Computer Science Journal of Moldova, vol.**28**, no.3(84), 2020

Semi-automated workflow for recognition of printed documents with heterogeneous content

Alexandru Colesnicov, Ludmila Malahov,
Svetlana Cojocaru, Lyudmila Burtseva

Abstract

The paper discusses problems of heterogeneous texts digitization. The archives of scanned printed documents grow dramatically by results of projects concerning cultural heritage preserving. Manual annotations of scanned document images and per page screen reading make the usage of these archives difficult and, sometimes, impossible. Existing document processing systems cannot automatically display content correctly due to the presence of heterogeneous content. We proposed a Web platform to maximize the support of semi-automated work of all used tools for recognition of heterogeneous documents. Maximizing support means both creating the convenient "single window" access to all tools, and reducing the manual part of the process as much as possible. For implementation, the convergent technology is used, which assembles complex software systems from ready-made heterogeneous modules on a single platform.

Keywords: platform for heterogeneous document recognition, page layout analysis, non-textual content recognition

1 Introduction

In most documents, whether old or contemporary, there are, along with the text, other elements: mathematical and chemical formulas, musical scores, diagrams, schemes, etc. In the process of digitizing such documents it is necessary to have an instrument, which would provide support not only for the recognition of the text itself, but also of other

components of a heterogeneous nature. For example, see collections of the scanned newspapers that are really huge archives [1] with heterogeneous content. Encyclopedia is a good example of a document with heterogeneous content, because we meet on its pages not only text, but a variety of content types including images, mathematical and chemical formulas, musical scores, technical drawings, chess notation, electronic circuits, etc.

Let's define the notion of heterogeneous content by associating it with the possibility of presentation in a scripting language [2]. The main features of such content are the following:

- the document is not exclusively in natural language;
- there is one or more scripting languages for presenting its components;
- the graphic representation can be rendered by script.

The modern active development of digitization of standard textual documents scans results in obtaining many robust and efficient solutions of this problem. Naturally, the researches employ already developed methods in solving the more general problem of heterogeneous content digitization. Nevertheless, heterogeneous content causes very specific problems which cannot be solved just by adapting of pure textual scans processing methods. Speaking generally about applied methods, the majority of them uses Deep Learning with additions or improvements which apply other techniques: dynamic programming [3], SVM [4], etc. A review of 29 techniques of script text mining with their accuracy estimation is presented in [5].

A number of latest research works present full described frameworks [4], [6]–[9]. The architecture of the presented frameworks depends on functionality: whole archive digitization or partial processing by request. These two groups frameworks are constructed inside almost identically. Basing on this, we can in general consider the architecture and the workflow of such framework as having defined pattern. Online services can either be included in the general framework or they create own online framework [10], [11]. The set of services requested by researchers of focus domain can also be considered as defined.

Recognition of documents with heterogeneous content is quite com-

plex and requires complex analysis of the image with division into homogeneous segments. The research uses already developed methods to solve the general problem of recognizing homogeneous content. Heterogeneous content causes very specific problems that cannot be solved simply by adapting methods for handling pure text scans. Generally speaking on applied methods, most of them use Deep Learning with additions or improvements.

Therefore, there is a need to create a framework that would provide support in digitizing heterogeneous texts, automating processes where it is possible and interacting with the user when the manual intervention or an expert opinion is necessary.

2 Scripting languages

We investigate formal presentation of graphical content with the aim of automated optical recognition. We need a term with wide coverage that includes formal languages for presentation of graphical content. The term "scripting languages" is used narrowly referring to a subclass of programming languages. Another term is "markup languages" but it is restricted by several specific areas.

Heterogeneous content includes graphical elements, in addition to the pure text. As to the graphical content, we exclude from our investigations pure images presented by bitmaps. The object of our research is the graphical content that may be presented by a description (script) in a formal language without loss of information, and may be restored from the said script. To be shorter, we will use the term "scripting languages" instead of more precise "formal languages for presentation of graphical content" in the rest of our paper.

Scripting languages were developed mainly in specific application areas, like mathematics and physics, chemistry, music, building and architecture, technical design, chess, etc.

Many features of scripting languages are defined by specific application area. For example, an important issue for chemistry is the unequivocal identification of a chemical for production and trade. The assortment of chemicals is huge (millions). Each of them can have

dozens of trade names, and, for structurally complex chemicals, tens and hundreds image variants. See [2] for details.

An approach to universal scripting language is LaTeX. It was developed with orientation to mathematics but was extended to many areas including vector graphics, chemistry, music, chess, etc. due to its unrestricted extensibility. This language is constantly being developed and supplemented, in particular in terms of the presentation of non-textual elements. The output from a LaTeX compiler may be used in any device provided with the corresponding driver.

Many standardized scripting languages are extensions of the XML markup language, for example: MathML[1] for mathematics, and MusicXML[2] for music [12].

A standardized language used to exchange CAD projects in the electronic form is STEP (ISO 10303, Standard for the Exchange of Product model data). The covered areas are mechanical engineering, architecture, and building.

3 On digitization of heterogeneous content

3.1 Best practices

Heterogeneous content digitization problem consists of several subproblems, each of which is enough complicated to be a subject of autonomic research. Main groups of subproblems corresponding to processing steps are: pre-processing; recognition/layout analysis; assembling and saving to digital form. Only the first group of subproblems is well studied today.

The pre-processing of scans before content analysis is a usual step of digitization [13]. The main stages of pre-processing are the enhancement and binarization of document images. Document image enhancement here is the improving of the perceptual quality for maximum restoring of document initial look, the binarization is the separation of texts and background. Pre-processing has the specific aspect that can

[1] https://www.w3.org/Math/
[2] https://www.musicxml.com/

be simply formulated as: what is the noise in particular case. We have an experience of using the specialized tools for image preparation (for example, Scan Tailor). OCR programs have internal features for this.

Full range of image processing methods is used to make easier the following recognition stage. For example, work [14] applied the edge enhancement to increase visibility of content elements borders.

The problem of saving in digital form of textual historical documents was mostly resolved. The heterogeneous content digitization has to result in saving digital representation of layout and all its elements. Today digital representation exists for majority of type of content. Moreover, there are well described rules of such digital representation creating. Even very specific content can be supplied by digital representation without new developing. For example, a detailed study of digital representation of equations can be found in [15].

3.2 Document structure and layout analysis

The main problem of digitization is layout analysis. Document pages are divided into areas with the same type of content. The complex page physical layout is converted into logical structure. Layout analysis supposes region segmentation and region classification.

Region segmentation was thoroughly researched. Several segmentation algorithms were introduced being classified as top-down and bottom-up algorithms [16]. Later hybrid algorithms were proposed combining both approaches plus split-and-merge strategy [14], [17].

The problem of the identification and analysis of segment content for labeling heterogeneous components in different types of documents was investigated. Many variants of page structure analysis that provides possibilities to automate recognition of heterogeneous content were proposed. See [18]–[20].

Starting point of general text recognition is the defining of lines of text. For calligraphy of Arabic type all signs located between lines are significant [21].

For standard layout case, the digitization of heterogeneous documents applies text/non-text separation only for specific use case: news-

papers [6]. The newspaper has text and illustrations organized in strict, usually rectangular, shapes. In the contrary, historical documents often have overlapping elements [22].

The classification of elements in heterogeneous documents was proposed by [23].

Second specific feature of heterogeneous content digitization is recognition of pure non-textual elements. One can find studies related to specific elements: figures [10],[24],[25]; molecular structures [20],[26]; tables [27],[28], etc.

However, the general solution for the whole range of problems was not found and not implemented. It is necessary to develop a specific platform combining automated, semi-automated, and manual work.

3.3 Systems for recognizing homogeneous content

There are many OCR tools, commercial and open source, offline and online. OCR for text-based print content is the most advanced in terms of automatic recognition capabilities, user services, and validation of recognized texts. Examples are commercial ABBYY Finereader(AFR)[3], and open source free OCRopus [4] and Tesseract[5].

Currently, OCR tools are based on two technologies, character patterns and neural networks. Development of neural networks permitted to evolve text recognition from recognition of separate characters to the intelligent methods of recognition of words, lines, or phrases at once.

These OCR systems support restricted layout analysis of page images. For example, OCRopus has such features as: binary morphology; adaptive thresholding; deep learning based skew correction; page frame detection. Least square baseline finding algorithm is for constrained text line finding and modeling.

AFR provides the entire document recognition cycle, starting with scanning. It provides complex image correction, page segmentation with automatic detection of the segment type (restricted by text, ta-

[3]https:/www.abbyy.com/en-eu/finereader/tech-specs/
[4]https://github.com/tmbdev/ocropy/
[5]https://github.com//tesseract-ocr/

ble, or picture), analysis of the table structure, hyphenation detection, correction against dictionary, manual editing after recognition, etc.

OCRopus and Tesseract are packages of separate programs managed from the command line. Tesseract OCR may be used as a standalone program, or through API as a subsystem extracting text from images. Tesseract is compatible with many popular programming languages (C++, JavaScript, Python). Newer versions of Tesseract return not only recognized text but co-ordinates of recognized words on the page [6]. Tesseract.js[7] is a tool to extract text from images using JavaScript from HTML pages. Python-tesseract[8] is a Python wrapper for Google's Tesseract-OCR.

There are other text recognition systems but many of them are no longer supported. For example, CuneiForm was discontinued since 2008. Examples of supported OCR systems are AFR, Tesseract, OCRopus, CIB OCR[9], OCR.space (online)[10], Infty Reader[11], etc.

Recognition of heterogeneous content needs to use different software depending on the problem being solved. To recognize pure non-textual elements (figures, multi-chart, equation, diagram, photo, plot, table, art, and technical drawings), it is necessary to develop techniques, create tests collections, and to develop an integrated platform.

3.4 Practices selected as basis of implementation

Presented analysis produces the basis for our decisions about semi-automated workflow organization. For implementation, the technology of assembling complex software systems from ready-made heterogeneous modules on a single platform was tested; each of modules performs a small part of the task, and modules are combined using

[6] https://medium.com/jaafarbenabderrazak.info/ocr-with-tesseract-opencv-and-python-d2c4ec097866

[7] https://proglib.io/p/tesseract-js-izvlekaem-tekst-iz-kartinok-s-pomoshchyu-javascript-2020-04-22

[8] https://pypi.org/project/pytesseract/

[9] https://ocr.team/

[10] https://ocr.space/

[11] https://www.sciaccess.net/en/InftyReader/

Docker[12].

Deep learning is selected as the execution method firstly because of existing software solutions for many necessary subtasks. The analysis discovers the widespread practice to combine previously developed methods and techniques with standard deep learning for obtaining better solution, which is very suitable for our goals.

Architecture and the workflow of the platform we intend to develop can be inferred from analyzed works.

We selected Python as the language of implementation because it provides a lot of ready solutions in its rich libraries. Python with orientation to the Natural Language Processing is presented in [29], [30]. Python packages are used to develop modules for: preliminary preparation of scanned documents (cleaning, alignment, resolution, etc.); the analysis and scanning process; parsing requests to generate processing sequences; image analysis to obtain a page template + metadata using manual intervention to determine the type of context (images can contain formulas, graphs, notes, etc.); recognition of individual components and obtaining a pattern of the restored page; page assembling; manual correction; obtaining results with their XML description.

4 Semi-automated workflow for recognition of heterogeneous documents

4.1 Problems and challenges

From the above, we can draw the following conclusions:

- It is very difficult to recognize many kinds of heterogeneous content,
- Analysis of page structure is a complicated task,
- We need to integrate different kinds of scripts in a unified script presentation of the document.

Therefore, a need for a tool, supposedly, a Web platform to support semi-automated work over heterogeneous documents, arises. Despite a

[12] https://www.docker.com/

lot of achievements, automated recognition of the heterogeneous content remains a difficult problem. Our goal is to maximize the support of semi-automated work.

We see the process in execution of required actions: download input scan; read request; run cleaner; execute OCR including page layout analysis; execute request processing; publish answer.

To solve this problem, we propose a platform for recognition of heterogeneous documents, which uses the previously described and newly developed software, and can perform all stages of processing.

4.2 The functionality of a platform for recognition of heterogeneous documents

4.2.1 The design

We propose a design of a platform for recognition of heterogeneous documents to maximize the support of all processing steps. Maximizing support means both creating the convenient "single window" access to all tools, and reducing the manual part.

The recognition of heterogeneous documents involves many subtasks. Some of them may be performed automatically using specialized software. Other subtasks need slight manual intervention or manual control. If the specialized software does not exist, the processing is executed manually under the general purpose software.

Therefore, we can group all involved subtasks as follows:

Automated: scan; segment recognition according to types of segments; assembling of script presentation of pages with metadata integration; reconstruction of page images from scripts; automated verification.

Semi-automated: image quality improvement; page layout analysis; task distribution for manual verification.

Manual: expert verification and manual correction.

The platform is the Web framework with backend and frontend (Figure 1). Each of processes that constitute the platform functionality is executed by programs of different types. Corresponding to their

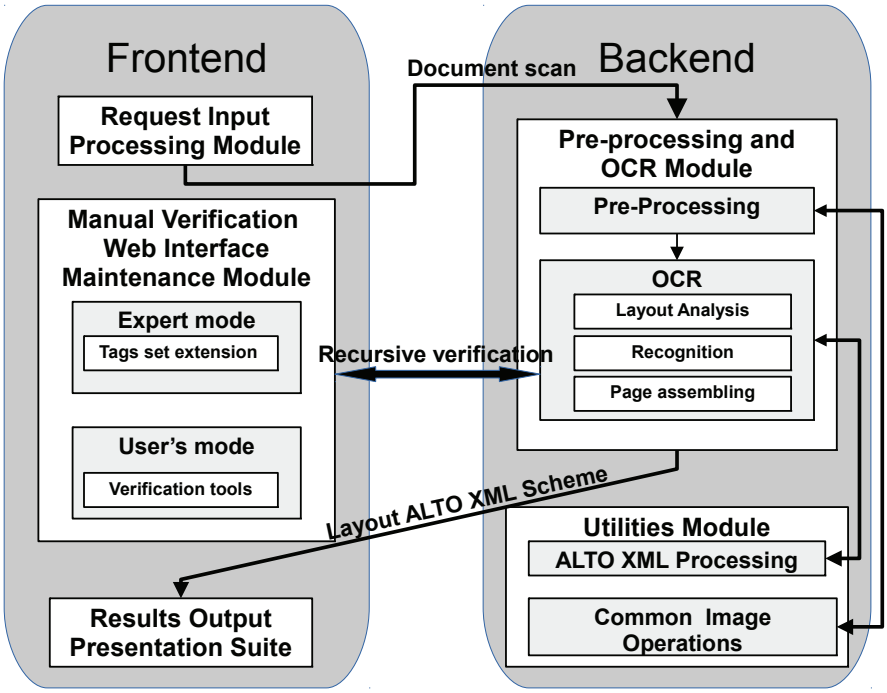


Figure 1. Main platform workflow

functionality, software is organized in container-based modules.The details of the platform functionality are described below.

4.2.2 Utility (backend)

The platform functionality supposes that modules are connected only by data transfer. For data representation, we chose ALTO (Analyzed Layout and Text Object) XML. The main valuable feature of ALTO XML is a good support of extensions, which would provide set of tags for new layout block types. Utilities include many commonly used tools like XML processing, image processing (ImageMagick), etc.

4.2.3 Input request (frontend)

This module aims to obtain user's input data. This may be only the document scan, but additional information like metadata can be supplied too. The module provides uploading and warehousing of the scans, and scheduling them for the next stage.

4.2.4 Pre-processing and OCR (backend)

Pre-processing of heterogeneous documents has specific features. Scans of such documents come mainly from Big Data archives, for example, created by projects in cultural heritage. All kinds of document images have the usual problems: noise, stains, shadings, distortions, etc. Pages can be damaged and can have different textures. During enhancement, we should not lose information.

Then the document scan is processed by a developed module described in details in 4.3. This script analyses the image and prepares page segment files and page maps. Page segments are recognized if possible. The next stage is the verification (automatic, semi-automatic, or manual). Finally, the assembling of the page scripts is performed. The result is a script presentation of pages of the document. This workflow is recursive as verification may unveil errors.

4.2.5 Manual verification (frontend)

Manual verification can be performed either by experts in the corresponding areas or directly by user depending on the particular requirements. Expert correction joins frontend and backend processes. It implies that the platform will be Web-based.

4.2.6 Results output (frontend)

This module executes a set of processes charged with the results presentation. At the previous stage, the recognized layout of the document scan was saved digitally as ALTO XML file. The module reconstructs source image and rebuilds layout scheme from ALTO XML

description. The result is generated to present the reconstructed document with availability of other details like ALTO XML page maps, metadata, annotations, etc.

4.3 Partitioning and mapping of heterogeneous documents

As discussed in 3.2, parsing a heterogeneous document is challenging. We tried some of the OCRs listed in 3.3. It turned out that the most complete tools are provided by ABBYY FineReader Engine (FRE). The set includes a ready-made command line interface (FRE CLI) utility with full recognition cycle. Unlike the standard AFR, the FRE CLI utility processes only one page at a time. The result is returned in XML format and contains the coordinates of the page segments, their type (text, image, table, separator), and, for text segments, also the recognized text. Thus, it is possible to cut the page into segments by type.

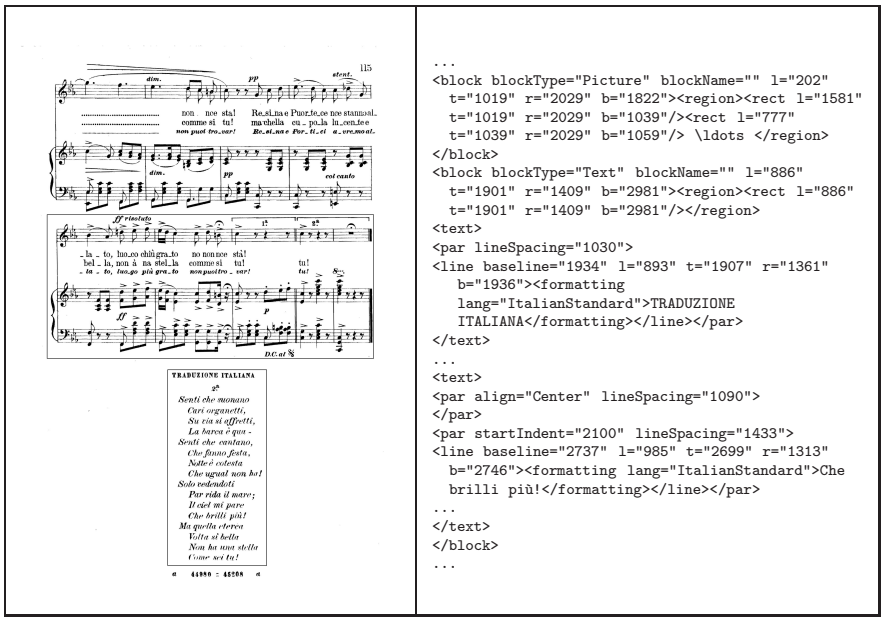
To process many pages, the utility can be called in a loop from a command script in any suitable language, including Python. FRE can be included in a Docker container.

A Python program was developed and coded to cut a scan of a document into segments with the same content. The process algorithm is as follows:

1. Using the FRE CLI utility over a page scan, we get an XML file with the coordinates of the upper and lower corners of the segment rectangles and the segment types (text, picture, table, etc.).
2. A program selects from the XML file the segment metadata (coordinates, segment types) and calls the image cutting utility.
3. A batch of scans of a multi-page document is processed in a cycle. File names with page images are set on the command line, with the ability to use regular expression placeholders “*” and “?”. For each image, a directory is created with a name derived from the image name, into which the XML file and page segments are placed in a format that matches the page image format. Sepa-

rators are excluded from processing. Operations over images are performed by the ImageMagick batch utility.

Thus, the program developed on the basis of FRE generates files with scans of document segments and XML files with page maps for further processing by the platform. Figure 2 on page 235 presents an example of a page scan and a fragment of an XML file after processing by the FRE CLI utility.




```
...
<block blockType="Picture" blockName="" l="202"
  t="1019" r="2029" b="1822"><region><rect l="1581"
  t="1019" r="2029" b="1039"/><rect l="777"
  t="1039" r="2029" b="1059"/> \ldots </region>
</block>
<block blockType="Text" blockName="" l="886"
  t="1901" r="1409" b="2981"><region><rect l="886"
  t="1901" r="1409" b="2981"/></region>
<text>
<par lineSpacing="1030">
<line baseline="1934" l="893" t="1907" r="1361"
   b="1936"><formatting
   lang="ItalianStandard">TRADUZIONE
   ITALIANA</formatting></line></par>
</text>
...
<text>
<par align="Center" lineSpacing="1090">
</par>
<par startIndent="2100" lineSpacing="1433">
<line baseline="2737" l="985" t="2699" r="1313"
  b="2746"><formatting lang="ItalianStandard">Che
  brilli più!</formatting></line></par>
...
</text>
</block>
...
```

Figure 2. A page scan and a fragment of an XML file after processing by the FRE CLI utility (the blocks framed on the left are described on the right)

5 Conclusions

Despite a lot of achievements, automated recognition of the heterogeneous content remains a difficult problem. We proposed a design of

Web platform to maximize the support of semi-automated work with all available tools for recognition of heterogeneous documents.

For implementation, the technology of assembling complex software systems from ready-made modules on a single platform is used. Each of modules performs a small part of the task inside its container. Python's rich libraries dedicated to natural language processing include solutions to the segmentation problem, but without identifying the type of "image". Specifically, only certain types of heterogeneous content are identified, for example, text vs. musical notes. The problem of segmentation, labeling (establishing coordinates), and complete classification of the type of heterogeneous content is not solved completely yet both in scientific publications and in the existing software.

Acknowledgement. This article was written as part of the research project 20.80009.5007.22 "Intelligent information systems for solving ill-structured problems, processing knowledge and big data".

References

[1] B. C. Lee, J. Mears, E. Jakeway, M. M. Ferriter, C. Adams, N. Yarasavage, D. Thomas, K. Zwaard, and D. S. Weld, "The Newspaper Navigator Dataset: Extracting And Analyzing Visual Content from 16 Million Historic Newspaper Pages in Chronicling America". *ArXiv*, abs/2005.01583 (2020).

[2] A. Colesnicov, S. Cojocaru, and L. Malahov, "Recognition of heterogeneous documents: problems and challenges," in *Proceedings of the 5th Conference on Mathematical Foundations of Informatics,* (3-6 July 2019, Iasi, Romania), Iasi: "Alexandru Ion Cuza" University Publishers, 2019, pp. 231–245.ISBN:978-606-714-481-9.

[3] Z. Ziran, X. Pic, S. U. Innocenti, D. Mugnai, and S. Marinai, "Text alignment in early printed books combining deep learning and dynamic programming," *Pattern Recognition Letters*, vol. 133, pp. 109–115, 2020.

[4] W. Qin, R. I. Elanwar, and M. Betke, "LABA: Logical Layout Analysis of Book Page Images in Arabic Using Multiple Support Vector Machines," in *2018 IEEE 2nd International Workshop on Arabic and Derived Script Analysis and Recognition (ASAR)*, 2018, pp. 35–40.

[5] S. Chadha, S. Mittal, and V. Singhal, "An Insight of Script Text Extraction Performance using Machine Learning Techniques," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 9, no. 1, November 2019. ISSN: 2278-3075.

[6] M. Alberti, M. Bouillon, R. Ingold, and M. Liwicki, "Open Evaluation Tool for Layout Analysis of Document Images," in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, (Kyoto), 2017, pp. 43–47.

[7] P. Li, X. Jiang, and H. Shatkay, "Extracting Figures and Captions from Scientific Publications," in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 2018.

[8] C. Clausner and A. Antonacopoulos, "Ontology and Framework for Semantic Labelling of Document Data and Software Methods," in *2018 13th IAPR International Workshop on Document Analysis Systems (DAS)*, (Vienna), 2018, pp. 73–78.

[9] L. Ma, C. Long, L. Duan, X. Zhang, Y. Li, and Q. Zhao, "Segmentation and Recognition for Historical Tibetan Document Images," in *IEEE Access,* 2020, vol. 8, pp. 52641–52651.

[10] N. Siegel et al., "Extracting Scientific Figures with Distantly Supervised Neural Networks," in *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries*, 2018.

[11] H. M. Al-Barhamtoshy and A. S. Alghamdi, "A Comprehensive Framework for OCR Web Services System for Arabic Calligraphy Documents," *International Journal of Engineering and Technology*, vol. 8, No. 1.11 (2019). DOI: 10.14419/ijet.v8i1.11.28084.

[12] A. Colesnicov, S. Cojocaru, and L. Malahov, "On Digitization of Documents with Script Presentable Content," in *Proceedings of the 5th Conference on Mathematical Society of the Rep. Moldova IMCS-55,* (Chisinau), September 28 – October 01, 2019, pp. 321–324. ISBN: 978-9975-68-378-4.

[13] I. V. Safonov, I. V. Kurilin, M. N. Rychagov, and E. V. Tolstaya, *Document Image Processing for Scanning and Printing*, Springer, 2019, 314 p.

[14] S. S. Kumar, P. Rajendran, P. Prabaharan, and K. P. Soman, "Text/Image Region Separation for Document Layout Detection of Old Document Images Using Non-linear Diffusion and Level Set," *Procedia Computer Science*, vol. 93, pp. 469–477, 2016. ISSN 1877-0509.

[15] P. Sojka, V. Novotný, E. F. Ayetiran, D. Lupták, and M. Štefánik, "Quo Vadis, Math Information Retrieval," in *Proceedings of the Thirteenth Workshop on Recent Advances in Slavonic Natural Language Processing, RASLAN 2019.* Brno: Tribun EU, 2019, pp. 117–128. ISBN 978-80-263-1517-9.

[16] A. M. Namboodiri and A. K. Jain, "Document structure and layout analysis," in *Digital Document Processing*, Springer, 2007, pp. 29–48.

[17] Yu. A. Bolotova, V. G. Spitsyn, and P. M. Osina, "A Review of Algorithms for Text Detection in Images and Videos," *Компьютерная оптика (Computational Optics)*, vol. 41, no. 3, pp. 441–452, 2017.

[18] M. Polyakova, A. Ishchenko, N. Volkova, and O. Pavlov, "Combined Method For Scanned Documents Images Segmentation Using Sequential Extraction of Regions," *Eastern-European Journal of Enterprise Technologies*, vol. 5/2 (95), 2018. ISSN 1729-3774.

[19] S. F. Rashid, A. Akmal, M. Adnan, A. A. Aslam, and A. Dengel, "Table Recognition in Heterogeneous Documents Using Machine

Learning," in *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR, vol. 1*, 2018, pp. 777-–782.

[20] J. Staker, K. Marshall, R. Abel, and C. McQuaw, "Molecular Structure Extraction From Documents Using Deep Learning," [Online]. Available: https://www.x-mol.com/paper/5382953.

[21] B. Kiessling, D. S. B. Ezra, and M. T. Miller, "BADAM: A Public Dataset for Baseline Detection in Arabic-script Manuscripts," in *HIP'19: Proceedings of the 5th International Workshop on Historical Document Imaging and Processing*, 2019, pp. 13-–18.

[22] S. Capobianco, and S. Marinai, "DocEmul: A Toolkit to Generate Structured Historical Documents," in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, (Kyoto), 2017, pp. 1186–1191.

[23] F. D. Julca-Aguilar, A. L. L. M. Maia, and N. S. T. Hirata, "Text/Non-Text Classification of Connected Components in Document Images," in *2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, (Niteroi), 2017, pp. 450–455. DOI: 10.1109/SIBGRAPI.2017.66.

[24] P. Li, X. Jiang, and H. Shatkay, "Figure and caption extraction from biomedical documents," *Bioinformatics,* vol. 35, no. 21, pp. 4381–4388, November 2019.

[25] C. Roullet, D. Fredrick, J. Gauch, and R. Vennarucci, "An Automated Technique to Recognize and Extract Images from Scanned Archaeological Documents," in *2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)*, (Sydney, Australia), 2019, pp. 20–25.

[26] A. S. Ashour and F. Shi, *Machine Learning in Bio-Signal Analysis and Diagnostic Imaging,* Academic Press, 2019, pp. 247–272. ISBN 9780128160862.

[27] N. L. Vine, M. Zeigenfuse, and M. Rowan, "Extracting Tables from Documents using Conditional Generative Adversarial Networks and Genetic Algorithms," in *2019 International Joint Conference on Neural Networks (IJCNN)*, (Budapest, Hungary), 2019, pp. 1–8.

[28] S. Paliwal, D. Vishwanath, R. Rahul, M. Sharma, and L. Vig, "TableNet: Deep Learning Model for End-to-end Table Detection and Tabular Data Extraction from Scanned Document Images," in *2019 International Conference on Document Analysis and Recognition (ICDAR)*, 2019, pp. 128–133.

[29] H. Lane, H. Hapke, and C. Howard, *Natural Language Processing in Action: Understanding, Analysing and Generating Text with Python*, March 2019, 544 p. ISBN: 9781617294631.

[30] D. Chopra, N. Joshi, and I. Mathur, *Mastering Natural Language Processing with Python*, Packt Publishing, June 2016, 237 p.

A. Colesnicov[1,2], L. Malahov[1,3],
S. Cojocaru[1,4], L. Burtseva[1,5]

Received November 05, 2020

[1] "Vladimir Andrunachievici" Institute of Mathematics and Computer Science
5 Academiei str., MD-2028, Chisinau
Republic of Moldova
[2] E–mail: acolesnicov@gmx.com
[3] E–mail: ludmila.malahov@math.md
[4] E–mail: svetlana.cojocaru@math.md
[5] E–mail: luburtseva@gmail.com

Computer Science Journal of Moldova, vol.**28**, no.3(84), 2020

On Alignment of Textual Elements in a Parallel Diachronic Corpus

Tudor Bumbu

Abstract

This paper presents the description of some tools and resources for aligning diachronic parallel texts. We try to emphasize the idea of improving the lexicographical similarity between words and expressions used in old and modern texts, using a special technique of natural language processing namely the BLEU score. The overall result of the research is a package of language tools and resources, which will serve to the automatic alignment of the old Romanian text with the modern text.

Keywords: text alignment, parallel diachronic corpus, word alignment tools, BLEU score

1 Introduction

The aligning of an old text to its modern representation means translating it into a contemporary language by replacing outdated lexical variants with modern expressions.

Parallel texts are valuable linguistic resources in many fields of research, but also in practical applications. The best known application is statistical machine translation, or more recently, neural machine translation (using neural networks). Also, parallel texts are researched in disambiguating the meaning of the word, recognizing proper names, learning the language model, but also in the diachronic analysis of natural languages.

A set of parallel texts can form a parallel corpus. The central part in working with a parallel corpus is the alignment task. Alignment in

this sense is the process of linking the corresponding textual parts of parallel texts. An important feature of parallel texts is the property of having in itself a correspondence between two or more texts, for example, the equivalence of translation or paraphrasing. We assume that they connect with each other via their meaning.

In computational linguistics, the term parallel text, also refers to pairs of collections of texts in the same field that include translations of one and the same document. But in most cases, parallel texts refer to bilingual corpora [10].

A diachronic parallel corpus refers to a parallel corpus, where the parallel texts share the same language but their time periods are different.

The purpose of our research is to translate historical texts into texts that use expressions and words from the modern dictionary of the Romanian language. Thus, the first beginning of the work is the elaboration of a parallel diachronic corpus with text in Romanian written hundreds of years ago, aligned with its modern variant. We took as primary resources text from the New Testament of the 17th century and New Testament of the 20th century, but an extension of the corpus follows with other texts from our cultural thesaurus.

2 Textual resources

Our main resource is the parallel diachronic corpus developed in Bumbu, 2019 [1], placed on the web with open access. One might say that this resource cannot be called a parallel corpus but only a parallel text. We are working to expand this resource by digitizing and iteratively adding new texts, so that in the end we get a large collection of parallel texts and we prefer to call it a parallel corpus.

The parallel corpus contains text from the book „Noulu Testament sau Înpacarea, au Leagea noao a lui Is. Hs.", printed in 1648 at the Belgrade Fortress, Transylvania [2] and text from the electronic version of the New Testament (written and annotated by Bartolomeu Valeriu Anania, archbishop of the Archdiocese of Vadul, Feleac and Cluj in 1990) [3]. The volume of text in the corpus expressed in sentences is

currently approx. 10,000 sentences.

Given the fact that the alphabet of the book printed in the seventeenth century is old and unused (the Romanian Cyrillic alphabet), it takes several steps to reach the editable text with the modern alphabet of the Romanian language. The first step is the optical character recognition (OCR). The OCR applied to the old book has an accuracy of up to 80%, and errors are corrected manually. Some issues that worsen the accuracy of OCR applied to the old book are: the pages in the book are worn (they have brownish spots and underlines below text lines); the text in the book has a lot of features specific to that period, such as: letters over other letters, words written together, references written in Slavonic with a smaller font, abbreviations, widespread use of accents, etc (Figure 1).

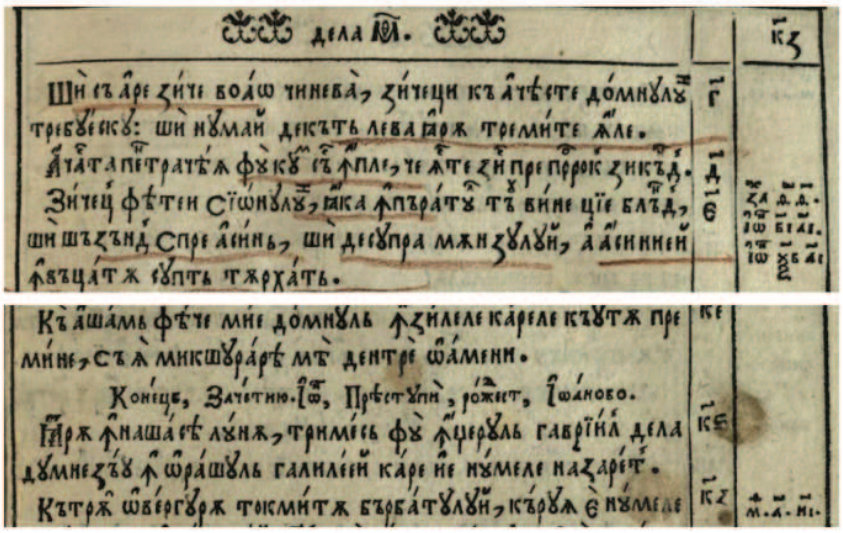


Figure 1. Two passages from the New Testament book of 1648

The next step after OCR is to transliterate the text from the Cyrillic alphabet into the Latin alphabet. At this stage we use a transliteration tool developed at the Institute of Mathematics and Computer Science "V. Andrunachievici" [5]. An example of transliterated text is shown in Table 1.

Table 1. Fragment of transliterated text of the New Testament of 1648

OCRed text	Text after transliteration
Ну џюдекарець касъ ну фиць џюдекаџ	Nu giudecareț casă nu fiț giudecaț
Къ ку че џюдекатъ вещи џюдека, џюдекавъвець	Că cu ce giudecată veți giudeca, giudecavăveț

After transliterating the text we proceed to verse alignment. The verses from the Old New Testament (NTV) were aligned with the verses from the Modern New Testament (NTM) (Table 2) using the Levenshtein distance. The verses are placed in the corpus according to their order in the book. Respectively, 1006 verses from the Gospel according to Matthew, 675 verses from the Gospel according to Mark, 1151 and 668 verses from the Gospel according to Luke and John.

Table 2. Aligned verses from NTV and NTM

NTV verses	NTM verses
Și fu îzilele acelia, veni Iisus de în nazareful galileei, şi să boteză dela Ioan în Iordan.	Și în zilele acelea, Iisus a venit din Nazaretul Galileii și S'a botezat în Iordan de către Ioan.
Și aciiş eşind de în apă, văzu deşchise ceriurele, şi duhul ca un porumbu, pogorănd spre El.	Și îndata, ieșind din apă, a vazut cerurile deschise și Duhul ca un porumbel pogorându-Se peste El.

In the next section some word alignment tools are briefly described, including a tool we are currently working on.

3 Alignment Tools

To reach the automatic alignment of old texts to the modern texts, there are still an indefinite number of steps to be done. An intermediate

step that appears is the alignment of old words to modern ones. The text elements here are words. Word alignment tools are usually special software that helps an expert to map source text words to target text words. One of these tools is the Berkeley Word Aligner (BWA) [7], a program written in Java, used in many cases for machine translation.

BWA aligns the words in a parallel corpus aligned at the sentence level, using a Hidden Markov Model (HMM) [8]. To train an HMM syntactic alignment model, a third parallel text is also needed for each pair of aligned text. All texts should be annotated with syntactic information. Considering the fact that we do not have any third parallel text, nor the annotation of the text with the syntactic information, this tool is useless to us at this stage.

Another tool pack for word alignment is GIZA ++ [9]. This tool also makes extensive use of hidden Markovian models to align texts. Giza ++ works directly with the aligned sentences from two parallel texts, without asking for any additional linguistic information attached to the text. Using an expectation maximization algorithm, the results of the alignment of the final words can be obtained after the software trains itself with a parallel corpus, several iterations, from source text to target text and vice-versa.

Given the fact that we are dealing with parallel diachronic texts, we decided to create an alignment tool that will satisfy special needs. Some examples of necessity will be: the calculation of the BLEU score between texts, sentences, expressions; interactive viewing of n-grams; viewing the word/expression coverage tree; working with more than two parallel texts at the same time etc.

Therefore, a parallel word alignment editor, which is still under construction, helps us to prepare the results of this paper.

The parallel word alignment editor is a WEB application, designed within Django framework, the Python programming language. The application consists of 3 general modules: the parallel text editing and parallel corpus formation module, the word processing module, and the word weighted graph creation module.

The text editing module is the gateway in the alignment application. It offers the possibility to view and edit several parallel texts

simultaneously, and the changes are automatically propagated in the corpus. At the same time, in this module we can add texts directly to the corpus.

Some functionalities already defined in the word processing module are: computing of 1-, 2-, 3-, 4-grams, counting and visualization of n-grams [Figure 2], division into tokens (words and punctuation), associating tokens with numeric identifiers (each word is assigned an ID and the ID is an integer), and calculating the BLEU score between sentences or texts.

The text graph creation module involves mapping the word graphs in the parallel text. This module does not have fully defined functionalities, but we focus on a powerful visualization and interactivity device.

Figure 2. Visualization of 3-grams of 2 parallel texts, ordered descending by frequency

One of the basic functionalities we are going to use to improve the similarity between parallel texts is the BLEU score (Bilingual Evaluation Understanding), a metric used to evaluate a sentence generated

in machine translation compared to a reference sentence (translated by an expert). The score was invented by Kishore Papineni et al., 2002 [6] to evaluate the predictions made by machine translation systems. This method has many advantages: it is fast and easy to calculate; easy to understand; language independent; it correlates very much with human evaluation of the translation and it can be used in diachronic analysis. The approach works by counting n-grams in the translated text and the reference text that match character by character, where 1-gram or uni-gram would be a word and bi-gram would be each pair of words. The comparison is made regardless of the order of the words and can reach up to 4 grams.

We assume that our translation is the source text or NTV in our corpus and the reference text is the target text or NTM in our corpus.

4 Conclusions

This paper presents 3 tools for aligning words in a parallel text. One of these tools is developed by us to frame special functionalities needed to align words in a parallel diachronic corpus. A specific functionality that is integrated by default in our application is the calculation of the BLEU score between sentences. We have included it as an approach to improve the similarity between the texts in a parallel corpus.

Acknowledgement. This work was supported by research project 20.80009.5007.22 "Intelligent information systems for solving ill-structured problems, processing knowledge and big data".

References

[1] T. Bumbu, "Building a Diachronic Parallel Corpus for the Alignment of the Old Romanian Texts," in *Proceedings of the Conference on Mathematical Foundations of Informatics MFOI-2019, July 3-6, 2019, Iasi, Romania*, pp. 263–269.

[2] *Noulu Testamentu sau Înpacarea, au Leagea noao a lui Is. Hs. Domnului nostru*, CR XVII V 13, Gheorghie Rakoti, Alba-Iulia, Romania, 1648, 676 p.

[3] *Biblia. Noul Testament*, Bartolomeu Valeriu Anania, Cluj-Napoca, Romania, 2001.

[4] A. Colesnicov, L. Malahov, and T. Bumbu, "Digitization of Romanian Printed Texts of the 17th Century," in *Proceedings of the 12th International Conference Linguistic Resources and Tools for Processing the Romanian Language. Alexandru Ioan Cuza University Press, 2016*, pp. 1–11.

[5] S. Cojocaru, A. Colesnicov, L. Malahov, T. Bumbu, and S. Ungur, "On Digitization of Romanian Cyrillic Printings of the 17th-18th Centuries," *Computer Science Journal of Moldova*, vol. 25, no. 2(74), pp. 217–225, 2017.

[6] K. Papineni, S. Roukos, T. Ward, and W. J. Zhu, "BLEU: a method for automatic evaluation of machine translation," in *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, July 6-12, 2002, Philadelphia, PA, USA*, pp. 311–318.

[7] P. Liang, B. Taskar, and D. Klein, "Alignment by Agreement," in *Proceedings of the Human Language Technology Conference of the NAACL, Main Conference, 2006*, pp. 104–111.

[8] R. L. Stratonovich, "Conditional Markov Processes," *Theory of Probability and Its Applications*, vol. 5, no. 2, pp. 156–178, 1960.

[9] Franz J. Och, "Giza++: Training of statistical translation models," 1999-2000. Available: http://www.fjoch.com/GIZA++.html.

[10] R. Rapp, "Identifying word translations in non-parallel texts," in *Proceedings of the 33rd Meeting of the Association for Computational Linguistics, Cambridge, MA, 1995*, pp. 320–322.

Tudor Bumbu

State University "Dimitrie Cantemir"
E–mail: bumbutudor10@gmail.com

Vladimir Andrunachievici Institute of Mathematics and Computer Science
E–mail: tudor.bumbu@math.md

Technical University of Moldova
E–mail: tudor.bumbu@iis.utm.md

Received October 28, 2020

Computer Science Journal of Moldova, vol.**28**, no.3(84), 2020

On one type of stability for multiobjective integer linear programming problem with parameterized optimality

Vladimir A. Emelichev, Yury V. Nikulin

Abstract

A multiobjective problem of integer linear programming with parametric optimality is addressed. The parameterization is introduced by dividing a set of objectives into a family of disjoint subsets, within each Pareto optimality is used to establish dominance between alternatives. The introduction of this principle allows us to connect such classical optimality sets as extreme and Pareto. The admissible perturbation in such problem is formed by a set of additive matrices, with arbitrary Hölder's norms specified in the solution and criterion spaces. The lower and upper bounds for the radius of strong stability are obtained with some important corollaries concerning previously known results.

Keywords: Multiobjective problem, integer programming, Pareto set, a set of extreme solutions, stability radius, Hölder's norms.

1 Introduction

Under certain restrictions on the type of space and the properties of the norm, it may turn out that an efficient solution of a specific optimization problem is preserved as a solution for all problems within some nonzero neighborhood in a metric space. Such conservation can be interpreted as the stability of the solution, and the non-existence of such a nonzero neighborhood can be considered as its instability. Quantitative characteristic of such a neighborhood can be called the stability

radius. The widespread use of discrete optimization models has attracted the attention of many experts to the study of various aspects of stability, as well as the problems of parametric and post-optimal analysis of both scalar (single-criterion) and vector (multicriteria) discrete optimization problems (see, for example, the monograph [1], the review [2] as well as the bibliography therein).

The main purpose of works based on the qualitative approach is to obtain conditions guaranteeing the problem possesses some beforehand given property of stability to small changes of the initial data. In the framework of the qualitative direction, the authors focus on identifying various types of stability of the problem [3, 4, 5, 6, 7, 8, 9], establishing a relationship between different types of stability [10], as well as on searching and describing the stability region of the optimal solution [11]. In some recent papers [12, 13], the proximity of some approaches is analyzed at the level of both problem statements and interpreting the common results.

The quantitative direction, described in sufficient detail in [12] (see also, [2] and [13]), is associated with obtaining estimates of permissible changes in the initial data of the problem, preserving a certain predetermined property of optimal solutions. For multiobjective problems this direction is developed in series of papers of V. Emelichev et. al [14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25]. Attempts to elaborate algorithms for calculating (and approximating) such estimates have been made in [26, 27, 28]. The need for such researches is caused by two basic reasons. First, for checking correctness of a concrete optimization model it is important to know borders of change of the input parameters, for which the solution of an optimization problem is not misrepresented. Secondly, there is an opportunity to build algorithms for solving discrete optimization problems, which are based on procedures of finding a stability radius. For example, such procedures can be useful for constructing algorithms solving a sequence of problems of similar type with initial data varying insignificantly.

The concept of stability radius was introduced and investigated for the first time by V. Leontev [29, 30] for the linear scalar trajectory problem, i.e. for the problem on a system of subsets of a finite set with

the linear objective function. Obviously, the most discrete optimization problems may be formulated as a particular case of integer linear programming. Therefore, the concept of stability radius naturally arises therein. Stability radius of integer linear programming problem is defined as the limiting level of independent perturbations of the vector criterion parameters for which new efficient solutions do not appear. Relaxing the demand of nonappearance of new efficient solutions we come to the concept of the strong stability introduced earlier for some scalar and vector discrete optimization problem [2, 3, 16]. This type of stability is understood as existence of a small neighborhood of problem parameters such that for any perturbation there exists an efficient solution preserving its Pareto optimality, although appearance of new efficient optima is not prohibited.

The paper is organized as follows. In Section 2, we formulate parametric optimality and introduce basic concepts. Section 3 contains some auxiliary statements about norms and several lemmas used later for the proof of the main result. In Section 4, we formulate and prove the main result regarding the lower and upper bounds for the strong stability radius. Section 5 lists most important corollaries.

2 Main definitions and problem formulation

Consider a multicriteria integer linear programming problem (ILP) in the following formulation. Let $C = [c_{ij}] \in \mathbf{R}^{m\times n}$ be a matrix whose rows are denoted by $C_i = (c_{i1}, c_{i2}, ..., c_{in}) \in \mathbf{R}^n$, $i \in N_m = \{1, 2, ..., m\}$, $m \geq 1$. Let $x = (x_1, x_2, ..., x_n)^T \in X \subset \mathbf{Z}^n$, $n \geq 2$, and the number of elements of the set X is finite and greater than one. On the set of (admissible) solutions X, we define a vector linear criterion

$$Cx = (C_1x, C_2x, \ldots, C_mx)^T \to \min_{x\in X}. \tag{1}$$

In the space $\mathbf{R}^k$ of arbitrary dimension $k \in \mathbf{N}$, we introduce a binary relation that generates the Pareto optimality principle [31]:

$$y \succ y' \quad \Leftrightarrow \quad y \geq y' \ \ \& \ \ y \neq y',$$

where $y=(y_1, y_2, ..., y_k)^T \in \mathbf{R}^k$, $y'=(y'_1, y'_2, ..., y'_k)^T \in \mathbf{R}^k$.
The symbol $\overline{\succ}$, as usual, denotes the negation of the relation $\succ$.

Let $\emptyset \neq I \subseteq N_m, |I| = v$, and let C_I denote the submatrix of the matrix $C \in \mathbf{R}^{m\times n}$ consisting of rows of this matrix with the numbers of the subset I, i.e.

$$C_I=(C_{i_1}, C_{i_2}, ..., C_{i_v})^T, \quad I=\{i_1, i_2, ..., i_v\},$$

$$1 \leq i_1<i_2< ... <i_v \leq m, \quad C_I \in \mathbf{R}^{v\times n}.$$

Let $s \in N_m$, and let $N_m=\cup_{k\in N_s} I_k$ be a partition of the set N_m into s nonempty sets, i.e. $I_k \neq \emptyset$, $k \in N_s$, and $i \neq j \Rightarrow I_i \cap I_j=\emptyset$. For this partition, we introduce a set of $(I_1, I_2, ..., I_s)$-efficient solutions according to the formula:

$$G^m(C, I_1, I_2, ..., I_s) =\{x \in X : \ \exists k \in N_s \ \ \forall x' \in X \ \ (C_{I_k}x \ \overline{\succ} \ C_{I_k}x')\}. \tag{2}$$

Sometimes for brevity we denote this set by $G^m(C)$.

Obviously, any N_m-efficient solution $x \in G^m(C, N_m)$ $(s=1)$ is Pareto optimal, i.e. efficient solution to problem (1). Therefore, the set $G^m(C, N_m)$ is the Pareto set [31]:

$$P^m(C) =\{x \in X : \ \forall x' \in X \ \ (Cx \ \overline{\succ} \ Cx')\}.$$

We also use the following set:

$$X(x, C) =\{x' \in X : \ Cx \succ Cx'\},$$

which is a set of solutions $x' \in X$ such that x' dominates x in Pareto sense in problem (1). Therefore,

$$P^m(C) =\{x \in X : \ X(x, C) =\emptyset\}.$$

In the other extreme case, when $s=m$, $G^m(C, \{1\}, \{2\}, ..., \{m\})$ is a set of extreme solutions (see e.g. [32]). This set is denoted by $E^m(C)$. Thereby, we have:

$$E^m(C) =\{x \in X : \ \exists k \in N_m \ \ \forall x' \in X \ \ (C_k x \ \overline{\succ} \ C_k x')\}=$$

$$\{x \in X : \ \exists k \in N_m \ \ \forall x' \in X \ \ (C_k x \leq C_k x')\}.$$

It is easy to see that the set of extreme solutions is composed of the best solutions for each of the m criteria. So, in this context, the parameterization of the optimality principle refers to the introduction of such a characteristic of the binary preference relation that allows us to connect the well-known choice functions, parameterizing them from the Pareto to the extreme.

Denoted by $Z^m(C, I_1, I_2, \ldots, I_s)$, the multicriteria ILP problem consists in finding the set $G^m(C, I_1, I_2, \ldots, I_s)$. Sometimes, for the sake of brevity, we use the notation $Z^m(C)$ for this problem.

It is easy to see that the set $P^1(C) = E^1(C)$ is the set of optimal solutions to the scalar (single-criterion) problem $Z^1(C, N_1)$, where $C \in \mathbf{R}^n$.

For any nonempty subset $I \subseteq N_m$ we introduce the notation

$$P(C_I) = \{x \in X : \ \forall x' \in X \ \ (C_I x \ \overline{\succ} \ C_I x')\},$$

$$X(x, C_I) = \{x' \in X : \ C_I x \ \succ \ C_I x')\},$$

i.e.

$$P(C_I) = \{x \in X : \ X(x, C_I) = \emptyset\}.$$

Then, by virtue of (2), we obtain

$$G^m(C, I_1, I_2, \ldots, I_s) = \{x \in X : \ \exists k \in N_s \ \ (x \in P(C_{I_k}))\}. \qquad (3)$$

Therefore, we have

$$G^m(C, I_1, I_2, \ldots, I_s) = \bigcup_{k \in N_s} P(C_{I_k}), \ \ N_m = \bigcup_{k \in N_s} I_k.$$

It is obvious that all the sets given here are nonempty for any matrix $C \in \mathbf{R}^{m \times n}$.

In the space of solutions $\mathbf{R}^n$, we define an arbitrary Hölder's norm l_p, $p \in [1, \infty]$, i.e. by the norm of a vector $a = (a_1, a_2, ..., a_n)^T \in \mathbf{R}^n$ we mean the number

$$\|a\|_p = \begin{cases} \left(\sum_{j \in N_n} |a_j|^p\right)^{1/p} & \text{if} \ \ 1 \leq p < \infty, \\ \max\{|a_j| : \ j \in N_n\} & \text{if} \ \ p = \infty. \end{cases}$$

In the space of criteria $\mathbf{R}^m$, we define an arbitrary Hölder's norm l_q, $q \in [1,\infty]$, and $l_p \neq l_q$. By the norm of the matrix $C \in \mathbf{R}^{m\times n}$ with the rows C_i, $i \in N_m$, we mean the norm of a vector whose components are the norms of the rows of the matrix. By that, we have

$$\|C\|_{pq}=\|(\|C_1\|_p,\|C_2\|_p,\ldots,\|C_m\|_p)\|_q.$$

Obviously,

$$\|C_i\|_p \leq \|C_I\|_{pq} \leq \|\mathrm{C}\|_{pq},\ i \in I \subseteq N_m. \tag{4}$$

So, it is easy to see that for any $a=(a_1,a_2,\ldots,a_n)^T \in \mathbf{R}^n$ with

$$|a_j| =\alpha,\quad j \in N_n,$$

the following equality holds

$$\|a\|_p=\alpha n^{1/p} \tag{5}$$

for any $p \in [1,\infty]$.

In the solution space $\mathbf{R}^n$ along with the norm l_p, $p \in [1,\infty]$, we will use the conjugate norm l_{p^*}, where the numbers p and p^* are connected, as usual, by the equality

$$\frac{1}{p}+\frac{1}{p^*}=1,$$

assuming $p^*=1$ if $p=\infty$, and $p^*=\infty$ if $p=1$. Therefore, we further suppose that the range of variation of the numbers p and p^* is the closed interval $[1,\infty]$, and the numbers themselves are connected by the above conditions.

Further we use the well-known Hölder's inequality

$$|a^Tb| \leq \|a\|_p\|b\|_{p^*} \tag{6}$$

that is true for any two vectors $a=(a_1,a_2,\ldots,a_n)^T \in \mathbf{R}^n$ and $b=(b_1,b_2,\ldots,b_n)^T \in \mathbf{R}^n$ and for any $p \in [1,\infty]$.

Perturbation of the elements of the matrix C is imposed by adding matrices C' from $\mathbf{R}^{m\times n}$ to it. Thus, the perturbed problem $Z^m(C+C')$ has the form

$$(C+C')x \to \min_{x\in X},$$

and the set of its $(I_1, I_2, ..., I_s)$-efficient solutions is $G^m(C+C', I_1, I_2, \ldots, I_s)$.

For an arbitrary number $\varepsilon > 0$, we define the set of perturbing matrices

$$\Omega_{pq}(\varepsilon) = \left\{ C' \in \mathbf{R}^{m\times n} : \ \|C'\|_{pq} < \varepsilon \right\}$$

with rows $C'_i,\ i \in N_m$.

Following [16, 34], the *strong stability radius* of the ILP problem $Z^m(C, I_1, I_2, \ldots, I_s)$, $m \in \mathbf{N}$, (called T_1-stability radius in the terminology of [1, 2, 8, 9]) is the number

$$\rho = \rho_s^m(p,q) = \begin{cases} \sup\Xi & \text{if } \ \Xi \neq \emptyset, \\ 0 & \text{if } \ \Xi = \emptyset, \end{cases}$$

where

$$\Xi = \left\{ \varepsilon > 0 : \ \forall C' \in \Omega_{pq}(\varepsilon) \qquad \left(G^m(C) \cap G^m(C + C') \neq \emptyset \right) \right\}.$$

Thus, the strong stability radius of the problem $Z^m(C)$ determines the limit level of perturbations of the elements of the matrix C that preserve optimality of at least one (not necessarily the same) solution of the set $G^m(C)$ of the original problem. For any $C' \in \Omega_{pq}(\varepsilon)$ and $\varepsilon > 0$, it is obvious that $G^m(C) \cap G^m(C + C') \neq \emptyset$ if $G^m(C) = \mathrm{X}$. Therefore, the problem $Z^m(C)$ with $\bar{G}^m(C) = G^m(C) \backslash X = \emptyset$ is called *non-trivial*.

The problem $Z^m(C)$ is called *degenerated* if the following formula holds

$$\forall x \notin G^m(C)\ \forall a \in \mathbf{R}^n\ \exists x^0 \in G^m(C)\ (a^T(x - x^0) \geq 0).$$

If the negation of the formula above is true, i.e.

$$\exists x^0 \notin G^m(C)\ \exists a \in \mathbf{R}^n\ \forall x \in G^m(C)\ (a^T(x^0 - x) < 0), \qquad (7)$$

then the problem $Z^m(C)$ is called *non-degenerated*.

It is easy to see that non-trivial problem is also non-degenerated if and only if there exists a solution $x^0 \notin G^m(C)$ such that a system containing $|G^m(C)|$ strict inequalities with n variables has a solution. In particular, as we show later (see the proof of Theorem 1), the Boolean problem $Z_B^m(C)$ has solutions.

3 Lemmas

Before formulating the main result regarding the strong stability radius bounds in the next section, we need to prove five supplementary statements presented in this section as lemmas.

Lemma 1. *A solution $x \notin G^m(C, I_1, I_2, \ldots, I_s)$ if and only if for any index $k \in N_s$ the solution $x \notin P(C_{I_k})$.*

Hereinafter, a^+ is a projection of a vector $a = (a_1, a_2, \ldots, a_k) \in \mathbf{R}^k$ on a positive orthant, i.e. $a^+ = [a]^+ = (a_1^+, a_2^+, \ldots, a_k^+)$, where superscript $+$ implies positive cut of vector a. That is, we have

$$a_i^+ = [a_i]^+ = \max\{0, a_i\},\ i \in N_k.$$

Lemma 2. *Given $p, q \in [1, \infty]$, $x^0 \in G^m(C, I_1, I_2, \ldots, I_s)$, $k \in N_s$ and $\varphi > 0$ such that for any $x \notin G^m(C)$, the inequality*

$$\left\| \left[C_{I_k}(x - x^0) \right]^+ \right\|_q \geq \varphi \|x - x^0\|_{p^*} > 0 \tag{8}$$

holds. Then the following formula is true:

$$\forall\ x \notin G^m(C) \quad \forall C' \in \Omega_{pq}(\varphi)\ \left(x \notin X(x^0, C_{I_k} + C'_{I_k}) \right). \tag{9}$$

Proof. Assume there exists a solution $\tilde{x} \notin G^m(C)$ and a perturbing matrix $\tilde{C} \in \Omega_{pq}(\varphi)$ such that

$$\tilde{x} \in X(x^0, C_{I_k} + \tilde{C}'_{I_k}).$$

Then for any index $i \in I_k$, the following inequality is true:

$$\left(C_i + \tilde{C}_i \right) x^0 \geq \left(C_i + \tilde{C}_i \right) \tilde{x}.$$

Hence, we have

$$\tilde{C}_i \left(x^0 - \tilde{x} \right) \geq C_i \left(x^0 - \tilde{x} \right), \quad i \in I_k.$$

From the above we derive

$$\left| \tilde{C}_i \left(\tilde{x} - x^0 \right) \right| \geq \left[C_i \left(\tilde{x} - x^0 \right) \right]^+, \quad i \in I_k.$$

Taking into consideration Hölder's inequality (6), we obtain

$$\|\tilde{C}_i\|_p \|\tilde{x} - x^0\|_{p^*} \geq \left[C_i\left(\tilde{x} - x^0\right)\right]^+, \quad i \in I_k.$$

Due to inequalities (4), we get a contradiction with (8):

$$\varphi\|\tilde{x} - x^0\|_{p^*} > \|\tilde{C}\|_{pq}\|\tilde{x} - x^0\|_{p^*} \geq \|\tilde{C}_{I_k}\|_{pq}\|\tilde{x} - x^0\|_{p^*} \geq$$
$$\geq \left\| \left[C_{I_k}\left(\tilde{x} - x^0\right)\right]^+ \right\|_q,$$

so formula (9) is valid. □

Lemma 3. *For any non-degenerated ILP problem $Z^m(C)$, there exists a non-zero matrix $C^* \in \mathbf{R}^{m\times n}$ such that*

$$G^m(C) \cap G^m(C^*) = \emptyset.$$

Proof. According to the definition of non-degenerated problem $Z^m(C)$, the equation (7) holds, i.e. for all $x \in G^m(C)$ the inequality

$$a^T(x^0 - x) < 0 \tag{10}$$

is true for any $x^0 \notin G^m(C)$. Obviously, $a \neq \mathbf{0} = (0, 0, \ldots, 0)^T \in \mathbf{R}^m$. Let rows $C_i^*,\ i \in N_m$ of the matrix $C^* \in \mathbf{R}^{m\times n}$ be defined as:

$$C_i^* = a^T,\ i \in N_m\ .$$

Then taking into account (10), we get

$$C_i^*\left(x^0 - x\right) < 0,\ i \in N_m.$$

Thus for any index $k \in N_s$, the solution $x \notin P(C_{I_k}^*)$ if $x \in G^m(C)$. Therefore, due to Lemma 1, we have $x \notin G^m(C^*)$. The last implies

$$G^m\left(C\right) \cap G^m\left(C^*\right) = \emptyset. \quad \square$$

Lemma 4. *Let $x^0 \in G^m\left(C, I_1, I_2, \ldots, I_s\right)$. For any non-trivial ILP problem $Z^m(C)$ and perturbing matrix $C' = (C'_{I_1}, C'_{I_2}, \ldots, C'_{I_s})^T \in \mathbf{R}^{m\times n}$ such that for some index $k \in N_s$, the equality*

$$X(x^0, C_{I_k} + C'_{I_k}) \cap \bar{G}^m\left(C\right) = \emptyset \tag{11}$$

holds. Then we have

$$G^m\left(C\right) \cap G^m\left(C + C'\right) \neq \emptyset \tag{12}$$

Proof. If $x^0 \in G^m(C + C')$, the statement of lemma is obvious. Assume $x^0 \notin G^m(C + C')$. Then according to Lemma 1, for any index $k \in N_s$ we have $x^0 \notin P(C_{I_k} + C'_{I_k})$. Due to the property of external stability of the Pareto set $P(C_{I_k} + C'_{I_k})$ (see e.g., [33]), there exists a solution $x^* \in P(C_{I_k} + C'_{I_k})$ such that $x^* \in X(x^0, C_{I_k} + C'_{I_k})$, and due to (3) $x^* \in G^m(C + C')$. Using (11), we get $x^* \in G^m(C)$. Hence, (12) holds. □

Lemma 5. *If $\rho_s(p, q) < \infty$, then the following formula holds:*

$$\exists a \in \mathbf{R}^n \ \forall x \in G^m(C) \ \exists x^0(x) \notin G^m(C) \ (a^T(x^0(x) - x) < 0), \quad (13)$$

□

Proof. Assume that (13) does not hold. Then we have

$$\forall a \in \mathbf{R}^n \ \exists x^0 \in G^m(C, I_1, I_2, \ldots, I_s)$$

$$\forall x \notin G^m(C, I_1, I_2, \ldots, I_s) \ (a^T(x - x^0) \geq 0).$$

Let $C' = (C'_{I_1}, C'_{I_2}, \ldots, C'_{I_s})^T \in \mathbf{R}^{m \times n}$ be any perturbing matrix. Then for any chosen index $k \in N_s$ there exists $x^0 \in G^m(C)$ such that for any $x \notin G^m(C)$ the inequality

$$\left(C_i + C'_i\right)(x - x^0) \geq 0, \ i \in I_k$$

holds.

Therefore, $x \notin X(x^0, C_{I_k} + C'_{I_k})$. Further, applying Lemma 4, we get that (12) is valid for any matrix $C' \in \mathbf{R}^{m \times n}$, i.e. $\rho_s(p, q) = \infty$. This contradiction ends the proof. □

4 Main result

For the multicriteria non-trivial ILP problem $Z^m(C, I_1, I_2, \ldots, I_s)$, $m \in \mathbf{N}$, for any $p, q \in [1,\infty]$ and $s \in N_m$ we define:

$$\varphi_s^m(p, q) = \max_{x' \in G^m(C)} \max_{k \in N_s} \min_{x \notin G^m(C)} \frac{\|[C_{I_k}(x - x')]^+\|_q}{\|x - x'\|_{p^*}},$$

$$\psi_s^m(p,q)=n^{\frac{1}{p}}m^{\frac{1}{q}} \min_{x\notin G^m(C)} \max_{x'\in G^m(C)} \max_{k\in N_s} \max_{i\in I_k} \frac{C_i(x-x')}{\|x-x'\|_1}.$$

We are now ready to formulate the main result.

Theorem 1. *For any $m \in \boldsymbol{N}$, $p,q \in [1,\infty]$ and $s \in N_m$, the strong stability radius of the multicriteria non-trivial ILP problem $Z^m(C, I_1, I_2, \ldots, I_s)$ has the following lower and upper bounds:*

$$0 < \varphi_s^m(p,q) \le \rho_s^m(p,q) \begin{cases} \le \|C\|_{pq}, & \text{if the problem is non-degenerated;} \\ = \infty, & \text{otherwise.} \end{cases}$$

If the problem is Boolean, i.e. $Z^m(C) = Z_B^m(C)$, then

$$0 < \varphi_s^m(p,q) \le \rho_s^m(p,q) \le \min\{\psi_s^m(p,q),\ \|C\|_{pq}\}.$$

Proof. Due to (3), the formula is true:

$$\forall\ x' \in G^m(C) \quad \exists\ k \in N_s\ \left(x' \in P(C_{I_k})\right).$$

Therefore, due to Lemma 1, for any index $k \in N_s$, we get $x \notin P(C_{I_k})$ if $x \notin G^m(C)$. From there we conclude that the lower bound is positive, i.e. $\varphi_s^m(p,q) > 0$.

Now we prove that $\rho_s^m(p,q) \ge \varphi_s^m(p,q)$. We choose an arbitrary perturbing matrix $C' \in \mathbf{R}^{m\times n}$ such that it belongs to $\Omega_{pq}(\varphi_s^m(p,q))$. In order to prove the lower bound for strong stability radius, it suffices to demonstrate that there exists a solution $x^* \in G^m(C) \cap G^m(C+C')$. According to the definition of the number $\varphi_s^m(p,q)$, there exist a solution $x^0 \in G^m(C)$ and an index $k \in N_s$ such that for any solution $x \notin G^m(C)$ we have:

$$\|[C_{I_k}(x-x^0)]^+\|_q \ge \varphi_s^m(p,q)\,\|x-x^0\|_{p^*} > 0.$$

From the above, by Lemma 2, we get that the following formula is true:

$$\forall\ x \notin G^m(C)\ \ \forall C' \in \Omega_{pq}(\varphi_s^m(p,q))$$
$$\forall\ x \notin G^m(C)\ \ \forall C' \in \Omega_{pq}(\varphi_s^m(p,q))\ \ \left(x \notin X(x^0, C_{I_k} + C'_{I_k})\right). \quad (14)$$

Further, we define a way of selecting a necessary solution

$$x^* \in G^m(C) \cap G^m\left(C+C'\right),$$

where $C' \in \Omega_{pq}(\varphi_s^m(p,q))$. If $x^0 \in G^m(C+C')$, then we select $x^* = x^0$. Otherwise, due to Lemma 1 we have $x^0 \notin P(C_{I_k}+C'_{I_k})$. Thus due to the property of outer stability for the Pareto set $P(C_{I_k}+C'_{I_k})$ (see e.g. [33]), we can chose a solution $x^* \in P(C_{I_k}+C'_{I_k})$ such that $x^* \in X(x^0, C_{I_k}+C'_{I_k})$. Taking into account the proven formula (14), $x^* \in G^m(C)$. Since, due to (3), we have $x^* \in G^m(C+C')$, that involves $\rho_s^m(p,q) \geq \varphi_s^m(p,q)$.

Further, we prove that inequality $\rho_s^m(p,q) \leq \|C\|_{pq}$ is valid for any non-degenerated problem $Z^m(C)$. Let $\varepsilon > \|C\|_{pq}$. According to Lemma 3, for any such problem, there exists a non-zero matrix $C^* \in \mathbf{R}^{m\times n}$ such that

$$G^m(C) \cap G^m(C^*) = \emptyset. \tag{15}$$

We consider a perturbing matrix $C^0 \in \mathbf{R}^{m\times n}$ defined as:

$$C^0 = \xi C^* - C,$$

where $0 < \xi < \frac{\varepsilon - \|C\|_{pq}}{\|C^*\|_{pq}}$. Then we easily derive

$$\left\|C^0\right\|_{pq} = \|\xi C^* - C\|_{pq} \leq \xi \|C^*\|_{pq} + \|C\|_{pq} < \varepsilon.$$

Therefore due to (15) we obtain

$$\forall \varepsilon > \|C\|_{pq} \quad \exists C^0 \in \Omega_{pq}(\varepsilon) \quad \left(G^m(C) \cap G^m\left(C+C^0\right) = \emptyset\right).$$

Thus, $\rho_s^m(p,q) < \varepsilon$ for any $\varepsilon > \|C\|_{pq}$. Hence, $\rho_s^m(p,q) \leq \|C\|_{pq}$.

Further, we show that for degenerate problem $Z^m(C)$, the strong stability radius is equal to infinity. Assume the opposite, i.e. assume that degenerated problems have a finite strong stability radius. Then, according to Lemma 5, formula (13) is valid. Thus letting

$$x^* = arg\min\{a^T(x^0(x)) : \ x \in G^m(C)\},$$

we get that the following inequality

$$a^T(x^0(x^*) - x) < 0$$

is true for any $x \in G^m(C)$. Thus, formula (7) is true, i.e. the problem $Z^m(C)$ is non-degenerated. The obtained contradiction proves that $\rho_s^m(p,q) = \infty$.

Further, we consider non-trivial Boolean problem $Z_B^m(C, I_1, I_2, \ldots, I_s)$, $C \in \mathbf{R}^{m\times n}$, $m \in \mathbf{N}$, $s \in N_m$, $X \subseteq \mathbf{E}^n = \{0,1\}^n$, $n \geq 2$. Clearly, the lower bounds proven above for ILP problem stay valid in Boolean case.

First, we prove that $\rho_s^m(p,q) \leq \psi_s^m(p,q)$.

According to the definition of number $\psi_s^m(p,q)$, there exists a solution $x^0 = (x_1^0, x_2^0, \ldots, x_n^0) \notin G^m(C)$ such that for any solution $x \in G^m(C)$ and any index $k \in N_s$ the following inequalities hold:

$$\psi_s^m(p,q)\,\|x^0 - x\|_1 \geq n^{\frac{1}{p}} m^{\frac{1}{q}} C_i\left(x^0 - x\right), \quad i \in I_k. \tag{16}$$

Let $\varepsilon > \psi_s^m(p,q)$. We choose a perturbing matrix $C^0 = [c_{ij}^0] \in \mathbf{R}^{m\times n}$ with rows C_i^0, $i \in N_m$ and elements defined as follows:

$$c_{ij}^0 = \begin{cases} -\delta \text{ if } & i \in N_m \text{ and } x_j^0 = 1, \\ \delta \text{ if } & i \in N_m \text{ and } x_j^0 = 0, \end{cases}$$

where

$$\psi_s^m(p,q) < \delta n^{\frac{1}{p}} m^{\frac{1}{q}} < \varepsilon. \tag{17}$$

Therefore, due to (5) we have

$$\left\|C_i^0\right\|_p = \delta n^{\frac{1}{p}}, i \in N_m,$$

$$\left\|C^0\right\|_{pq} = \delta n^{\frac{1}{p}} m^{\frac{1}{q}},$$

$$C^0 \in \Omega_{pq}(\varepsilon).$$

Moreover, the following inequalities are obvious:

$$C_i\left(x^0 - x\right) = -\delta\|x^0 - x\|_1 < 0, \ i \in I_k.$$

Using (16) and (17), we conclude that the following inequalities hold for any solution $x \in G^m(C)$:

$$\left(C_i + C_i^0\right)\left(x^0 - x\right) \leq \left(\frac{\psi_s^m(p,q)}{n^{\frac{1}{p}} m^{\frac{1}{q}}} - \delta\right)\|x^0 - x\|_1 < 0, \ i \in I_k.$$

Thus for any index $k \in N_s$ we have $x \in G^m(C)$ and $x \notin P\left(C_{I_k} + C^0_{I_k}\right)$, and hence, due to Lemma 1, $x \notin G^m\left(C + C^0\right)$. Summarizing, for any $\varepsilon > \psi^m_s(p,q)$ there exists the perturbing matrix $C^0 \in \Omega_{pq}(\varepsilon)$ such that $G^m(C) \cap G^m\left(C + C^0\right) = \emptyset$, i.e. $\rho^m_s(p,q) < \varepsilon$. Thus, we have just proven that $\rho^m_s(p,q) \le \psi^m_s(p,q)$.

Finally, we prove that $\rho^m_s(p,q) \le \|C\|_{pq}$ is valid for any Boolean problem $Z^m_B(C, I_1, I_2, \ldots, I_s)$. In order to do this, it suffices to show that any non-trivial Boolean problem is also non-degenerated. Let $\alpha > 0$ and $x^0 = (x^0_1, x^0_2, \ldots, x^0_n) \notin G^m(C)$. We choose a vector $a = (a_1, a_2, \ldots, a_n)^T$ with elements defined as follows:

$$a_j = \begin{cases} -\alpha \text{ if } & x^0_j = 1, \\ \alpha \text{ if } & x^0_j = 0. \end{cases}$$

Then for any $x \in G^m(C)$ $(x \ne x^0)$, we have

$$a^T(x^0 - x) < 0.$$

Thus, (7) is true, and hence $Z^m_B(C)$ is non-degenerated. Therefore, collecting all the proven above, we get $\rho^m_s(p,q) \le \|C\|_{pq}$. This ends the proof of the main result. □

5 Corollaries

From Theorem 1 we get the following well-known result:

Corollary 1. *[16] For any $m \in \boldsymbol{N}$ and any $p = q = \infty$, the strong stability radius of the multicriteria non-trivial ILP problem $Z^m(C, N_m)$, $C \in \mathbf{R}^{m \times n}$ of finding the Pareto set $P^m(C)$ has the following bounds:*

$$0 < \varphi^m_1(\infty, \infty) \le \rho^m_1(\infty, \infty).$$

Moreover,

$$0 < \varphi^m_1(\infty, \infty) \le \rho^m_1(\infty, \infty) \le \psi^m_1(\infty, \infty),$$

if the problem $Z^m(C, N_m)$ *is Boolean, where*

$$\varphi_1^m(\infty,\infty) = \max_{x' \in P^m(C)} \quad \min_{x \notin P^m(C)} \quad \max_{i \in N_m} \frac{C_i(x-x')}{\|x-x'\|_1},$$

$$\psi_1^m(\infty,\infty) = \min_{x \notin P^m(C)} \max_{x' \in P^m(C)} \quad \max_{i \in N_m} \frac{C_i(x-x')}{\|x-x'\|_1}.$$

The *stability radius* of an efficient solution $x^0 \in P^m(C)$ of the ILP problem $Z^m(C, N_m)$, $m \in \mathbf{N}$, is called the number

$$\rho_s^m(x^0,p,q) = \begin{cases} \sup \Theta_{\mathrm{pq}} & \text{if } \Theta_{\mathrm{pq}} \neq \emptyset, \\ 0 & \text{if } \Theta_{\mathrm{pq}} = \emptyset, \end{cases}$$

where

$$\Theta_{pq} = \Big\{ \varepsilon > 0 : \quad \forall C' \in \Omega_{pq}(\varepsilon) \qquad \big(x^0 \in P^m(C + C')\big) \Big\}.$$

In [35] it was shown that for any $m \in \boldsymbol{N}$ and $p, q \in [1, \infty]$, the stability radius of an efficient solution $x^0 \in P^m(C)$ of the multicriteria non-trivial ILP problem $Z^m(C, N_m)$ is expressed by the formula:

$$\rho^m(x^0,p,q) = \min_{x \in X \setminus \{x^0\}} \frac{\|[C(x-x^0)]^+\|_q}{\|x-x^0\|_{p^*}}.$$

It is evident that $\rho_1^m(p,q) = \rho^m(x^0,p,q)$ if $P^m(C) = \{x^0\}$. Therefore, from Theorem 1 we conclude the following result.

Corollary 2. *If* $P^m(C) = \{x^0\}$*, then the strong stability radius of the multicriteria ILP problem* $Z^m(C, N_m)$*,* $C \in \mathbf{R}^{m \times n}$*, of finding the Pareto set* $P^m(C)$ *is expressed by the formula for any* $m \in \boldsymbol{N}$ *and* $p, q \in [1, \infty]$*:*

$$\rho_1^m(p,q) = \varphi_1^m(p,q) = \min_{x \in \mathrm{X} \setminus \{x^0\}} \frac{\|[C(x-x^0)]^+\|_q}{\|x-x^0\|_{p^*}}.$$

In scalar case (single criterion), we have $P^1(C) = G^1(C, N_1)$, $C \in \mathbf{R}^n$, i.e. the Pareto set constricts to a set of optimal solutions in $Z^1(C, N_1)$. It is easy to see that the problem $Z^1(C, N_1)$ with condition $P^1(C) \neq X$ is non-degenerated. Therefore, Theorem 1 transforms into the following result for $m = 1$.

Corollary 3. *Let x^0 be an optimal solution for scalar ILP problem $Z^1(C, N_1)$, $C \in \mathbf{R}^n$. Then for any $p, q \in [1, \infty]$, the strong stability radius has the following bounds:*

$$0 < \min_{x \notin P^1(C)} \frac{C^T(x - x^0)}{\|x - x^0\|_{p^*}} \leq \rho_1^1(p.q) \leq \| C \|_{pq} .$$

From Theorem 1, we get the following known results.

Corollary 4. *[21] For any $m \in \boldsymbol{N}$, $p, q \in [1, \infty]$, the strong stability radius of the multicriteria non-trivial Boolean problem $Z_B^m(C, N_m)$ consisting in finding the Pareto set $P^m(C)$ has the following lower and upper bounds:*

$$0 < \max_{x' \in P^m(C)} \min_{x \notin P^m(C)} \frac{\|[C(x - x')]^+\|_q}{\|x - x'\|_{p^*}} \leq \rho_1^m(p, q) \leq$$

$$n^{\frac{1}{p}} m^{\frac{1}{q}} \min_{x \notin P^m(C)} \max_{x' \in P^m(C)} \max_{i \in N_m} \frac{C_i(x - x')}{\|x - x'\|_1}.$$

Corollary 5. *[36] For any $m \in \boldsymbol{N}$, $p, q \in [1, \infty]$, the strong stability radius of the multicriteria non-trivial Boolean problem $Z_B^m, (C, \{1\}, \{2\}, \ldots \{n\})$, $C \in \mathbf{R}^{n \times m}$ consisting in finding the extreme set $E^m(C)$ has the following lower and upper bounds:*

$$0 < \max_{x' \in E^m(C)} \max_{i \in N_m} \min_{x \notin E^m(C)} \frac{C_i(x - x')}{\|x - x'\|_{p^*}} \leq \rho_m^m(p, q) \leq$$

$$n^{\frac{1}{p}} m^{\frac{1}{q}} \min_{x \notin E^m(C)} \max_{x' \in E^m(C)} \max_{i \in N_m} \frac{C_i(x - x')}{\|x - x'\|_1}.$$

References

[1] I. Sergienko and V. Shilo, *Discrete Optimization Problems. Problems, methods, research*, Kiev: Naukova dumka, 2003.

[2] V. Emelichev, V. Kotov, K. Kuzmin, T. Lebedeva, N. Semenova, and T. Sergienko, "Stability in the combinatorial vector optimization problems," *Journal Automation and Information Sciences*, vol. 26, no. 2, pp. 27–41, 2004.

[3] V. Emelichev and K. Kuzmin, "On the strong stability of a vector problem of threshold Boolean functions minimization," *Informatics*, vol. 1, pp. 16–27, 2008.

[4] Y. Nikulin, O. Karelkina, and M. Mäkelä, "On accuracy, robustness and tolerances in vector Boolean optimization," *European Journal of Operational Research*, vol. 224, pp. 449–457, 2013.

[5] Y. Nikulin, "Stability and accuracy functions in a coalition game with bans, linear payoffs and antagonistic strategies," *Annals of Operations Research*, vol. 172, pp. 25–35, 2009.

[6] V. Emelichev, O. Karelkina, and K. Kuzmin, "Qualitative stability analysis of multicriteria combinatorial minimin problems," *Control and Cybernetics*, vol. 41, no. 1, pp. 57–79, 2012.

[7] E. Gurevsky, O. Battaia, and A. Dolgui, "Balancing of simple assembly lines under variations of task processing times," *Annals of Operation Research*, vol. 201, pp. 265–286, 2012.

[8] T. Lebedeva, N. Semenova, and T. Sergienko, "Stability of vector problems of integer optimization: relationship with the stability of sets of optimal and nonoptimal solutions," *Cybernetics and Systems Analysis*, vol. 41, no. 4, pp. 551–558, 2005.

[9] T. Lebedeva and T. Sergienko, "Different types of stability of vector integer optimization problem: general approach," *Cybernetics and Systems Analysis*, vol. 44, no. 3, pp. 429–433, 2008.

[10] M. Libura, E. Van der Poort, G. Sierksma, and J. Van der Veen, "Stability aspects of the traveling salesman problem based on k-best solutions," *Discrete Applied Mathematics*, vol. 87, no. 1–3, pp. 159–185, 1998.

[11] Y. Sotskov, N. Sotskova, T. Lai, and F. Werner, *Scheduling under uncertainty: theory and algorithms*, Minsk: Belaruskaya nauka,

2010.

[12] E. Gordeev, "Comparison of three approaches to studying stability of solutions to problems of discrete optimization and computational geometry," *Journal of Applied and Industrial Mathematics*, vol. 9, no. 3, pp. 358–366, 2015.

[13] E. Gordeev, "Comparison of three approaches to studying the stability of solutions to discrete optimization and computational geometry problems," *Discrete Analysis and Operations Research*, vol. 22, no. 3, pp. 18–35, 2015. (in Russian)

[14] V. Emelichev and Y. Nikulin, "Numerical measure of strong stability and strong quasistability in the vector problem of integer linear programming," *Computer Science Journal of Moldova*, vol. 7, no. 1(19), pp. 105–117, 1999.

[15] V. Emelichev and D. Podkopaev, "Stability and regularization of vector problems of integer linear programming," *Discrete Analysis and Operations Research. Ser. 2*, vol. 8, no. 1, pp. 47–69, 2001.

[16] V. Emelichev, E. Girlich, Y. Nikulin, and D. Podkopaev, "Stability and regularization of vector problem of integer linear programming," *Optimization*, vol. 51, no. 4, pp. 645–676, 2002.

[17] V. Emelichev, Y. Nikulin, and K. Kuzmin, "Stability analysis of the Pareto optimal solution for some vector Boolean optimization problem," *Optimization*, vol. 54, no. 4. pp. 545–561, 2005.

[18] V. Emelichev and K. Kuzmin, "Estimating the stability radius of the vector max-cut problem," *Discrete Mathematics and Applications*, vol. 23, no. 1, pp. 145–152, 2013.

[19] V. Emelichev and K. Kuzmin, "Stability in multicriteria max-cut problem," *Transaction of National Academy of Sciences of Azerbaijan. Physics and Mathematics Sciences*, vol. 33, no. 6, pp. 90–99, 2013.

[20] V. Emelichev and D. Podkopaev, "Quantitative stability analysis for vector problems of 0-1 programming," *Discrete Optimization*, vol. 7, no. 1–2, pp. 48–63, 2010.

[21] V. Emelichev and K. Kuzmin, "On the T1-stability radius of a multicriteria linear Boolean problem with Hölder's norms in parameter spaces," *Taurida Journal of Computer Science Theory*

and Mathematics, vol. 30, no. 1, pp. 49–64, 2016.

[22] V. Emelichev, Y. Nikulin, and V. Korotkov, "Stability analysis of efficient portfolios in a discrete variant of multicriteria investment problem with Savage's risk criteria," *Computer Science Journal of Moldova*, vol. 25, no.3(75), pp. 303–328, 2017.

[23] V. Emelichev and E. Gurevsky, "On quasi-stability of the vector Boolean problem of minimizing absolute deviations of linear functions from zero," *Computer Science Journal of Moldova*, vol. 14, no.2(41), 2006.

[24] V. Emelichev and E. Gurevsky, "On stability of a Pareto-optimal solution under perturbations of the parameters for a multicriteria combinatorial partition problem," *Computer Science Journal of Moldova*, vol. 16, no.2(47), 2008.

[25] V. Emelichev, E. Gurevsky, and A. Platonov, "Measure of stability for a finite cooperative game with a generalized concept of equilibrium," *Buletinul Academiei de Stiinte a Republicii Moldova. Matematica*, vol. 52, pp. 17–26, 2006.

[26] N. Chakravarti and A. Wagelmans, "Calculation of stability radius for combinatorial optimization problems," *Operations Research Letters*, vol. 23, no. 1, pp. 1–7, 1998.

[27] S. Van Hoesel and A. Wagelmans, "On the complexity of postoptimality analysis of 0-1 programs," *Discrete Applied Mathematics*, vol. 91, no. 1-3, pp. 251–263, 1999.

[28] J. Roland, Y. De Smet, and J. Rui Figuera, "On the calculation of stability radius for multi-objective combinatorial optimization problems by inverse optimization," *4OR-Q J Oper Res.*, vol. 10, no. 4, pp. 379–389, 2012.

[29] V. Leontev and E. Gordeev, "Qualitative analysis of trajectory problems," *Kibernetika*, no. 5, pp. 82–90, 1986.

[30] V. Leontiev, "Stability in linear discrete problems," *Problems of Cybernetics*, vol. 35, pp. 169–184, 1979.

[31] V. Pareto, *Manuel D'economie Politique*, Paris: V. Giard & E. Briere, 1909.

[32] K. Miettinen, *Nonlinear Multiobjective Optimization*, Boston: Kluwer, 1999.

[33] V. Podinovskii and V. Noghin, *Pareto-Optimal Solutions of Multicriteria Problems*, Moscow: Fizmatlit., 2007.

[34] V. Emelichev and Y. Nikulin, "Strong stability measures for multicriteria quadratic integer programming problem of finding extremum solutions," *Computer Science Journal of Moldova*, vol. 26, no.2(77), pp. 115–125, 2018.

[35] V. Emelichev and K. Kuzmin, "A general approach to studing the stability of a Pareto optimal solution of a vector integer linear programming problem," *Discrete Math. Appl.*, vol. 17, no. 4, pp. 340–354, 2007.

[36] S. Bukhtoyarov and V. Emelichev, "Aspects of stability for multicriteria integer linear programming problem," *Discrete Analysis and Operations Research*, vol. 26, no. 1. pp. 5–19, 2019.

Vladimir Emelichev, Yury Nikulin Received August 21, 2020

Vladimir Emelichev
Belarusian State University
Nezavisimosti 4, 220030 Minsk, Belarus
E–mail: vemelichev@gmail.com

Yury Nikulin (corresponding author)
University of Turku
Vesilinnantie 5, 20014 Turku, Finland
E–mail: yury.nikulin@utu.fi

Computer Science Journal of Moldova, vol.**28**, no.3(84), 2020

Analysis of Faults in Cyber-Physical Systems by Finite Discrete-Time Markov Chains

Volodymyr G. Skobelev, Volodymyr V. Skobelev

Abstract

In the given paper the problem of Cyber-Physical Systems behavior analysis in the occurrence of faults is investigated. To present the fault-free behavior of the investigated Cyber-Physical System as well as its behaviors in the presence of admissible faults, some Finite Discrete-Time Markov Chain is proposed and analyzed. It is shown that a single stationary probability distribution exists for the proposed model. This distribution is applied for characterization the behaviors of the investigated Cyber-Physical System in terms of the distance from the current fault to the set of critical faults. Besides, the algorithm for bounded probabilistic reachability analysis of the target set of faults is proposed.

Keywords: Cyber-Physical Systems, faults, Discrete-Time Markov Chains, bounded probabilistic reachability, probabilistic counterexamples.

1 Introduction

In the design and analysis of Cyber-Physical Systems (CPS), hybrid automata [1, 2, 3] are usually used as mathematical models for which automated verification procedures are performed by using these or the other Model Checkers.

The need to analyze some significant attributes, such as reliability, safety, and behaviors associated with the decrease in the performance has lead to the necessity to use stochastic models.

A stochastic hybrid automaton [4, 5] generalizes a deterministic hybrid automaton by assigning probabilities to the arcs of the graph as well as by allowing the values of continuous variables to be reset according to given continuous probability distributions. These generalizations result in the high complexity of both the model itself and its analysis methods. Besides, only a small amount of information presented in a stochastic hybrid automaton can be required for solving a sufficiently wide variety of specific problems of CPS analysis.

An important non-trivial sub-class of stochastic hybrid automata consists of probabilistic automata [6, 7], i.e. the generalization of deterministic hybrid automata only by assigning probabilities to the arcs of the graph. This model is much more simple than a stochastic hybrid automaton. Nevertheless, it has been shown in [8] that probabilistic automata can be successfully used for modeling the behavior of the analyzed CPS in the presence of faults.

When the decrease in the performance of the given CPS is analyzed, it is important to estimate the probability of the occurrences for the behaviors in the presence of these or the other faults.

In the given paper, for solving this problem the Finite Discrete-Time Markov Chain (FDTMC) associated with the analyzed CPS is proposed and investigated. The essential advantage of this model is that it is much simpler than a probabilistic automaton. Besides, it gives the possibility to generate probabilistic counterexamples [9], i.e. the sets of finite paths with a critical probability mass.

2 Proposed model

A fault in the analyzed CPS $\mathfrak{S}$ is called admissible if at its occurrence the CPS $\mathfrak{S}$ can continue to operate, possibly with some loss of its functionality.

Let $F_{\mathfrak{S}}^{(0)}$ ($|F_{\mathfrak{S}}^{(0)}| \geq 2$) be some finite set of all admissible single faults that can occur in the analyzed CPS $\mathfrak{S}$. As it is usually accepted, the set of all admissible faults that can occur in the analyzed CPS $\mathfrak{S}$ can be defined as some subset $F_{\mathfrak{S}}$ ($F_{\mathfrak{S}}^{(0)} \subset F_{\mathfrak{S}} \subseteq \mathcal{B}(F_{\mathfrak{S}}^{(0)}) \setminus \{\emptyset\}$). Everywhere

further it is assumed that the set $F_{\mathfrak{S}}$ satisfies to the following condition:

$$(\forall f \in F_{\mathfrak{S}})(\forall f' \in \mathcal{B}(F_{\mathfrak{S}}^{(0)}) \setminus \{\emptyset\})(f' \subset f \Rightarrow f' \in F). \tag{1}$$

Let $S_{\mathfrak{S}} = \{\emptyset\} \cup F_{\mathfrak{S}}$. Everywhere further it is supposed that the elements of the set $S_{\mathfrak{S}}$ are enumerated according to their cardinality non-decreasing, i.e.,

$$S_{\mathfrak{S}} = \{s_1, \ldots, s_k\},$$

where

$$|s_1| \leq |s_2| \leq \cdots \leq |s_k|.$$

The elements of the set $S_{\mathfrak{S}}$ can be interpreted as follows: the element $s_1 = \emptyset$ is associated with the fault-free CPS $\mathfrak{S}$, while any element s_i $(i = 2, \ldots, k)$ is associated with the CPS $\mathfrak{S}$ in the presence of the $|s_i|$-multiple fault s_i.

We define the set $S_{\mathfrak{S}}^{crtcl}$ of critical faults in the CPS $\mathfrak{S}$ as the set of all maximal due to the inclusion relation elements of the set $S_{\mathfrak{S}}$, i.e.

$$S_{\mathfrak{S}}^{crtcl} = \{s \in S_{\mathfrak{S}} | (\forall s' \in S_{\mathfrak{S}})(s \not\subset s')\}. \tag{2}$$

We can associate with the analyzed CPS $\mathfrak{S}$ some FDTMC $\mathfrak{M}_{\mathfrak{S}}$ defined by the stochastic $(k \times k)$-matrix

$$P_{\mathfrak{S}} = \begin{array}{c|cccc} & s_1 & s_2 & \ldots & s_k \\ \hline s_1 & p_{11} & p_{12} & \cdots & p_{1k} \\ s_2 & p_{21} & p_{22} & \cdots & p_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s_k & p_{k1} & p_{k2} & \cdots & p_{kk} \end{array}$$

such that the following two formulae are true:

$$(\forall i = 1, \ldots, k)(0 < p_{ii} < 1), \tag{3}$$

$$(\forall i = 1, \ldots, k)(\forall j \in \{1, \ldots, k\} \setminus \{i\})(p_{ij} > 0 \Leftrightarrow$$

$$\Leftrightarrow s_i \subset s_j \& |s_j| = |s_i| + 1 \vee s_j \subset s_i \& |s_i| = |s_j| + 1). \tag{4}$$

Due to (1)-(4):

1. The analyzed CPS $\mathfrak{S}$ can operate with positive probability either it is fault-free or if any admissible fault s_i $(i = 2, \ldots, k)$ occurs.

2. If $\mathfrak{S}$ is fault-free, then with positive probability only any single fault can occur.

3. If $\mathfrak{S}$ operates in the presence of any critical fault $s \in S_{\mathfrak{S}}^{crtcl}$, then no additional faults can occur, and with positive probability only any single fault $f \in s$ can be repaired.

4. If $\mathfrak{S}$ operates in the presence of any fault $s \in S_{\mathfrak{S}} \backslash (\{\emptyset\} \cup S_{\mathfrak{S}}^{crtcl})$, then with positive probability only either any single fault $f \in s$ can be repaired or any single fault $f \in F_{\mathfrak{S}}^{(0)} \backslash s$ can occur.

Example 1. Let the analyzed CPS $\mathfrak{S}_1$ be the water tanks system considered in [2]. It consists of two tanks, namely the left tank and the right tank. Both tanks are leaking at a constant rate. Water is added to the system at a constant rate through a hose. At any point in time the hose is dedicated either to the left tank or to the right tank. The hose can switch between the tanks instantaneously.

Let $F_{\mathfrak{S}_1}^{(0)} = \{f_1, f_2\}$, where f_1 means that the water-level sensor in the left tank is faulty, and f_2 means that the water-level sensor in the right tank is faulty.

We set $F_{\mathfrak{S}_1} = \{\{f_1\}, \{f_2\}, \{f_1, f_2\}\}$. Therefore,

$$S_{\mathfrak{S}_1} = \{s_1, s_2, s_3, s_4\},$$

where $s_1 = \emptyset$, $s_2 = \{f_1\}$, $s_3 = \{f_2\}$, $s_4 = \{f_1, f_2\}$.

With the CPS $\mathfrak{S}_1$ we can associate some FDTMC $\mathfrak{M}_{\mathfrak{S}_1}$, defined by the following stochastic (4×4)-matrix

$$P_{\mathfrak{S}_1} = \begin{array}{c|cccc} & s_1 & s_2 & s_3 & s_4 \\ \hline s_1 & p_1 & a_1p_1 & a_1p_1 & 0 \\ s_2 & a_2p_2 & p_2 & 0 & a_3p_2 \\ s_3 & a_2p_2 & 0 & p_2 & a_3p_2 \\ s_4 & 0 & a_4p_3 & a_4p_3 & p_3 \end{array},$$

where $0 < p_i < 1$ $(i = 1, 2, 3)$. The digraph of the FDTMC $\mathfrak{M}_{\mathfrak{S}_1}$ is shown in Figure 1.

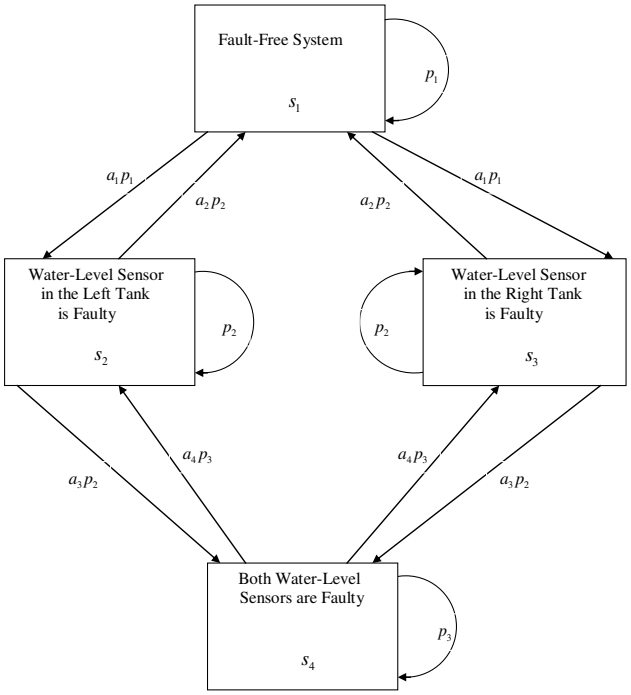


Figure 1. The digraph of the FDTMC $\mathfrak{M}_{\mathfrak{S}_1}$

Since $P_{\mathfrak{S}_1}$ is a stochastic matrix, we get

$$p_1 + a_1p_1 + a_1p_1 = 1 \Rightarrow a_1 = 0.5(p_1^{-1} - 1),$$
$$a_2p_2 + p_2 + a_3p_2 = 1 \Rightarrow a_2 + a_3 = p_2^{-1} - 1,$$
$$a_4p_3 + a_4p_3 + p_3 = 1 \Rightarrow a_4 = 0.5(p_3^{-1} - 1).$$

□

We denote $G_{\mathfrak{M}_{\mathfrak{S}}}$ the digraph associated with the FDTMC $\mathfrak{M}_{\mathfrak{S}}$.

Theorem 1. *The FDTMC $\mathfrak{M}_{\mathfrak{S}}$ is aperiodic and irreducible.*

Proof. Due to (3), $p_{ii} > 0$ for all $i = 1, \ldots, k$. Therefore, each state s_i $(i = 1, \ldots, k)$ of the FDTMC $\mathfrak{M}_{\mathfrak{S}}$ is aperiodic, i.e. $\mathfrak{M}_{\mathfrak{S}}$ is an aperiodic FDTMC.

Due to (4), in the digraph $G_{\mathfrak{M}_{\mathfrak{S}}}$ of the FDTMC $\mathfrak{M}_{\mathfrak{S}}$ for the state $s_1 = \emptyset$ and any state $s = \{f_1, \ldots, f_r\} \in S_{\mathfrak{S}} \backslash \{s_1\}$ there are pathes

$$s_1 = \emptyset \to \{f_1\} \to \{f_1, f_2\} \to \cdots \to s = \{f_1, \ldots, f_r\}$$

and

$$s = \{f_1, \ldots, f_r\} \to \{f_1, \ldots, f_{r-1}\} \to \cdots \to \{f_1\} \to s_1 = \emptyset,$$

where each transition occurs with positive probability, i.e. the state $s_1 = \emptyset$ communicates with any state $s = \{f_1, \ldots, f_r\} \in S_{\mathfrak{S}} \backslash \{s_1\}$.

This factor implies that the digraph $G_{\mathfrak{M}_{\mathfrak{S}}}$ is strongly connected, i.e. in the FDTMC $\mathfrak{M}_{\mathfrak{S}}$ the set of states $S_{\mathfrak{S}}$ forms the single communicating class.

Therefore, $\mathfrak{M}_{\mathfrak{S}}$ is an irreducible FDTMC.

□

Theorem 1 implies that for the FDTMC $\mathfrak{M}_{\mathfrak{S}}$ there exists exactly one stationary distribution

$$\overrightarrow{\psi} = (\psi_{s_1}, \ldots, \psi_{s_k}),$$

where the component ψ_{s_i} $(i = 1, \ldots, k)$ is the long-term proportion of transitions that the FDTMC $\mathfrak{M}_{\mathfrak{S}}$ makes into the state s_i, i.e. the stationary probability for the FDTMC $\mathfrak{M}_{\mathfrak{S}}$ to transit to the state s_i.

The vector $\overrightarrow{\psi}$ can be computed as the solution of the system of equations

$$\begin{cases} \overrightarrow{\psi} P_{\mathfrak{S}} = \overrightarrow{\psi} \\ \sum\limits_{i=1}^{k} \psi_{s_i} = 1 \end{cases}. \tag{5}$$

It is worth to note that each component ψ_{s_i} $(i = 1, \ldots, k)$ of the vector $\overrightarrow{\psi}$ is a strictly positive number.

Example 2. Solving the system of equations (5) for the CPS $\mathfrak{S}_1$ (see Example 1), we compute the stationary distribution

$$\overrightarrow{\psi} = (\psi_{s_1}, \psi_{s_2}, \psi_{s_3}, \psi_{s_4}),$$

where:

$$\psi_{s_1} = \frac{a_2 p_2}{1-p_1}\left(\frac{a_2 p_2}{1-p_1} + \frac{a_3 p_2}{1-p_3} + 1\right)^{-1},$$

$$\psi_{s_2} = \psi_{s_3} = 0.5\left(\frac{a_2 p_2}{1-p_1} + \frac{a_3 p_2}{1-p_3} + 1\right)^{-1},$$

$$\psi_{s_4} = \frac{a_3 p_2}{1-p_3}\left(\frac{a_2 p_2}{1-p_1} + \frac{a_3 p_2}{1-p_3} + 1\right)^{-1}.$$

□

Dealing with the digraph $G_{\mathfrak{M}_{\mathfrak{S}}}$ of the FDTMC $\mathfrak{M}_{\mathfrak{S}}$, we can define the pair-wise disjoint subsets $S_{\mathfrak{S}}^{(i)}$ $(i = 0, 1, \dots)$ of the set $S_{\mathfrak{S}}$, such that

$$S_{\mathfrak{S}}^{(i)} = \{s \in S_{\mathfrak{S}} \,|\, \text{Distance}(s, S_{\mathfrak{S}}^{crtcl}) = i\} \ (i = 0, 1, \dots).$$

Therefore, $S_{\mathfrak{S}}^{(0)} = S_{\mathfrak{S}}^{crtcl}$ and the set $S_{\mathfrak{S}}^{(i)}$ $(i = 1, 2, \dots)$ consists of all states such that i is the least number for sequential occurring of additional single faults that lead the CPS $\mathfrak{S}$ to operate in the presence of critical faults.

The stationary probability $\mathbb{P}_{st}(S_{\mathfrak{S}}^{(i)})$ $(i = 0, 1, \dots)$ that the CPS $\mathfrak{S}$ operates in some state that is an element of the set $S_{\mathfrak{S}}^{(i)}$ can be computed as follows:

$$\mathbb{P}_{st}(S_{\mathfrak{S}}^{(i)}) = \sum_{s \in S_{\mathfrak{S}}^{(i)}} \psi_s \ (i = 0, 1, \dots).$$

In particular, the stationary probability that the CPS $\mathfrak{S}$ operates in the presence of some critical fault can be computed as follows:

$$\mathbb{P}_{st}(S_{\mathfrak{S}}^{crtcl}) = \mathbb{P}_{st}(S_{\mathfrak{S}}^{(0)}) = \sum_{s \in S_{\mathfrak{S}}^{crtcl}} \psi_s.$$

The probability distribution

$$\mathbb{P}_{st}(S^{(0)}_{\mathfrak{S}}), \mathbb{P}_{st}(S^{(1)}_{\mathfrak{S}}), \ldots . \tag{6}$$

characterizes situations in which the analyzed CPS $\mathfrak{S}$ operates in the presence of some fault that is an element of the set of faults determined by its "distance" to the set of critical faults.

It should be noted that an essential characteristic of the CPS $\mathfrak{S}$ behavior is the study of the changes of the probability distribution (6) under variations of the probabilities p_{ij} $(i, j = 1, \ldots, k)$.

Example 3. For the CPS $\mathfrak{S}_1$ (see Examples 1 and 2) we get

$$S^{(0)}_{\mathfrak{S}_1} = S^{crtcl}_{\mathfrak{S}_1} = \{s_4\},$$

$$S^{(1)}_{\mathfrak{S}_1} = \{s_2, s_3\},$$

and

$$S^{(2)}_{\mathfrak{S}_1} = \{s_0\}.$$

Therefore,

$$\mathbb{P}_{st}(S^{(0)}_{\mathfrak{S}_1}) = \psi_{s_4} = \frac{a_3 p_2}{1-p_3}\left(\frac{a_2 p_2}{1-p_1} + \frac{a_3 p_2}{1-p_3} + 1\right)^{-1},$$

$$\mathbb{P}_{st}(S^{(1)}_{\mathfrak{S}_1}) = \psi_{s_2} + \psi_{s_3} = \left(\frac{a_2 p_2}{1-p_1} + \frac{a_3 p_2}{1-p_3} + 1\right)^{-1},$$

$$\mathbb{P}_{st}(S^{(2)}_{\mathfrak{S}_1}) = \psi_{s_1} = \frac{a_2 p_2}{1-p_1}\left(\frac{a_2 p_2}{1-p_1} + \frac{a_3 p_2}{1-p_3} + 1\right)^{-1}.$$

□

3 Bounded probabilistic analysis of the CPS $\mathfrak{S}$

For the analyzed CPS $\mathfrak{S}$ estimation the occurrences of behaviours associated with the decreasing in the performance in the presence of faults can be reduced to computation the probability $\mathbb{P}_{st}(s_1, S^{trgt}_{\mathfrak{S}})$ to reach

this or the other target set $S_{\mathfrak{S}}^{trgt}$ ($\emptyset \neq S_{\mathfrak{S}}^{trgt} \subset S_{\mathfrak{S}}$) of states starting in the state s_1, as follows.

Let $\Pi_{s_1,S_{\mathfrak{S}}^{trgt}}$ be the set of all strings $\pi = s_{i_0}s_{i_1}\dots s_{i_r} \in S_{\mathfrak{S}}^{+}$ such that

$$s_{i_0} = s_1,$$

$$s_{i_j} \notin S_{\mathfrak{S}}^{trgt} \ (j = 0, 1, \dots, r-1),$$

$$s_{i_r} \in S_{\mathfrak{S}}^{trgt},$$

and

$$p_{i_j i_{j+1}} > 0 \ (j = 0, 1, \dots, r-1).$$

Due to [10, 11], with a string $\pi = s_{i_0}s_{i_1}\dots s_{i_r}$ it is associated the probability

$$\mathbb{P}(\pi) = \mathbb{P}(\mathrm{Cyl}(\pi)) = \prod_{j=0}^{r-1} p_{i_j i_{j+1}}.$$

Therefore,

$$\mathbb{P}(s_1, S_{\mathfrak{S}}^{trgt}) = \sum_{\pi \in \Pi_{s_1,S_{\mathfrak{S}}^{trgt}}} \mathbb{P}(\pi).$$

For real CPS, it is computationally infeasible to deal with the infinite set $\Pi_{s_1,S_{\mathfrak{S}}^{trgt}}$. Instead, bounded reachability properties [9, 12] can be analysed as follows.

For the given positive number λ ($\lambda < 1,$) and positive integer h we denote $\mathcal{P}(s_1, S_{\mathfrak{S}}^{trgt}, \lambda, h)$ the property that for the analysed CPS $\mathfrak{S}$ the probability to reach a state in $S_{\mathfrak{S}}^{trgt}$ starting in the state s_1 by at most h steps is not greater then λ.

Let

$$\Pi^{(l)}_{s_1,S_{\mathfrak{S}}^{trgt}} = \Pi_{s_1,S_{\mathfrak{S}}^{trgt}} \cap S_{\mathfrak{S}}^{l} \ (l = 2, \dots, h+1).$$

It is evident that the sets $\Pi^{(l)}_{s_1,S_{\mathfrak{S}}^{trgt}}$ ($l = 2, \dots, h+1$) can be computed sufficiently easily than the set $\Pi_{s_1,S_{\mathfrak{S}}^{trgt}}$.

Example 4. Let us consider the CPS $\mathfrak{S}_1$ (see Example 1). Setting

$$S^{trgt}_{\mathfrak{S}_1} = S^{crtcl}_{\mathfrak{S}_1} = \{s_4\},$$

we get

$$\Pi^{(2)}_{s_1, S^{crtcl}_{\mathfrak{S}_1}} = \emptyset,$$

$$\Pi^{(3)}_{s_1, S^{crtcl}_{\mathfrak{S}_1}} = \{s_1 s_2 s_4, s_1 s_3 s_4\},$$

$$\Pi^{(4)}_{s_1, S^{crtcl}_{\mathfrak{S}_1}} = \{s_1^2 s_2 s_4, s_1^2 s_3 s_4, s_1 s_2^2 s_4, s_1 s_3^2 s_4\},$$

$$\Pi^{(5)}_{s_1, S^{crtcl}_{\mathfrak{S}_1}} = \{s_1^3 s_2 s_4, s_1 s_2 s_1 s_2 s_4, s_1^2 s_2^2 s_4, s_1^3 s_3 s_4,$$

$$s_1 s_3 s_1 s_3 s_4, s_1^2 s_3^2 s_4, s_1 s_2 s_1 s_3 s_4, s_1 s_3 s_1 s_2 s_4\}$$

and so on.

□

It is evident, that the analyzed CPS $\mathfrak{S}$ satisfies to the property $\mathcal{P}(s_1, S^{trgt}_{\mathfrak{S}}, \lambda, h)$ if and only if the following inequality holds:

$$\sum_{l=2}^{h+1} \sum_{\pi \in \Pi^{(l)}_{s_1, S^{trgt}_{\mathfrak{S}}}} \mathbb{P}(\pi) \leq \lambda.$$

Therefore, the property $\mathcal{P}(s_1, S^{trgt}_{\mathfrak{S}}, \lambda, h)$ fails for the analyzed CPS $\mathfrak{S}$ if and only if for some subset

$$\mathcal{C} \subseteq \bigcup_{l=2}^{h+1} \Pi^{(l)}_{s_1, S^{trgt}_{\mathfrak{S}}}$$

the following inequality holds:

$$\mathbb{P}(\mathcal{C}) = \sum_{\pi \in \mathcal{C}} \mathbb{P}(\pi) > \lambda.$$

This subset $\mathcal{C}$ is called a counterexample.

An attempt to design a counterexample can be used to reduce computations in the process of checking the property $\mathcal{P}(s_1, S_{\mathfrak{S}}^{trgt}, \lambda, h)$ for the analyzed CPS $\mathfrak{S}$.

Indeed, let us suppose that the elements of any non-empty set of strings $\Pi^{(l)}_{s_1, S_{\mathfrak{S}}^{trgt}}$ $(l = 2, \ldots, h+1)$ are enumerated according to their probabilities non-increasing. Then the following algorithm can be applied for checking the property $\mathcal{P}(s_1, S_{\mathfrak{S}}^{trgt}, \lambda, h)$ for the analyzed CPS $\mathfrak{S}$.

Algorithm 1. (Checking the property $\mathcal{P}(s_1, S_{\mathfrak{S}}^{trgt}, \lambda, h)$).

Step 1. $l := 2$, $\mathcal{C} := \emptyset$, $\mathbb{P}(\mathcal{C}) := 0$.

Step 2. If $\Pi^{(l)}_{s_1, S_{\mathfrak{S}}^{trgt}} \neq \emptyset$, then go to Step 3, else go to Step 6.

Step 3. Select the first element $\pi \in \Pi^{(l)}_{s_1, S_{\mathfrak{S}}^{trgt}}$, $\mathcal{C} := \mathcal{C} \cup \{\pi\}$, $\Pi^{(l)}_{s_1, S_{\mathfrak{S}}^{trgt}} := \Pi^{(l)}_{s_1, S_{\mathfrak{S}}^{trgt}} \backslash \{\pi\}$, $\mathbb{P}(\mathcal{C}) := \mathbb{P}(\mathcal{C}) + \mathbb{P}(\pi)$.

Step 4. If $\mathbb{P}(\mathcal{C}) > \lambda$, then go to Step 5, else go to Step 2.

Step 5. Print "For the system $\mathfrak{S}$ the property $\mathcal{P}(s_1, S_{\mathfrak{S}}^{trgt}, \lambda, h)$ is false", print the designed counterexample $\mathcal{C}$ in the explicit form, and HALT.

Step 6. $l := l + 1$.

Step 7. If $l \leq h + 1$, then go to Step 2, else go to Step 8.

Step 8. Print "For the system $\mathfrak{S}$ the property $\mathcal{P}(s_1, S_{\mathfrak{S}}^{trgt}, \lambda, h)$ is true", and HALT.

Theorem 2. *Algorithm 1 is complete and sound.*

Proof. Due to Steps 1-4, 6, and 7, in the process of sequential generation of the sets of strings

$$\Pi^{(2)}_{s_1, S_{trgt}}, \Pi^{(3)}_{s_1, S_{trgt}}, \ldots$$

the validity of the properties

$$\mathcal{P}(s_1, S_{\mathfrak{S}}^{trgt}, \lambda, 1), \mathcal{P}(s_1, S_{\mathfrak{S}}^{trgt}, \lambda, 2), \ldots$$

are checked sequentially.

Due to Step 5, if for some $l \in \{2, \ldots, h+1\}$ the property $\mathcal{P}(s_1, S_{\mathfrak{S}}^{trgt}, \lambda, l-1)$ is false, then Algorithm 1 prints "For the system $\mathfrak{S}$ the property $\mathcal{P}(s_1, S_{\mathfrak{S}}^{trgt}, \lambda, h)$ is false", prints the designed counterexample $\mathcal{C}$ in the explicit form, and halts.

Due to Step 8, if all properties $\mathcal{P}(s_1, S_{\mathfrak{S}}^{trgt}, \lambda, l-1)$ $(l = 2, \ldots, h+1)$ are true, then Algorithm 1 prints "For the system $\mathfrak{S}$ the property $\mathcal{P}(s_1, S_{\mathfrak{S}}^{trgt}, \lambda, h)$ is true", and halts.

Therefore Algorithm 1 is complete and sound.

□

Example 5. Let us consider the CPS $\mathfrak{S}_1$ (see Example 1). Setting $p_1 = 0.80$, $p_2 = p_3 = 0.40$, and $a_2 = a_3$, we get $a_1 = 0.125$, and $a_2 = a_3 = a_4 = 0.75$. Therefore, the FDTMC $\mathfrak{M}_{\mathfrak{S}_1}$ is defined by the following stochastic (4×4)-matrix

$P_{\mathfrak{S}_1} =$

	s_1	s_2	s_3	s_4
s_1	0.80	0.10	0.10	0
s_2	0.30	0.40	0	0.30
s_3	0.30	0	0.40	0.30
s_4	0	0.30	0.30	0.40

.

Let us check the property $\mathcal{P}(s_1, S_{\mathfrak{S}_1}^{crtcl}, 0.1, 3)$, i.e. $\lambda = 0.1$, $h = 3$, and $S_{\mathfrak{S}_1}^{trgt} = S_{\mathfrak{S}_1}^{crtcl} = \{s_4\}$.

The elements of the sets $\Pi^{(3)}_{s_1, S_{\mathfrak{S}_1}^{crtcl}}$ and $\Pi^{(4)}_{s_1, S_{\mathfrak{S}_1}^{crtcl}}$ (see Example 4) are enumerated according to their probabilities non-increasing. Indeed,

$$\mathbb{P}(s_1 s_2 s_4) = \mathbb{P}(s_1 s_3 s_4) = 0.030,$$

$$\mathbb{P}(s_1^2 s_2 s_4) = \mathbb{P}(s_1^2 s_3 s_4) = 0.024,$$

and

$$\mathbb{P}(s_1 s_2^2 s_4) = \mathbb{P}(s_1 s_3^2 s_4) = 0.012.$$

Applying Algorithm 1, we get:

Step 1. $l := 2$, $\mathcal{C} := \emptyset$, $\mathbb{P}(\mathcal{C}) := 0$.
Step 2. $\Pi^{(2)}_{s_1, S^{trgt}_{\mathfrak{S}_1}} = \emptyset$. We go to Step 6.
Step 6. $l := 2 + 1 = 3$.
Step 7. $3 \leq 3 + 1$. We go to Step 2.
Step 2. $\Pi^{(3)}_{s_1, S^{trgt}_{\mathfrak{S}_1}} \neq \emptyset$. We go to Step 3.
Step 3. $\pi := s_1 s_2 s_4$, $\mathcal{C} := \{s_1 s_2 s_4\}$, $\Pi^{(3)}_{s_1, S^{trgt}_{\mathfrak{S}_1}} := \{s_1 s_3 s_4\}$,
$\mathbb{P}(\mathcal{C}) := \mathbb{P}(s_1 s_2 s_4) = 0.030$.
Step 4. $\mathbb{P}(\mathcal{C}) = 0.030 \leq 0.1$. We go to Step 2.
Step 2. $\Pi^{(3)}_{s_1, S^{trgt}_{\mathfrak{S}_1}} \neq \emptyset$. We go to Step 3.
Step 3. $\pi := s_1 s_3 s_4$, $\mathcal{C} := \{s_1 s_2 s_4, s_1 s_3 s_4\}$, $\Pi^{(3)}_{s_1, S^{trgt}_{\mathfrak{S}_1}} := \emptyset$,
$\mathbb{P}(\mathcal{C}) := \mathbb{P}(\mathcal{C}) + \mathbb{P}(s_1 s_3 s_4) = 0.030 + 0.030 = 0.060$.
Step 4. $\mathbb{P}(\mathcal{C}) = 0.60 \leq 0.1$. We go to Step 2.
Step 2. $\Pi^{(3)}_{s_1, S^{trgt}_{\mathfrak{S}_1}} = \emptyset$. We go to Step 6.
Step 6. $l := 3 + 1 = 4$.
Step 7. $4 \leq 3 + 1$. We go to Step 2.
Step 2. $\Pi^{(4)}_{s_1, S^{trgt}_{\mathfrak{S}_1}} \neq \emptyset$. We go to Step 3.
Step 3. $\pi := s_1^2 s_2 s_4$, $\mathcal{C} := \{s_1 s_2 s_4, s_1 s_3 s_4, s_1^2 s_2 s_4\}$,
$\Pi^{(4)}_{s_1, S^{trgt}_{\mathfrak{S}_1}} := \{s_1^2 s_3 s_4, s_1 s_2^2 s_4, s_1 s_3^2 s_4\}$,
$\mathbb{P}(\mathcal{C}) := \mathbb{P}(\mathcal{C}) + \mathbb{P}(s_1^2 s_2 s_4) = 0.060 + 0.024 = 0.084$.
Step 4. $\mathbb{P}(\mathcal{C}) = 0.084 \leq 0.1$. We go to Step 2.
Step 2. $\Pi^{(4)}_{s_1, S^{trgt}_{\mathfrak{S}_1}} \neq \emptyset$. We go to Step 3.
Step 3. $\pi := s_1^2 s_3 s_4$, $\mathcal{C} := \{s_1 s_2 s_4, s_1 s_3 s_4, s_1^2 s_2 s_4, s_1^2 s_3 s_4\}$,
$\Pi^{(4)}_{s_1, S^{trgt}_{\mathfrak{S}_1}} := \{s_1 s_2^2 s_4, s_1 s_3^2 s_4\}$,
$\mathbb{P}(\mathcal{C}) := \mathbb{P}(\mathcal{C}) + \mathbb{P}(s_1^2 s_3 s_4) = 0.084 + 0.024 = 0.108$.
Step 4. $\mathbb{P}(\mathcal{C}) = 0.108 > 0.1$. We go to Step 5.

Step 5. Print "For the system $\mathfrak{S}_1$ the property $\mathcal{P}(s_1, S^{trgt}_{\mathfrak{S}_1}, 0.1, 3)$ is false", print the designed counterexample

$$\mathcal{C} := \{s_1 s_2 s_4, s_1 s_3 s_4, s_1^2 s_2 s_4, s_1^2 s_3 s_4\},$$

and HALT.

□

It is evident that Algorithm 1 can be easily applied for the study of the violation of the property $\mathcal{P}(s_1, S^{trgt}_{\mathfrak{S}}, \lambda, h)$ under variations of the probabilities p_{ij} $(i, j = 1, \ldots, k)$, as well as under variations of the values of λ and h.

4 Conclusions

The proposed FDTMC is intended for characterization behaviors of the analyzed CPS in the presence of admissible faults. For this FDTMC there exists the single stationary distribution (see Theorem 1). Therefore, probabilities of decreasing in performance of the analyzed CPS in the presence of faults can be estimated. Besides, the proposed FDTMC can be used for the bounded probabilistic analysis of the reachability of the target set of faults for the analyzed CPS (see Algorithm 1).

The essential characteristic of the proposed FDTMC is that it can be used for analysis of the effect of variations of faults probabilities on the probability distribution (6), as well as on the reachability of the target set of faults. Moreover, the proposed FDTMC can be used for symbolic modeling of the analyzed CPS by using these or the other suitable software tools.

References

[1] R. Alur, C. Courcoubetis, N. Halbwachs, T.A. Henzinger, P.H. Ho, X. Nicollin, A. Olivero, J. Sifakis, and S. Yovine, "The Algorithmic analysis of hybrid systems," *Theoretical Computer Science*, vol. 138, no. 1, pp. 3–34, 1995.

[2] J. Lygeros, "Lecture notes on hybrid systems," *Notes for an ENSIETA short course*, vol. 2–6, no. 2, 2004, Available: https://people.eecs.berkeley.edu/ sastry/ee291e/lygeros.pdf

[3] J.-F. Raskin, "An introduction to hybrid automata," in *Handbook of Networked and Embedded Control Systems. Control Engineering*, D. Hristu-Varsakelis and W.S. Levine, Eds. Boston, Basel, Berlin: Birkhäuser, pp. 491–518, 2005.

[4] G.A. Pérez Castañeda, J.-.F Aubry, and N. Brinzei, "Stochastic hybrid automata model for dynamic reliability assessment," *Journal of Risk and Reliability*, vol. 225, no. 1, pp. 28–41, 2011.

[5] M. Franzle, E.M. Hahn, H. Hermanns, N. Wolovick, and L. Zhang, "Measurability and safety verification for stochastic hybrid systems," in *Proc. of The* 14^{th} *International Conference on Hybrid Systems: Conmputation and Control - HSCC'11*, (April 12–14, 2011, Chicago, IL, USA), 2011, pp. 43–52.

[6] J. Sproston, "Decidable Model Checking of Probabilistic Hybrid Automata," in *Formal Techniques in Real-Time and Fault-Tolerant Systems. FTRTFT 2000* (Lecture Notes in Computer Science, vol. 1926), M. Joseph, Ed., 2000, pp. 31–45.

[7] L. Zhang, Z. She, S. Ratschan, H. Hermanns, and E. M. Hahn, "Safety Verification for Probabilistic Hybrid Systems," in *Computer Aided Verification. CAV 2010* (Lecture Notes in Computer Science, vol. 6174), T. Touili, B. Cook, and P. Jackson, Eds., 2010, pp. 196–211.

[8] J. Sproston, "Exact and approximate abstraction for classes of stochastic hybrid systems," *Electronic Communication of the European Association of Software Science and Technology*, vol. 70, 2014, pp. 79–88.

[9] E. Abraham, B. Becker, C. Dehnert, N. Jansen, J.-P. Katoen, and R. Wimmer, "Counterexample generation for Discrete-Time Markov Models: An introductory survey," in *Formal Methods for*

Executable Software Models. SFM 2014 (Lecture Notes in Computer Science, vol. 8483), M. Bernardo, F. Damiani, R. Hähnle, E.B. Johnsen, and I. Schaefer, Eds., 2014, pp. 65–121.

[10] J.G. Kemeney, J.L. Snell, and A. W. Knapp, *Denumerable Markov Chains*, NY, USA: Springer-Verlag, 1976.

[11] M. Kwiatkowska, G. Norman, and D. Parker, "Stochastic model checking," in *Formal Methods for Performance Evaluation. SFM 2007* (Lecture Notes in Computer Science, vol. 4486), M. Bernardo and J. Hillston, Eds., 2007, pp. 220–270.

[12] A. Biere, A. Cimatti, E. M. Clarke, O. Strichman, and Y. Zhu, "Bounded model checking," *Advances in Computers*, vol. 58, 2003, pp. 118–149.

Volodymyr G. Skobelev, Volodymyr V. Skobelev Received May 27, 2020

Volodymyr G. Skobelev
V.M. Glushkov Institute of Cybernetics of NAS of Ukraine
40 Glushkova ave., Kyiv, Ukraine, 03187
Phone: +38 063 431 86 05
E–mail: skobelevvg@gmail.com

Volodymyr V. Skobelev
V.M. Glushkov Institute of Cybernetics of NAS of Ukraine
40 Glushkova ave., Kyiv, Ukraine, 03187
Phone: +38 066 276 85 72
E–mail: volodimirvskobelev@gmail.com

Computer Science Journal of Moldova, vol.**28**, no.3(84), 2020

Loser-out multi metaheuristic framework for multi-objective optimization

Jamshid Tamouk, Nasser Lotfi

Abstract

This paper proposes a multi metaheuristic framework consisting of four multi-objective optimization (MOO) algorithms in which they compete with each other along four phases to be surviving in the next phases. Likewise, it is assumed that number of phases is equal to the number of metaheuristics. The proposed method, named as Loser-Out-Framework (LOF) from this point on, runs in consecutive sessions so that a session starts with dividing global population into several subpopulations. Thereafter in the first phase, entire set of metaheuristics is assigned to each subpopulation and then metaheuristics are performed over subpopulations to modify and improve them. In continuation of each phase, non-dominated solutions extracted by all metaheuristic sets are stored in global archive, and then the most ineffective metaheuristic of each subpopulation is eliminated. The proposed method is evaluated and tested over the well-known DTLZ and WFG benchmarks. Comparative evaluations against several state-of-the-art algorithms exhibits that the proposed framework outperforms others in terms of extracted Pareto front quality.

Keywords: Multi-Objective Optimization, Metaheuristic, Metaheuristic Based Framework.

1 Introduction

Even though metaheuristics have been widely applied for solving NP-hard optimization problems, no individual metaheuristic performing well enough over all optimization problems can be found. Therefore, having an appropriate framework consisting of various metaheuristics became significantly important when high quality solutions for

a vast variety of optimization problems are needed. In this paper, four recently published metaheuristic algorithms, namely NSGA-III, R-NSGA-III, U-NSGA-III and the well-known MOEA/D, are implemented and used in the proposed framework for solving multi-objective optimization (MOO) problems. Four MOO algorithms in the proposed framework compete with each other along four phases to be surviving in the next phases. The proposed method, named as Loser-Out-Framework (LOF) from this point on, runs in consecutive sessions so that a session starts with dividing global population into several subpopulations. Afterwards in the first phase, all metaheuristics are assigned to each subpopulation, and then metaheuristics are performed over subpopulations to improve them. Later on, the non-dominated solutions found by all metaheuristic sets are kept in global archive, and then the weakest metaheuristic is eliminated from each subpopulation. The proposed method is tested over three-objective DTLZ and WFG benchmarks by following the process described for these benchmark problems. Success of the proposed framework on most of the test cases, in comparison to state-of-the-art algorithms puts it in the first position among its well-known competitors. A detailed description of the proposed architecture is given in Section 4. The rest of this paper is organized as follows: next section (Section 2) indicates a brief discussion on state-of-the-art algorithms in metaheuristic-based approaches for MOO problems. In Section 3, MOO metaheuristics used in the proposed framework are discussed and then the proposed framework for MOO is introduced in Section 4. In Section 5 experimental results and their evaluations are given, and finally, the conclusion and feature research directions are presented in Section 6.

2 State-of-the-art in metaheuristic-based approaches for MOO

Pretty rich amount of state-of-the-art methods for solving multi-objective optimization problems can be found in the literature [8]-[12]. Some recently proposed algorithms, namely NSGA-III-WA, NSGA-III, VAEA, RVEA, MOEA/D and MOEA/D-M2M, are selected from state-

of-the-art methods to be used for evaluation of our proposed method. NSGA-III-WA algorithm was proposed by Wang et al. in order to improve standard NSGA-III. Their proposed NSGA-III-WA improves NSGA-III in terms of evolution strategy (using combination of the new differential evolution strategy proposed in this article together with genetic evolution strategy) and weight vector adjustment (to decompose the objective space into several subspaces) [7]. The proposed algorithm speeds up the method without reducing the performance by adding a discriminating condition to the algorithm. Authors tested NSGA-II-WA over DTLZ benchmark set and found that it is effective in terms of convergence and distribution [7]. Generation framework of NSGA-III-WA is presented in Algorithm 2.1, where in the algorithm: N is population size, P_t is the initial population, Q_t is the population result from applying the differential operator on population and W_{unit} is the weight vector.

Input: N structured reference points W_unit, parent population P_t
Output: P_{t+1}

(1) Initialization (P_t , W_unit)
(2) Gen = 1
(3) **While** Gen $\leq$ Gen_max **do**
(4) Qt = Evolutionary strategy (P_t)
(5) Rt = $P_t \cup Q_t$
(6) $P_t + 1$ = Environmental_selection (R_t)
(7) W_unit = Weight_Adjustment (W_unit)
(8) Gen++
(9) **End While**
(10) **Return** P_t+1

Algorithm 2.1. Generation framework of the proposed NSGA-III-WA[7]

The population size of R_t is set to 2N (due to combining two populations P_t and Q_t together with the size of N) and differential op-

erator is taken into account as an evolutionary strategy. Also for selecting the best individuals as a next generation, the environmental selection operation (which includes: normalization of objective values, associate of each individual resulted from normalization with the reference points, Compute the Niche Count of the Reference Point, Niche Preservation Operation) is employed. Kalyanmoy and Himanshu [15] proposed NSGA-III in 2014 which differs from NSGA-II in terms of individual selections. NSGA-III is a robust technique that eliminates the weaknesses of NSGA-II [7],[15], e.g., it solves the issue of lacking uniform diversity. The diversity is maintained by computing crowding distance in NSGA-II, but in NSGA-III it is realized by using reference direction niching. NSGA-III attempts to find Pareto-front optimal solutions that are close to reference points. The algorithm has been tested over different benchmarks and outcomes proved its robustness. Algorithm 2.2 shows the general steps of NSGA-III in search process. More details of NSGA-III can be found in [2], [7], [15].

1. Generate N reference points regarding to the population size.
2. Generate initial population.
3. Repeat until termination criterion is satisfied:
   3.1. Apply related operators to generate new population
   3.2. Combine the populations and apply non-dominated sorting.
   3.3. Assign points to the reference points.
4. END

Algorithm 2.2. NSGA-III steps

VAEA (Vector Angle-Based Evolutionary Algorithm) was proposed by Xiang et al. in 2017 which is based on angle decomposition [14]. VAEA tunes convergence and diversity of search space without needing reference points and it uses maximum-vector-angle-first principle to guarantee the wideness and uniformity of the solution set. It has a similar environmental selection to the NSGA-II and NSGA-III, but with

different niche preservation operation. The authors evaluated VAEA over many-objective benchmarks, and the obtained results illustrated that VAEA tunes convergence and diversity in a good manner. Algorithm 2.3 shows the steps of VAEA framework, where: P is the initial population with the size of N, Q is the population result from applying the Mating_selection and Variation operators on P, S is the union set of P and Q, G is the number of generations, and G_{max} is the maximum number of generations.

```
1: Initialization (P)
2: G = 1
3: While G leq G_max do
4:     P' = Mating_selection(P)
5:     Q = Variation(P')
6:     S = P ∪ Q
7:     P = Environmental_selection(S)
8:     G + +
9: End While
10: Return P
```

Algorithm 2.3. Framework of the VAEA [14]

Mating selection is the operation to select more potential solutions for mating pool (i.e.$P^{'}$) based on the fitness value of each individual. Variation is the operation to generate a set of offspring solutions (i.e. Q) by applying crossover and mutation. Environmental selection is a procedure to select N solutions from union set of P and Q.

Ran, Yaochu, Markus and Bernhard [13] suggested RVEA (Reference Vector Guided Evolutionary Algorithm) in 2016 in which the novelty of this approach is to employ two components, namely, reference vector guided selection and the reference vector adaptation (which is used to dynamically tune the weight vectors in accordance to objective functions) to improve the performance. The authors tested RVEA against 5 state-of-the-art methods and found that RVEA is effective and cost-efficient. Algorithm 2.4 indicates the RVEA frame-

work, where: P_0 is the initial population, P_t is the population resulted after t iterations, Q_t is the population resulted from applying offspring-creation operation on P_t, N is the population size, V_0 is the initial reference vectors, V_t is the reference vectors resulted after t iterations, P_{t+1} is the new population resulted from applying reference-vector-guided-selection operator on the current population, and V_{t+1} is the new reference vectors resulted from applying reference-vector-adaptation operator on the current reference vectors.

1: **Input**: the maximal number of generations t_{max}, a set of unit reference vectors $V_0 = \{v_{0,1}, v_{0,2}, \ldots, v_{0,N}\}$;

2: **Output**: final population $P_{t_{max}}$;
3: /*Initialization*/

4: **Initialization**: create the initial population P_0 with N randomized individuals;
5: /*Main Loop*/
6: **while** t ¡ t_{max} **do**
7: Q_t = offspring-creation(P_t);
8: $P_t = P_t \cup Q_t$;
9: P_{t+1} = reference-vector-guided-selection(t, P_t, V_t);
10: V_{t+1}= reference-vector-adaptation(t, P_{t+1}, V_t, V_0);
11: t = t + 1;
12: **end while**

Algorithm 2.4. RVEA framework

In the algorithm, offspring-creation is the operation used to generate offspring using crossover, mutation and elitism selection.

MOEA/D-M2M was proposed by Hai-Lin, Fangqing and Quingfu [22] in 2014 which is based on the divide and conquer technique. The proposed algorithm divides the Pareto-front and search space into segments and sub-spaces, then each sub-problem regarding to its own segment and sub-space is solved separately. This way the distribution of extracted solutions is increased. Algorithm 2.5 shows the steps of MOEA/D-M2M.

Input :
- MOP (1);
- A stopping criterion;
- K: the number of the subproblems;
- $v^1, \ldots, v^K$: K unit direction vectors;
- S: the size of subpopulation;
- Genetic operators and their associated parameters.

Output: Ψ: a set of nondominated solutions
Initialization: Uniformly randomly choose K $\times$ S points from $[a, b]^n$, compute their F-values and then use them to set $P_1, \ldots, P_K$.
while the stopping criterion is not met **do**
  Generation of New Solutions:
  Set R = ϕ;
  for k $\leftarrow$ 1 to K **do**
    foreach x $\in P_K$ **do**
      Randomly choose y from P_K;
      Apply genetic operators on x and y to generate a new solution z;
      Compute F(z);
      R := R $\cup$ $\{z\}$;
  end
  Q := R $\cup(\cup_{k=1}^{K} P_K)$;
  use Q to set $P_1, \ldots., P_K$; **end**
  Find all the nondominated solutions in $\cup_{k=1}^{K} P_K$ and output them.
End

Algorithm 2.5. MOEA/D-M2M

Quingfu and Hui [3] proposed MOEA/D (Multi-objective Evolutionary Algorithm) in 2007 which is based on decomposition. MOEA/D divides a problem into sub-problems and then optimizes them at the same time. MOEA/D speeds up the optimization process without affecting the quality of results. Meanwhile, it provides a better distribution over the extracted objectives. The MOEA/D became very popular after earning the first place (among the 13 competitors) in CEC2009 contest. Back then, diverse versions of MOEA/D have been proposed by applying different decomposition methods, e.g. MOEA/D-DE [16],

MOEA/D-DRA [18], MOEA/D-XBS [19]. More details of MOEA/D can be found in [3].

3 MOO metaheuristics used within the proposed framework

3.1 NSGA-III (Non-dominated Sorting Genetic Algorithm)

NSGA-III was introduced by Kalyanmoy and Himanshu [15] in 2014 (designed for problems with many objectives) so that its implementation is based on NSGA- and NSGA-II. NSGA-III and NSGA-II are basically similar (in both of them, nondominated sorting is applied to rank the population into a number of fronts), but mostly different in terms of mechanism, e.g., in NSGA-III, the diversity and convergence metrics are enhanced by using a set of reference points and directions for selecting the nondominated solutions for the next generation (i.e., it is the combination of Pareto-based evolutionary multiobjective optimization/EMO algorithm and decomposition). Using a reference direction, it would be possible to start from a reference point and pass over reference direction. Meanwhile in each epoch, a population member is found for each reference direction. Likewise, diversity is maintained by computing crowding distance in NSGA-II, but in NSGA-III it is realized by using reference direction niching. NSGA-III attempts to find Pareto-front optimal solutions that are close to reference points, but in the most of optimization problems optimal Pareto-front is not known.

NSGA-III works robustly against the problems with many objectives (three or more); and, similar to NSGA-II, there is no need for any parameter setting other than population size, termination criteria, crossover, and mutation probabilities. There is a none-algorithmic parameter for number of reference points in which population size is dependent on it (they should be approximately equal). The extracted Pareto-front guarantees a good distribution of points. NSGA-III follows the steps shown in Algorithm 3.1 in search process. More details of NSGA-III can be found in [2], [7], [15].

1. Generate N reference points regarding the population size.
2. Generate initial population.
3. Repeat until termination criterion is satisfied:
   3.1 Apply related operators to generate new population
   3.2. Combine the populations and apply non-dominated sorting.
   3.3. Assign points to the reference points.
4. END

Algorithm 3.1. NSGA-III steps

3.2 R-NSGA-III

R-NSGA-III was proposed by Yash, Kalyanmoy and Julian [4] in 2018 (based on recently proposed NSGA-III and R-NSGA-II method) in order to find a desired part of optimal Pareto-front. To do so, R-NSGA-III defines some reference points and carries out multi-criterion decision-making method. Also, the method uses an epsilon parameter to represent minimum distance between neighbors. When the epsilon value is larger, large amount of solutions around reference points are selected. R-NSGA-III applies clearing based niching instead of crowding distance based niching. In this algorithm the authors extend the R-NSGA-II method to get a more uniform emphasis in finding solutions for all supplied aspiration points. Moreover, they extend the NSGA-III with reference point concept for the same purpose [4] by taking into account a novel reference point generation method based on user-supplied aspiration points (z^K) in which K aspiration points ($z_i^{1,\ldots,K}$) are provided by the user in M-dimensional objective space:

$$r^{(K)} = \left(z_1^{(K)}, z_2^{(K)}, \ldots, z_M^{(k)}\right) \quad k = 1, 2, \ldots, K,$$

where, M is the number of objectives, K is the number of aspiration points, and z^K is the k_{th} aspiration point. R-NSGA-III employs the same genetic operators and survival selection process as NSGA-III. More details of RNSGA-III can be found in [4].

3.3 MOEA/D (decomposition-based multi-objective evolutionary algorithm)

MOEA/D (multi-objective evolutionary algorithm based on decomposition) was proposed by Quingfu and Hui [3] in 2007 which is one of the most popular multi-objective evolutionary algorithms. MOEA/D breaks down the original multi-objective problem into several single-objective sub-problems using decomposition method and then optimizes them concurrently (with regard to optimization, a number of weight vectors with good distribution are often required) by taking into account the neighborhood relationships among sub-problems (for selecting mating parents and population replacement).

Diverse versions of MOEA/D have been proposed by applying different decomposition methods, e.g. MOEA/D-DE [16], MOEA/D-DRA [18], MOEA/D-XBS [19], and MOEA/D-GR [20]. Figure 3.1 represents the flowchart of MOEA/D algorithm [21]. Steps of this algorithm are illustrated in Figure 3.1, where N is the population size and weight vector size $(\lambda_1, \ldots, \lambda_N)$; x is the initial population $(x_1, \ldots, x_n)$, in which x_i is considered as the i_{th} solution; f_i is the i_{th} objective; z_i in formula $Z = (z_1, \ldots, z_m)$ is the best value found so far for objective f_i; λ_i is the weight vector; T is the number of the closest weight vectors to each weight vector (λ_i); and EP is the external population for storing the found NDSs.

Due to optimizing the sub-problems using some neighborhood sub-problems, MOEA/D is not time consuming method. Even though MOEA/D is faster than NSGA-II, the obtained results are similar or even better than NSGA-II. Likewise MOEA/D achieves better distribution for three-objective problems than other methods. Detailed information can be found in [3].

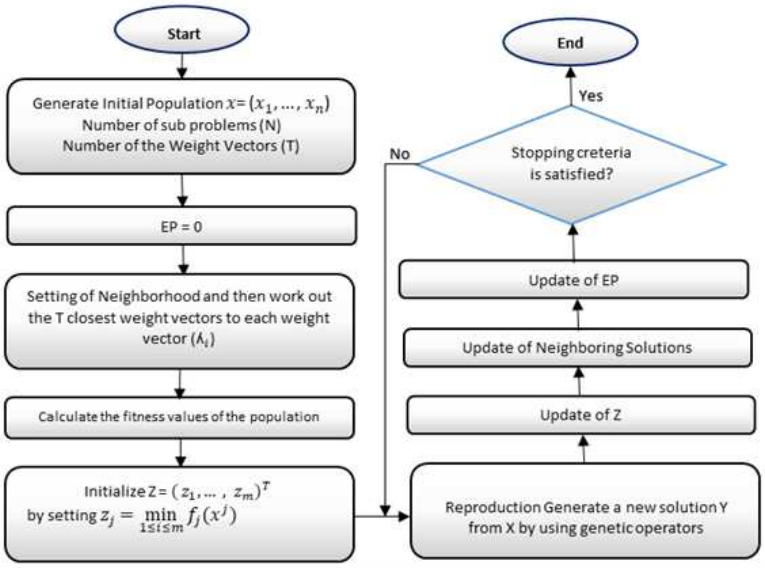


Figure 3.1. MOEA/D Flowchart [21]

3.4 U-NSGA-III

U-NSGA-III is a unified evolutionary optimization algorithm based on NSGA-III. As described in Section 3.1, NSGA-III was proven to work well on many objectives problems. But it has difficulties when applying it to single or two-objective problems; therefore, U-NSGA-III comes up with certain changes in NSGA-III algorithm to overcome this shortage by unifying all three types of optimization problems together (namely, single-objective, multi-objective and many-objective problems). Some changes are:

- Unlike NSGA-III, N (population size) and H (number of reference points) are different parameters with a condition that $N \geq H$ and N is a multiple of four.

- It increases the performance of NSGA-III by replacing tournament pressure (niching-based tournament selection) with random selection.
- In multi-objective optimization, U-NSGA-III has multiple population members for each reference direction, and its non-dominated sorting method is used to divide population into multiple non-dominated fronts. It focuses on non-dominated solutions and solutions closer to reference points.

Input: Two parents: p_1 and p_2
Output: Selected individual, p_s
1: **if** $\pi(p_1) = \pi(p_2)$ **then**
2: **if** p_1.rank ¡ p_2.rank **then**
3: $p_s = p_1$
4: **else** 5: **if** p_2.rank ¡ p_1.rank **then**
6: $p_s = p_2$
7: **else**
8: **if** $d_\perp(p_1) < d_\perp(p_2)$**then**
9: $p_s = p_1$
10: **else**
11: $p_s = p_2$
12: **end if**
13: **end if**
14: **end if**
15: **else**
16: p_s = random Pick (p_1, p_2)
17: **end if**

Algorithm 3.2. Niching based selection

But, similar to NSGA-II and NSGA-III, this algorithm also doesn't need additional parameters than usual optimization algorithms [5], and the rest of process is similar to standard NSGA-III. Algorithm 3.2 indicates the pseudo-code for niching-based selection in U-NSGA-III, where $\pi(s)$ denotes the closest reference point, p_1 and p_2 are the parents

passed to the algorithm as an input, and p_s is the best individual selected so far (i.e., winner). As it is shown in the pseudo-code in Algorithm 3.2, if two solutions under consideration belong to different associated niches (i.e., reference directions), in that case one of them should be selected randomly. Otherwise, the solution that belongs to a better non-dominated rank should be selected. However, if both solutions belong to the same niche and the same non-dominated front, the closest one to the reference direction should be selected.

4 The proposed multi metaheuristic framework for MOO

Loser-Out-Framework (LOF) consists of four different robust metaheuristics cooperating together in collecting Non-Dominated Solutions (NDS) in a global archive and also updating subpopulations with the improved ones during execution. All metaheuristics compete with each other to survive and no to be eliminated. The proposed process starts with dividing the global population to 'p' sub-populations and then, as the first phase, applying entire set of metaheuristics (which contains 'n' number of metaheuristics) to all 'p' sub-populations for fitness evaluations of '$\left(\frac{\alpha}{s-n+i}\right) t^{s-n+i}+\beta$' (which totally should not exceed the maximum number of fitness evaluations allowed for each test instance) times. For example, let's assume that $\alpha = 100, t = 10, s = 5, \beta = 100$, and $i = 0, \ldots, n$. So, the first phase will run for $\left(\frac{100}{5-4-0}\right) t^{5-4-0} + 100 = 1100$ times to find the first loser of each sub-population and remove it from their metaheuristic set. Non-dominated sets, Pareto front, found by all metaheuristics including the loser will be stored in a global archive while the evolved-populations by the loser are discarded (to avoid trapping into local optimum) and this process will continue until the number of survived metaheuristics in each set becomes one. However, the rest of the fitness evaluations (which is a considerably big number) will be assigned to the winner of the competition in each subpopulation. At the end, all Non-Dominated Solutions found by all metaheuristics are combined and then a constant number of non-dominated solutions

is selected as a final result. Thereafter, for comparison, the IGD values are computed. Process and steps of the proposed framework are shown in Figure 4.1.

There is a simple but efficient logic behind having smaller number of fitness evaluations in earlier phases and increasing it in the next phases, that is, one metaheuristic (Loser) is eliminated from metaheuristic set of each sub-population in each step, therefore, in the next phases more effective metaheuristics remain in sets. This is how to give more chances to effective ones and increase the quality of extracted solutions.

Figure 4.2 illustrates an example of LOF approach containing four metaheuristics (m1, m2, m3, and m4) applied to population which has been divided into four sub-populations (sub-pop1, sub-pop2, sub-pop3, and sub-pop4). The following indicates the calculation of fitness evaluations for all four phases under the assumption of total fitness evaluations of 300,000.

$$\left(\frac{\alpha}{s-n+i}\right) t^{s-n+i} + \beta$$

$$\left(\frac{\alpha}{s-n+0}\right) t^{s-n+0} + \left(\frac{\alpha}{s-n+1}\right) t^{s-n+1} + \left(\frac{\alpha}{s-n+2}\right) t^{s-n+2} + \left(\frac{\alpha}{s-n+3}\right) t^{s-n+3} + \beta = 300,000$$

$$\left(\frac{100}{5-4+0}\right) 10^{5-4+0} + \left(\frac{100}{5-4+1}\right) 10^{5-4+1} + \left(\frac{100}{5-4+2}\right) 10^{5-4+2} + \left(\frac{100}{5-4+3}\right) 10^{5-4+3} + \beta = 300,000$$

$$100 \times 10 + 50 \times 100 + 33 \times 1000 + 25 \times 10000 + \beta = 1000 + 5000 + 33000 + 250000 + \beta \approx 300,000$$

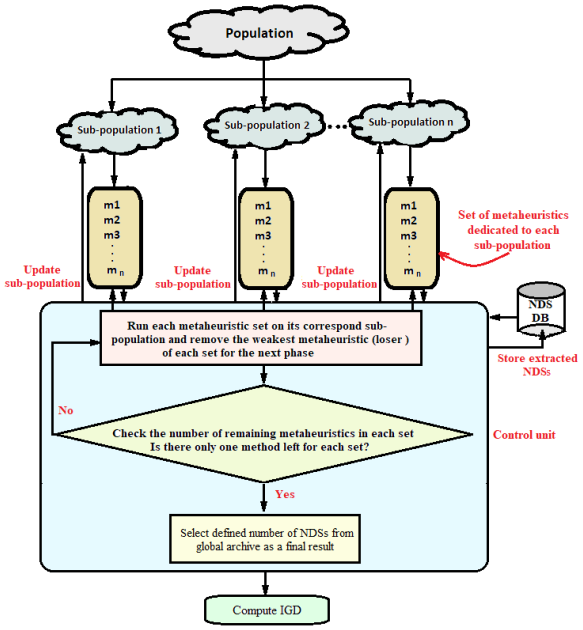


Figure 4.1. Loser-Out-Framework steps and components

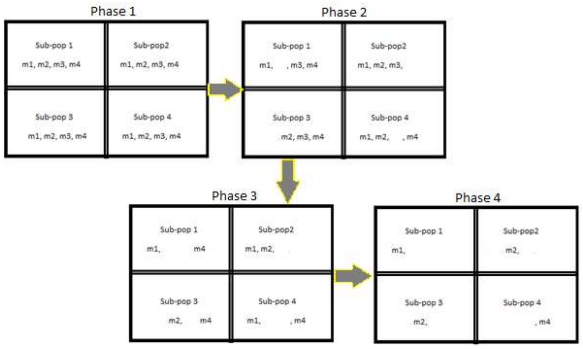


Figure 4.2. Process of dropping the loser metaheuristics of each sub-population

5 Evaluations and Experimental results

The proposed method is evaluated over the well-known three-objectives DTLZ and WFG benchmarks [6], and inverse generational distance (IGD) values are computed over 30 independent runs of each problem. Thereafter, the proposed framework is compared to six state-of-the-art algorithms [7]. For each of the test functions, the number of fitness evaluations is set according to values reported in references [7].

Table 5.1 illustrates the rank of each metaheuristic used in Loser-Out-Framework for DTLZ test cases. As an example, for DTLZ1 test case, MOEA/D was the worst one, and it has been removed at the end of first phase. Similarly, NSGA-III was the best one which survives until the last phase and earns a big portion of fitness evaluations. Based on Table 5.1, R-NSGA-III is the best performing method winning 3 out of 6 problems and NSGA-III takes the second position with 2 out of 6 problems.

Table 5.1. Rank of the metaheuristics used in Loser-Out-Framework applied to three-objectives DTLZ

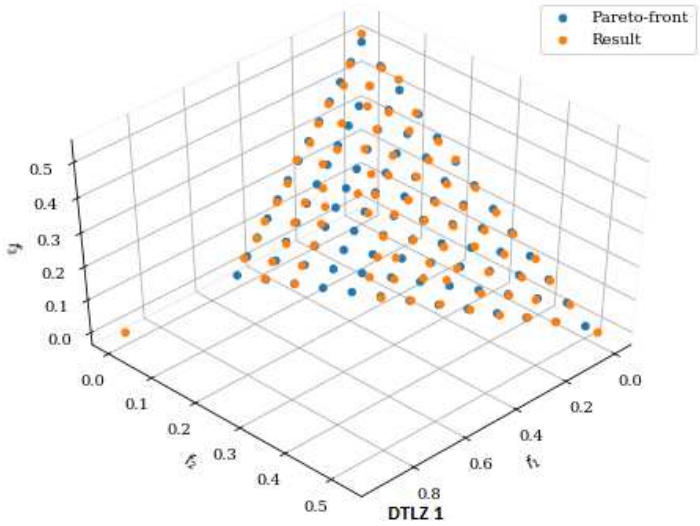
Function	MOEA/D	NSGA-III	R-NSGA-III	U-NSGA-III
DTLZ1	4	1	3	2
DTLZ2	4	2	3	1
DTLZ3	2	4	1	3
DTLZ4	4	1	3	2
DTLZ5	4	2	1	3
DTLZ6	3	4	1	2

Table 5.2 illustrates the average IGD scores obtained by LOF and its six recently published state-of-the-art competitors [8]-[12] over three objective DTLZ instances.

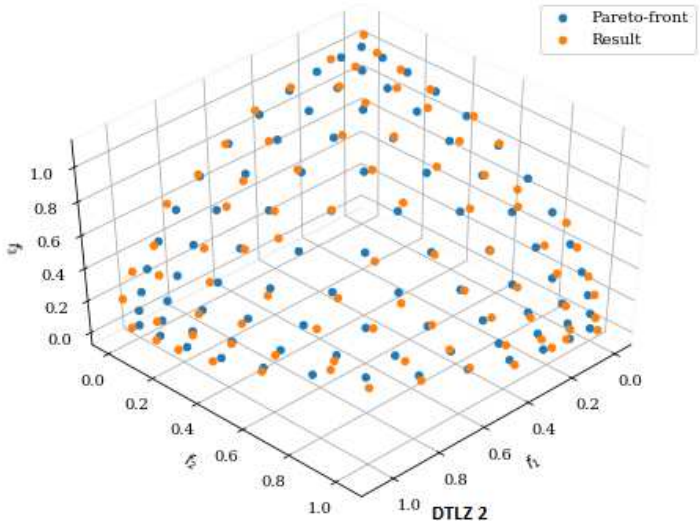
Table 5.2. IGD values obtained by LOF and its six competitors for three-objectives DTLZ

Function	NSGA-III-WA	NSGA-III	VAEA	RVEA	MODA/D	MODA/D-M2M	LOF	LOF Rank out of 7
DTLZ1	0.0314 ±0.0006	0.0209 ±0.0006	0.0777 ±0.0008	0.0620 ±0.0027	0.0408 ±0.0071	0.0431 ±0.0055	0.01787 ±0.0021	1
DTLZ2	0.0547 ±0.0002	0.05457 ±0.0004	0.0563 ±0.0008	0.0549 ±0.0001	0.0639 ±0.0007	0.0941 ±0.002	0.04353 ±0.0044	1
DTLZ3	0.0589 ±0.0007	0.0993 ±0.0008	0.0559 ±0.0019	0.0660 ±0.0044	0.0638 ±0.0014	0.0949 ±0.001	0.05673 ±0.0039	2
DTLZ4	0.0029 ±0.0001	0.0036 ±0.0007	0.0553 ±0.1937	0.0033 ±0.0002	0.0643 ±0.1009	0.0793 ±0.0316	0.00344 ±0.0050	3
DTLZ5	0.1281 ±0.0158	0.1143 ±0.0056	0.1674 ±0.0570	0.2057 ±0.0032	0.4196 ±0.0023	0.0432 ±0.0088	0.09230 ±0.0078	2
DTLZ6	0.9766 ±0.0252	1.516 ±0.0912	1.656 ±0.0509	1.303 ±0.0202	1.515 ±0.0075	1.826 ±0.0036	0.94340 ±0.0044	1
DTLZ7	NA	NA	NA	NA	NA	NA	0.08722 ±0069	NA

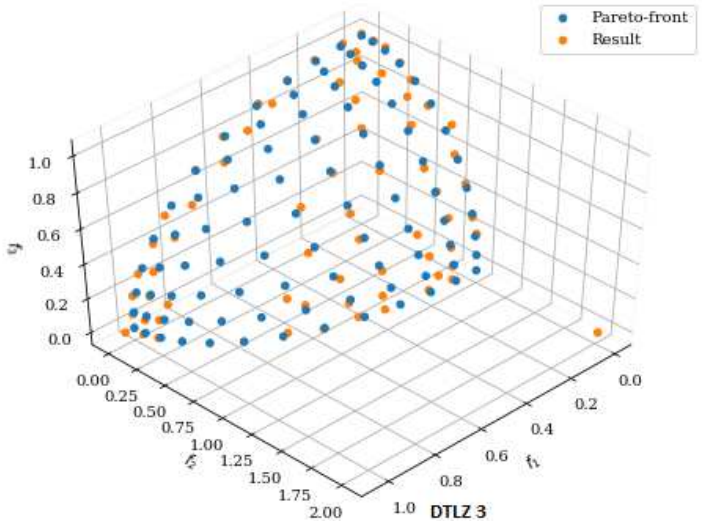
Based on Table 5.2, LOF takes the first position for DTLZ1, DTLZ2 and DTLZ6, the second position – for DTLZ3 and DTLZ5, and the third position – for DTLZ4 test instances. It means that in 50% of the test cases, LOF takes the first position and for the rest of the test cases it performs very close to the best ones. Likewise, Pareto-Fronts extracted by LOF for problems DTLZ1 to DTLZ7 are visually represented in Figure 5.1.



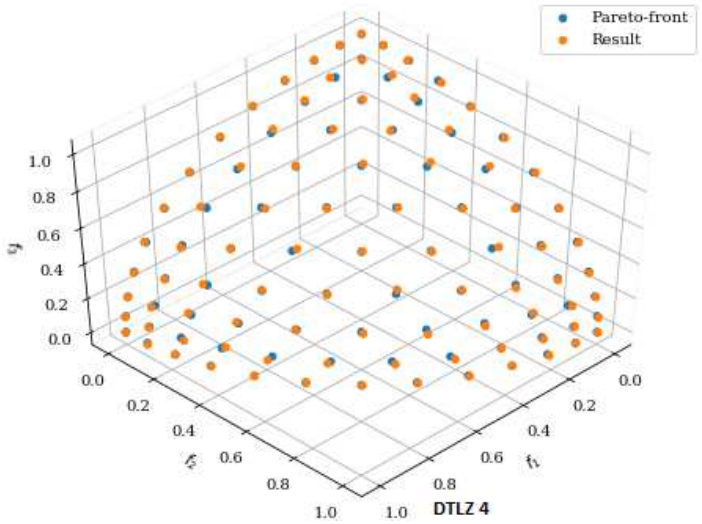
(a) **Figure 5.1.a.** Pareto front and optimal solution set for DTLZ1



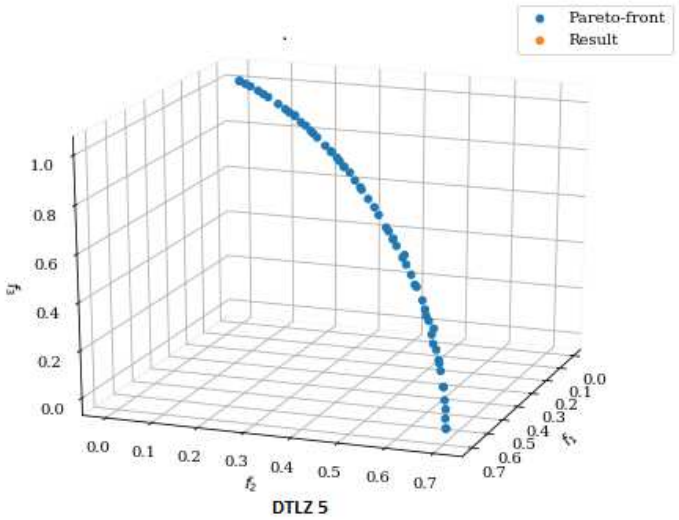
(b) **Figure 5.1.b.** Pareto front and optimal solution set for DTLZ2



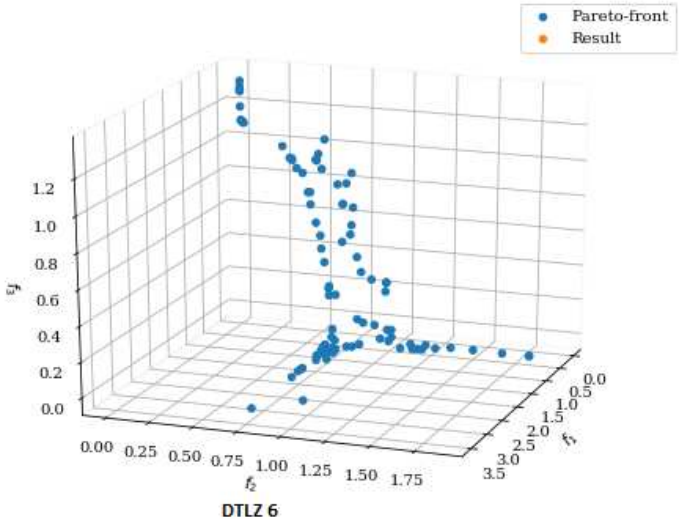
(a) **Figure 5.1.c.** Pareto front and optimal solution set for DTLZ3



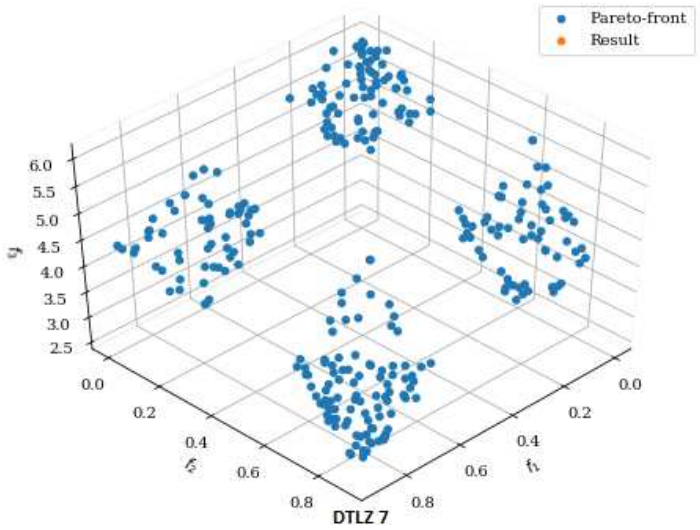
(b) **Figure 5.1.d.** Pareto front and optimal solution set for DTLZ4



(a) **Figure 5.1.e.** Pareto front and optimal solution set for DTLZ5



(b) **Figure 5.1.f.** Pareto front and optimal solution set for DTLZ6



(a) **Figure 5.1.g.** Pareto front and optimal solution set for DTLZ7

Figure 5.1. Pareto-Fronts extracted by LOF for problems DTLZ1 to DTLZ7

Below, in Figure 5.2, the average IGD values obtained by LOF and its competitors for DTLZ1 to DTLZ6 test cases are presented. It can be seen that the proposed method, shown in black, is either the best one or very close to the best one.

Table 5.3 illustrates the Friedman Aligned Ranks Test values calculated based on the IGD scores of 6 state-of-the-art algorithms and the new framework to find the order of the LOF among its competitors. Likewise, calculating the statistical similarity of results obtained by LOF to competitors [16], [17] is presented. Consequently, Table 5.4 shows the average of rank values, FAR and p-values for all algorithms.

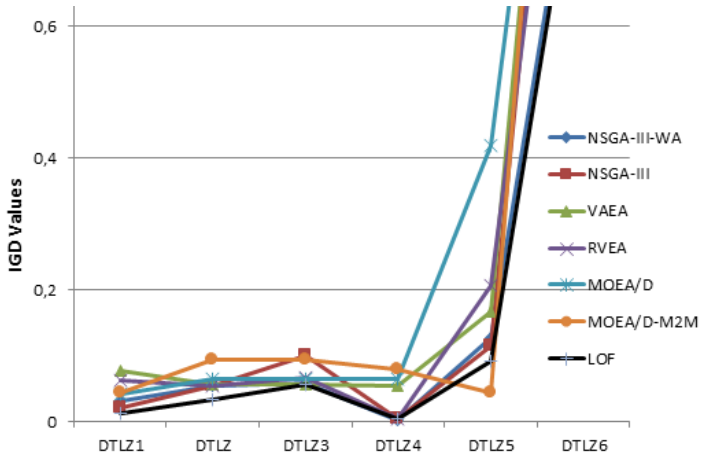


Figure 5.2. Average IGD values provided by LOF and its competitors for DTLZ test cases

Table 5.3. Calculated ranks of Friedman aligned for all pairs of DTLZ problems and Methods

Function	NSGA-III-WA	NSGA-III	VAEA	RVEA	MODA/D	MODA/D-M2M	LOF
DTLZ1	18	14	35	29	25	27	8
DTLZ2	22	21	24	23	28	34	13
DTLZ3	17	32	15	20	19	30	16
DTLZ4	9	12	31	10	33	37	11
DTLZ5	7	6	26	36	40	3	5
DTLZ6	2	39	41	4	38	42	1
Sum	75	124	172	122	183	173	54
AVG	12.50	20.67	28.67	20.33	30.50	28.83	9.00

According to Table 5.3, the average rank value of LOF is 9.00 which is the smallest one (i.e., the best performing algorithm). On the other hand, the p-value of LOF is also very close to zero which shows the remarkable statistical difference between the results of LOF and its competitors.

Table 5.4. Friedman Aligned Rank and p-value computed for all the methods for DTLZ problems

Algorithms	**Average values of Friedman Aligned Ranks overall problem instances**
NSGA-III-WA	12.50
NSGA-III	20.67
VAEA	28.67
RVEA	20.33
MODA/ D	30.50
MODA/ D-M2M	28.83
LOF	9.00
$\mathbf{F_{AR}}$	**15.054483**
P-value	**0.012837**

Finally, the proposed framework is evaluated over some three-objective WFG instances to prove the superiority of LOF when compared to its competitors.

Table 5.5 illustrates the rank of each metaheuristic used in Loser-Out-Framework for WFG test cases. According to Table 5.5, R-NSGA-III is the best performed method winning 2 out of four problems, and both NSGA-III and U-NSGA-III take the second position winning 1 out of 4 problems.

Table 5.5. Rank of the metaheuristics used in Loser-Out-Framework applied on three-objectives WFG

Function	MOEA/D	NSGA-III	R-NSGA-III	U-NSGA-III
WFG1	2	4	3	1
WFG2	4	1	2	3
WFG3	2	4	1	3
WFG4	4	3	1	2

Table 5.6 illustrates the average IGD scores for four three-objective WFG instances of LOF and its six recently published state-of-the-art competitors [8]-[12]. Based on Table 5.6, LOF takes the first position for WFG3 and WFG4, the second position – for WFG1 and the third position – for WFG3 test instances. It means that again in 50% of the test cases LOF takes the first position and for the rest of the cases it performs very close to the best ones.

Table 5.6. IGD values obtained by LOF and its six competitors for three-objectives WFG

Function	NSGA-III-WA	NSGA-III	VAEA	RVEA	MODA/D	MODA/D-M2M	LOF	LOF Rank out of 7
WFG1	1.171 ±0.2727	1.370 ±0.3356	1.324 ±0.2315	1.047 ±0.2417	1.216 ±0.2173	1.211 ±0.3725	1.157 ±0.2820	2
WFG2	0.2149 ±0.0613	0.2839 ±0.1040	0.3218 ±0.0893	0.3157 ±0.0436	1.317 ±0.0701	0.3714 ±0.0482	0.293 ±0.0451	3
WFG3	0.2163 ±0.0264	0.3791 ±0.0816	0.1489 ±0.0069	0.1977 ±0.0328	0.1793 ±0.0303	0.2361 ±0.0462	0.144 ±0.0272	1
WFG4	0.2043 ±0.0022	0.2147 ±0.0003	0.2317 ±0.0073	0.2272 ±0.0037	0.2475 ±0.0037	0.3581 ±0.0031	0.198 ±0.0008	1

Table 5.7 and Table 5.8 show Friedman Aligned Ranks Test, average of rank values, FAR and p-values for the proposed framework and its six competitors. According to Table 5.7, the average rank value of LOF is 5.00 which is the smallest one (i.e., the best performing algorithm). Small p-value of LOF also shows the remarkable statistical difference between the results of LOF and its competitors.

Table 5.7. Calculated ranks of Friedman aligned for all pairs of WFG problems and Methods

Function	NSGA-III-WA	NSGA-III	VAEA	RVEA	MODA/D	MODA/D-M2M	LOF
WFG1	11	7	27	21	18	19	1
WFG2	15	14	17	16	20	26	6
WFG3	10	24	8	13	12	22	9
WFG4	2	5	23	3	25	28	4
Sum	38	50	75	53	75	95	20
AVG	9.50	12.50	18.75	13.25	18.75	23.75	5.00

6 Conclusions and future works

This paper presents a new framework, called the Loser-Out-Framework (LOF), to solve MOO problems. LOF consists of four different and robust metaheuristics cooperating with each other to collect Non-Dominated Solutions (NDS) in a global archive and also to improve subpopulations. These four metaheuristics are also in competition to survive. At the end, all Non-Dominated Solutions found by all metaheuristics are combined and then a constant number of non-dominated solutions are selected as a final result. Thereafter for comparison, the IGD values are computed. The effectiveness of the proposed method is tested over three-objective DTLZ and WFG benchmark instances, and its performance is comparatively evaluated against the well-known

modern MOO algorithms. The success of the proposed approach on the most of test problems in comparison to state-of-the-art algorithms indicates the efficiency of LOF in solving MOPs and puts it in the first position against its competitors. For the future, it is planned to extend this framework by adding more well-performed metaheuristics (consequently more phases) and evaluate it by applying to some real-world problems.

Table 5.8. Friedman Aligned Rank and p-value computed for all the methods for WFG problems

Algorithms	**Average values of Friedman Aligned Ranks overall problem instances**
NSGA-III-WA	9.5
NSGA-III	12.5
VAEA	18.75
RVEA	13.25
MODA/ D	18.75
MODA/ D-M2M	2375
LOF	5
F_{AR}	**9.5901**
P-value	**0.014300**

References

[1] K. Deb, S. Agrawal, A. Pratap, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. on Evolutionary Computation*, vol. 6, pp. 182–197, 2002.

[2] B. Julian, D. Kalyanmoy, and C. R. Proteek, "Investigating the normalization procedure of NSGA-III," in *Evolutionary Multi-Criterion Optimization, 10th International Conference, EMO 2019, East Lansing*, (MI, USA), 2019, pp. 229–240.

[3] Z. Qingfu and L. Hui Li, "A multi-objective evolutionary algorithm based on decomposition," *IEEE Transactions on Evolutionary Computation*, vol. 11, no. 6, pp. 712–731, 2007.

[4] Y. Vesikar, K. Deb, and J. Blank, "Reference point based NSGA-III for preferred solutions," *IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1587–1594, 2018.

[5] H. Seada and K. Deb, "A unified evolutionary optimization procedure for single, multiple, and many objectives," *IEEE Transactions on Evolutionary Computation*, pp. 358–369, 2016.

[6] S. Huband, P. Hingston, L. Barone, and L. While, "A Review of Multiobjective Test Problems and a Scalable Test Problem Toolkit," *IEEE Trans. On Evolutionary Computation*, vol. 10, pp.477–506, 2006.

[7] Y. Wang and X. Sun, "A Many-Objective Optimization Algorithm Based on Weight Vector Adjustment," *Computational Intelligence and Neuroscience*, vol. 2018, 2018. Article ID 4527968.

[8] M. Ojha, K. P. Singh, P. Chakraborty P, and S. Verm, "An Aggregation Based Approach with Pareto Ranking in Multiobjective Genetic Algorithm," in *Proceedings of Fifth International Conference on Soft Computing for Problem Solving. SocProS 2015, Volume 2* (Advances in Intelligent Systems and Computing, vol. 437), Springer Singapore, 2016, pp. 261–271. eBook ISBN: 978-981-10-0451-3.

[9] Q. Zhang, W. Liu, and H. Li, "The performance of a new version of MOEA/D on CEC09 unconstrained MOP test instances, CEC 2009," *Proc. of the IEEE Eleventh Conf. on Evolutionary Computation*, pp. 203–208, 2009.

[10] W. Li, L. Wang, Q. Jiang, X. Hei, and B. Wang, "Multiobjective cloud particle optimization algorithm based on decomposition," *Algorithms*, vol. 8, no. 2, pp.157-176, 2015.

[11] W. Mashwani and A. Salhi, "Multiobjective memetic algorithm based on decomposition," *Applied Soft Computing*, vol. 21, pp. 221–243, 2014.

[12] K. Li, R. Wang, T. Zhang and H. Ishibuchi, "Evolutionary Many-Objective Optimization: A Comparative Study of the State-of-the-Art," *IEEE Access*, vol. 6, pp. 26194–26214, 2018. DOI: 10.1109/ACCESS.2018.2832181.

[13] R. Cheng, Y. Jin, M. Olhofer, and B. Sendhoff, "A reference vector guided evolutionary algorithm for many-objective optimization," *IEEE Trans. Evol. Comput.*, vol. 20, no. 5, pp. 773–791, Oct. 2016.

[14] Y. Xiang, Y. Zhou, M. Li, and Z. Chen, "A vector angle-based evolutionary algorithm for unconstrained many-objective optimization," *IEEE Transactions on Evolutionary Computation*, vol. 21, pp. 131–152, 2017.

[15] K. Deb and H. Jain, "An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, Part I: solving problems with box constraints," *IEEE Transactions on Evolutionary Computation*, vol. 18, pp. 577–601, 2014.

[16] Q. Zhang, W. Liu, and H. Li, "The performance of a new version of MOEA/D on CEC2009 unconstrained MOP test instances," in *IEEE Congress on Evolutionary Computation – CEC2009*, pp. 203–208, 2009.

[17] J. Derrac, S. Garcia, D. Molina, and F. Herrera, "A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms," *Swarm and Evolutionary Computation*, vol. 1, pp. 3–18, 2011.

[18] H. Li and Q. Zhang, "Multiobjective Optimization Problems with Complicated Pareto Sets, MOEA/D and NSGA-II," *IEEE Trans. Evol. Comput*, vol. 13, pp. 284–302, 2009.

[19] Z. Wang, Q. Zhang, A. Zhou, M. Gong, and L. Jiao, "Adaptive Replacement Strategies for MOEA/D," *IEEE Trans. Cybern*, pp. 474–486, 2016.

[20] Q. Zhang, W. Liu, and H. Li, "The performance of a new version of MOEA/D on CEC09 unconstrained MOP test instances," in *IEEE Congr. Evol. Comput*, pp. 203–208, 2009.

[21] I. Ahmadianfar, A. Adib, and M. Taghian, "A Multi-Objective Evolutionary Algorithm Using Decomposition (MOEA/D) And Its Application In Multipurpose Multi-Reservoir Operations," *International Journal Of Optimization In Civil Engineering*, pp. 167–187, 2015.

[22] H. L. Liu, F. Gu, and Z. Quingfu, "Decomposition of a Multiobjective Optimization Problem into a Number of Simple Multiobjective Subproblems," *IEEE Transactions on Evolutionary Computation*, vol. 18, no. 3, pp. 450–455, 2014.

Jamshid Tamouk, Nasser Lotfi Received July 26, 2020

Jamshid Tamouk
Faculty of Engineering, Eastern Mediterranean University,
Famagusta, N. Cyprus via Mersin 10, Turkey
E–mail: jamshid.tamouk@cc.emu.edu.tr

Nasser Lotfi
Faculty of Engineering, Cyprus Science University, Girne, N.
Cyprus via Mersin 10, Turkey
E–mail: nasser.lotfi2020@gmail.com

Computer Science Journal of Moldova, vol.**28**, no.3(84), 2020

On groupoids with Bol-Moufang type identities

Grigorii Horosh, Victor Shcherbacov,
Alexandru Tcachenco, Tatiana Yatsko

Abstract

We present results about groupoids of small order with Bol-Moufang type identities both classical and non-classical which are listed in [7], [9].

2000 Mathematics Subject Classification: 08A35, 68R05, 20N05

Key words and phrases: groupoid, quasigroup, Bol-Moufang type identity.

1 Introduction

Binary groupoid $(G, *)$ is a non-empty set G together with a binary operation "$*$" which is defined on the set G.

An identity based on a single binary operation is of Bol-Moufang type if "both sides consist of the same three different letters taken in the same order but one of them occurs twice on each side" [9]. We use list of 60 Bol-Moufang type identities given in [13].

There exist other (more general) definitions of Bol-Moufang type identities and, therefore, other lists and classifications of such identities [1], [7]. An identity based on a single binary operation is of generalized Bol-Moufang type if "both sides consist of the same three different letters but one of them occurs twice on each side" [1], [7]. In this paper we use both classifications.

We shall name here classical Bol-Moufang type identities as identities from Fenyves's list. We shall name here non-classical type identities

as generalized Bol-Moufang type identities. It is clear that any classical type identity is also of non-classical type, but inverse is not true.

Quasigroups and loops, in which Bol-Moufang type identity is true, are central and classical objects of Quasigroup Theory. We recall, works of R. Moufang, G. Bol, R. H. Bruck, V. D. Belousov, K. Kunen, S. Gagola and of many other mathematicians are devoted to the study of quasigroups and loops with Bol-Moufang type identities [1]–[3], [7], [9]–[11], [14], [16], [19].

We continue the study of groupoids with Bol-Moufang type identities [17], [6] [4], [5], [12], [22].

Notice, research of groupoids of small order with some identities is a well-known problem in mathematical literature.

For groupoids the following natural problem is researched: how many groupoids with some identities of small order there exist? A list of numbers of semigroups of orders up to 8 is given in [20], of order 9 – in [8]; a list of numbers of quasigroups up to 11 is given in [15], [18].

2 Results

2.1 Some results on groupoids of order two

It is clear that there exist 16 groupoids of order 2 and there exist n^{n^2} of groupoids of order n. For example, $3^{3^2} = 19\,683$, $4^{4^2} = 4\,294\,967\,296$.

We list isomorphic pairs of groupoids of order two. If a groupoid does not have a pair, then this groupoid has an automorphism group of order two.

Below, the quadruple 22 12 means a groupoid of order 2 with the following Cayley table:

$*$	1	2
1	2	2
2	1	2

and so on. In such a record, a groupoid is commutative if and only if two elements of a quadruple, the second and the third, are equal.

Groupoid $(G, \cdot)$ is isomorphic to groupoid $(G, \circ)$ if there exists a permutation α of symmetric group S_G such that $x \circ y = \alpha^{-1}(\alpha x \cdot \alpha y)$ for all $x, y \in G$.

Groupoid $(G, \cdot)$ is anti-isomorphic to groupoid $(G, \circ)$ if there exists a permutation α of symmetric group S_G such that $x \circ y = \alpha^{-1}(\alpha y \cdot \alpha x)$ for all $x, y \in G$.

Remark 1. *If groupoid $(G, \cdot)$ is anti-isomorphic to groupoid $(G, \circ)$, then groupoid $(G, \circ)$ is anti-isomorphic to groupoid $(G, \cdot)$. Really $x \cdot y = \alpha(\alpha^{-1} y \circ \alpha^{-1} x)$ for all $x, y \in G$.*

Remark 2. *In commutative groupoid $(G, \cdot)$ any anti-isomorphism coincides with isomorphism.*

It is easy to check that the following propositions are fulfilled.

Proposition 1. *Only the following groupoids of order two are isomorphic in pairs:*

11 11 *and* 22 22*;*
11 12 *and* 12 22*;*
11 21 *and* 21 22*;*
11 22*;*
12 11 *and* 22 12*;*
12 12*;*
12 21 *and* 21 12*;*
21 11 *and* 22 21*;*
21 21*;*
22 11.

Proposition 2. *Only the following groupoids of order two are anti-isomorphic in pairs:*

11 21 *and* 22 12*;*
21 22 *and* 12 11*;*
11 22 *and* 12 12*;*
21 21 *and* 22 11.

Proposition 3. *Only the following groupoids of order two are isomorphic or anti-isomorphic:*

11 11 *and* 22 22*;*
11 12 *and* 12 22*;*
11 21 *and* 21 22 *and* 22 12 *and* 12 11*;*
11 22 *and* 12 12*;*
12 21 *and* 21 12*;*
21 11 *and* 22 21*;*
21 21 *and* 22 11*.*

Corollary 1. *The following groupoids of order two are non isomorphic and non anti-isomorphic in pairs:* 11 11*;* 11 12*;* 11 21*;* 11 22*;* 12 21*;* 21 11*;* 21 21*.*

Using the list of groupoids which is presented in Proposition 3, we can compose other lists of groupoids for Corollary 1. For example, instead of groupoid 11 11 we can write groupoid 22 22 and so on.

In the list presented in Corollary 1 semigroups of order two are underlined [23].

2.2 (12)-parastrophes of identities

We recall, (12)-parastrophe of groupoid $(G, \cdot)$ is a groupoid $(G, *)$ in which operation "$*$ " is obtained by the following rule:

$$x * y = y \cdot x. \tag{1}$$

It is clear that for any groupoid $(G, \cdot)$ there exists its (12)-parastrophe groupoid $(G, *)$.

Cayley table of groupoid $(G, *)$ is a mirror image of the Cayley table of groupoid $(G, \cdot)$ relative to main diagonal. Notice, for any binary quasigroup there exist five its parastrophes [2], [18], [22] more.

Suppose that an identity F is true in groupoid $(G, \cdot)$. Then we can obtain a (12)-parastrophic identity F^* of the identity F, replacing the operation "$\cdot$" with the operation "$*$" and changing the order of variables by using rule (1).

Remark 3. *In quasigroup case, similarly to* (12)*-parastrophe identity other parastrophe identities can be defined. See [21] for details.*

It is clear that an identity F is true in groupoid $(G,\cdot)$ if and only if in groupoid $(Q,*)$ the identity F^* is true.

Proposition 4. *The number of groupoids of a finite fixed order in which the identity F is true coincides with the number of groupoids in which the identity F^* is true.*

Example 1. *[12]. Way 1. We find* (12)*-parastrophe of the Bol-Moufang type identity F_1: $xy\cdot zx=(xy\cdot z)x$.*

We have $(x*z)*(y*x)=x*(z*(y*x))$. After renaming of variables $(y\leftrightarrow z)$ and operation $(*\to\cdot)$ we obtain the following Bol-Moufang type identity F_3: $xy\cdot zx=x(y\cdot zx)$.

Therefore $(F_1)^*=F_3$. The vice versa, $(F_3)^*=F_1$, is also true.

Way 2. We recall, left translation of a groupoid $(G,\cdot)$ is defined as follows: $L_ax=a\cdot x$ for all $x\in G$; right translation of a groupoid $(G,\cdot)$ is defined similarly: $R_ax=x\cdot a$ for all $x\in G$ and a fixed element $a\in G$.

Then we can re-write identity F_1 in the following form: $L_xy\cdot R_xz=R_x(L_xy\cdot z)$.

There exists the following connections between left and right translations of a groupoid $(G,\cdot)$ and its (12)*-parastrophe [18], [22]:*

$$L_a^*=R_a^{\cdot},R_a^*=L_a^{\cdot}. \tag{2}$$

Further, using rules (1) and (2), we have $L_xz\cdot R_xy=L_x(z\cdot R_xy)$, $xz\cdot yx=x(z\cdot yx)$. After renaming variables $(y\leftrightarrow z)$, we obtain the following Bol-Moufang type identity F_3: $xy\cdot zx=x(y\cdot zx)$, i.e., $(F_1)^*=F_3$.

Theorem 1. *For classical Bol-Moufang type identities over groupoids the following equalities are true:*

$(F_1)^*=F_3$, $(F_2)^*=F_4$, $(F_5)^*=F_{10}$, $(F_6)^*=F_6$, $(F_7)^*=F_8$, $(F_9)^*=F_9$, $(F_{11})^*=F_{24}$, $(F_{12})^*=F_{23}$, $(F_{13})^*=F_{22}$, $(F_{14})^*=F_{21}$, $(F_{15})^*=F_{30}$, $(F_{16})^*=F_{29}$, $(F_{17})^*=F_{27}$, $(F_{18})^*=F_{28}$, $(F_{19})^*=F_{26}$, $(F_{20})^*=F_{25}$, $(F_{31})^*=F_{34}$, $(F_{32})^*=F_{33}$, $(F_{35})^*=F_{40}$, $(F_{36})^*=F_{39}$, $(F_{37})^*=F_{37}$, $(F_{38})^*=F_{38}$, $(F_{41})^*=F_{53}$, $(F_{42})^*=F_{54}$, $(F_{43})^*=F_{51}$,$(F_{44})^*=F_{52}$, $(F_{45})^*=F_{60}$, $(F_{46})^*=F_{56}$, $(F_{47})^*=F_{58}$, $(F_{48})^*=F_{57}$, $(F_{49})^*=F_{59}$, $(F_{50})^*=F_{55}$.

For quasigroups, the analogue of Theorem 1 is given in [14].

Proposition 5. *Any of the following groupoids* 11 11*,* 22 22*,* 11 12*,* 12 22*,* 11 22*,* 12 12 *satisfies any of the identities* F_1–F_{60}.

Proof. It is possible to use direct calculations. □

2.3 Number of groupoids

Original algorithm is elaborated, the corresponding program is written for generating the groupoids of small (2, 3, and 4) orders with generalized Bol-Moufang identities.

The developed algorithm consists of two parts. In the first part we generate a groupoid. In the second part we check, if this groupoid satisfies a Bol-Moufang identity. And so on.

Usually we present a groupoid as a string or two-dimensional array.

Notice, number of groupoids of order 3 with mentioned in Table 2 identities are also given in [5].

We count number of groupoids of order two with classical Bol-Moufang type identities given in [13], including also the number of non-isomorphic ones and number of non-isomorphic and non-anti-isomorphic groupoids of order 2 (see Table 1). Notice, in some places Table 1 coincides with the corresponding table from [12].

Table 1 is organised as follows: in the first column there is given the name of identity in Fen'vesh list; in the second – the abbreviation of this identity, if this identity has a name; in the third – the identity is given; in the fourth column there is indicated the number of groupoids of order 2 with the corresponding identity; in the fifth column – the number of non-isomorphic groupoids; and, in the sixth column – the number of non-isomorphic and non-anti-isomorphic groupoids with the corresponding identity is given.

Table 1: Number of groupoids of order 2 with classical Bol-Moufang identities

Name	**Abb.**	**Ident.**	**2**	**n.-is.**	**n.-is., an.**
F_1		$xy \cdot zx = (xy \cdot z)x$	10	6	5
F_2		$xy \cdot zx = (x \cdot yz)x$	9	6	5
F_3		$xy \cdot zx = x(y \cdot zx)$	10	6	5
F_4	middle Mouf.	$xy \cdot zx = x(yz \cdot x)$	9	6	5
F_5		$(xy \cdot z)x = (x \cdot yz)x$	11	7	6
F_6	extra ident.	$(xy \cdot z)x = x(y \cdot zx)$	10	7	5
F_7		$(xy \cdot z)x = x(yz \cdot x)$	9	6	5
F_8		$(x \cdot yz)x = x(y \cdot zx)$	9	6	5
F_9		$(x \cdot yz)x = x(yz \cdot x)$	10	6	5
F_{10}		$x(y \cdot zx) = x(yz \cdot x)$	11	7	6
F_{11}		$xy \cdot xz = (xy \cdot x)z$	8	5	4
F_{12}		$xy \cdot xz = (x \cdot yx)z$	9	7	6
F_{13}	extra ident.	$xy \cdot xz = x(yx \cdot z)$	9	6	5
F_{14}		$xy \cdot xz = x(y \cdot xz)$	10	6	5
F_{15}		$(xy \cdot x)z = (x \cdot yx)z$	11	7	6
F_{16}		$(xy \cdot x)z = x(yx \cdot z)$	11	7	6
F_{17}	left Mouf.	$(xy \cdot x)z = x(y \cdot xz)$	10	7	5
F_{18}		$(x \cdot yx)z = x(yx \cdot z)$	8	5	4

F_{19}	left Bol	$(x \cdot yx)z = x(y \cdot xz)$	9	6	5
F_{20}		$x(yx \cdot z) = x(y \cdot xz)$	9	6	5
F_{21}		$yx \cdot zx = (yx \cdot z)x$	10	6	5
F_{22}	extra ident.	$yx \cdot zx = (y \cdot xz)x$	9	6	5
F_{23}		$yx \cdot zx = y(xz \cdot x)$	9	6	5
F_{24}		$yx \cdot zx = y(x \cdot zx)$	8	5	4
F_{25}		$(yx \cdot z)x = (y \cdot xz)x$	9	6	5
F_{26}	right Bol	$(yx \cdot z)x = y(xz \cdot x)$	9	6	5
F_{27}	right Mouf.	$(yx \cdot z)x = y(x \cdot zx)$	10	7	5
F_{28}		$(y \cdot xz)x = y(xz \cdot x)$	8	5	4
F_{29}		$(y \cdot xz)x = y(x \cdot zx)$	11	7	6
F_{30}		$y(xz \cdot x) = y(x \cdot zx)$	11	7	6
F_{31}		$yx \cdot xz = (yx \cdot x)z$	8	5	4
F_{32}		$yx \cdot xz = (y \cdot xx)z$	9	6	5
F_{33}		$yx \cdot xz = y(xx \cdot z)$	9	6	5
F_{34}		$yx \cdot xz = y(x \cdot xz)$	8	5	4
F_{35}		$(yx \cdot x)z = (y \cdot xx)z$	9	6	5
F_{36}	RC ident.	$(yx \cdot x)z = y(xx \cdot z)$	9	6	5
F_{37}	C ident.	$(yx \cdot x)z = y(x \cdot xz)$	10	7	5
F_{38}		$(y \cdot xx)z = y(xx \cdot z)$	8	5	4
F_{39}	LC ident.	$(y \cdot xx)z = y(x \cdot xz)$	9	6	5
F_{40}		$y(xx \cdot z) = y(x \cdot xz)$	9	6	5

F_{41}	LC ident.	$xx \cdot yz = (x \cdot xy)z$	9	6	5
F_{42}		$xx \cdot yz = (xx \cdot y)$ z	12	7	5
F_{43}		$xx \cdot yz = x(x \cdot yz)$	8	5	4
F_{44}		$xx \cdot yz = x(xy \cdot z)$	9	6	5
F_{45}		$(x \cdot xy)z = (xx \cdot y)z$	9	6	5
F_{46}	LC ident.	$(x \cdot xy)z = x(x \cdot yz)$	11	7	6
F_{47}		$(x \cdot xy)z = x(xy \cdot z)$	8	5	4
F_{48}	LC ident.	$(xx \cdot y)z = x(x \cdot yz)$	10	7	5
F_{49}		$(xx \cdot y)z = x(xy \cdot z)$	9	6	5
F_{50}		$x(x \cdot yz) = x(xy \cdot z)$	11	7	6
F_{51}		$yz \cdot xx = (yz \cdot x)x$	8	5	4
F_{52}		$yz \cdot xx = (y \cdot zx)x$	9	6	5
F_{53}	RC ident.	$yz \cdot xx = y(zx \cdot x)$	9	6	5
F_{54}		$yz \cdot xx = y(z \cdot xx)$	12	7	6
F_{55}		$(yz \cdot x)x = (y \cdot zx)x$	11	7	6
F_{56}	RC ident.	$(yz \cdot x)x = y(zx \cdot x)$	11	7	6
F_{57}	RC ident.	$(yz \cdot x)x = y(z \cdot xx)$	10	7	6
F_{58}		$(y \cdot zx)x = y(zx \cdot x)$	8	5	4
F_{59}		$(y \cdot zx)x = y(z \cdot xx)$	9	6	5
F_{60}		$y(zx \cdot x) = y(z \cdot xx)$	9	6	5

Identities Left Bol and Right Bol, LC and RC, LN and RN, L2

and L3, M1 and M3, M2 and M4, T1 and T3, T4 and T5, are (12)-parastrophic identitities. Therefore the numbers of groupoids of fixed order with these (12)-parastrophic identitities coincide.

Table 2: Number of groupoids of order 2, 3 and 4 with Bol-Moufang identities.

Name	**Abbr.**	**Ident.**	**#2**	**#3**	**#4**
Extra	EL	$x(y(zx)) = ((xy)z)x$	10	239	18744
Moufang	ML	$(xy)(zx) = (x(yz))x$	9	196	25113
Left Bol	LB	$x(y(xz)) = (x(yx))z$	9	215	22875
Right Bol	RB	$y((xz)x) = ((yx)z)x$	9	215	22875
C-loops	CL	$y(x(xz)) = ((yx)x)z$	10	209	26583
LC-loops	LC	$(xx)(yz) = (x(xy))z$	9	220	26583
RC-loops	RC	$y((zx)x) = (yz)(xx)$	9	220	26583
Middle Nuclear Square	MN	$y((xx)z) = (y(xx))z$	8	350	122328
Right Nuclear Square	RN	$y(z(xx)) = (yz)(xx)$	12	932	2753064
Left Nuclear Square	LN	$((xx)y)z = (xx)(yz)$	12	932	2753064
Comm. Moufang	CM	$(xy)(xz) = (xx)(zy)$	8	297	111640
Comm. C-loop	CC	$(y(xy))z = x(y(yz))$	8	169	12598
Comm. Alternative	CA	$((xx)y)z = z(x(yx))$	6	110	10416
Comm. Nuclear square	CN	$((xx)y)z = (xx)(zy)$	9	472	1321661
Comm. loops	CP	$((yx)x)z = z(x(yx))$	8	744	1078744

Continued on next page

Table 2 – *Continued from previous page*

Name	**Abbr.**	**Ident.**	**#2**	**#3**	**#4**
Cheban, 1	C1	$x((xy)z) = (yx)(xz)$	8	219	19846
Cheban, 2	C2	$x((xy)z) = (y(zx))x$	6	153	12382
Lonely, I	L1	$(x(xy))z = y((zx)x)$	6	117	6076
Cheban, I, Dual	CD	$(yx)(xz) = (y(zx))x$	8	219	19846
Lonely, II	L2	$(x(xy))z = y((xx)z)$	7	157	11489
Lonely, III	L3	$(y(xx))z = y((zx)x)$	7	157	11489
Mate, I	M1	$(x(xy))z = ((yz)x)x$	6	111	11188
Mate, II	M2	$(y(xx))z = ((yz)x)x$	7	196	26785
Mate, III	M3	$x(x(yz)) = y((zx)x)$	6	111	11188
Mate, IV	M4	$x(x(yz)) = y((xx)z)$	7	196	26785
Triad, I	T1	$(xx)(yz) = y(z(xx))$	6	162	67152
Triad, II	T2	$((xx)y)z = y(z(xx))$	6	180	53832
Triad, III	T3	$((xx)y)z = (yz)(xx)$	6	162	67152
Triad, IV	T4	$((xx)y)z = ((yz)x)x$	6	132	42456
Triad, V	T5	$x(x(yz)) = y(z(xx))$	6	132	42456
Triad, VI	T6	$(xx)(yz) = (yz)(xx)$	8	1419	9356968
Triad, VII	T7	$((xx)y)z = ((yx)x)z$	12	428	2914658
Triad, VIII	T8	$(xx)(yz) = y((zx)x)$	6	120	11580
Triad, IX	T9	$(x(xy))z = y(z(xx))$	6	102	6192
Frute	FR	$(x(xy))z = (y(zx))x$	6	129	16600
Crazy Loop	CR	$(x(xy))z = (yx)(xz)$	7	136	12545
Krypton	KR	$((xx)y)z = (x(yz))x$	9	268	93227

About computer. The corresponding program has run at 4 nodes with the following properties: System – Virtual machine; OS – Ubuntu

18.04; CPU – Virtual CPU, up to 2.40GHz; Memory – 8GB.

Acknowledgement. The authors thank Prof. C.V. Gaindric for his valuable suggestions.

References

[1] Reza Akhtar, Ashley Arp, Michael Kaminski, Jasmine Van Exel, Davian Vernon, and Cory Washington, "The varieties of Bol-Moufang quasigroups defined by a single operation," *Quasigroups Related Systems*, vol. 20, no. 1, pp. 1–10, 2012.

[2] V.D. Belousov, *Foundations of the Theory of Quasigroups and Loops*, Moscow: Nauka, 1967. (in Russian).

[3] R.H. Bruck, "Some theorems on Moufang loops," *Math. Z.*, vol. 73, pp. 59–78, 1960.

[4] Vladimir Chernov, Alexander Moldovyan, and Victor Shcherbacov, "On some groupoids of order three with Bol-Moufang type of identities," in *Proceedings of the Conference on Mathematical Foundations of Informatics MFOI2018, July 2-6, 2018, Chisinau*, (Chisinau, Moldova), 2018, pp. 17–20.

[5] Vladimir Chernov, Alexander Moldovyan, and Victor Shcherbacov, "On some groupoids of small orders with Bol-Moufang type of identities," 2018, arXiv:1812.02511.

[6] Vladimir Chernov, Nicolai Moldovyan, and Victor Shcherbacov, "On some groupoids of small order," in *The Fourth Conference of Mathematical Society of the Republic of Moldova dedicated to the centenary of Vladimir Andrunachievici (1917-1997), June 28 - July 2, 2017, Chisinau, Proceedings CMSM4'*, (Chisinau, Moldova), 2017, pp. 51–54.

[7] B. Cote, B. Harvill, M. Huhn, and A. Kirchman, "Classification of loops of generalized Bol-Moufang type," *Quasigroups Related Systems*, vol. 19, no. 2, pp. 193–206, 2011.

[8] Andreas Distler and Tom Kelsey, "The semigroups of order 9 and their automorphism groups," *Semigroup Forum*, vol. 88, no. 1, pp. 93–112, 2014.

[9] F. Fenyves, "Extra loops. II. On loops with identities of Bol-Moufang type," *Publ. Math. Debrecen*, vol. 16, pp. 187–192, 1969.

[10] Stephen M. Gagola, "Hall's theorem for Moufang loops," *J. Algebra*, vol. 323, no. 12, pp. 3252–3262, 2010.

[11] Stephen M. Gagola, "How and why Moufang loops behave like groups," *Quasigroups Related Systems*, vol. 19, no. 1, pp. 1–22, 2011.

[12] Grigorii Horosh, Victor Shcherbacov, Alexandru Tcachenco, and Tatiana Yatsko, "On some groupoids with Bol-Moufang type identities," in *Proceedings of the 5-th Conference on Mathematical Foundations of Informatics, MFOI 2019, July 3-6*, (Iasi, Romania), July 2019, pp. 149–158.

[13] T.G. Jaíyeola, E. Ilojide, M. O. Olatinwo, and F. Smarandache, "On the Classification of Bol-Moufang Type of Some Varieties of Quasi Neutrosophic Triplet Loop (Fenyves BCI-Algebras)," *Symmetry*, vol. 10, pp. 1–16, 2018. DOI:10.3390/sym10100427.

[14] K. Kunen, "Quasigroups, loops and associative laws," *J. Algebra*, vol. 185, no. 1, pp. 194–204, 1996.

[15] Brendan D. McKay and Ian M. Wanless, "On the number of latin squares," *Ann. Comb.*, vol. 9, no. 3, pp. 335–344, 2005.

[16] R. Moufang, "Zur Structur von Alternativ Körpern," *Math. Ann.*, vol. 110, pp. 416–430, 1935.

[17] B. V. Novikov, "On decomposition of Moufang groupoids," *Quasigroups and related systems*, vol. 16, no. 1, pp. 97–101, 2008.

[18] H.O. Pflugfelder, *Quasigroups and Loops: Introduction*, Berlin: Heldermann Verlag, 1990.

[19] J. D. Phillips and Petr Vojtechovsky, "The varieties of loops of Bol-Moufang type," *Algebra Universalis*, vol. 54, no. 3, pp. 259–271, 2005.

[20] S. Satoh, K. Yama, and M. Tokizawa, "Semigroups of order 8," *Semigroup forum*, vol. 49, pp. 7–29, 1994.

[21] A.V. Scerbacova and V.A. Shcherbacov, "On spectrum of medial T_2-quasigroups," *Bul. Acad. Ştiinţe Repub. Mold. Mat.*, no. 2, pp. 143–154, 2016.

[22] Victor Shcherbacov, *Elements of Quasigroup Theory and Applications*, Boca Raton: CRC Press, 2017.

[23] Wikipedia. "Semigroup with two elements," 2015. https://en.wikipedia.org/wiki/Semigroup_with_two_elements.

Grigorii Horosh[1], Victor Shcherbacov[2], Alexandr Tcachenco[3], Tatiana Yatsko[4]

[1]Vladimir Andrunachievici Institute of Mathematics and Computer Science, Moldova
E–mail: grigorii.horos@math.md

[2]Vladimir Andrunachievici Institute of Mathematics and Computer Science, Moldova
E–mail: victor.scerbacov@math.md

[3]Vladimir Andrunachievici Institute of Mathematics and Computer Science, Moldova
E–mail: atkacheno405@gmail.com

[4]Shevchenko Transnistria State University
E–mail: yaczkot@bk.ru

Computer Science Journal of Moldova, vol.**28**, no.3(84), 2020

Digital signature scheme set in a hidden cyclic group

D.N. Moldovyan A.A. Moldovyan N.A. Moldovyan

Abstract

A new form of the hidden discrete logarithm problem is proposed as cryptographic primitive for the development of the post-quantum signature schemes, which is characterized in performing two masking operations over each of two elements from a hidden finite cyclic group used to compute the public-key elements. The latter is contained in the set of non-invertible elements of the finite non-commutative associative algebra with a two-sided unit. One of the said masking operations represents the automorphism-map operation and the other one is the left-sided (right-sided) multiplication by a local right-sided (left-sided) unit acting on the said hidden group. Two 4-dimensional algebras are considered as possible algebraic supports of the developed signature schemes. The formulas describing the sets of local left-sided and right-sided units are derived. Periodic functions set on the base of the public parameters of the signature scheme contain periods depending on the discrete logarithm value, but every of them takes on the values relating to different finite groups contained in the algebraic support. Therefore one can expect that the computational difficulty of breaking the introduced signature schemes on a hypothetic quantum computer is superpolinomial.

Keywords: finite associative algebra, non-commutative algebra, global unit, local unit, right-sided unit, left-sided unit, discrete logarithm problem, public-key cryptoscheme, digital signature, post-quantum cryptosystem.

MSC 2010: 68P25, 68Q12, 68R99, 94A60, 16Z05, 14G50

1 Introduction

Currently, the development of the post-quantum (PQ) public key (PK) cryptographic algorithms and protocols is considered as one of challenges in the area of applied and theoretic cryptography [1], [2]. A response to this challenge is the world competition for the development of the PQ PK cryptoschemes, announced by NIST in 2016 [3],[4]. The problem of the development of the practical PQ PK cryptoschemes is connected with the following items: i) quantum computers can suddenly appear in practice in near future; ii) at present the most widely used PK cryptoschemes are based on computational difficulty of the discrete logarithm problem (DLP) and the factorization problem (FP), however, each of these problems can be solved on a hypothetic quantum computer in polynomial time [5]–[7].

Computationally hard problems, other than DLP and FP, are used as primitives of the PQ PK cryptoschemes [9], [10], one of which is the hidden DLP (HDLP) [8]. The HDLP seems to be a promising primitive for designing PQ signature schemes [11], [12], PQ public key-agreement protocols [13],[14], and PQ commutative ciphers [15]. For the first time the HDLP had been defined in finite algebra of quaternions using the automorphism-map as the operation masking the hidden cyclic group in which the basic exponentiation operation is performed [8]. However, that form of the HDLP can be reduced to the ordinary DLP in a finite field [16]. Therefore, in the design of the HDLP-based signature scheme [17], using the finite quaternion algebra as its algebraic support, a strengthened form of the HDLP was applied. In the signature scheme [17], the basic exponentiation operation N^x (where x is a private value) is performed in the hidden cyclic group generated by a non-invertible element N of the algebra and the PK includes two elements $Y = G \circ N^x \circ G^{-1}$ and $Z = Q \circ N \circ Q^{-1}$, where G and Q are two secret invertible elements that define two different automorphism-map operations each of which is mutually commutative with the exponentiation operation.

In the present paper we show that, setting the hidden cyclic group generated by a non-invertible element of the finite non-commutative

associative algebra (FNAA) with a global two-sided unit, provides possibility to design the HDLP-based signature schemes in which, during the process of generating the PK, the left-sided (rightleft-sided) multiplication by a local right-sided (left-sided) unit is used as additional masking operation performed on the element N^x (N) of the hidden group. Two 4-dimensional FNAAs are considered as algebraic support of the developed signature scheme. The formulas describing the sets of local left-sided and right-sided units are derived.

2 The used algebraic support

Usually the multiplication operation in a m-dimensional FNAA (denoted as $\circ$) is defined by the following formula

$$A \circ B = \left(\sum_{i=0}^{m-1} a_i \mathbf{e}_i \right) \circ \left(\sum_{j=0}^{m-1} b_j \mathbf{e}_j \right) = \sum_{j=0}^{m-1} \sum_{i=0}^{m-1} a_i b_j \left(\mathbf{e}_i \circ \mathbf{e}_j \right), \qquad (1)$$

where $A = \sum_{i=0}^{m-1} a_i \mathbf{e}_i$ and $B = \sum_{j=0}^{m-1} b_j \mathbf{e}_j$ are m-dimensional vectors; $\mathbf{e}_1, \mathbf{e}_2, \ldots \mathbf{e}_m$ are basis vectors. The product $\mathbf{e}_i \circ \mathbf{e}_j$ for all possible pairs of the integers i and j is to be replaced by some single-component vector $\lambda \mathbf{e}_k$ indicated in the cell at intersection of the ith row and the jth column of so called basis vector multiplication table (BVMT). The value $\lambda \neq 1$ is called structural coefficient.

2.1 A first 4-dimensional FNAA

One can easily show that the BVMT, shown as Table 1, defines over the ground finite field $GF(p)$ the 4-dimensional FNAA containing the global two-sided unit

$$E = \left(\frac{1}{\lambda - 1}, \frac{1}{1 - \lambda}, \frac{1}{1 - \lambda}, \frac{\lambda}{\lambda - 1} \right). \qquad (2)$$

The global means that the value E acts as unit element on every vector of the algebra. If each of the vector equations $A \circ X = E$ and $X \circ A =$

Table 1. The BVMT setting the 4-dimensional FNAA ($\lambda \neq 0$; $\lambda \neq 1$).

$\circ$	$\mathbf{e}_0$	$\mathbf{e}_1$	$\mathbf{e}_2$	$\mathbf{e}_3$
$\mathbf{e}_0$	$\lambda\mathbf{e}_0$	$\lambda\mathbf{e}_1$	$\mathbf{e}_0$	$\mathbf{e}_1$
$\mathbf{e}_1$	$\mathbf{e}_0$	$\mathbf{e}_1$	$\mathbf{e}_0$	$\mathbf{e}_1$
$\mathbf{e}_2$	$\lambda\mathbf{e}_2$	$\lambda\mathbf{e}_3$	$\mathbf{e}_2$	$\mathbf{e}_3$
$\mathbf{e}_3$	$\mathbf{e}_2$	$\mathbf{e}_3$	$\mathbf{e}_2$	$\mathbf{e}_3$

E has the same solution A^{-1}, then the vector A is called invertible and the vector A^{-1} is called inverse to the vector A. In the FNAA with the multiplication operation defined by Table 1 the vector $A = (a_0, a_1, a_2, a_3)$ is invertible, if the following invertibility condition holds true

$$a_1 a_2 \neq a_0 a_3. \tag{3}$$

Correspondingly, the non-invertibility condition is as follows

$$a_1 a_2 = a_0 a_3. \tag{4}$$

Using the condition (4) one can easily compute the number of the non-invertible vectors that is equal to $\eta = p^3 + p^2 - p$. Therefore, for the order Ω of the multiplicative group of the considered algebra (number of its invertible vectors) one can get $\Omega = p^4 - \eta$:

$$\Omega = p(p-1)\left(p^2 - 1\right). \tag{5}$$

In the algebra there exists no other element, except the global unit E, which acts as unit element on an invertible vector. On the contrary, there exists a variety of local left-sided and local right-sided units acting on some fixed non-invertible vector N and some subsets of non-invertible vectors connected with the vector N.

To derive the formula describing the set of the local left-sided units, one should consider the solutions of the vector equation $X \circ N = N$, where $N = (n_0, n_1, n_2, n_3)$ is a vector satisfying the non-invertibility condition $n_1 n_2 = n_0 n_3$, which can be reduced to the following two

independent systems of two linear equations:

$$\begin{cases} (\lambda n_0 + n_2)\, x_0 + (n_0 + n_2)\, x_1 = n_0; \\ (\lambda n_1 + n_3)\, x_0 + (n_1 + n_3)\, x_1 = n_1; \end{cases} \tag{6}$$

$$\begin{cases} (\lambda n_0 + n_2)\, x_2 + (n_0 + n_2)\, x_3 = n_2; \\ (\lambda n_1 + n_3)\, x_2 + (n_1 + n_3)\, x_3 = n_3. \end{cases} \tag{7}$$

The main determinant of each of the latter systems is equal to zero. The auxiliary determinants of the system (6) are

$$\Delta_0 = n_0\,(n_1 + n_3) - n_1\,(n_0 + n_2) = (n_0 n_3 - n_1 n_2) = 0.$$

$$\Delta_1 = n_1\,(\lambda n_0 + n_2) - n_0\,(\lambda n_1 + n_3) = n_1 n_2 - n_0 n_3 = 0.$$

For the system (6) we have p solutions described by the formula

$$x_1 = \frac{n_0 - (\lambda n_0 + n_2)\, x_0}{n_0 + n_2},$$

where $x_0 = 0, 1, \ldots, p-1$, if $n_0 + n_2 \neq 0$, or by the formula

$$x_0 = \frac{n_0 - (n_0 + n_2)\, x_1}{\lambda n_0 + n_2},$$

where $x_1 = 0, 1, \ldots, p-1$, if $\lambda n_0 + n_2 \neq 0$. The auxiliary determinants of the system (7) are also equal to zero:

$$\Delta_2 = n_2\,(n_1 + n_3) - n_3\,(n_0 + n_2) = n_1 n_2 - n_0 n_3 = 0.$$

$$\Delta_3 = n_3\,(\lambda n_0 + n_2) - n_2\,(\lambda n_1 + n_3) = \lambda\,(n_0 n_3 - n_1 n_2) = 0.$$

The system (7) has p solutions described by the formula $x_3 = \frac{n_2 - (\lambda n_0 + n_2) x_2}{n_0 + n_2}$, where $x_2 = 0, 1, \ldots, p-1$, if $n_0 + n_2 \neq 0$, or by the formula $x_2 = \frac{n_2 - (n_0 + n_2) x_3}{\lambda n_0 + n_2}$, where $x_3 = 0, 1, \ldots, p-1$, if $\lambda n_0 + n_2 \neq 0$.

Thus, for the non-invertible vector N that satisfies the condition $n_0 + n_2 \neq 0$ there exist p^2 different left-sided units $L = (l_0, l_1, l_2, l_3)$ described by the formula

$$L = \left(d, \frac{n_0 - (\lambda n_0 + n_2)\, d}{n_0 + n_2}, h, \frac{n_2 - (\lambda n_0 + n_2)\, h}{n_0 + n_2} \right), \tag{8}$$

where $d, h = 0, 1, \ldots, p-1$. Using the non-invertibility condition (5), one can easily derive the following formula describing all local left-sided units that are non-invertible vectors:

$$L' = \left(d, \frac{n_0 - (\lambda n_0 + n_2)\, d}{n_0 + n_2}, \frac{n_2}{n_0} d, \frac{n_0 n_2 - (\lambda n_0 + n_2)\, n_2 d}{n_0^2 + n_0 n_2} \right), \quad (9)$$

where $d = 0, 1, \ldots, p-1$. Since the set (9) includes p different non-invertible vectors, one can conclude that the set (8) contains $p^2 - p$ invertible and p non-invertible elements of the considered 4-dimensional FNAA.

To derive the formula describing all local right-sided units relating to the non-invertible vector N, one is to consider the solutions of the vector equation $N \circ X = N$ that can be reduced to the following two systems of equations:

$$\begin{cases} (\lambda n_0 + n_1)\, x_0 + (n_0 + n_1)\, x_2 = n_0; \\ (\lambda n_2 + n_3)\, x_0 + (n_2 + n_3)\, x_2 = n_2; \end{cases} \quad (10)$$

$$\begin{cases} (\lambda n_0 + n_1)\, x_1 + (n_0 + n_1)\, x_3 = n_1; \\ (\lambda n_2 + n_3)\, x_1 + (n_2 + n_3)\, x_3 = n_3. \end{cases} \quad (11)$$

Each of the systems (10) and (11) has the main determinant equal to zero. The auxiliary determinants of each of the systems (10) and (11) are equal to zero. Therefore, for the system (10) we have p solutions described by the formula $x_2 = \frac{n_0 - (\lambda n_0 + n_1) x_0}{n_0 + n_1}$, where $x_0 = 0, 1, \ldots, p-1$, if $n_0 + n_1 \neq 0$, or by the formula $x_0 = \frac{n_0 - (n_0 + n_1) x_2}{\lambda n_0 + n_1}$, where $x_1 = 0, 1, \ldots, p-1$, if $\lambda n_0 + n_1 \neq 0$.

For the system (11) we have p solutions described by the formula $x_3 = \frac{n_1 - (\lambda n_0 + n_1) x_1}{n_0 + n_1}$, where $x_2 = 0, 1, \ldots, p-1$, if $n_0 + n_1 \neq 0$, or by the formula $x_1 = \frac{n_1 - (n_0 + n_1) x_3}{\lambda n_0 + n_1}$, where $x_3 = 0, 1, \ldots, p-1$, if $\lambda n_0 + n_1 \neq 0$.

Thus, p^2 different right-sided units $R = (r_0, r_1, r_2, r_3)$ relate to the non-invertible vector N satisfying the condition $n_0 + n_1 \neq 0$, and the set of the R-units is described by the formula

$$R = \left(d, h, \frac{n_0 - (\lambda n_0 + n_1)\, d}{n_0 + n_1}, \frac{n_1 - (\lambda n_0 + n_1)\, h}{n_0 + n_1} \right), \quad (12)$$

where $d, h = 0, 1, \ldots, p-1$. One can easily derive the formula describing all of p local right-sided units that are non-invertible vectors and show that the set (12) includes $p^2 - p$ invertible and p non-invertible 4-dimensional vectors. The sets (8) and (12) contain p common vectors among which only one vector E_N is non-invertible. These p units are local two-sided units. The single local two-sided unit E_N relating to the vector N represents the unit of the cyclic group generated by N.

2.2 A second 4-dimensional FNAA

To obtain a higher performance of the signature scheme one can define the vector multiplication operation using a BVMT containing eight cells with the structural constant equal to zero. The appropriate BVMT defining the 4-dimensional FNAA with global two-sided unit $E = \left(\mu^{-1}, \lambda^{-1}, 0, 0\right)$ is shown as Table 2. Every vector $A = (a_0, a_1, a_2, a_3)$ is invertible, if the following invertibility condition holds true

$$a_0 a_1 \neq a_2 a_3. \tag{13}$$

Respectively, the vector $N = (n_0, n_1, n_2, n_3)$ is a non-invertible vector, if the following non-invertibility condition holds true

$$n_0 n_1 = n_2 n_3. \tag{14}$$

The algebra contains $p^3 + p^2 - p$ non-invertible vectors and

$$\Omega = p(p-1)\left(p^2 - 1\right)$$

invertible ones, exactly like in the case of FNAA defined by Table 1.

One can derive the following formulas describing the sets of the local left-sided units, local right-sided units, and local two-sided units relating to the non-invertible vector N:

$$L_N = \left(d,\ h,\ \frac{n_1}{\mu n_3}\left(1 - \lambda h\right),\ \frac{n_0}{\lambda n_2}\left(1 - \mu d\right)\right), \tag{15}$$

where $d, h = 0, 1, \ldots, p-1$;

Table 2. The BVMT setting the 4-dimensional FNAA ($\lambda \neq 0$; $\mu \neq 0$).

$\circ$	$\mathbf{e}_0$	$\mathbf{e}_1$	$\mathbf{e}_2$	$\mathbf{e}_3$
$\mathbf{e}_0$	$\mu\mathbf{e}_0$	0	0	$\mu\mathbf{e}_3$
$\mathbf{e}_1$	0	$\lambda\mathbf{e}_1$	$\lambda\mathbf{e}_2$	0
$\mathbf{e}_2$	$\mu\mathbf{e}_2$	0	0	$\mu\mathbf{e}_1$
$\mathbf{e}_3$	0	$\lambda\mathbf{e}_3$	$\lambda\mathbf{e}_0$	0

$$R_N = \left(d, h,\ \frac{n_0}{\lambda n_3}\left(1-\mu d\right),\ \frac{n_1}{\mu n_2}\left(1-\lambda h\right)\right), \tag{16}$$

where $d, h = 0, 1, \ldots, p-1$;

$$E_N = \left(d,\ \frac{\lambda n_1 - \mu n_0 + \mu^2 n_0 d}{\lambda^2 n_1},\ \frac{n_0}{\lambda n_3}\left(1-\mu d\right),\ \frac{n_0}{\lambda n_2}\left(1-\mu d\right)\right), \tag{17}$$

where $d = 0, 1, \ldots, p-1$.

Each of the sets (15) and (16) includes $p^2 - p$ invertible vectors and p non-invertible ones. The set (17) includes $p-1$ invertible vectors and one non-invertible vector.

3 The hidden DLP and a masked form of it

The DLP is set in a cyclic group Γ of prime order q as follows: $Y' = Z'^x$, where Z' is a generator of the group Γ and the value $x < q$ is unknown integer. Computation of the value x, when the group elements Z' and Y' are known, is called DLP. The HDLP is set so that at least one of the values Z' and Y' is masked (hidden). When setting the HDLP, the cyclic group Γ is set as a subset of elements of a FNAA. The exponentiation operation Z'^x contributes mainly to the hardness of both the DLP and the HDLP, therefore it is called the base operation. The auxiliary operations used to mask the values Z' and Y' are called the masking operations. When developing the HDLP-based PK cryptoscheme, one

should use the masking operations that are mutually commutative with the base exponentiation operation. The automorphism-map [8], [17] and homomorphism-map [14] operations are examples of masking operations. A particular form of the HDLP is defined by the used masking operations. Development of the PK cryptoschemes of different types is connected with applying different versions of the HDLP. When developing the PK cryptoschemes, to have possibility to select elements of order q having large size ($\geq$ 256 bits), usually there are used the FNAAs defined over the field $GF(p)$ with characteristic $p = 2q + 1$, where q is a prime.

For the first time HDLP was applied for development of the public key-agreement scheme [8]. That form of HDLP includes masking only one of two elements Z' and Y' and can be defined as follows:

HDLP: Given a FNAA, an algebra element Z' generating a cyclic group of prime order q, an invertible element Q of order q, which satisfies condition $Z' \circ Q \neq Q \circ Z'$, and an element $Y = \psi_{Q^w}(Z'^x) = Q^w \circ Z'^x \circ Q^{-w}$, where $w < q$ and $x < q$ are non-negative integers; ψ_{Q^w} is an automorphism-map operation; find the algebra element Q^w and integer x.

In the signature scheme [17] an enhanced form of the HDLP is used, which is characterized in masking both the element Z' and the element Y'.

Enhanced HDLP: Given a FNAA and non-invertible algebra elements $Z = \psi_G(Z'^x) = G \circ Z'^x \circ G^{-1}$, $Y = \psi_Q(Y'^x) = Q \circ Z'^x \circ Q^{-1}$, and invertible element $T = Q \circ G^{-1}$, where invertible elements Q and G have order q; the conditions $Z' \circ Q \neq Q \circ Z'$, $Z' \circ G \neq G \circ Z'$, and $G \circ Q \neq Q \circ G$ holds true; $1 < x < q$ is a non-negative integer; find the value of x.

One can note that in the latter definition only the value x is to be found, since the signature can be computed using the value x (discrete logarithm in the hidden cyclic group generated by the element Z') and the public parameters Z and Y. When using a non-invertible algebra element N as parameter Z', one can hide each of the elements $Z' = N$ and $Y' = N^x$ performing on it two different masking operations, namely, the ψ operation and an additional operation connected with

multiplication by a local single-sided unit.

For example, the first masking operation can be implemented in the form $\varphi_L(N^x) = N^x \circ L_N$ to transform the vector N^x and in the form $\varphi_R(N) = R_N \circ N$ to transform the vector N. The automorphism-map operations ψ_Q and ψ_G can be used as the second masking operation when transforming the vectors N^x and N correspondingly: $Y = \psi_Q(\varphi_L(N^x)) = Q \circ N^x \circ L_N \circ Q^{-1}$ and $Z = \psi_G((\varphi_R(N))) = G \circ R_N \circ N \circ G^{-1}$. One can easily show that the following equalities hold true for arbitrary non-negative integers k and x:

$$(\varphi_L(N^x))^k = \varphi_L\left((N^x)^k\right); \quad (\varphi_R(N))^k = \varphi_R\left(N^k\right).$$

Thus, each of the said additional masking operations is mutually commutative with ψ operation. Due to such property of the used masking operations the developed signature scheme (see next Section 4) performs correctly. The introduced signature scheme is based on the following form of the HDLP.

Masked HDLP: Given a FNAA, non-invertible algebra elements $Z = \psi_G(\varphi_L(N)) = G \circ N \circ L_N \circ G^{-1}$ and $Y = \psi_Q(\varphi_R(N^x)) = Q \circ R_N \circ N^x \circ Q^{-1}$, and invertible element $T = Q \circ L_N^{-1} \circ G^{-1}$, where N is a non-invertible algebra element generating a cyclic group of order q; L_N is local left-sided unit for N; R_N is local right-sided unit for N; the invertible algebra elements Q and G have order q and satisfy the conditions $N \circ Q \neq Q \circ N$, $N \circ G \neq G \circ N$, and $G \circ Q \neq Q \circ G$; $1 < x < q$ is non-negative integer; find the value of x.

In Subsection 4.4, it is described method for computing the signature using public parameters Z and Y and the value x, therefore, to break the proposed signature scheme, it is sufficient to find only the unknown value of discrete logarithm in the hidden cyclic group generated by the unknown element N. In Subsection 4.5, it is shown that an algorithm for forging a signature can be used to compute the value x, i.e. the proposed signature scheme is as secure as the masked HDLP is computationally difficult.

4 A new signature scheme

4.1 Private and public keys

Each of the FNAA described in Section 2 can be used as algebraic support of the proposed signature scheme. The algebras are to be defined over the field $GF(p)$, where $p = 2q + 1$ and q are two prime numbers having large size ($\geq$ 256 bits). One can easily generate a random non-invertible vector N that has order equal to the prime q. Such vector defines a finite cyclic group of the order q. The vector N is one of the elements of private key. The next element of private key is the non-negative integer $x < q$ which is used to compute the vector N^x.

Procedure of the PK generation is as follows:

1. Generate a random invertible vector Q of the order q and a random local left-sided unit L_N, that is an invertible vector, which satisfy the following conditions: $Q \circ N \neq N \circ Q$ and $R_N \circ N \neq N \circ R_N$.
2. Compute the first element Y of the PK:

$$Y = Q \circ N^x \circ L_N \circ Q^{-1}$$

3. Generate a random invertible vector G of the order q and a random local right-sided unit R_N, that is an invertible vector, which satisfy the following conditions: $G \circ N \neq N \circ G$ and $R_N \circ N \neq N \circ R_N$.
4. Compute the second element Z of the PK:

$$Z = G \circ R_N \circ N \circ G^{-1}.$$

5. Compute the third element T of the PK:

$$T = Q \circ L_N^{-1} \circ G^{-1}.$$

This procedure outputs the PK in form of the triple of the vectors (Y, Z, T).

All other elements used to generate the PK are secrete and part of them represent the private key in the form of the integer x and the triple of the vectors $\left(Q, N, G^{-1}\right)$. Other secret elements are not attributed to the private key, since they are not used in the signature computation procedure.

4.2 Signature generation and verification procedures

Generation of the signature (e, s) to the electronic document M is to be performed as follows:

1. Select a random integer $k < q$ and, using the elements Q and G^{-1} of the private key, compute the vector

$$V = Q \circ N^k \circ G^{-1}.$$

2. Using a specified hash-function F_h that satisfies the collision-resistance requirement, compute the value e of the hash function from the document M to which the vector V is attached: $e = F_h(M, V)$.

3. Compute the value $s = k - xe \bmod q$.

One should note that for signing a document M unique integer k is to be used. If two different documents are signed using the same value k, then one can compute the private value x from two signatures. Therefore the value of k is to be generated at random.

Signature verification procedure is executed as follows:

1. Using the PK (Y, Z, T) compute the vector

$$V' = Y^e \circ T \circ Z^s.$$

2. Compute the value $e' = F_h(M, V')$.

3. If $e' = e$, then the signature is accepted as genuine. Otherwise it is rejected.

4.3 Correctness of the signature scheme

Correctness proof of the signature scheme is as follows:

$$\begin{aligned} V' &= Y^e \circ (T) \circ Z^s = \\ &= \left(Q \circ N^x \circ L_N \circ Q^{-1}\right)^e \circ (T) \circ \left(G \circ R_N \circ N \circ G^{-1}\right)^s = \\ &= Q \circ N^{ex} \circ L_N \circ Q^{-1} \circ Q \circ L_N^{-1} \circ G^{-1} \circ G \circ R_N \circ N^s \circ G^{-1} = \\ &= Q \circ N^{ex+s} \circ G^{-1} = Q \circ N^{ex+k-ex} \circ G^{-1} = Q \circ N^k \circ G^{-1} = V \Rightarrow \\ &\qquad \Rightarrow F_h(M, V') = F_h(M, v) \Rightarrow e' = e. \end{aligned} \tag{18}$$

Thus, the signature (e, s) computed correctly will pass the verification procedure as genuine signature.

4.4 Alternative procedure for computing a signature

One can significantly reduce the size of the private key replacing the described signature verification procedure by the following one:

1. Select two random integers $k_1 < q$ and $k_2 < q$. Then, using the PK (Y, Z, T), compute the vector

$$V = Y^{k_1} \circ T \circ Z^{k_2}.$$

2. Compute the value $e = F_h(M, V)$.
3. Compute the value $s = k_2 + (k_1 - e)\, x \bmod q$.

Using the last version of the signature generation procedure one gets the private key in the form of one integer value x. However, after such modification, computing the signature will require performing one additional exponentiation operation.

The existence of an alternative signature generation procedure shows that it is sufficient to know only one secret value to forge a signature, namely, the value of x. In Subsection 4.5 this fact is used to perform formal security proof for the developed signature scheme.

4.5 On formal security proof

The method [19], proposed for providing formal security proof of the Schnorr DLP-based signature scheme [18], is well applicable to the developed HDLP-based signature scheme. Like in the Schnorr signature algorithm [18], in the developed signature scheme during the signature generation process the base exponentiation operation N^k is performed before computation of the first signature element that is the hash value $e = F_h(M, V)$, where $V = Q \circ N^k \circ G^{-1}$. In the formal security proof [19] one supposes that the hash function F_h is free of some properties that the signature forger can take advantage of [20]. Such assumption is reasonable in the case of using a collision-resistant hash-function.

In the method [19] it is considered a forger that can compute the signature element equally well for different hash functions F_h and F'_h. Suppose the forger runs two computer programs that use the same input data and the same random integer k, but different hash functions.

He will get two signatures (e, s) and (e', s') with the fixed value V and values $e = F_h(M, V)$ and $e' = F'_h(M, V)$. For some fixed integer k and fixed hash-function value e (e') there exists unique value of the second signature element s (s'). Therefore, for two signatures computed by hypothetic forging computer program one can write the following equalities: $s = k - ex \bmod q$ and $s' = k - e'x \bmod q$ from which the forger can easily compute the private value $x = (s - s')(e' - e)^{-1} \bmod q$.

The question of how a computer program can calculate the value of V when the values of Q, N, and G are unknown requires explanation. From the alternative procedure for calculating the signature, it can be seen that the calculation by the formula $V = Y^{k_1} \circ T \circ Z^{k_2}$ gives the same result as the calculation by the formula $V = Q \circ N^k \circ G^{-1}$ at $k = k_1 x + k_2 \bmod q$. Thus, fixing two values k_1 and k_2 results in fixing the value k. Actually, due to the existence of an alternative signature regeneration procedure, the reductionist security proof method [19] works well for the developed signature scheme.

4.6 Computational complexity of the signature scheme

Computational complexity of the procedures for i) generating private and public keys, ii) computing a signature, and iii) verifying a signature can be estimated in multiplication operations in the field $GF(p)$ and in exponentiation operations in the used FNAA (see Table 3) taking into account that i) one exponentiation in the FNAA used as algebraic support equals on the average to 384 multiplications $\circ$, ii) computation of the value U^{-1} for some invertible vector U is performed as solving the vector equation $U \circ X = E$, and iii) the local units R_N and L_N are computed with the use of formulas (8) and (12) for the case of the first 4-dimensional algebra and (15) and (16) for the case of the second 4-dimensional algebra.

The obtained estimate results show that the proposed signature scheme is sufficiently fast. For example, computational complexity of the signature generation (signature verification for the case of 64-bit public exponent) in the 2048-bit RSA cryptoscheme can be evaluated as $\approx 3 \cdot 2^{16}$ ($\approx 3 \cdot 2^{11}$) multiplications in $GF(p)$ with 256-bit prime p.

Table 3. A rough estimate of the implementation complexity.

Procedure for	# multiplications in $GF(p)$ for the first (second) FNAA	# exponentiations in FNAA
generating keys	$< 3 \cdot 2^{11}$ $(< 3 \cdot 2^{10})$	< 2
computing signature	$3 \cdot 2^{10}$ $(3 \cdot 2^{9})$	1
verifying signature	$3 \cdot 2^{11}$ $(3 \cdot 2^{10})$	2
alternative computing signature	$3 \cdot 2^{11}$ $(3 \cdot 2^{10})$	2

5 Discussion and conclusion

The expected PQ security of the proposed signature scheme is connected with the fact that a periodic function constructed on the basis of public parameters of the scheme takes on the values from many different groups contained in the FNAA used as algebraic support. For example, the function $f(i,j) = Y^i \circ T \circ Z^j$ contains a period with the length depending on the value x, however, the values of $f(i,j)$ are not limited to the values of any one group. Indeed, this function can be represented in the following form:

$$f(i,j) = Q \circ N^{ix+j} \circ G^{-1} = F(i,j) \circ V,$$

where $F(i,j) = Q \circ N^{ix+j} \circ Q^{-1}$ is the function taking on the values in frame of the cyclic group generated by the generator $Q \circ N \circ Q^{-1}$ and the vector $V = Q \circ G^{-1}$ is fixed. Due to multiplying different elements belonging to a fixed cyclic group by a fixed vector that has value out of this group, the function $f(i,j)$ takes on values belonging to different groups contained in the FNAA, whereas a quantum computer effectively finds the period lengths of a function whose values lie within a given finite group [6], [7].

Two different 4-dimensional FNAA with a global two-sided unit have been considered as algebraic supports of the proposed signature scheme. However, the 6-dimensional and 8-dimensional FNAAs repre-

sent significant interest for implementing other versions of the proposed signature scheme. Probably, using the FNAAs with dimension $m \geq 6$ it is reasonable to invent some other signature schemes such that their public parameters will not allow one to compose the periodic functions containing a period having the length depending on the private value x. Such potential signature schemes are particularly interesting as candidates for PQ PK cryptoschemes.

Acknowledgement. The authors sincerely thank the anonymous Referee for his comments and detailed advice on improving the content of the article.

This work was supported by the budget theme No. 0060-2019-010.

References

[1] *Post-Quantum Cryptography. 8th International Conference, PQCrypto 2017 Proceedings*, Utrecht, The Netherlands, June 26-28, 2017 (Lecture Notes in Computer Science, vol. 10346), 2017.

[2] *Post-Quantum Cryptography. 9th International Conference, PQCrypto 2018 Proceedings*, Fort Lauderdale, FL, USA, April 9-11, 2018 (Lecture Notes in Computer Science, vol. 10786), 2018.

[3] Submission Requirements and Evaluation Criteria for the Post-Quantum Cryptography Standardization Process. NIST PQCrypto project. https://csrc.nist.gov/CSRC/media/Projects/Post-Quantum-Cryptography/documents/call-for-proposals-final-dec-2016.pdf

[4] *Post-Quantum Cryptography. Proceedings of the 10th International Conference, PQCrypto 2019, Chongqing, China, May 8−10, 2019*, (Lecture Notes in Computer Science, vol. 11505), 2019.

[5] P. W. Shor, "Polynomial-time algorithms for prime factorization and discrete logarithms on quantum computer," *SIAM Journal of Computing*, vol. 26, pp. 1484–1509, 1997.

[6] A. Ekert and R. Jozsa, "Quantum computation and Shorś factoring algorithm," *Rev. Mod. Phys.*, vol. 68, p. 733, 1996.

[7] R. Jozsa, "Quantum algorithms and the fourier transform," *Proc. Roy. Soc. London Ser A*, vol. 454, pp. 323 – 337, 1998.

[8] D.N. Moldovyan, "Non-Commutative Finite Groups as Primitive of Public-Key Cryptoschemes," *Quasigroups and Related Systems*, vol. 18, no. 2, pp. 165–176, 2010.

[9] Q. Alamelou, O. Blazy, S. Cauchie, and Ph. Gaborit, "A code-based group signature scheme," *Designs, Codes and Cryptography*, vol. 82, no. 1–2, pp. 469–493, 2017.

[10] P. Hiranvanichakorn, "Provably Authenticated Group Key Agreement based on Braid Groups: The Dynamic Case," *International Journal of Network Security*, vol. 19, no. 4, pp. 517–527, 2017.

[11] A. A. Moldovyan and N. A. Moldovyan, "Post-quantum signature algorithms based on the hidden discrete logarithm problem," *Computer Science Journal of Moldova*, vol. 26, no. 3(78), pp. 301–313, 2018.

[12] N. A. Moldovyan, "Finite Non-commutative Associative Algebras for Setting the Hidden Discrete Logarithm Problem and Post-quantum Cryptoschemes on Its Base," *Buletinul Academiei de Stiinte a Republicii Moldova. Matematica*, no. 1(89), pp. 71–78, 2019.

[13] N. A. Moldovyan and A. A. Moldovyan, "Finite Non-commutative Associative Algebras as carriers of Hidden Discrete Logarithm Problem," *Bulletin of the South Ural State University. Ser. Mathematical Modelling, Programming & Computer Software (Bulletin SUSU MMCS)*, vol. 12, no. 1, pp. 66–81, 2019.

[14] D. N. Moldovyan, "Post-quantum public key-agreement scheme based on a new form of the hidden logarithm problem," *Computer Science Journal of Moldova*, vol. 27, no. 1(79), pp. 56–72, 2019.

[15] D. N. Moldovyan, A. A. Moldovyan, and N. A. Moldovyan, "Post-quantum commutative encryption algorithm," *Computer Science Journal of Moldova*, vol. 27, no. 3(81), pp. 299–317, 2019.

[16] A. S. Kuzmin, V. T. Markov, A. A. Mikhalev, A. V. Mikhalev, and A. A. Nechaev, "Cryptographic Algorithms on Groups and

Algebras," *Journal of Mathematical Sciences*, vol. 223, no. 5, pp. 629–641, 2017.

[17] N. A. Moldovyan and I. K. Abrosimov, "Post-quantum electronic digital signature scheme based on the enhanced form of the hidden discrete logarithm problem," *Vestnik of Saint Petersburg University. Applied Mathematics. Computer Science. Control Processes*, vol. 15, no. 2, pp. 212–220, 2019. https://doi.org/10.21638/11702/spbu10.2019.205 (In Russian)

[18] C. P. Schnorr, "Efficient signature generation by smart cards," *Journal of Cryptology*, vol. 4, pp. 161–174, 1991.

[19] D. Pointcheval and J. Stern, "Security Arguments for Digital Signatures and Blind Signatures," *Journal of Cryptology*, vol. 13, pp. 361–396, 2000.

[20] N. Koblitz and A. J. Menezes, "Another Look at "Provable Security"," *Journal of Cryptology*, vol. 20, pp. 3–38, 2007.

D. N. Moldovyan, A. A. Moldovyan,
N. A. Moldovyan

Received January 18, 2020
Revised August 30, 2020
Accepted September 19, 2020

St. Petersburg Federal Research Center of
the Russian Academy of Sciences (SPC RAS),
St. Petersburg Institute for Informatics and
Automation of the Russian Academy of Sciences
14 Liniya, 39, St.Petersburg, 199178
Russia
E–mail: nmold@mail.ru

Computer Science Journal of Moldova, vol.**28**, no.3(84), 2020

Investigation of Some Cryptographic Properties of the 8x8 S-boxes Created by Quasigroups

Aleksandra Mileva, Aleksandra Stojanova,
Dušan Bikov, Yunqing Xu

Abstract

We investigate several cryptographic properties in 8-bit S-boxes obtained by quasigroups of order 4 and 16 with several different algebraic constructions. Additionally, we offer a new construction of N-bit S-boxes by using different number of two layers – the layer of bijectional quasigroup string transformations, and the layer of modular addition with N-bit constants. The best produced 8-bit S-boxes so far are regular and have algebraic degree 7, nonlinearity 98 (linearity 60), differential uniformity 8, and autocorrelation 88. Additionally we obtained 8-bit S-boxes with nonlinearity 100 (linearity 56), differential uniformity 10, autocorrelation 88, and minimal algebraic degree 6. Relatively small set of performed experiments compared with the extremly large set of possible experiments suggests that these results can be improved in the future.

Keywords: 8-bit S-boxes, nonlinearity, differential uniformity, autocorrelation.

MSC 2010: 20N05, 94A60.

1 Introduction

The main building blocks for obtaining confusion in all modern block ciphers are so called substitution boxes, or S-boxes. Usually, they work with much less data (for example, 4 or 8 bits) than the block size, so they need to be highly nonlinear. Two of the most successful attacks against modern block ciphers are linear cryptanalysis (introduced by

Matsui [21]), which exploits input-output correlation, and differential cryptanalysis (introduced by Biham and Shamir [2]), which exploits difference propagation.

Designers of block ciphers very often choose S-boxes with special cryptographic properties, which means high nonlinearity (or low linearity), low differential uniformity, high algebraic degree, low autocorrelation and regularity (balance). AES S-box is the example of the best found 8x8 S-boxes, which is optimal with respect to most of these cryptographic properties. It has nonlinearity 112 (or linearity 32), algebraic degree 7, differential uniformity 4, and autocorrelation 32.

Mihajloska and Gligoroski [26] constructed optimal 4x4 S-boxes from quasigroups of order 4, by using four e quasigroup string transformations. Motivated by their work, we offer two constructions of the 8x8 S-boxes from quasigroups of order 4 and 16, by using different number of e quasigroup string transformations. Main contribution of this paper is a new construction of N-bit S-boxes which uses different number of two layers – the layer of bijectional quasigroup string transformations, and the layer of modular addition with N-bit constants. Specifically, we demonstrated this construction method with quasigroups of order 4 and 16 and modular addition with 8-bit constants. Quasigroups of order 4 can be seen as 4x2 S-boxes, while quasigroups of order 16 can be seen as 8x4 S-boxes, so, we offer an algebraic construction of 8x8 S-boxes from smaller ones. We investigate some of the cryptographic properties of the obtained S-boxes, without looking at the cost of their implementation in hardware.

This paper has the following structure: Section 2 is about mathematical preliminaries for quasigroup string transformations, basics about n-ary Boolean functions and Boolean maps, and definition of some cryptographic properties for them. Some existing methods for the generation of 8-bit S-boxes, together with the best obtained values of the cryptographic properties of these S-boxes are presented in Section 3. Section 4 presents the two constructions of 8-bit S-boxes by using e quasigroup string transformations produced by quasigroups of order 4 and 16, together with the experimental results. The new construction of N-bit S-boxes and obtained experimental results are

presented in the Section 5. Finally, concluding remarks are given in Section 6.

2 Mathematical Preliminaries

2.1 Quasigroup String Transformations

A quasigroup $(Q, *)$ is a groupoid, i.e., a pair of nonempty set Q and a binary operation $*$, such that for all $a, b \in Q$ there exist unique $x, y \in Q$ satisfying the equalities $a * x = b$ and $y * a = b$ [1]. In the case when Q is finite, the multiplication table of $(Q, *)$ is a Latin square of order $|Q|$, where all rows and columns are permutations of Q. For the quasigroup operation $*$ on the set Q, another operation, a right division $\backslash$ can be derived by:

$$x \backslash y = z \Longleftrightarrow x * z = y.$$

Given a finite quasigroup $(Q, *)$, consider the set Q as an alphabet with word set $Q^+ = \{x_1 x_2 \dots x_t \mid x_i \in Q, t \geq 1\}$. For the fixed letter $l \in Q$ (called a leader), the transformations $e_l, d_l : Q^+ \to Q^+$ are defined in [20], as follows:

$$e_l(x_1 \dots x_t) = (z_1 \dots z_t) \Longleftrightarrow z_j = \begin{cases} l * x_1, & j = 1 \\ z_{j-1} * x_j, & 2 \leq j \leq t \end{cases}, \qquad (1)$$

$$d_l(z_1 \dots z_t) = (x_1 \dots x_t) \Longleftrightarrow x_j = \begin{cases} l \backslash z_1, & j = 1 \\ z_{j-1} \backslash z_j, & 2 \leq j \leq t \end{cases}. \qquad (2)$$

Any combination of these elementary quasigroup string transformations is a permutation and d_l is an inverse to e_l. Linear quasigroups produce linear e_l and d_l quasigroup string transformations [25]. Additionally, some non-linear quasigroups always produce linear e_l and d_l transformations. For example, there is a set of 48 non-linear quasigroups of order 4 that always produce linear e_l and d_l transformations [25]. In the rest of the paper we will use e and d instead e_l and d_l, respectively.

Definition 1. *[14] A finite quasigroup $(Q, *)$ of order r is said to be shapeless if and only if it is non-idempotent, non-commutative, non-associative, it does not have neither left nor right unit, it does not contain proper sub-quasigroups, and there is no $k < 2r$ such that identities of the kinds*

$$\underbrace{x * (x \cdots * (x}_{k} * y)) = y, \;\; y = ((y * \underbrace{x) * \cdots x) * x}_{k}$$

are satisfied in $(Q, *)$.

2.2 Some Cryptographic Properties of n-ary Boolean Functions and Boolean Maps

Let $\mathbb{F}_2$ denote the Galois field with two elements, and let $\mathbb{F}_2^n$ denote the vector space of binary n-tuples over $\mathbb{F}_2$ with respect to addition $\oplus$ (Boolean function XOR) and scalar multiplication (Boolean function conjunction). There is a correspondence between $\mathbb{F}_2^n$ and $\mathbb{Z}_{2^n}$ via

$$\varphi_1 : \mathbb{F}_2^n \to \mathbb{Z}_{2^n} : \mathbf{x} = (x_1, \ldots x_n) \to x = \sum_{i=1}^{n} x_i 2^{i-1},$$

and there is a correspondence between $\mathbb{F}_2^n$ and $\mathbb{F}_{2^n}$ via

$$\varphi_2 : \mathbb{F}_2^n \to \mathbb{F}_{2^n} : \mathbf{x} = (x_1, \ldots x_n) \to x = \sum_{i=1}^{n} x_i \beta_i,$$

where $\{\beta_1, \ldots, \beta_n\}$ is a basis of $\mathbb{F}_{2^n}$ over $\mathbb{F}_2$.

An $n-$*ary Boolean function* is a function $f : \mathbb{F}_2^n \to \mathbb{F}_2$. A *Boolean map* (or vector valued Boolean function or vectorial Boolean function) is a map $S : \mathbb{F}_2^n \to \mathbb{F}_2^m$, $(m \geq 1)$. Every Boolean map S can be represented by m $n-$ary Boolean functions $f_i : {\mathbb{F}_2}^n \to \mathbb{F}_2$, called *coordinate functions* of S, as follows:

$$S(x_1, \ldots, x_n) = (f_1(x_1, \ldots, x_n), f_2(x_1, \ldots x_n), \ldots, f_m(x_1, \ldots, x_n)).$$

Each $n-$ary Boolean function f_i can be represented in Algebraic Normal Form as

$$f_i(x_1, x_2, \ldots, x_n) = \bigoplus_{I \subseteq \{1,2,\ldots,n\}} \alpha_I (\prod_{i \in I} x_i), \tag{3}$$

where $\alpha_I \in \mathbb{F}_2$. The right-hand side of (3) can be interpreted as a polynomial in the field $(\mathbb{F}_2, \oplus, \cdot)$ and the *algebraic degree* of f_i, $deg(f_i)$, is taken to be the degree of the polynomial.

Definition 2. *The (minimal) algebraic degree of a Boolean map S is defined as the minimum of the algebraic degrees of its non-trivial coordinate functions $(f_1, f_2, \ldots, f_m)$, and it can be expressed as:*

$$deg(S) = \min_{\boldsymbol{u} \in \mathbb{F}_2^m \setminus \{\boldsymbol{0}\}} deg(u_1 f_1 \oplus u_2 f_2 \oplus \ldots \oplus u_m f_m).$$

If $deg(f_i) \leq 1, \forall i \in \{1, 2, \ldots, m\}$, S is an *affine function*. A *linear function* is a non-constant affine function S for which $S(\mathbf{0}) = 0$.

The *(Hamming) weight* of a vector $\mathbf{x} \in \mathbb{F}_2^n$ is equal to the number of components equal to 1 and is denoted by $wt(\mathbf{x})$. The *(Hamming) distance* between two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{F}_2^n$, denoted by $d(\mathbf{x}, \mathbf{y})$ is the number of positions in which they differ. The *(Hamming) weight* of a Boolean function f, $wt(f)$, is the number of function values equal to 1. A Boolean function f is *balanced* if and only if $wt(f) = 2^{n-1}$.

For two vectors $\mathbf{x} = (x_1, \ldots, x_n), \mathbf{y} = (y_1, \ldots, x_y) \in \mathbb{F}_2^n$, the *inner product* or *scalar product* is defined as $\mathbf{x} \cdot \mathbf{y} = \bigoplus_{i=1}^{n} x_i y_i$. A *selection vector* $\mathbf{a}$ is a binary vector that selects all components i of a vector that have $a_i = 1$. By $\mathbf{a} \cdot \mathbf{x}$ (or $\mathbf{a}^T \mathbf{x}$) the linear combination of the components of a vector $\mathbf{x}$ selected by $\mathbf{a}$, analogous to vector inner product, can be represented. A linear Boolean function $\varphi_{\mathbf{a}} = \mathbf{a} \cdot \mathbf{x}$ is completely specified by its corresponding selection vector $\mathbf{a}$.

A *bias* of an $n-$ary Boolean function f is defined as

$$\varepsilon(f) = \sum_{\mathbf{x} \in \mathbb{F}_2^n} (-1)^{f(\mathbf{x})} = 2^n - 2wt(f).$$

So, balanced Boolean functions have $\varepsilon(f) = 0$.

An $n-$ary Boolean function f on $\mathbb{F}_2^n$ is uniquely determined by its Walsh-Hadamard transform (WHT). The Walsh-Hadamard transform $W_f : \mathbb{F}_2^n \to \mathbb{R}$ of f is defined for all $\mathbf{x} \in \mathbb{F}_2^n$ as

$$W_f(\mathbf{x}) = \sum_{\mathbf{a} \in \mathbb{F}_2^n} (-1)^{f(\mathbf{a}) \oplus \mathbf{a} \cdot \mathbf{x}} (= \varepsilon(f \oplus \varphi_{\mathbf{a}})), \tag{4}$$

where $W_f(\mathbf{x}) \in [-2^n, 2^n]$ and is known as a spectral Walsh coefficient. The real-valued vector of all spectral Walsh coefficients is known as a WHT Spectrum. The WHT spectrum of f corresponds to the biases of all approximations of f by a linear function.

The Walsh transform $W_S : \mathbb{F}_2^n \times \mathbb{F}_2^m \to \mathbb{R}$ of a Boolean map S is defined for all pairs $(\mathbf{u}, \mathbf{v}) \in (\mathbb{F}_2^n, \mathbb{F}_2^m)$ as

$$W_S(\mathbf{u}, \mathbf{v}) = \sum_{\mathbf{a} \in \mathbb{F}_2^n} (-1)^{\mathbf{u} \cdot \mathbf{a} \oplus \mathbf{v} \cdot S(\mathbf{a})}. \tag{5}$$

Definition 3. *The nonlinearity of an $n-$ary Boolean function f (introduced in [22]), denoted by $NL(f)$, is defined as the distance to the nearest affine function on $\mathbb{F}_2^n$. It can be expressed in terms of the spectral Walsh coefficients by*

$$NL(f) = 2^{n-1} - \frac{1}{2} \max_{\boldsymbol{x} \in \mathbb{F}_2^n} |W_f(\boldsymbol{x})|. \tag{6}$$

$NL(f) = 0$ iff f is affine function.

Definition 4. *The linearity of an $n-$ary Boolean function f, denoted by $L(f)$, is defined as*

$$L(f) = \max_{\boldsymbol{x} \in \mathbb{F}_2^n} |W_f(\boldsymbol{x})|. \tag{7}$$

Linearity and nonlinearity of a given $n-$ary Boolean function f are connected by the following equation:

$$L(f) + 2NL(f) = 2^n. \tag{8}$$

For $L(f)$, the inequality $2^{\frac{n}{2}} \leq L(f) \leq 2^n$ holds. $L(f) = 2^n$ iff f is affine function. Boolean functions for which $L(f) = 2^{\frac{n}{2}}$ are called *bent*

functions (introduced by Rothaus [31]), and they exist only for even n. Because bent functions are highly biased ($\varepsilon(f) = \pm 2^{\frac{n}{2}}$), they are of little use in cryptography.

Linear approximation table for Boolean map S is a $2^n \times 2^m$ table whose entries are defined for all pairs $(\mathbf{u}, \mathbf{v}) \in (\mathbb{F}_2^n, \mathbb{F}_2^m)$ as

$$LAT_S(\mathbf{u}, \mathbf{v}) = W_S(\mathbf{u}, \mathbf{v}).$$

Definition 5. *The nonlinearity and linearity of a Boolean map S [29] are defined as*

$$NL(S) = \min_{\boldsymbol{v} \in \mathbb{F}_2^n \setminus \{\boldsymbol{0}\}} NL(\boldsymbol{v} \cdot S) = 2^{n-1} - \frac{1}{2} \max_{\boldsymbol{a} \neq \boldsymbol{0}, \boldsymbol{b} \in \mathbb{F}_2^m} |W_S(\boldsymbol{a}, \boldsymbol{b})|, \quad (9)$$

$$L(S) = \max_{\boldsymbol{v} \in \mathbb{F}_2^m \setminus \{\boldsymbol{0}\}} L(\boldsymbol{v} \cdot S), \quad (10)$$

where $\boldsymbol{v} \cdot S = \bigoplus_{i=1}^{n} v_i f_i$ is the linear combination of the coordinate functions of S.

$L(S) \geq 2^{\frac{n}{2}}$, and Nyberg [27] showed that equality can hold only if $n \geq 2m$ and n is even. For $n = m$, $L(S) \geq 2^{\frac{n+1}{2}}$ with equality for odd n only (Chabaud-Vaudenay theorem [6]) . The functions achieving this bound are called *almost bent functions*. Because for even n and $n = m$, some $n \times n$ S-boxes with $L(S) = 2^{\frac{n+2}{2}}$ are known, Dobbertin [12] conjectured that this value is the minimum.

From linear approximation table, one can easily calculate *linear probability bias* $\varepsilon_S(\mathbf{u}, \mathbf{v})$, which is amount by which the probability of a linear expression holding deviates from $\frac{1}{2}$. The formula is $\varepsilon_S(\mathbf{u}, \mathbf{v}) = LAT_S(\mathbf{u}, \mathbf{v})/2^n - \frac{1}{2}$.

Difference distribution table for Boolean map S is a $2^n \times 2^m$ table whose entries are defined for all pairs $(\mathbf{u}, \mathbf{v}) \in (\mathbb{F}_2^n, \mathbb{F}_2^m)$ as

$$DDT_S(\mathbf{u}, \mathbf{v}) = \sharp\{\mathbf{x} \in \mathbb{F}_2^n | S(\mathbf{x}) \oplus S(\mathbf{x} \oplus \mathbf{u}) = \mathbf{v}\}.$$

Definition 6. *The differential uniformity of a Boolean map S [28], denoted by $\Delta(S)$, is defined as*

$$\Delta(S) = \max_{\boldsymbol{u} \in \mathbb{F}_2^n \setminus \{\boldsymbol{0}\}, \boldsymbol{v} \in \mathbb{F}_2^m} DDT_S(u, v). \quad (11)$$

For differential uniformity, $\Delta(S) \geq max\{2, 2^{n-m}\}$ holds, and for $n \geq m$, $\Delta(S)$ takes only even values in $[2^{n-m}, 2^n]$. Nyberg [27] showed that for $n > m$, $\Delta(S) = 2^{n-m}$ if and only if $n \geq 2m$ and n is even. This kind of functions are known as *perfect nonlinear* functions and they are the same as the bent functions. For $n \leq m$, $\Delta(S) = 2$, and this kind of functions are known as *almost perfect nonlinear*. So, bijective S-boxes can have the smallest differential uniformity of 2, and there are examples for odd n.

For the n-ary Boolean function f on $\mathbb{F}_2^n$ one can define an Autocorrelation transform (ACT) $ACT_f : \mathbb{F}_2^n \to \mathbb{R}$ for all $\mathbf{x} \in \mathbb{F}_2^n$ as

$$ACT_f(\mathbf{x}) = \sum_{\mathbf{a} \in \mathbb{F}_2^n} (-1)^{f(\mathbf{a}) \oplus f(\mathbf{a} \oplus \mathbf{x})}, \tag{12}$$

where $ACT_f(\mathbf{x}) \in [-2^n, 2^n]$ is known as a spectral autocorrelation coefficient and $ACT_f(\mathbf{0}) = 2^n$. The real-valued vector of all spectral autocorrelation coefficients is known as an ACT Spectrum.

Definition 7. *The Absolute indicator of an n-ary Boolean function f, denoted by $AC(f)$, is defined as the maximal non-trivial absolute spectral autocorrelation coefficient, or*

$$AC(f) = \max_{\boldsymbol{x} \in \mathbb{F}_2^n \setminus \{\boldsymbol{0}\}} |ACT_f(\boldsymbol{x})|. \tag{13}$$

Definition 8. *The Absolute indicator of a Boolean map S is defined as*

$$AC(S) = \max_{\boldsymbol{v} \in \mathbb{F}_2^m \setminus \{\boldsymbol{0}\}} AC(\boldsymbol{v} \cdot S). \tag{14}$$

Two n-ary Boolean functions f and g belong to the same *equivalence class* (or are *affine equivalent*) if and only if there exist some non-singular binary matrix D, vectors $\mathbf{a}, \mathbf{b} \in \mathbb{F}_2^n$ and a scalar $c \in \mathbb{F}_2$, such that $g(\mathbf{x}) = f(D\mathbf{x} \oplus \mathbf{a}) \oplus \mathbf{b} \cdot \mathbf{x} \oplus c$. (Two n-ary Boolean functions f and g are affine equivalent if there exists an affine permutation A of $\mathbb{F}_2^n$ such that $g(\mathbf{x}) = f(A(\mathbf{x}).)$

The algebraic degree, nonlinearity and absolute indicator are invariant under affine equivalence [22], [30]. Any Boolean map S and its

inverse have same linearity and differential uniformity (Nyberg [29]). In the same paper, Nyberg proved that differential uniformity is invariant under affine permutations onto the input space and the output space. Two S-boxes S_1 and S_2 are *affinely equivalent* if there exist two affine permutations A_1 and A_2, such that $S_2 = A_2 \circ S_1 \circ A_1$. So, the affinely equivalent S-boxes have same differential uniformity (and same algebraic degree, nonlinearity and absolute indicator).

Definition 9. *A Boolean map S is regular if and only if all non-zero its coordinate functions are balanced.*

With other words, this means that when $n \geq m$, for each output $y \in \mathbb{F}_2^m$ there are exactly 2^{n-m} inputs that are mapped to y. The well known fact is that the bijective S-boxes (permutations) are always regular.

S-boxes need to be:

- with high minimal algebraic degree to resist low order approximation attacks and higher order differential attacks
- with high nonlinearity (low linearity) to resist linear attacks
- with low differential uniformity to resist differential attacks
- with low absolute indicator (autocorrelation) to improve the avalanche effect of the cipher
- regular to resist trivial statistical attacks.

3 Some Existing Methods for Generation of 8-bit S-Boxes

Existing methods for generation of S-boxes can be divided mainly in three groups: algebraic constructions, pseudo-random generation and heuristic generation.

The first group of S-boxes are constructed by applying some mathematical operations and transformations, like finite field inversion

Table 1. A comparison between some of the cryptographic properties of the bijective 8x8 S-boxes produced by different generation methods (− stands for "missing data")

Method	NL(S)	L(S)	$\Delta(S)$	AC(S)	deg(S)
Finite Field Inversion [28] (AES S-box)	112	32	4	32	7
3-round Feistel [5]	96	64	8	–	–
4-uniform permutations method [33]	98	60	4	–	–
Finite Field Multiplication [11]	106 108	44 40	6	56 64	7
Hill climbing method [24]	100	56	–	–	–
Tweaking method [13]	106	44	6	56	7
Simulated annealing method [7]	102	52	–	80	–
GaT [32]	104	48	–	–	–
Gradient descent method [18]	104	48	8	80	7
Hybrid heuristic method [17]	102 104	52 48	6	96	4
Spectral-linear and spectral-difference methods [23]	104	48	6	–	7
GA1 [15]	106 108	44 40	6	56 48	6
GA2 [15]	110 112	36 32	6	40 32	7
SpImmAlg [16]	104	48	6	88	7
Quasigroup 4 method [this paper]	98	60	8	88	7
Quasigroup 16 method [this paper]	100 98	56 60	10 8	88	6 7

method (e.g., AES S-box), or using a smaller S-box as starting point (e.g., finite field multiplication method [11]). In this group one can find 8-bit S-boxes with the best known cryptographic properties. The second group of S-boxes are obtained by pseudo-random generation, and usually they do not have very good cryptographic properties, because of the large input space and very small number of strong S-boxes. The third group of S-boxes are generated by iteratively improving given S-box with respect to one or more cryptographic properties, with the help of the heuristic algorithms (e.g., simulated annealing method [7]). The main advantage of the last generation method is the large number of S-boxes close to the best known.

Our constructions are algebraic constructions, and Table 1 presents the comparisons of our results with the results of some existing methods for generating bijective 8-bit S-boxes. However, our results are based on the experiments made so far, and there is a big possibility these results to be improved by more performed experiments. The set of quasigroups of order 16 is extremly large, and here only one quasigroup with specific features is used, so the number of performed experiments is relatively small.

Mihajloska and Gligoroski [26] constructed optimal 4x4 S-boxes from quasigroups of order 4, by using four e quasigroup string transformations alternating in normal and reverse mode (in a sense that they apply the string in reverse order), on 4 bits. They obtained 9216 optimal Q-S-boxes - with nonlinearity 4 (linearity 8), differential uniformity 4, autocorrelation of 16, maximal algebraic degree of 3 and minimal algebraic degree of 2.

4 Constructions of 8-bit S-boxes with Quasigroups of order 4 and 16

We investigate several cryptographic properties of the 8x8 S-boxes obtained by constructions similar to the one presented in [26], by using quasigroup string transformations, produced by quasigroups of order 4 and 16. The reverse mode of e and d quasigroup string transformations

we will denote here as *oe* and *od*, respectively (Figure 1). The argument of the quasigroup string transformations is the 8-bit string of 4 elements for quasigroups of order 4 and of 2 elements for quasigroups of order 16.

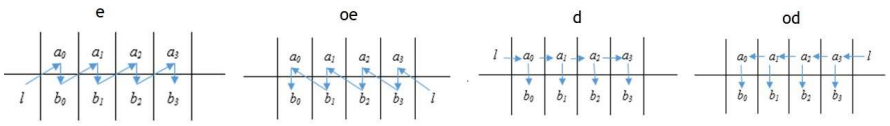


Figure 1. Application of e and d transformations in normal and reverse mode on 8-bit string of 4 elements

4.1 Construction with Quasigroups of order 4

For experiments with quasigroups of order 4, we use only 384 quasigroups out of the total 576. We excluded all linear quasigroups and 48 other non-linear quasigroups that always produce linear e and d transformations. In the representation of the type of used quasigroup string transformations, $n$eoe type means that there are total of n e quasigroup string transformations, used alternately in normal and reverse mode, while $n$dod type means that there are total of n d quasigroup string transformations, used alternately in normal and reverse mode. For each used type and used n, we generated all 8-bit S-boxes and calculated their cryptographic properties. The best obtained results are written in the tables given below, and the best S-boxes are memorized by the lexicographic order of the used quasigroup and the leaders.

Method 1 – In the first method we use different number of e transformations generated by quasigroups of order 4, alternately in normal and reverse mode, on the 8-bit string of 4 2-bit elements, as follows (see Algorithm 1):

Algorithm 1. Construction of 8-bit S-box by Method 1

Input: Q - quasigroup of order 4, *neoe* type and
vector of leaders $L = (l_1, l_2, \dots l_n)$

Output: S

For all possible input blocks x_1, x_2, x_3, x_4 in lexicographic ordering
  $(p_1, p_2, p_3, p_4) = (x_1, x_2, x_3, x_4)$
  For $i = 1$ to n
    If i is odd
      $(t_1, t_2, t_3, t_4) = e_{l_i}(p_1, p_2, p_3, p_4)$
    else
      $(p_4, p_3, p_2, p_1) = e_{l_i}(t_4, t_3, t_2, t_1)$
Use all output blocks from the last round to generate S

Taking into account that block ciphers that use S-boxes, need their inversions for decryption process, the construction of S^{-1} is given below (see Algorithm 2):

Algorithm 2. Construction of S^{-1}

Input: Q - quasigroup of order 4, *neoe* type and
vector of leaders $L = (l_1, l_2, \dots l_n)$

Output: S^{-1}

For all possible input blocks x_1, x_2, x_3, x_4 in lexicographic ordering
  $(p_1, p_2, p_3, p_4) = (x_1, x_2, x_3, x_4)$
  For $i = n$ down to 1
    If n is even
      If i is even
        $(t_4, t_3, t_2, t_1) = d_{l_i}(p_4, p_3, p_2, p_1)$
      else
       $(p_1, p_2, p_3, p_4) = d_{l_i}(t_1, t_2, t_3, t_4)$
    else
      If i is odd
        $(t_1, t_2, t_3, t_4) = d_{l_i}(p_1, p_2, p_3, p_4)$
      else
        $(p_4, p_3, p_2, p_1) = d_{l_i}(t_4, t_3, t_2, t_1)$
Use all output blocks from the last round to generate S^{-1}

The results for 8-bit S-boxes with best examined cryptographic properties for each $n = \{1, \dots, 13\}$ are given in Table 2, while the graphical visualisation of 4*eoe* type is given on Figure 2.

The best produced 8x8 S-boxes are obtained by 13 e quasigroup transformations, alternating in normal and reverse mode, and they have differential uniformity 8, nonlinearity 98 (linearity 60), autocorrelation

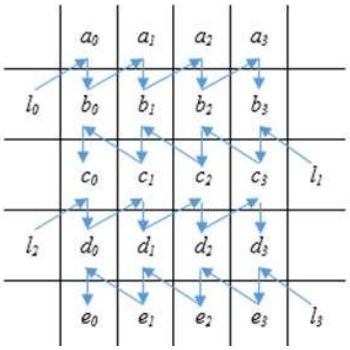


Figure 2. 4*eoe* type – four *e* transformations, alternating in normal and reverse mode, on 8-bit string of 4 2-bit elements

Table 2. The best results for S-boxes obtained by Method 1

Type	NL(S)	L(S)	$\Delta(S)$	AC(S)	$max\{deg(f_i)\}$	deg(S)	No. of S
1e	0	256	256	256	4	1	1152
2eoe	0	256	128	256	6	1	768
3eoe	64	128	64	256	6	3	9216
4eoe	64	128	24	256	7	4	192
5eoe	92	72	16	128	7	6	192
6eoe	96	64	10	104	7	6	96
7eoe	98	60	10	96	7	7	192
8eoe	98	60	10	88	7	6	288
9eoe	98	60	10	88	7	7	480
10eoe	98	60	10	88	7	7	8352
11eoe	98	60	8	112	7	6	96
12eoe	98	60	8	88 96	7	6 7	768 1632
13eoe	98	60	8	88	7	7	96

88 and maximal and minimal algebraic degree 7. There are 96 such S-boxes, obtained by 24 different quasigroups of order 4. One example is the S-box S1 (Table 4) obtained from the quasigroup 34 of order 4, with consecutive leaders (0, 3, 2, 3, 3, 3, 1, 1, 0, 2, 0, 0, 0).

4.2 Construction with quasigroups of order 16

All the experiments with quasigroups of order 16 are done with one, specifically chosen quasigroup of order 16. The requirements for choosing were quasigroup to be shapeless and with highest nonlinearity and lowest differential uniformity if you represent it as 8×4 S-box in a randomly generated set. From 1,000 different randomly tested quasigroups of order 16, we have chosen the best obtained, such as the quasigroup, presented in Figure 3, with differential uniformity 38, nonlinearity 100 (linearity 56), autocorrelation 64 and maximal and minimal algebraic degree 6. The best S-boxes are memorized by the used leaders.

9	2	5	0	15	12	8	6	10	4	13	3	11	1	7	14
15	0	2	9	4	10	14	12	11	3	7	8	5	6	13	1
13	6	4	11	12	9	0	1	14	10	8	5	3	15	2	7
4	9	0	14	5	8	6	7	2	15	10	13	1	11	12	3
1	8	11	15	13	3	10	5	4	2	6	12	7	9	14	0
7	11	8	5	10	1	13	4	12	6	9	0	14	2	3	15
12	10	3	1	9	2	4	13	7	11	5	14	6	0	15	8
11	7	12	3	6	15	9	14	8	0	2	4	10	5	1	13
5	3	1	2	8	13	15	9	0	14	11	10	12	7	6	4
0	14	10	12	7	6	11	2	15	5	3	1	13	4	8	9
14	13	9	6	11	5	2	8	1	7	12	15	4	3	0	10
10	12	7	8	14	0	1	3	6	9	4	11	15	13	5	2
3	4	13	10	0	7	12	15	5	8	1	2	9	14	11	6
2	15	6	7	1	4	5	11	3	13	14	9	0	8	10	12
6	1	14	4	2	11	3	0	13	12	15	7	8	10	9	5
8	5	15	13	3	14	7	10	9	1	0	6	2	12	4	11

Figure 3. Shapeless quasigroup of order 16

Method 2 – The second method is like Method 1, but with the randomly generated shapeless quasigroup of order 16 (Figure 3) and

using 8-bit string of 2 4-bit elements, as follows (see Algorithm 3):

Algorithm 3. Construction of 8-bit S-box by Method 2

Input: Q - quasigroup of order 16, *neoe* type and vector of leaders $L = (l_1, l_2, \dots l_n)$

Output: S

For all possible input blocks x_1, x_2 in lexicographic ordering
  $(p_1, p_2) = (x_1, x_2)$
  For $i = 1$ to n
    If i is odd
      $(t_1, t_2) = e_{l_i}(p_1, p_2)$
    else
      $(p_2, p_1) = e_{l_i}(t_2, t_1)$
Use all output blocks from the last round to generate S

The construction of S^{-1} is given below (see Algorithm 4):

Algorithm 4. Construction of S^{-1}

Input: Q - quasigroup of order 16, *neoe* type and vector of leaders $L = (l_1, l_2, \dots l_n)$

Output: S^{-1}

For all possible input blocks x_1, x_2 in lexicographic ordering
  $(p_1, p_2) = (x_1, x_2)$
  For $i = n$ down to 1
    If n is even
      If i is even
        $(t_2, t_1) = d_{l_i}(p_2, p_1)$
      else
        $(p_1, p_2) = d_{l_i}(t_1, t_2)$
    else
      If i is odd
        $(t_1, t_2) = d_{l_i}(p_1, p_2)$
      else
        $(p_2, p_1) = d_{l_i}(t_2, t_1)$
Use all output blocks from the last round to generate S^{-1}

The results for 8-bit S-boxes with best examined cryptographic properties for each $n = \{1, \dots, 7\}$ are presented in Table 3, while one of the best produced 8x8 S-boxes is S2 (Table 4), obtained by 5 e quasigroup transformations, alternating in normal and reverse mode, with consecutive leaders (3, 2, 0, 10, 9). This S-box is with differential uniformity 8, nonlinearity 98 (linearity 60), autocorrelation 88, maximal and minimal algebraic degree 7, and without fixed points.

Table 3. The best results for S-boxes obtained by Method 3

Type	NL(S)	L(S)	$\Delta(S)$	AC(S)	$max\{deg(f_i)\}$	deg(S)	No. of S
1e	32	192	34	256	6	3 2	5 11
1oe	32	192	34	256	6	3 2	5 11
2eoe	96 94	64 68	10	96 88	7	6	2 1
3eoe	98 96	60 64	10	96 88	7	7 (6)	0 (2) 2 (12)
4eoe	98	60	10	88	7	7 6	1 5
5eoe	98	60	8	88	7	7	1
6eoe	98	60	8 10	96 80	7	7 (6)	1 (3) 2 (0)
7eoe	98	60	8	88	7	7 6	2 5

Table 4. Some best S-boxes obtained by Method 1 and Method 2

S-box S1	S-box S2
Q34, 13eoe, $L = (0, 3, 2, 3, 3, 3, 1, 1, 0, 2, 0, 0, 0)$ NL (L) = 98 (60), $\Delta = 8$, AC = 88, deg = 7	5eoe, $L = (3, 2, 0, 10, 9)$ NL (L) = 98 (60), $\Delta = 8$, AC = 88, deg = 7
47 04 69 8D 30 CF E0 2B D6 1C C2 E8 79 56 32 6A FC F5 52 10 19 8A 9C 01 AB 7A 9E 7B 55 EC 63 8F 26 BA 21 AE E2 D3 A9 C5 0E C3 45 44 FA 31 DD D5 DE 48 39 DF 2E C9 D9 46 77 B2 BC CA 17 FE A8 BE CC 94 11 0C 97 E4 A5 25 8E F9 CB 14 3B 99 12 FB 8C 54 AA E1 6E F0 38 87 7D BB 08 EA 73 A1 5E 96 A0 3F D8 6C 05 1B 34 A6 1E AF 7F E9 37 AC 4E 23 A2 15 80 9A DA A4 0B 6D C8 81 E7 5D FD 9D 68 67 51 93 3A 03 71 C4 A3 BD 50 C7 36 5B 86 9B 06 6B 66 E3 62 85 1D 3D EE 0F D2 40 A7 C6 B8 91 29 B1 B3 5A D4 84 B5 49 28 B7 3E 27 CD 18 F4 2F 02 B4 2C D0 F7 74 70 64 EB 92 D7 F2 78 65 C1 4B E6 2D E5 98 7C BF 4F 20 4D 60 1F 4A 07 0A 61 22 89 F6 7E 4C DB 5C 0D 75 6F 76 83 9F EF 95 C0 00 59 41 90 5F 1A 42 2A 53 F3 DC 35 16 88 33 F1 B6 FF AD B0 F8 B9 24 8B 43 58 57 13 09 D1 82 ED 3C CE 72	0B A2 4C FA 57 C2 87 27 9C D9 A9 5E 96 31 32 59 B7 9A DF 6E 04 E1 A6 23 6B 6A 4D 12 ED 74 0C 43 A7 14 B0 B9 A1 CA 81 EA 8D FB E8 6D 13 52 3B 8E 35 9B 39 2F EE E2 42 BC 11 F5 BD E9 90 4E 1E 1B C1 DB 97 34 68 FD 2B 05 FF A4 01 0F C6 8B 49 53 88 66 E7 F7 55 A0 26 AD A3 69 22 28 EB CF 70 37 95 D7 FE 78 92 FC 93 83 5C 7A B4 F2 41 1F 38 67 85 D1 60 D5 D4 BA BE 25 C4 40 F4 48 8A 58 AE 5A C7 B1 1A 51 AB D6 0A 9E 3A 6C 16 19 F8 8F 62 84 AA 2A F3 E3 1D 79 7C 45 E6 C8 4F 33 80 B5 76 7D A8 06 18 46 5D 3D 10 E0 2C D0 2E 03 CC 24 7F DD 61 72 00 20 E5 9D D8 75 02 EC 2D 73 64 8C 9F 09 B8 15 0E C5 56 C9 CE 17 5B B2 7B 3F 07 44 CD A5 08 21 3E AC E4 B3 89 82 1C 0D D2 30 BF BB 47 C3 B6 F9 65 63 50 91 99 DC 36 98 EF 54 6F 29 F0 F6 3C 5F CB 71 7E AF DA 77 F1 4B C0 94 86 D3 4A DE

5 New Construction

Our new construction of N-bit S-boxes is a generalization of the constructions with e quasigroup string transformations (and d quasigroup string transformations used for the inverse S-boxes) (see Algorithm 5). We mixed two different layers – the layer of bijectional quasigroup string transformations, and the layer of modular addition with N-bit constants. We choose a quasigroup of order q, such that $N = w \cdot log2(q)$. In any other case, we have $q = N$ and $w = 1$. We choose a vector of n bijectional quasigroup string transformations $T = (qst_1, qst_2, \ldots, qst_n)$, which has a corresponding vector of inverse quasigroup string transformations $T^{-1} = (qst_1^{-1}, qst_2^{-1}, \ldots, qst_n^{-1})$.

Algorithm 5. New construction of N-bit S-box

Input: Q - quasigroup of order q, vector of n bijectional quasigroup string transformations $T = (qst_1, qst_2, \ldots, qst_n)$, vector of leaders $L = (l_1, l_2, \ldots l_n)$ and vector of n-bit constants $C = (c_0, c_1, c_2, \ldots c_n)$

Output: S

For all possible input blocks $x_1, x_2, \ldots, x_w$ in lexicographic ordering
  $(p_1, p_2, \ldots, p_w) = (x_1, x_2, \ldots, x_w) + c_0 (mod\ \ 2^N)$
  For $i = 1$ to n
    If i is odd
      $(t_1, t_2, \ldots, t_w) = (qst_i)_{l_i}(p_1, p_2, \ldots, p_w) + c_i (mod\ \ 2^N)$
    else
      $(p_w, \ldots, p_2, p_1) = (qst_i)_{l_i}(t_w, \ldots, t_2, t_1)$
      $(p_1, p_2, \ldots, p_w) = (p_1, p_2, \ldots, p_w) + c_i (mod\ \ 2^N)$
Use all output blocks from the last round to generate S

The construction of the S^{-1} is given below (Algorithm 6):

Algorithm 6. Construction of S^{-1}

Input: Q - quasigroup of order q, vector of n bijectional quasigroup string transformations $T = (qst_1, qst_2, \ldots, qst_n)$, vector of leaders $L = (l_1, l_2, \ldots l_n)$ and vector of n-bit constants $C = (c_0, c_1, c_2, \ldots c_n)$

Output: S^{-1}

For all possible input blocks $x_1, x_2, \ldots, x_w$ in lexicographic ordering
$(p_1, p_2, \ldots, p_w) = (x_1, x_2, \ldots, x_w) - c_n (mod\ \ 2^N)$
For $i = n$ down to 1
If n is even
If i is even
$(t_w, \ldots, t_2, t_1) = (qst_i^{-1})_{l_i}(p_w, \ldots, p_2, p_1) - c_{i-1} (mod\ \ 2^N)$
else
$(p_1, p_2, \ldots, p_w) = (qst_i^{-1})_{l_i}(t_1, t_2, \ldots, t_w) - c_{i-1} (mod\ \ 2^N)$
else
If i is odd
$(t_1, t_2, \ldots, t_w) = (qst_i^{-1})_{l_i}(p_1, p_2, \ldots, p_w) - c_{i-1} (mod\ \ 2^N)$
else
$(p_w, \ldots, p_2, p_1) = (qst_i^{-1})_{l_i}(t_w, \ldots, t_2, t_1) - c_{i-1} (mod\ \ 2^N)$
Use all output blocks from the last round to generate S^{-1}

Method 3 – The third method generates 8-bit S-boxes by using our new construction with parameters $N = 8$, $q = 4$, $w = 4$ and by using only e quasigroup string transformations generated by quasigroups of order 4. Method 1 can be seen as a special case of the Method 3, where all used constants are zeros. The results for 8-bit S-boxes with best examined crypographic properties for each $n = \{1, \ldots, 4\}$ and several different constant vectors, are presented in Table 5.

Method 4 – The fourth method generates 8-bit S-boxes by using our new construction with parameters $N = 8$, $q = 16$, $w = 2$ and by using only e quasigroup string transformations generated by the quasigroup of order 16 (Figure 2). Method 2 can be seen as a special case of Method 4, where all used constants are zeros. The results for 8-bit S-boxes with best examined crypographic properties for each $n = \{1, \ldots, 5\}$ and several different constant vectors, are presented in Table 6. The best produced 8x8 S-boxes are obtained by only 3 e quasigroup transformations, alternating in normal and reverse mode. There are two different best groups. One group has differential uniformity 8, nonlinearity 98 (linearity 60), autocorrelation 88, and minimal algebraic degree 7. One representative of this group is S3 (Table 7), ob-

Table 5. Method 3 – part of the results

Type	NL(S)	L(S)	$\Delta(S)$	AC(S)	$max\{deg(f_i)\}$	deg(S)	No. of S
1e, $C = (0, c_1)$	4	248	132	256	7	3	6144
	32	192	164				1536
1e, $C = (c_0, c_1)$	64	128	64	216	7	6	640
2eoe, $C = (0, 0, c_2)$	64	128	128	256	7	3	3968
3eoe, $C = (0, 0, 0, c_3)$	80	96	32	152	7	6	32
3eoe, $C = (0, c_1, 0, c_3)$	88	80	18	128	7	6	64
4eoe, $C = (0, 0, 0, 0, c_4)$	88	80	24	160	7	5	32
5eoe,	96	64	10	96	7	6	32
$C = (0, 0, 0, 0, 0, c_5)$			12	88			32
6eoe,	98	60	10	88	7	6	32
$C = (0, 0, 0, 0, 0, 0, c_6)$							
7eoe,	98	60	10	88	7	7	3056
$C = (0, 0, 0, 0, 0, 0, 0, c_7)$							

tained by using leaders $(5, 1, 5)$ and constants $(0, 1, 90, 9)$. The second group has differential uniformity 10, nonlinearity 100 (linearity 56), autocorrelation 88, minimal algebraic degree 6, and maximal algebraic degree 7. One representative of this group is S4 (Table 7), obtained by using leaders $(13, 3, 7)$ and constants $(0, 8, 136, 70)$.

6 Conclusion

The main contribution of this paper is a new generic construction of N-bit S-boxes, presented with mixed layers of bijective quasigroup string transformations and modular addition with N-bit constants. Special case of 8-bit S-boxes are investigated, together with their main cryptographic properties. The results are very promising, and further experiments are needed to obtain 8-bit S-boxes with even better cryptographic properties. This can be done, because only a very small subset of S-boxes produced by only one specially selected quasigroup of order 16, and also, very small subset of S-boxes produced by quasigroups of order 4 with several constant vectors are investigated.

References

[1] V. D. Belousov, *Foundations of quasigroups and loops*, Moscow: Nauka, 1967.

[2] E. Biham and A. Shamir, "Differential cryptanalysis of DES-like cryptosystems," *J. Cryptology*, vol. 4, no. 1, pp. 3–72, 1991.

[3] A. Bogdanov, L.R. Knudsen, G. Le, C. Paar, A. Poschmann, M.J.B. Robshaw, Y. Seurin, and C. Vikkelsoe, "PRESENT: An Ultra-Lightweight Block Cipher," in *Cryptographic Hardware and Embedded Systems - CHES 2007. CHES 2007* (Lecture Notes in Computer Science, vol 4727), P. Paillier and I. Verbauwhede, Eds. Berlin Heidelberg: Springer, 2007, pp. 450–466. https://doi.org/10.1007/978-3-540-74735-2_31.

Table 6. Method 4 – part of the results

Type	NL(S)	L(S)	$\Delta(S)$	AC(S)	$max\{deg(f_i)\}$	deg(S)	No. of S
1e, $C = (0, c_1)$	64	128	26	240	7	6	32
1e, $C = (c_0, 0)$	64	128	26	232	7	6	24
1e, $C = (c_0, c_1)$	84	88	18	200	7	6	16
2eoe, $C = (0, 0, c_2)$	98	60	10	96	7	6	12
2eoe, $C = (c_0, c_1, c_2)$	100	56	10	104	7	6	24
	98	60	8	96			56
	98	60	10	88			72540
3eoe, $C = (0, 0, 0, c_3)$	98	60	10	88	7	7	12
3eoe, $C = (0, c_1, c_2, 0)$	98	60	8	96	7	7	256
3eoe, $C = (0, 93, c_2, c_3)$	98	60	8	96	7	7	1024
3eoe, $C = (0, 0..12, c_2, c_3)$	100	56	10	88	7	6	4
	98	60	8			7	20
4eoe, $C = (0, 0, 0, 0, c_4)$	98	60	10	88	7	7	368
						6	856
4eoe, $C = (0, 0, c_2, 0, 0)$	98	60	8	96	7	7 (6)	0 (2)
			10	88			327 (860)
4eoe, $C = (0, 0, c_2, 0, c_4)$	98	60	8	88	7	7	28
5eoe, $C = (0, 0, 0, 0, 0, c_5)$	98	60	8	88	7	7	4

Table 7. Some best S-boxes obtained by Method 4

S-box S3	S-box S4
3eoe, $L = (5, 1, 5)$, $C = (0, 1, 90, 9)$ NL (L) = 98 (60), $\Delta = 8$, AC = 88, deg = 7	3eoe, $L = (13, 3, 7)$, $C = (0, 8, 136, 70)$ NL (L) = 100 (56), $\Delta = 10$, AC = 88, deg = 6
71 6B 60 A1 27 FA B6 C4 F3 16 D3 3E 08 DC D9 BC 85 B7 B9 90 2E CA A6 BE D5 A7 1D 1A D1 03 37 58 65 78 67 22 52 76 4B 42 F9 7D BB DF 54 F4 11 56 DD E5 38 BF 61 3A 9B 18 CC 29 19 BD FF ED D4 33 F0 B1 CD 53 3C 24 6A 05 2C 9F 12 F5 91 10 94 C0 1C 4F AF 2F 63 6C 21 84 86 2B 96 44 4D A4 25 14 A5 9E AB 5A AD 72 81 49 45 40 0C 0B E7 7E 2D DB C8 9D 04 B0 4E 8F 39 CE C7 34 DE 01 AE EA 77 B5 F2 A0 E4 8A 7B 23 EF 97 3B E1 D2 5E 80 9A 68 00 99 1B FD 1E FE C1 C9 0F 4C 8D 70 59 98 31 79 0A 6E CB B2 C3 07 3F 17 09 E6 89 57 FB 5D 35 AC E8 C5 E0 F6 43 0D 8E 69 6D 0E B4 D8 E3 EE 95 88 B8 BA 50 93 DA 47 CF 5C F8 87 28 C2 15 C6 F7 51 02 E9 66 FC 30 75 8C 9C 6F 55 54 1F AA 7A 5F 7C 32 EB 73 20 46 A8 5B 82 41 8B 74 D6 2A 83 A2 4A A9 D7 13 48 92 26 E2 D0 A3 64 F1 62 EC 06 3D 7F B3	87 DD 70 A3 98 3B F2 CF 4D C6 2E FE 1A 86 13 D0 51 95 79 65 44 B5 92 BE 48 B3 24 17 76 1E 4A 1B E9 61 60 F9 8E 91 DB 09 5E 4E 71 3A 39 D1 DF A4 E1 1F A6 A8 CE F8 C9 20 EA 53 C2 B1 0B EC 7D 4F AE 5D E0 7F D4 A2 94 03 29 D3 BA 77 43 A0 73 62 EE 8D AA 2B D6 6F AD 2A ED A9 55 47 96 25 68 AB 01 E8 0F 46 FC 26 CB 0E 42 04 B0 B8 E6 58 97 C3 78 DA 3C 16 63 9F 21 0D 93 E4 FF BF 0C 3E A7 C7 B6 02 80 B7 B9 38 7B D7 1C 57 33 C1 D8 90 14 F6 EF 54 28 83 40 59 E2 12 C5 18 2F BD 35 E3 19 8F AF 9A FB 37 4B 64 7E 6E 9D 30 41 B4 D2 E5 0A 5F 07 99 3F BB 2D BC D5 FD 6D 10 DC 67 B2 88 5B F3 27 34 84 36 22 81 05 F1 74 32 F4 89 66 C4 F5 11 1D EB 15 6A 69 2C CA 56 D9 82 3D E7 5C 6B 5A 85 DE 7A 6C 72 F7 9B 45 AC 4C 08 CD F0 23 31 FA 8A 52 9C 00 50 8C 49 75 A5 C8 C0 8B 9E A1 CC 7C 06

[4] A. Breaken, "Cryptographic properties of Boolean Functions and S-boxes," Ph.D. dissertation, Katholieke Universiteit Leuven, 2006.

[5] A. Canteaut, S. Duval and G. Leurent, "Construction of Lightweight S-Boxes using Feistel and MISTY structures (Full Version)," *Cryptology ePrint Archive*, Report 2015/711, 2015.

[6] F. Chabaud and S. Vaudenay, "Links between differential and linear cryptanalysis," in *Advances in Cryptology - EUROCRYPT 1994* (Lecture Notes in Computer Science, vol. 950), Springer-Verlag, 1995, pp. 356–365.

[7] J.A. Clark, J.L. Jacob, and S. Stepney, "The design of s-boxes by simulated annealing," *New Generation Computing*, vol. 23, no. 3, pp. 219–231, 2004.

[8] J. Daemen, R. Govaerts, and J. Vandewalle, "Correlation matrices," in *FSE 1994* (Lecture Notes in Computer Science, vol. 1008), B. Preneel, Ed. Berlin Heidelberg: Springer, 1994, pp. 275–285.

[9] J. Daemen, "Cipher and Hash Function Design. Strategies based on Linear and Differential Cryptanalysis," Ph.D. dissertation, Katholieke Universiteit Leuven, 1995.

[10] J. Daemen and V. Rijmen, *The Design of Rijndael: AES*, The Advanced Encryption Standard. Springer-Verlag, 2002.

[11] R.A. de la Cruz Jimenez, "Generation of 8-Bit S-Boxes Having Almost Optimal Cryptographic Properties Using Smaller 4-Bit S-Boxes and Finite Field Multiplication," in *Progress in Cryptology - LATINCRYPT 2017* (Lecture Notes in Computer Science, vol.11368), T. Lange, O. Dunkelman, Eds. Cham: Springer, 2017.

[12] H. Dobbertin, "One-to-one highly nonlinear power functions on GF(2n)," *Appl. Algebra Engrg. Comm. Comput.*, vol. 9, pp. 139–152, 1998.

[13] J. Fuller and W. Millan, "Linear redundancy in S-boxes," in *Fast Software Encryption 2003 (FSE'03)* (Lecture Notes in Computer Science, vol. 2887), T. Johansson, Ed. Berlin Heidelberg: Springer, 2003, pp. 74–86.

[14] D. Gligoroski, S. Markovski, and L. Kocarev, "Edon-R, an Infinite Family of Cryptographic Hash Functions," in *The Second NIST Cryptographic Hash Workshop*, (UCSB, Santa Barbara, CA), 2006.

[15] G. Ivanov, N. Nikolov, and S. Nikova, "Reversed genetic algorithms for generation of bijective S-boxes with good cryptographic properties," *Cryptogr. Commun.*, vol. 8, no. 2, pp. 247–276, 2016.

[16] G. Ivanov, N. Nikolov, and S. Nikova, "Cryptographically Strong S-Boxes Generated by Modified Immune Algorithm," in *Cryptography and Information Security in the Balkans* (Lecture Notes in Computer Science, vol. 9540), 2016, pp. 31–42.

[17] H. Isa, N. Jamil, and M. Z'aba, "Hybrid Heuristic Methods in Constructing Cryptographically Strong S-boxes," *International Journal of Cryptology Research*, vol. 6, no. 1, pp. 247–276, 2016.

[18] O.V. Kazymyrov, V.N. Kazymyrova, and R.V. Oliynykov, "A method for generation of high-nonlinear S-Boxes based on gradient descent," *Mat. Vopr. Kriptogr.*, vol. 5, no. 2, pp. 71–78, 2014.

[19] G. Leander and A. Poschmann, "On the Classification of 4 Bit S-Boxes," in *Arithmetic of Finite Fields* (Lecture Notes in Computer Science, vol. 4547), C. Carlet, B. Sunar, Eds. Berlin Heidelberg: Springer, 2007, pp. 159–176.

[20] S. Markovski, D. Gligoroski, and S. Andova, "Using quasigroups for one-one secure encoding," in *VIII Conf. Logic and Computer Science LIRA 1997*, (Novi Sad, Serbia), 1997, pp. 157–162.

[21] M. Matsui, "Linear Cryptanalysis Method for DES Cipher," in *Advances in Cryptology, EUROCRYPT 1993* (Lecture Notes in

Computer Science, vol. 765), T. Helleseth, Ed. Berlin Heidelberg: Springer, 1993, pp. 386–397.

[22] W. Meier and O. Staffelbach, "Nonlinearity criteria for cryptographic functions," in *Workshop on the theory and application of cryptographic techniques on Advances in cryptology*, New York, USA: Springer-Verlag New York, Inc., 1990, pp. 549–562.

[23] A. Menyachikhin, "Spectral-linear and spectral-difference methods for generating cryptographically strong S-Boxes," in *Preproceedings of CTCrypt'16*, (Yaroslavl, Russia), 2016, pp. 232–252.

[24] W. Millan, "How to improve the nonlinearity of bijective S-boxes," in *Third Australian Conference on Information Security and Privacy 1998* (Lecture Notes in Computer Science, vol. 1438), Springer-Verlag, 1998, pp. 181–192.

[25] A. Mileva, "Analysis of Some Quasigroup Transformations as Boolean Functions," *Math. Balkanica*, vol. 26, Fasc. 3-4, 359–368, 2012.

[26] H. Mihajloska and D. Gligoroski, "Construction of Optimal 4-bit S-boxes by Quasigroups of Order 4," in *SECURWARE 2012*, 2012.

[27] K. Nyberg, "Perfect nonlinear S-boxes," in *Eurocrypt 1991* (Lecture Notes in Computer Science, vol. 547), D.W. Davies, Ed. Springer, 1991, pp. 378–385.

[28] K. Nyberg, "Differentially uniform mappings for cryptography," in *Advances in Cryptology - EUROCRYPT 1993* (Lecture Notes in Computer Science, vol. 765), Springer-Verlag, 1994, pp. 55–64.

[29] K. Nyberg, "S-boxes and round functions with controllable linearity and differential uniformity," in *FSE 1995* (Lecture Notes in Computer Science, vol. 1008), B. Preneel, Ed., Berlin Heidelberg: Springer, 1995, pp. 111–130.

[30] B. Preneel, "Analysis and Design of Cryptographic Hash Functions," Ph.D. dissertation, Katholieke Universiteit Leuven, 1994.

[31] O.S. Rothaus, "On "bent" functions," *J. Comb. Theory, Ser. A*, vol. 20, no. 3, pp. 300–305, 1976.

[32] P. Tesař, "A new method for generating high non-linearity S-boxes," *Radioengineering*, vol. 19, no. 1, pp. 23–26, 2010.

[33] L. Qu, Y. Tan, C. Tan, and C. Li, "Constructing differentially 4-uniform permutations over $\mathbb{F}_2^{2^k}$ via the switching method," *IEEE Transactions on Inform. Theory*, vol. 59, no. 7, pp. 4675–4686, 2013.

Aleksandra Mileva, Aleksandra Stojanova,
Dušan Bikov, Yunqing Xu

Received June 25, 2020
Accepted August 23, 2020

Aleksandra Mileva
Faculty of Computer Science, University "Goce Delcev"
Stip, Republic of N. Macedonia
Phone: ++38932550106
E–mail: aleksandra.mileva@ugd.edu.mk

Aleksandra Stojanova
Faculty of Computer Science, University "Goce Delcev"
Stip, Republic of N. Macedonia
Phone: ++38932550123
E–mail: aleksandra.stojanova@ugd.edu.mk

Dušan Bikov
Faculty of Computer Science, University "Goce Delcev"
Stip, Republic of N. Macedonia
E–mail: dusan.bikov@ugd.edu.mk

Yunqing Xu
Ningbo University, Peoples Republic of China
Ningbo, Peoples Republic of China
E–mail: xuyunqing@nbu.edu.cn

Computer Science Journal of Moldova, vol.**28**, no.3(84), 2020

A sharp upper bound on the independent 2-rainbow domination in graphs with minimum degree at least two

Rana Khoeilar, Mahla Keibari, Mustapha Chellali,

Seyed Mahmoud Sheikholeslami*

Abstract

An independent 2-rainbow dominating function (I2-RDF) on a graph G is a function f from the vertex set $V(G)$ to the set of all subsets of the set $\{1,2\}$ such that $\{x \in V \mid f(x) \neq \emptyset\}$ is an independent set of G and for any vertex $v \in V(G)$ with $f(v) = \emptyset$ we have $\bigcup_{u \in N(v)} f(u) = \{1,2\}$. The *weight* of an I2-RDF f is the value $\omega(f) = \sum_{v \in V} |f(v)|$, and the independent 2-rainbow domination number $i_{r2}(G)$ is the minimum weight of an I2-RDF on G. In this paper, we prove that if G is a graph of order $n \geq 3$ with minimum degree at least two such that the set of vertices of degree at least 3 is independent, then $i_{r2}(G) \leq \frac{4n}{5}$.

Keywords: independent k-rainbow dominating function, independent k-rainbow domination number.

MSC 2010: 05C69.

1 Introduction

In this paper, G is a simple graph with vertex set $V(G)$ and edge set $E(G)$ (briefly V and E). For every vertex $v \in V$, the *open neighborhood* $N(v)$ is the set $\{u \in V(G) \mid uv \in E(G)\}$ and the *closed neighborhood* of v is the set $N[v] = N(v) \cup \{v\}$. Similarly, the *open neighborhood* of a set $S \subseteq V$ is the set $N(S) = \cup_{v \in S} N(v)$, and the *closed neighborhood*

* Seyed Mahmoud Sheikholeslami: s.m.sheikholeslami@azaruniv.ac.ir

of S is the set $N[S] = N(S) \cup S$. The *degree* of a vertex $v \in V$ is $\deg_G(v) = \deg(v) = |N(v)|$ and the *minimum degree* of a graph G is denoted by $\delta = \delta(G)$.

For a positive integer k, a *k-rainbow dominating function* (k-RDF) of a graph G is a function f from $V(G)$ to the set of all subsets of the set $\{1, 2, \ldots, k\}$ such that for every vertex $v \in V(G)$ with $f(v) = \emptyset$ the condition $\bigcup_{u \in N(v)} f(u) = \{1, 2, \ldots, k\}$ is fulfilled. The *weight* of a k-RDF f is the value $\omega(f) = \sum_{v \in V} |f(v)|$. The *$k$-rainbow domination number* of a graph G, denoted by $\gamma_{rk}(G)$, is the minimum weight of a k-RDF of G. Note that $\gamma_{r1}(G)$ is the classical domination number $\gamma(G)$. The k-rainbow domination number was introduced by Brešar, Henning, and Rall [4] and has been studied by several authors (see for example [1], [5], [6], [8], [12], [13], [15], [16]).

If we additionally require that the set of vertices $x \in V(G)$ with $f(x) \neq \emptyset$ induces an independent set of G, then the situation is very different. Let $k \geq 1$ be an integer, and let G be a graph. A k-RDF f of G is an *independent k-rainbow dominating function* (Ik-RDF) if $\{x \in V(G) \mid f(x) \neq \emptyset\}$ is an independent set of G. The minimum weight $w(f)$ of an Ik-RDF f of G is called the *independent k-rainbow domination number* of G, and is denoted by $i_{rk}(G)$. Clearly, $\gamma_{rk}(G) \leq i_{rk}(G)$ holds for every graph G. Moreover, independent k-rainbow domination can be seen as a generalization of independent domination, since the number $i_{r1}(G)$ is precisely the *independent domination number* $i(G)$ of G which has been widely studied (see a survey [10]). The independent rainbow domination number was studied in, for example, [3], [7], [9], [14]. Recently, Shao et al. [14] have shown that the independent k-rainbow domination problem is NP-complete. Therefore, it is natural to look for good bounds on the independent k-rainbow domination number of graphs, especially for the case $k = 2$ which is the most studied. Our main contribution in this paper is the following upper bound on the independent 2-rainbow domination number for a class of graphs with minimum degree at least two.

Theorem 1. Let G be a connected graph of order n with $\delta(G) \geq 2$ such that the set of vertices with degree at least three is an independent set. Then $i_{r2}(G) \leq \frac{4}{5}n$. This bound is sharp for the cycle C_5.

We recall that for trees, Amjadi et al. [2] showed that if T is a tree of order $n \geq 3$, then $i_{r2} \leq \frac{3n}{4}$. This result has recently been extended to connected bipartite graphs by Fujita et al. [9] who also gave other results on $i_{rk}(G)$ when $k \geq 3$. Moreover, it has been noticed in [7], that for any graph G, $i_{r2}(G) \leq 2i(G)$. Therefore, the $\frac{4}{5}n$ upper bound on the independent 2-rainbow domination number holds for any connected graph G of order n with $i(G) \leq \frac{2}{5}n$.

2 Proof of Theorem 1

For the proof of Theorem 1, we need three preparatory results.

Proposition 1. For $n \geq 5$, the path P_n has an I2-RDF f that assigns $\emptyset$ to the end-vertices of P_n and $\omega(f) \leq \frac{4n}{5}$.

Proof. Let $P_n = v_1v_2 \dots v_n$ and define $f : V(P_n) \to \mathcal{P}(\{1,2\})$ as follows. If $n \equiv 0 \pmod 3$, then $f(v_{3i+2}) = \{1,2\}$ for $0 \leq i \leq \frac{n}{3} - 1$, and $f(x) = \emptyset$ otherwise; if $n \equiv 2 \pmod 3$, then $f(v_{n-1}) = \{1,2\}$, $f(v_{3i+2}) = \{1,2\}$ for $0 \leq i \leq \frac{n-5}{3}$, and $f(x) = \emptyset$ otherwise, and if $n \equiv 1 \pmod 3$, then $f(v_2) = \{1,2\}, f(v_4) = \{1\}$, $f(v_{3i+6}) = \{1,2\}$ for $0 \leq i \leq \frac{n-4}{3}$, and $f(x) = \emptyset$ otherwise. Clearly, $\omega(f) \leq \frac{4n}{5}$ and f is an I2-RDF of P_n assigning $\emptyset$ to the end-vertices of P_n. □

For integers $r \geq 3$ and $s \geq 1$, let $C_{r,s}$ be the connected graph obtained from a cycle $C_r = (u_1u_2 \dots u_r)$ and a path $P_s = v_1v_2 \dots v_s$ by adding the edge u_1v_1.

Proposition 2. For integers $r \geq 3$ and $s \geq 1$ with $r+s \geq 4$, the graph $C_{r,s}$ has *an I2-RDF* f that assigns $\emptyset$ to v_s and $\omega(f) \leq \frac{4(r+s)}{5}$.

Proof. Define $f : V(C_{r,s}) \to \mathcal{P}(\{1,2\})$ by:

$f(u_{3i+1}) = f(v_{3j+2}) = \{1,2\}$ for $0 \leq i \leq \frac{r-3}{3}$, $0 \leq j \leq \frac{s-3}{3}$, and $f(x) = \emptyset$ otherwise, when $r \equiv 0 \pmod 3$ and $s \equiv 0 \pmod 3$,

$f(u_{3i+1}) = f(v_{3j+3}) = \{1,2\}$ for $0 \leq i \leq \frac{r-3}{3}$, $0 \leq j \leq \frac{s-4}{3}$, and $f(x) = \emptyset$ otherwise, when $r \equiv 0 \pmod 3$ and $s \equiv 1 \pmod 3$,

$f(v_s) = \{1,2\}$, $f(u_{3i+2}) = f(v_{3j+2}) = \{1,2\}$ for $0 \leq i \leq \frac{r-3}{3}$, $0 \leq j \leq \frac{s-5}{3}$, and $f(x) = \emptyset$ otherwise, when $r \equiv 0 \pmod 3$ and $s \equiv 2 \pmod 3$,

$f(u_1) = \{1\}$, $f(u_{3i+3}) = f(v_{3j+2}) = \{1,2\}$ for $0 \leq i \leq \frac{r-4}{3}$, $0 \leq j \leq \frac{s-3}{3}$, and $f(x) = \emptyset$ otherwise, when $r \equiv 1 \pmod 3$ and $s \equiv 0 \pmod 3$,

$f(u_1) = \{1,2\}$, $f(u_{3i+3}) = f(v_{3j+3}) = \{1,2\}$ for $0 \leq i \leq \frac{r-4}{3}$, $0 \leq j \leq \frac{s-4}{3}$, and $f(x) = \emptyset$ otherwise, when $r \equiv 1 \pmod 3$ and $s \equiv 1 \pmod 3$,

$f(u_{3i+3}) = f(v_{3j+1}) = \{1,2\}$ for $0 \leq i \leq \frac{r-4}{3}$, $0 \leq j \leq \frac{s-2}{3}$, and $f(x) = \emptyset$ otherwise, when $r \equiv 1 \pmod 3$ and $s \equiv 2 \pmod 3$,

$f(u_{3i+1}) = f(v_{3j+2}) = \{1,2\}$ for $0 \leq i \leq \frac{r-2}{3}$, $0 \leq j \leq \frac{s-3}{3}$, and $f(x) = \emptyset$ otherwise, when $r \equiv 2 \pmod 3$ and $s \equiv 0 \pmod 3$,

$f(u_{3i+1}) = f(v_{3j+3}) = \{1,2\}$ for $0 \leq i \leq \frac{r-2}{3}$, $0 \leq j \leq \frac{s-4}{3}$, and $f(x) = \emptyset$ otherwise, when $r \equiv 2 \pmod 3$ and $s \equiv 1 \pmod 3$,

$f(u_r) = \{1\}$, $f(u_{3i+3}) = f(v_{3j+1}) = \{1,2\}$ for $0 \leq i \leq \frac{r-5}{3}$, $0 \leq j \leq \frac{s-2}{3}$, and $f(x) = \emptyset$ otherwise, when $r \equiv 2 \pmod 3$ and $s \equiv 2 \pmod 3$.

In either case, f is an I2-RDF of $C_{r,s}$ of weight at most $\frac{4(r+s)}{5}$ with the desired property. □

Let $\mathcal{F}$ be the family of all simple graphs obtained from some connected multigraph H without loops with $\delta(H) \geq 3$ by subdividing every edge of H at least once and at most four times. By definition, the smallest graph of $\mathcal{F}$ has order at least 5. Also, we note that if $G \in \mathcal{F}$, then the set $\{x \in V(G) \mid \deg(x) \geq 3\}$ is an independent set of G.

Proposition 3. If $G \in \mathcal{F}$, then G has an *I2-RDF* f that assigns a non-empty set to every vertex of degree at least 3 and $\omega(f) \leq \frac{4n(G)}{5}$.

Proof. Let $G \in \mathcal{F}$ be a graph of order n. We proceed by induction on n. Clearly, the result is immediate for $n = 5$. Let $n \geq 6$ and assume that the result holds for all graphs in $\mathcal{F}$ of order less than n. Let $G \in \mathcal{F}$ be a graph of order n. Set $A = \{x \in V(G) \mid \deg(x) \geq 3\}$ and $B = V(G) - A$. Since $G \in \mathcal{F}$, A is independent. In the sequel, we will call an induced path P of G an *A-ear path* if $V(P) \subset B$ and P is connected to A by either its unique vertex (if $|V(P)| = 1$) or by

each of its two end-vertices. For each $i \in \{1,2,3,4\}$, let Q_i be the set of all A-ear paths P of G of order i and let $\mathcal{Q} = \bigcup_{i=1}^{4} Q_i$. Clearly, $B = \bigcup_{P\in\mathcal{Q}} V(P)$. Moreover, for each A-ear path P, let $X_P = \{u \in A \mid u \text{ is adjacent to a vertex of } P\}$. Hence, $A = \bigcup_{P\in\mathcal{Q}} X_P$ and since $G \in \mathcal{F}$, we have $|X_P| = 2$ for each $P \in \mathcal{Q}$. Therefore, $|A| \geq 2$.

Assume first that $|A| = 2$ and let $A = \{u, v\}$. Note that $n = |A|+m_1+2m_2+3m_3+4m_4$ and $m_1+m_2+m_3+m_4 \geq 3$, where $m_i = |Q_i|$ for $i \in \{1,2,3,4\}$. Let $Q_4 = \{v_1^i v_2^i v_3^i v_4^i \mid 1 \leq i \leq m_4\}$ if $Q_4 \neq \emptyset$, $Q_3 = \{m_1^j m_2^j m_3^j \mid 1 \leq j \leq m_3\}$ if $Q_3 \neq \emptyset$, $Q_2 = \{w_1^k w_2^k \mid 1 \leq k \leq m_2\}$ if $Q_2 \neq \emptyset$ and $Q_1 = \{z_1^l \mid 1 \leq l \leq m_1\}$ if $Q_1 \neq \emptyset$. Suppose that $uv_1^i, um_1^j, uw_1^k, uz_1^l, v_4^i v, m_3^j v, w_2^k v, z_1^l v \in E(G)$ for each i, j, k, l. Define $g : V(G) \to \mathcal{P}(\{1,2\})$ by $g(u) = g(v) = \{1,2\}$, $g(v_2^i) = \{1,2\}$ for $1 \leq i \leq m_4$ and $g(m_2^j) = \{2\}$ for $1 \leq j \leq m_3$, and $g(x) = \emptyset$ otherwise. Obviously, g is an I2-RDF of G that assigns a non-empty set to every vertex in A. In addition, we have

$$i_{r2}(G) \leq 4 + 2m_4 + m_3 \leq \frac{4(2 + 4m_4 + 3m_3 + 2m_2 + m_1)}{5} = \frac{4n}{5}.$$

Next, we assume that $|A| \geq 3$. Suppose first, there are two vertices $u, v \in A$ such that $\deg(u)$, $\deg(v) \geq 4$ and there is an A-ear path $P = v_1 \dots v_k$ with $k \geq 3$ connecting u and v. Let $G' = G - V(P)$. Since $G' \in \mathcal{F}$, by the induction hypothesis G' has an I2-RDF f such that $|f(u)|, |f(v)| \geq 1$ and $\omega(f) \leq \frac{4(n-k)}{5}$. Assume, without loss of generality, that $1 \in f(u)$. Then f can be extended to an I2-RDF g of G as follows: assign $\{1,2\}$ to v_{3i+2} for $0 \leq i \leq \frac{k}{3} - 1$ and $\emptyset$ to other vertices, when $k \equiv 0 \pmod 3$; assign $\{2\}$ to v_2, $\{1,2\}$ to v_{3i+1} for $1 \leq i \leq \frac{k-2}{3}$ and $\emptyset$ to other vertices, when $k \equiv 2 \pmod 3$; assign $\{1,2\}$ to u, v_{3i} for $1 \leq i \leq \frac{k-1}{3}$ and $\emptyset$ to other vertices, when $k \equiv 1 \pmod 3$. Clearly, g is an I2-RDF of G of weight at most $\frac{4n(G)}{5}$ with the desired property. Hence, we can assume that there is no two vertices of degree at least four connected by A-ear path P of order at least three. Consider the following cases.

Case 1. $Q_4 \neq \emptyset$.
Let $P_1 = x_1^1 x_2^1 x_3^1 x_4^1 \in Q_4$ and let $ux_1^1, x_4^1 v_1 \in E(G)$, where $u, v_1 \in A$.

By assumption, u or v_1 has degree three, say $\deg(u) = 3$. Consider also the following situations.

(I) u is adjacent to two A-ear paths in Q_4.
Let $P_2 = x_1^2x_2^2x_3^2x_4^2 \in Q_4 - \{P_1\}$ such that ux_1^2, $v_2x_4^2 \in E(G)$. Let G' be the graph obtained from G by removing vertices $\{x_1^1, x_2^1, x_3^1, x_1^2, x_2^2\}$ and adding edges ux_4^1, ux_3^2. Clearly, $G' \in \mathcal{F}$, and by the induction hypothesis, there exists an I2-RDF f of G' of weight at most $\frac{4(n-5)}{5}$ assigning a non-empty set to every vertex of degree at least 3. It follows that $f(x_3^2) = \emptyset$ and $f(x_4^2) = \emptyset$, and, thus, $f(u) = f(v_2) = \{1, 2\}$ (to 2-rainbow dominate x_3^2, x_4^2). Define $g : V(G) \to \mathcal{P}(\{1, 2\})$ by $g(x_3^1) = g(x_3^2) = \{1, 2\}$, $g(x_1^1) = g(x_2^1) = g(x_1^2) = g(x_2^2) = \emptyset$, and $g(x) = f(x)$ otherwise. Clearly, g is an I2-RDF of G that assigns a non-empty set to all vertices of A and has weight $\omega(g) = \omega(f) + 4 \leq \frac{4(n-5)}{5} + 4 = \frac{4n}{5}$.

According to the previous situation, we may assume that P_1 is the unique A-ear path in Q_4 adjacent to u.

(II) u is adjacent to an A-ear path in Q_2.
Let $P_2 = x_1^2x_2^2$ be an A-ear path in Q_2 such that $ux_1^2, v_2x_2^2 \in E(G)$. Let G' be the graph obtained from G by removing vertices x_1^1, x_2^1, x_3^1 and adding the edge ux_4^1. Clearly, $G' \in \mathcal{F}$, and by the induction hypothesis, there exists an I2-RDF f of G' of weight at most $\frac{4(n-3)}{5}$ assigning a non-empty set to every vertex of degree at least 3. Likewise to situation (I), one can see that $f(u) = f(v_2) = \{1, 2\}$. Now define $g : V(G) \to \mathcal{P}(\{1, 2\})$ by $g(x_3^1) = \{1, 2\}$, $g(x_1^1) = g(x_2^1) = \emptyset$, and $g(x) = f(x)$ otherwise. Clearly, g is an I2-RDF of G that assigns a non-empty set to every vertex of A and has weight $\omega(g) = \omega(f) + 3 \leq \frac{4(n-3)}{5} + 2 < \frac{4n}{5}$.

(III) u is adjacent to an A-ear path in Q_3.
Let $P_2 = x_1^2x_2^2x_3^2$ be an A-ear path in Q_3 such that ux_1^2, $v_2x_3^2 \in E(G)$. Let G' be the graph obtained from G by removing $x_1^1, x_2^1, x_3^1, x_1^2$ and adding edges ux_4^1, ux_2^2. Since $G' \in \mathcal{F}$, by the induction hypothesis there is an I2-RDF f of G' of weight at most $\frac{4(n-4)}{5}$ that assigns non-empty sets to every vertex of degree at least 3. As above, $f(u) = f(v_2) = \{1, 2\}$. Define $g : V(G) \to \mathcal{P}(\{1, 2\})$ by $g(x_1^2) = \{1\}$, $g(x_3^1) = \{1, 2\}$, $g(x_1^1) = g(x_2^1) = g(x_1^2) = \emptyset$, and $g(x) = f(x)$ otherwise. Clearly, g is an

I2-RDF of G of weight $\omega(g) = \omega(f) + 3 \leq \frac{4(n-4)}{5} + 3 < \frac{4n}{5}$ having the desired property.

(IV) u is adjacent to two A-ear paths in Q_1.
Let $P_2 = x_1^2$ and $P_3 = x_1^3$ be A-ear paths in $\mathcal{Q}_1$ such that $ux_1^2, ux_1^3, v_2x_1^2, v_3x_1^3 \in E(G)$, where $v_2, v_3 \in A$.

- $v_1 \notin \{v_2, v_3\}$.
  Let G' be the graph obtained from G by removing $x_1^1, x_2^1, x_3^1, x_4^1, u$ and adding the edges $v_1x_1^2$ and $v_1x_1^3$. Since $G' \in \mathcal{F}$, by the induction hypothesis there is an I2-RDF f of G' of weight at most $\frac{4(n-5)}{5}$ assigning non-empty sets to every vertex of degree at least 3. Since $|f(v_2)|, |f(v_3)| \geq 1$, we must have $f(x_1^2) = f(x_1^3) = \emptyset$. Now the function $g : V(G) \to \mathcal{P}(\{1,2\})$ defined by $g(u) = g(x_3^1) = \{1,2\}$, $g(x_1^1) = g(x_2^1) = g(x_4^1) = \emptyset$, and $g(x) = f(x)$ otherwise, is an I2-IRDF of G with the desired property and weight $\omega(g) = \omega(f) + 4 \leq \frac{4(n-5)}{5} + 4 \leq \frac{4n}{5}$.

- $v_1 \in \{v_2, v_3\}$.
  Without loss of generality, assume that $v_1 = v_2$. Suppose first that $v_1 \neq v_3$ and let G' be the graph obtained from G by removing x_1^1, x_2^1, x_1^3, u and adding the edges $v_3x_1^2$, $v_3x_3^1$. Clearly, $G' \in \mathcal{F}$ and, thus, by the induction hypothesis, there is an I2-RDF f of G' of weight at most $\omega(f) \leq \frac{4(n-4)}{5}$ that assigns a non-empty set to every vertex of degree at least 3. As above, one can easily see that $f(v_1) = f(v_3) = \{1,2\}$ and, thus, $f(x_3^1) = f(x_4^1) = f(x_1^2) = \emptyset$. Now define the function $g : V(G) \to \mathcal{P}(\{1,2\})$ by $g(x_2^1) = \{1,2\}$, $g(x_1^1) = g(x_1^3) = \emptyset$, $g(u) = \{1\}$, and $g(x) = f(x)$ otherwise. Clearly, g is an I2-RDF of G with the desired property and weight $\omega(g) = \omega(f) + 3 \leq \frac{4(n-4)}{5} + 3 < \frac{4n}{5}$.

  Now assume that $v_1 = v_2 = v_3$. Since $|A| \geq 3$ and G is connected, we have $\deg(v_1) \geq 4$. Let $w \in A - \{u, v_1\}$ and let G' be the graph obtained from G by removing x_1^1, x_2^1, u, x_1^3 and adding the edges wx_3^1, wx_1^2. Since $G' \in \mathcal{F}$, by the induction hypothesis, there is an I2-RDF f of G' of weight at most $\omega(f) \leq \frac{4(n-4)}{5}$ such that f assigns a non-empty set to every vertex of degree at least

3. As above, we must have $f(v_1) = \{1, 2\}$. Now the function $g : V(G) \to \mathcal{P}(\{1, 2\})$ defined by $f(x_2^1) = \{1, 2\}$, $g(u) = \{1\}$, $g(x_1^1) = g(x_1^3) = \emptyset$, and $g(x) = f(x)$ otherwise, is an I2-RDF of G such that g assigns a non-empty set to every vertex in A and $\omega(g) = \omega(f) + 3 \leq \frac{4(n-4)}{5} + 3 < \frac{4n}{5}$.

Seeing Case 1, we can assume from now on that $Q_4 = \emptyset$.

Case 2. $Q_3 \neq \emptyset$.

Let $P = x_1^1x_2^1x_3^1 \in Q_3$ and let $ux_1^1, x_3^1v_1 \in E(G)$, where $u, v_1 \in A$. By assumption, we may assume, without loss generality, that $\deg(u) = 3$. Consider the following situations.

(V) u is adjacent to three A-ear-paths in Q_3.

Let $P_2 = x_1^2x_2^2x_3^2$ and $P_3 = x_1^3x_2^3x_3^3$ be two A-ear-paths in $Q_3 - \{P\}$ such that $ux_1^3, ux_1^2, v_2x_3^2, v_3x_3^3 \in E(G)$. Let G' be the graph obtained from G by removing $x_1^1, x_2^1, x_1^2, x_2^2, x_1^3, x_2^3$ and by adding edges ux_3^1, ux_3^2, ux_3^3. Then $G' \in \mathcal{F}$, and by the induction hypothesis, there is an I2-RDF f of G' of weight at most $\frac{4(n-6)}{5}$ that assigns non-empty sets to vertices of degree at least 3. Without loss of generality, assume that $1 \in f(u)$. Define $g : V(G) \to \mathcal{P}(\{1, 2\})$ by $g(u) = \{1, 2\}$, $g(x_1^i) = \emptyset$ for $i \in \{1, 2, 3\}$, $g(x_2^i) = \{1\}$ if $f(v_i) = \{1, 2\}$, $g(x_2^i) = \{1, 2\} - f(v_i)$ if $|f(v_i)| = 1$ for $i \in \{1, 2, 3\}$, and $g(x) = f(x)$ otherwise. Clearly, g is an I2-RDF of G with the desired property and weight $\omega(g) = \omega(f) + 4 \leq \frac{4(n-6)}{5} + 4 < \frac{4n}{5}$.

(VI) u is adjacent to an A-ear path in Q_2.

Let $P_2 = x_1^2x_2^2 \in Q_2$ such that $ux_1^2, v_2x_1^2 \in E(G)$. Let G' be the graph obtained from G by removing x_1^1, x_2^1 and adding the edge ux_3^1. Then $G' \in \mathcal{F}$ and by the induction hypothesis, there is an I2-RDF f of G' of weight at most $\frac{4(n-2)}{5}$ such that $f(x) \neq \emptyset$ for every $x \in A$. Since $f(x_1^2) = f(x_2^2) = \emptyset$, we deduce that $f(u) = \{1, 2\}$. Define now the function $g : V(G) \to \mathcal{P}(\{1, 2\})$ by $g(x_1^1) = \emptyset$, $g(x_2^1) = \{1\}$ if $f(v_1) = \{1, 2\}$ or $g(x_2^1) = \{1, 2\} - f(v_1)$ if $|f(v_1)| = 1$, and $g(x) = f(x)$ otherwise. Clearly, g is an I2-RDF of G of weight $\omega(g) = \omega(f) + 1 \leq \frac{4(n-2)}{5} + 1 < \frac{2n}{3}$ and such that $g(x) \neq \emptyset$ for every $x \in A$.

(VII) u is adjacent to two A-ear paths in Q_3 and to an A-ear path in Q_1.
Let $P_2 = x_1^2x_2^2x_3^2 \in Q_3$ and $P_3 = x_1^3 \in Q_1$ such that $ux_1^3, ux_1^2, v_2x_3^2$, $v_3x_1^3 \in E(G)$. Let G' be the graph obtained from G by removing $x_1^1, x_2^1, x_1^2, x_2^2$ and adding edges ux_3^1, ux_3^2. Clearly, $G' \in \mathcal{F}$ and by the induction hypothesis, there is an I2-RDF f of G' of weight at most $\frac{4(n-4)}{5}$ such that $f(x) \neq \emptyset$ for every $x \in A$. Define the function $g : V(G) \to \mathcal{P}(\{1,2\})$ by $g(u) = \{1,2\}$, $g(x_1^i) = \emptyset$ for $i \in \{1,2\}$, $g(x_2^i) = \{1\}$ if $f(v_i) = \{1,2\}$, $g(x_2^i) = \{1,2\} - f(v_i)$ if $|f(v_i)| = 1$, and $g(x) = f(x)$ otherwise. Note that $|g(u)| - |f(u)| \leq 1$. Clearly, g is an I2-RDF of G of weight $\omega(g) = \omega(f) + 3 \leq \frac{4(n-4)}{5} + 3 < \frac{4n}{5}$ such that $g(x) \neq \emptyset$ for every $x \in A$.

(VIII) u is adjacent to an A-ear-path in Q_3 and to two A-ear paths in Q_1.
Let $P_2 = x_1^2$ and $P_3 = x_1^3$ be A-ear paths in Q_1 such that $ux_1^2, ux_1^3 \in E(G)$. Suppose $v_2x_1^2, v_3x_1^3 \in E(G)$, where $v_2, v_3 \in A$.

- $v_1 \notin \{v_2, v_3\}$.
  Let G' be the graph obtained from G by removing x_1^1, x_2^1, x_3^1, u and adding the edges $v_1x_1^2$ and $v_2x_1^3$. Clearly, $G' \in \mathcal{F}$ and by the induction hypothesis, there is an I2-RDF f of G' of weight at most $\omega(f) \leq \frac{4(n-4)}{5}$ such that $f(x) \neq \emptyset$ for every $x \in A - \{u\}$. Without loss of generality, we assume that $1 \in f(v_1)$ and define the function g on $V(G)$ by $g(u) = \{1,2\}$, $g(x_1^1) = g(x_3^1) = \emptyset$, $g(x_2^1) = \{2\}$, and $g(x) = f(x)$ otherwise. Clearly, g is an I2-RDF of G of weight $\omega(g) = \omega(f) + 3 \leq \frac{4(n-4)}{5} + 3 < \frac{2n}{3}$. In addition, $g(x) \neq \emptyset$ for every $x \in A$.

- $v_1 \in \{v_2, v_3\}$ and $v_1 = v_2 = v_3$.
  Since $|A| \geq 3$, we have $\deg(v_1) \geq 4$. First, let there exist a path v_3zv_4 in G, where $v_4 \in A$ and $z \notin \{x_1^3, x_1^2\}$. Since $\deg(v_4) \geq 3$, we deduce that $A - \{u, v_3, v_4\} \neq \emptyset$. Let $w \in A - \{u, v_3, v_4\}$. Assume that $\deg(v_1) \geq 5$, and let G' be the graph resulting from the deletion of vertices u, x_1^1, x_1^2, x_1^3 and the addition of the edge wx_2^1. Then $G' \in \mathcal{F}$ and by the induction hypothesis, there is an

I2-RDF f of G' of weight at most $\frac{4(n-4)}{5}$ such that $f(x) \neq \emptyset$ for every $x \in A-\{u\}$. We also have $f(v_3) = f(w) = \{1,2\}$. But then the function g defined on $V(G)$ by $f(u) = \{1,2\}$, $f(x_2^1) = \{1\}$, $f(x_1^2) = f(x_1^3) = \emptyset$, and $g(x) = f(x)$ otherwise is an I2-RDF of G of weight $\omega(f) + 3 \leq \frac{4(n-4)}{5} + 3 < \frac{2n}{3}$ and such that $g(x) \neq \emptyset$ for every $x \in A$. Now assume that $\deg(v_1) = 4$, and let G' be the graph obtained from G by removing $u, x_1^1, x_2^1, x_3^1, x_1^2, x_1^3$ and by adding the edge wv_3. Note that v_3 has degree two in G' and, thus, belongs to an A'-ear path joining v_4 and w. Since $G' \in \mathcal{F}$, by the induction hypothesis, there is an I2-RDF f of G' of weight at most $\frac{4(n-6)}{5}$ such that $f(x) \neq \emptyset$ for every $x \in A - \{u, v_3\}$. Clearly, $f(v_4) = f(w) = \{1,2\}$ and $f(z) = f(v_3) = \emptyset$. Now f can be extended to an 2-IRDF of G with the desired property by assigning $\{1\}$ to u, $\{2\}$ to v_3, $\{1,2\}$ to x_2^1 and $\emptyset$ to $x_1^1, x_3^1, x_1^2, x_1^3$ and, thus, $\omega(g) = \omega(f) + 4 < \frac{4n}{5}$.

Now let there exist a path v_3zyv_4 in G, where $v_4 \in A$ and $z \notin \{x_3^1\}$. As above, we have $A - \{u, v_3, v_4\} \neq \emptyset$, and so pick a vertex $w \in A - \{u, v_3, v_4\}$. If $\deg(v_1) \geq 5$, then the result follows, as above, by considering the same graph G' obtained from the removal of vertices u, x_1^1, x_1^2, x_1^3 and the addition of the edge wx_2^1. Hence, we assume that $\deg(v_1) = 4$. Then delete vertices $u, x_1^1, x_2^1, x_3^1, v_3, x_1^2, x_1^3$ and add the edge zw, and let G' be the resulting graph. Clearly, $G' \in \mathcal{F}$ and by the induction hypothesis, there is an I2-RDF f of G' of weight at most $\frac{4(n-7)}{5}$ such that $f(x) \neq \emptyset$ for every $x \in A - \{u, v_3\}$. Since $f(v_4) = f(w) = \{1,2\}$ and $f(z) = f(y) = \emptyset$, function f can be extended to an I2-RDF of G by assigning $\{1,2\}$ to u, v_3, $\{1\}$ to x_2^1 and $\emptyset$ to other vertices. It follows that $\omega(g) = \omega(f) + 5 < \frac{4n}{5}$ and $g(x) \neq \emptyset$ for every $x \in A$.

- $v_1 \in \{v_2, v_3\}$ and $|\{v_1, v_2, v_3\}| = 2$.
  Suppose, without loss of generality, that $v_1 = v_2$ and $v_1 \neq v_3$. Let G' be the graph obtained from G by removing x_1^1, x_3^1, u and by adding the edges $v_3x_1^2, v_3x_2^1$. Then $G' \in \mathcal{F}$ and by the induction hypothesis, there is an I2-RDF f of G' of weight at most $\frac{4(n-3)}{5}$ such that $g(x) \neq \emptyset$ for every $x \in A - \{u\}$. As above, we

must have $f(v_1) = f(v_3) = \{1, 2\}$ and so $f(x_2^1) = f(x_3^1) = \emptyset$. In this case, the function g defined by $g(u) = \{1\}$, $= g(x_2^1) = \{2\}$, $g(x_1^1) = g(x_1^3) = \emptyset$, and $g(x) = f(x)$ otherwise, is an I2-RDF of G of weight $\omega(g) = \omega(f) + 2 \leq \frac{4(n-3)}{5} + 2 < \frac{4n}{5}$. Moreover, $g(x) \neq \emptyset$ for every $x \in A - \{u\}$.

Considering Case 2, we may assume that $Q_3 = \emptyset$.

Case 3. $Q_2 \neq \emptyset$.

Let $P_1 = x_1^1 x_2^1 \in Q_2$ with $ux_1^1, x_2^1 v_1 \in E(G)$. Without loss of generality, we assume that $\deg(u) \leq \deg(v_1)$. Consider the following situations.

(IX) $\deg(u) \geq 4$ and u is adjacent to at least two A-ear paths in Q_2.

By assumption we have $\deg(v_1) \geq 4$. Let $P_2 = x_1^2 x_2^2$ be a second A-ear path in Q_2 such that $ux_1^2 \in E(G)$ and let $v_2 x_2^2 \in E(G)$ for some $v_2 \in A$. Remove vertices x_1^1, x_2^1 and let G' be the resulting graph. Then $G' \in \mathcal{F}$ and by the induction hypothesis, there is an I2-RD-function f of G' of weight at most $\frac{4(n-2)}{5}$ such that $f(x) \neq \emptyset$ for every $x \in A$. Clearly, $f(u) = \{1, 2\}$ and $|f(v_1)| \geq 1$. Define the function g by $g(x_1^1) = g(x_2^1) = \emptyset$, $g(v_1) = \{1, 2\}$, and $g(x) = f(x)$ otherwise. Then g is an 2-IRDF of G of weight $\omega(f) + 1 \leq \frac{4(n-2)}{5} + 1 < \frac{4n}{5}$ having the property that $g(x) \neq \emptyset$ for every $x \in A$.

(X) $\deg(u) = 3$ and u is adjacent to at least two A-ear paths in Q_2.

Let $P_2 = x_1^2 x_2^2 \in Q_2$ be an A-ear path in G such that $ux_1^2 \in E(G)$ and $v_2 x_2^2 \in E(G)$, where $v_2 \in A$.

- $v_1 = v_2$.
  Let G' be the graph obtained from G by removing x_1^1 and adding the edge ux_2^1. Then $G' \in \mathcal{F}$ and by the induction hypothesis, there is an I2-RD-function f of G' of weight $\frac{4(n-1)}{5}$ with the desired property, in particular $f(u) = f(v_1) = \{1, 2\}$. In this case, f can be extended to an I2-IRDF of G by assigning $\emptyset$ to x_1^1 such that f satisfies the conditions.

- $v_1 \neq v_2$.
  Since $\deg(u) = 3$, let uPv_3 be a path in G such that $P \in (Q_1 \cup Q_2) - \{P_1, P_2\}$. Let $z \in V(P)$ be the vertex adjacent to u. We may assume that $v_1 \notin \{v_2, v_3\}$. Let G' be the resulting graph after removing vertices x_1^1, x_2^1, u and adding edges $v_1x_1^2$ and v_1z. Then $G' \in \mathcal{F}$ and by the induction hypothesis, there exists an I2-RDF f of G' satisfying our conditions. Since $f(v_1) = \{1, 2\}$, we can define the function g on $V(G)$ by $g(u) = \{1, 2\}$, $g(x_1^1) = g(x_1^2) = \emptyset$, and $g(x) = f(x)$ otherwise. Then g is an I2-RDF of G of weight $\omega(f) + 2 \leq \frac{4(n-3)}{5} + 2 < \frac{4n}{5}$ such that $g(x) \neq \emptyset$ for every $x \in A$.

(XI) The other neighbors of u belong to ear-paths in Q_1.

Considering the above cases and subcases, we may assume that $\mathcal{Q} = Q_1 \cup Q_2$ and that each vertex in A is adjacent to at most one A-ear path in Q_2. In that case, since $G \in \mathcal{F}$, it is obtained from connected multigraph H without loops with $\delta(H) \geq 3$ by subdividing any edge at most twice so that the set of edges of H subdivided twice is independent (in H). Hence, let $u_1v_1, \ldots, u_kv_k$ be the edges of H subdivided twice and let A'' be the set of all vertices in H for which all edges that are incident are subdivided once. Therefore, we have $|V(H)| = 2k + |A''|$ and $|E(H)| = \frac{1}{2}\sum_{v \in V(H)} \deg(v) \geq \frac{3}{2}|V(H)| = 3k + \frac{3}{2}|A''|$ (because $\delta(H) \geq 3$, k edges of H are subdivided twice and the remaining edges are subdivided once). Hence, the order of G is

$$n = |V(H)| + |E(H)| + k \geq 6k + \frac{5}{2}|A''|.$$

It is easy to see that the function g defined on $V(G)$ by $g(x) = \{1, 2\}$ for $x \in V(H)$, and $g(x) = \emptyset$ otherwise, is an I2-RDF of G that assigns non-empty sets to vertices of A and $\omega(g) = 2|V(H)| = 4k + 2|A''| < \frac{4(6k+\frac{5}{2}|A''|)}{5} \leq \frac{4n}{5}$. This completes the proof. □

Now, we can proceed to the proof of Theorem 1.

Proof of Theorem 1. We use an induction on the order n. If $n \leq 5$,

then, clearly, G is connected having at most two vertices of degree at least three. More precisely, $G \in \{C_3, C_4, C_5\}$ or G is either obtained from two cycles C_3 sharing the same vertex or G is the complete bipartite graph $K_{2,3}$. In this case, it can be easily checked that $i_{r2}(C_n) \leq \frac{4}{5}n$, establishing the base case. Let $n \geq 6$, and assume that the result holds for all graphs G' of order less than n with minimum degree at least two such that the set of vertices with degree at least three is independent. Let G be a graph of order n such that $\delta(G) \geq 2$ and the set of vertices with degree at least three is independent. We can assume that G is connected for otherwise the result follows by applying the induction hypothesis on each component of G.

If $\Delta(G) = 2$, then $G = C_n$. Since for the cycle C_n, $i(C_n) = \lceil n/3 \rceil$, we obtain that $i_{r2}(C_n) \leq 2\lceil n/3 \rceil$ and, clearly, $2\lceil n/3 \rceil \leq \frac{4}{5}n$ for all $n \geq 8$. Since $i_{r2}(C_7) = 5 < \frac{4}{5}n$, we deduce that $i_{r2}(C_n) \leq \frac{4}{5}n$. Hence, assume that $\Delta(G) \geq 3$, and let $A = \{v \in V(G) \mid \deg(v) \geq 3\}$ and $B = V(G) - A$. Consider the A-ear paths and keep the same notations as defined in the proof of Proposition 3. Note that $A = \bigcup_{P \in \mathcal{Q}} X_P$, $V(G) = A \cup \bigcup_{P \in \mathcal{Q}} V(P)$ and $1 \leq |X_P| \leq 2$ for each $P \in Q$. Assume first that there exists an A-ear path P such that $\delta(G - V(P)) = 1$. Since G is simple, this means that $|V(P)| \geq 2$ and some vertex of G of degree three is adjacent to the end-vertices of P. Thus, $|X_P| = 1$. In that case, let $X_P = \{a\}$ and $N_G(a) - V(P) = \{b\}$. Clearly, $b \in B$ (since A is independent) and, thus, there is a unique A-ear path P' in which b is an end-vertex of P'. Let c be the other end-vertex of P' (possibly $b = c$). Let G' be the graph resulting from the deletion of vertex a and all vertices of P and P'. Then $\delta(G') \geq 2$ and by the induction hypothesis, $i_{r2}(G') \leq \frac{4|V(G')|}{5}$. On the other hand, since $G'' = G[V(P) \cup V(P') \cup \{a\}]$ is isomorphic to $C_{|V(P)|+1,|V(P')|}$, by Proposition 2, G'' has an I2-RDF g such that $\omega(g) \leq \frac{4n(G'')}{5}$ and $g(c) = \emptyset$. Now, for any $i_{r2}(G')-$function, the function h defined on $V(G)$ by $h(x) = f(x)$ for all $x \in V(G')$ and $h(x) = g(x)$ for all $x \in V(G'')$ is an I2-RDF of G. Therefore,

$$\begin{aligned} i_{r2}(G) &\leq i_{r2}(G') + i_{r2}(G'') \\ &\leq \frac{4|V(G')|}{5} + \frac{4|V(P) \cup V(P') \cup \{a\}|}{5} = \frac{4n}{5}. \end{aligned}$$

In the next, we can assume that $\delta(G - V(P)) \geq 2$ for each A-ear path $P \in \mathcal{Q}$. It follows that $|X_P| = 2$ for each A-ear path $P \in \mathcal{Q}$. Assume that $\mathcal{Q} - (Q_1 \cup Q_2 \cup Q_3 \cup Q_4) \neq \emptyset$, and let $P \in \mathcal{Q} - (Q_1 \cup Q_2 \cup Q_3 \cup Q_4)$. By Proposition 1, P has an I2-RDF g such that $\omega(g) \leq \frac{4|V(P)|}{5}$ and g assigns $\emptyset$ to the end-vertices of the path P. Now, let G' be the graph obtained from G' by removing all vertices of P. By the induction hypothesis, we have $i_{r2}(G') \leq \frac{4|V(G')|}{5}$. Clearly, for every $i_{r2}(G')$-function f, the function h defined on $V(G)$ by $h(x) = f(x)$ for all $x \in V(G')$ and $h(x) = g(x)$ for all $x \in V(P)$, is an I2-RDF of G and, thus, $i_{r2}(G) \leq i_{r2}(G') + i_{r2}(P) \leq \frac{4}{5}n$. Assume now that $\mathcal{Q} = Q_1 \cup Q_2 \cup Q_3 \cup Q_4$. Then $G \in \mathcal{F}$ and the result follows from Proposition 3. □

References

[1] H. Abdollahzadeh Ahangar, J. Amjadi, N. Jafari Rad, and V.D. Samodivkin, "Total k-Rainbow domination numbers in graphs," *Commun. Comb. Optim.*, vol. 3, pp. 37–50, 2018.

[2] J. Amjadi, N. Dehgardi, N. Mohammadi, S.M. Sheikholeslami, and L. Volkmann, "Independent 2-rainbow domination in trees," *Asian-Eur. J. Math.*, vol. 8, ID: 1550035, 2015.

[3] J. Amjadi, M. Falahat, N.J. Rad, and S.M. Sheikholeslami, "Strong equality between the 2-rainbow domination and independent 2-rainbow domination numbers in trees," *Bull. Malays. Math. Sci. Soc.*, vol. 39, pp. 205–218, 2016.

[4] B. Brešar, M. A. Henning, and D. F. Rall, "Rainbow domination in graphs," *Taiwanese J. Math.*, vol. 12, pp. 213–225, 2008.

[5] B. Brešar and T. K. Šumenjak, "On the 2-rainbow domination in graphs," *Discrete Appl. Math.*, vol. 155, pp. 2394–2400, 2007.

[6] G. J. Chang, J. Wu, and X. Zhu, "Rainbow domination on trees," *Discrete Appl. Math.*, vol. 158, pp. 8–12, 2010.

[7] M. Chellali and N.J. Rad, "Independent 2-rainbow domination in graphs," *J. Combin. Math. Combin. Comput.*, vol. 94, pp. 133–148, 2015.

[8] T. Chunling, L. Xiaohui, Y. Yuansheng, and L. Meiqin, "2-rainbow domination of generalized Petersen graphs $P(n, 2)$," *Discrete Appl. Math.*, vol. 157, pp. 1932–1937, 2009.

[9] S. Fujita, M. Furuya, and C. Magnant, "General upper bounds on independent k-rainbow domination," *Discrete Appl. Math.*, vol. 258, pp. 105–113, 2019.

[10] W. Goddard and M.A. Henning, "Independent domination in graphs, A survey and recent results," *Discrete Math.*, vol. 313, pp. 839—854, 2013.

[11] N. Jafari Rad, "Critical concept for 2-rainbow domination in graphs," *Australas. J. Combin.*, vol. 51, pp. 49–60, 2011.

[12] Q. Kang, V. Samodivkin, Z. Shao, S.M. Sheikholeslamiæ, and M. Soroudi, "Outer-independent k-rainbow domination," *Journal of Taibah University for Science*, vol. 13, pp. 883–891, 2019.

[13] D. Meierling, S. M. Sheikholeslami, and L. Volkmann, "Nordhaus-Gaddum bounds on the k-rainbow domatic number of a graph," *Appl. Math. Lett.*, vol. 24, pp. 1758–1761, 2011.

[14] Z. Shao, Z. Li, A. Peperko, J. Wan, and J. Žerovnik, "Independent rainbow domination of graphs," *Bull. Malays. Math. Sci. Soc.*, 2018.

[15] S. M. Sheikholeslami and L. Volkmann, "The k-rainbow domatic number of a graph," *Discuss. Math. Graph Theory*, vol. 32, pp. 129–140, 2012.

[16] Y. Wu and N. Jafari Rad, "Bounds on the 2-rainbow domination number of graphs," *Graphs Combin.*, vol. 29, pp. 1125–1133, 2013.

Rana Khoeilar, Mahla Kheibari,
Mustapha Chellali, Seyed Mahmoud Sheikholeslami

Received March 09, 2020
Accepted October 04, 2020

Rana Khoeilar
Department of Mathematics
Azarbaijan Shahid Madani University
Tabriz, I.R. Iran
E–mail: khoeilar@azaruniv.ac.ir

Mahla Kheibari
Department of Mathematics
Azarbaijan Shahid Madani University
Tabriz, I.R. Iran
E–mail: m.kheibari@azaruniv.ac.ir

Mustapha Chellali
LAMDA-RO Laboratory, Department of Mathematics
University of Blida
B.P. 270, Blida, Algeria
E–mail: m_chellali@yahoo.com

Seyed Mahmoud Sheikholeslami
Department of Mathematics
Azarbaijan Shahid Madani University
Tabriz, I.R. Iran
E–mail: s.m.sheikholeslami@azaruniv.ac.ir