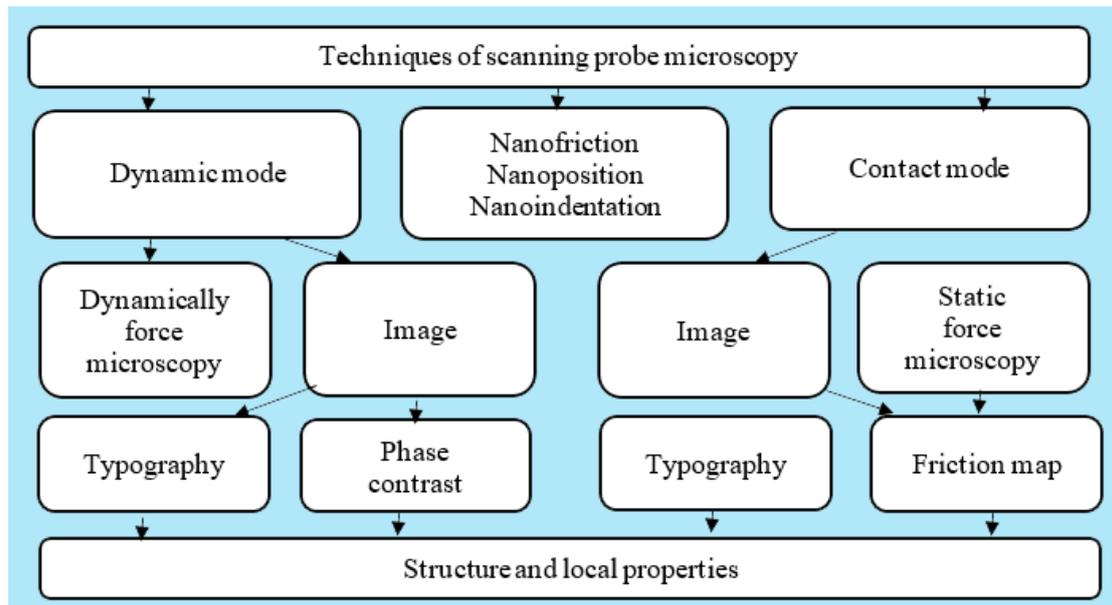


Figure 4: Methods of scanning probe microscopy at elastic and not the elastic effect of the probe

follows (figure 4).

Despite the variety of types and applications of modern scanning microscopes, their work is based on similar principles, and their designs differ little from each other.

Another common feature that unites not only sounding instruments, but also other measuring equipment presented in the article for measuring in the nanomaterial range is their increased sensitivity to external destabilizing factors, such as the composition and properties of the environment, lighting, potential difference, and magnetic field, temperature, etc. Because the properties of certain nanomaterials are different from conventional ones, the effect of destabilizing factors on the measurement results can be completely unpredictable.

Therefore, when choosing an existing or developing a new measurement method, the features of the kinetics and mechanism of the interaction of the measuring nanoobject with the environment should be thoroughly studied, the properties of nanostructures, process parameters, and the model of their mathematical description should be determined.

A necessary condition for the choice of measurement conditions, such assessments, and subsequent interpretation of the data is the reproducibility of the results and the invariance of the sample for a set of properties.

When changing the size can change not only the values of certain values but also the nature of their dependencies on properties, environments, and influences.

The main directions of creation and use of methods of measurement of micro- and nanosystems are defined:

- creation of model representations of connections of topological characteristics and reaction of the object to external influences

- establishing links between the properties of the element or system and the topology of the composition, structure, charges, and fields
- establishing links between technological factors, the kinetics of formation, and the properties of nanostructures
- creation of models that link the functional parameters of devices with the properties of structures that are determined during their manufacture

Therefore, to obtain more reliable results, the method of nano measurements should take into account the effects of external destabilizing factors on the nanoobject and ensure the possibility of correcting the deviations caused by them.

3. Discussion

Based on the analysis, a method for measuring nanoobjects in static and dynamic modes is proposed. This method will take into account the influence of destabilizing factors for nanostructures of different types. The method consists of the continuous determination of the coordinates of the points on the surface of the nanoobject, their processing, and presented in the form of a three-dimensional image. A cluster is introduced to determine the compensating link that corresponds to the tolerance field according to the spectrum of the color image.

The color value of the discrete points of the digital image forms the area, which is represented as isolines. Mathematical software includes geometric parameters of the reference nanostructure. As a result, a region of the color image is formed, which corresponds to the deviations from the shape and location of the surfaces and differs from each other in color (figure 5).

In addition to measuring flat nanostructures, a big task is to determine the topographic features of the surface and the angle of rotation of the plane of the nanoobject. This dependence is formed in the form of a three-dimensional array, in which one column occupies the angle of inclination of the plane, the second – the topographic features of the surface, the third – the coordinate. When scanning the surface of the part, a function based on a mathematical model of the process of traversing the measuring object is used.

The obtained values of the angle of inclination of the plane and topographic features, as well as the values of the coordinates, are compared with the tolerance field, which is presented in the color image (figure 6).

The differential-digital method makes it possible to study errors, determine the parameters of the nanoobject and obtain cross-sections using the developed methods of algorithmic error compensation, which provides zero offsets for the operating conditions of the measurement after the training procedure to obtain a statistical result. The development of the differential-digital method involves the use of an additional information parameter in the mathematical model in the form of a compensating link. The introduction of additional (a priori) information leads the problem to the correct according to the method of regularization for the functional

$$\Omega(x, \lambda) = |Ax - b|^2 + \lambda|x - x_0|,$$

where $\lambda \rightarrow \infty$, x_0 – is the a priori solution vector that coincides with the regularization coefficient.

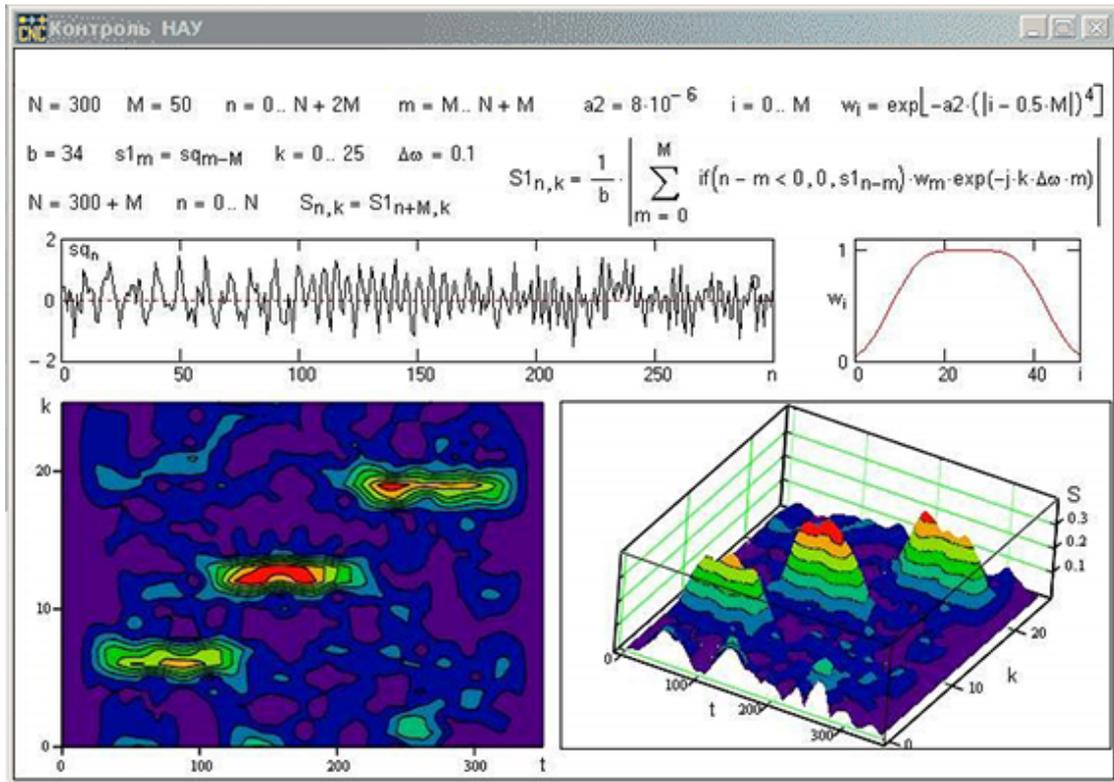


Figure 5: The result of measuring the nanostructure using the differential-digital method

Using this approach to the problem of determining the geometric parameters of the aviation part of a complex spatial surface in small segments, using Delaunay triangulation, using the operation of minimizing the sum of the squares of the deviation, we obtain the following expression:

$$\Phi(\alpha) = \sum_{(i=1)}^n \beta_i^2 + \text{reg}(a, w),$$

where $\sum_{(i=1)}^n \beta_i^2$ – the sum of the squares of the deviation of n measured points, then from the constructed surface; $a = a_0, a_1, \dots, a_i, \dots, a_k$ – the required geometric parameters of the aviation part; $\text{reg}(a, w)$ - regulatory member, which includes information about the nominal value of the geometric parameter and the regularization coefficient w :

$$\text{reg}(a, w) = w \sum_{(i=1)}^k (a_i - a_{i_{y_j v}})^2.$$

Thus, we obtain the problem of the regularizing link, which characterizes the deviation of the parameters of nanoobjects from their nominal values, we can obtain a clear point of minimum in the minimization function. The obtained results were stable and satisfied the value of the tolerance field of the measuring part.

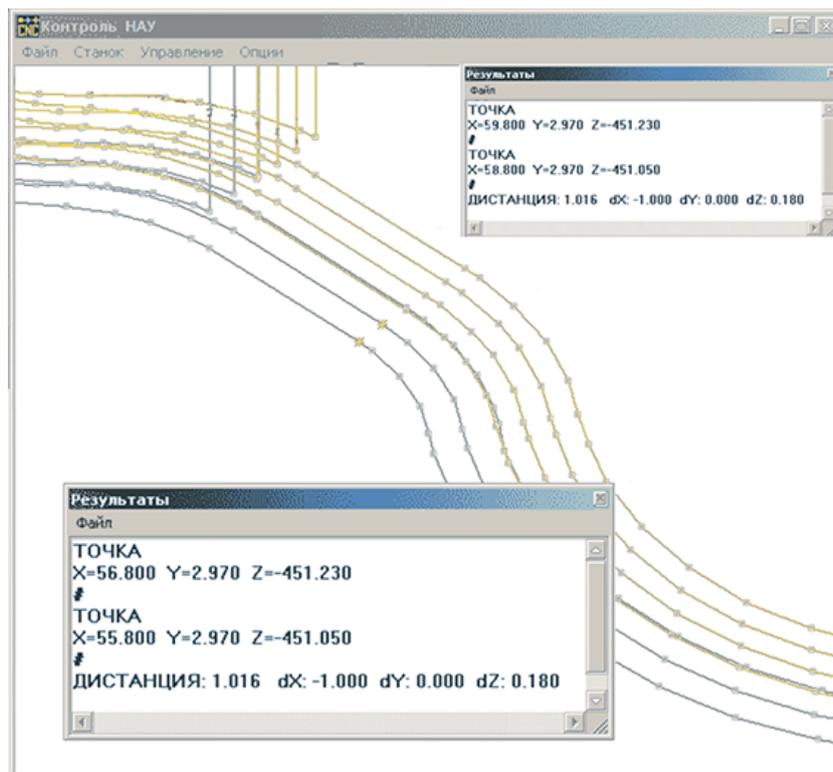


Figure 6: Program window with the results of scanning the surface of the nanoobject

4. Conclusions

There are many advantages and disadvantages in choosing a specific measurement method to measure the topography of non-objects. There is no best general method. The use of integration of information provided by different methods is proposed, but the results of different methods may be contradictory in some cases, namely:

1. Each method is based on the need to study the different properties of nanoobjects
2. The steps of sample preparation can modify the results by further unintentional movements of the particles in the matrix, changing the average diameter detected by the SPM
3. Uncertainty that is not taken into account may also affect the final measurements
4. The use of different weighing in determining the average diameter of the size distribution. Comparing the results of different measurements of the same methodology will be less problematic in this sense because the existing error is constant

Comparing the results of different measurements of the same methodology will be less problematic in this sense because the existing error is constant. Therefore, to test the method, the error must be accurately measured, and the results must be consistent with other methods. Therefore, increasing the number of measurement methods used in the metrological analysis of nanoobjects will increase the reliability and accuracy of measurement results and each

method will provide additional information. The main condition for correct comparison of the result is knowledge of the specific parameters on which each method is based. The analysis was conducted to obtain the most complete information for the development of a computer program based on a discrete-digital measurement method, which would take into account the measurement information obtained by different measurement methods, taking into account the impact of destabilizing factors.

References

- [1] R. Balabai, Electronic properties of functionalized graphene nanoribbons, *Ukrainian Journal of Physics* 58 (2013) 389–397. doi:10.15407/ujpe58.04.0389.
- [2] E. Glushko, A. Stepanyuk, Electromagnetic modes inside the island kind 2d photonic crystal resonator, *Springer Proceedings in Physics* 195 (2017) 263–274. doi:10.1007/978-3-319-56422-7_18.
- [3] V. Zenou, A. Kiv, D. Fuks, V. Ezerski, N. Moiseenko, The microscopic mechanism of silicon precipitation in al/si system, *Materials Science and Engineering A* 435-436 (2006) 556–563. doi:10.1016/j.msea.2006.07.095.
- [4] P. Todua, Metrology and standardization in nanotechnologies and the nanoindustry, *Measurement Techniques* 51 (2008) 462–469.
- [5] J. Garnæs, L. Nielsen, K. Dirscherl, J. Jørgensen, J. Rasmussen, P. Lindelof, C. Sørensen, Two-dimensional nanometer-scale calibration based on one-dimensional gratings, *Applied Physics A* 66 (1998) S831–S835.
- [6] W. Fritzsche, L. Takac, E. Henderson, Application of atomic force microscopy to visualization of dna, chromatin, and chromosomes, *Critical Reviews™ in Eukaryotic Gene Expression* 7 (1997).
- [7] S. Roth, L. Dellmann, G. Racine, N. De Rooij, High aspect ratio uv photolithography for electroplated structures, *Journal of Micromechanics and Microengineering* 9 (1999) 105.
- [8] AZOnano, 2021. URL: <https://www.azonano.com/>.
- [9] R. Said, Microfabrication by localized electrochemical deposition: experimental investigation and theoretical modelling, *Nanotechnology* 14 (2003) 523.
- [10] H. Iwasaki, T. Yoshinobu, K. Sudoh, Nanolithography on sio₂/si with a scanning tunnelling microscope, *Nanotechnology* 14 (2003) R55.
- [11] A. Majumdar, P. Oden, J. Carrejo, L. Nagahara, J. Graham, J. Alexander, Nanometer-scale lithography using the atomic force microscope, *Applied Physics Letters* 61 (1992) 2293–2295.
- [12] C. Hyon, S. Choi, S.-H. Song, S. Hwang, M. Son, D. Ahn, Y. Park, E. Kim, Application of atomic-force-microscope direct patterning to selective positioning of inas quantum dots on gaas, *Applied Physics Letters* 77 (2000) 2607–2609.
- [13] M. Falvo, The nanomanipulator: A teleoperator for manipulating materials at the nanometer scale, in: *Proc. Int. Symp. on the Science and Technology of Atomically Engineered Materials*, 1995, pp. 579–586.
- [14] M. Sitti, H. Hashimoto, Tele-nanorobotics using atomic force microscope, in: *Proceedings. 1998 IEEE/RSJ International Conference on Intelligent Robots and Systems. Innovations*

in Theory, Practice and Applications (Cat. No. 98CH36190), volume 3, IEEE, 1998, pp. 1739–1746.

[15] M. Sitti, H. Hashimoto, Two-dimensional fine particle positioning under an optical microscope using a piezoresistive cantilever as a manipulator, *Journal of Micromechatronics* 1 (2000) 25–48.

[16] Digital Instruments, 2021. URL: <https://www.digital-instruments.com/>.