



ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

2022, ТОМ 22, НОМЕР 4 (ИЮЛЬ–АВГУСТ)

ISSN 2226-1494 (PRINT), 2500-0373 (ONLINE)

ФОТОНИКА И ОПТОИНФОРМАТИКА

Исследование дисперсионных свойств конгруэнтного кристалла ниобата лития в терагерцовом диапазоне частот

Шумидай В.С., Опарин Е.Н., Набилкова А.О.,
Мельник М.В., Цыпкин А.Н., Козлов С.А. 635

ОПТИЧЕСКИЕ СИСТЕМЫ И ТЕХНОЛОГИИ

Исследование влияния наведенных механических напряжений, вызванных герметизацией стеклоприпоем, на значение коэффициента поляризационной экстинкции в двулучепреломляющем оптическом волокне

Калугин Е.Э., Мухтубаев А.Б., Мешиковский И.К. 643

Метод дистанционного контроля радиационных параметров космических аппаратов на основе рентгенофлуоресцентного анализа

Лукьянова Л.А., Свитнев И.В., Харитонов Е.А.,
Гаврилов И.Е. 650

Волоконно-оптический амплитудный датчик направления и величины изгиба

Дмитриев А.А., Гребнев К.В., Смирнов Д.С.,
Варжель С.В. 659

АВТОМАТИЧЕСКОЕ УПРАВЛЕНИЕ И РОБОТОТЕХНИКА

Компенсация внешних возмущений для многоканальных систем с запаздыванием в управлении

Нгуен Х.Т., Власов С.М., Пыркин А.А., Скобелева А.В. 666

КОМПЬЮТЕРНЫЕ СИСТЕМЫ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

Построение криптографических схем, основанных на эллиптических кривых над рациональными числами

Давыдов В.В., Дакуо Ж.-М.Н., Иогансон И.Д.,
Хуцаева А.Ф. 674

Алгоритм генерации проектных решений для управления данными и проектно-производственными процедурами на этапах жизненного цикла электронного изделия

Донецкая Ю.В. 681

Способ повышения эффективности комплексной обработки данных дистанционного зондирования Земли при решении задач мониторинга пространственных объектов

Карин С.А., Карин А.И. 691

Разработка модели обнаружения сетевых аномалий трафика в беспроводных распределенных самоорганизующихся сетях

Легашев Л.В., Гришина Л.С., Парфенов Д.И.,
Жигалов А.Ю. 699

Применение FN-корректора с целью повышения качества классификации аудиособытий

Голубков А.М., Шуранов Е.В. 708

Усиление роли микроархитектурных этапов проектирования встраиваемых систем

Кольчурин М.В., Пинкевич В.Ю., Платунов А.Е. 716

Многомерный двоичный классификатор дерева решений на основе неглубокой нейронной сети (на англ. яз.)

Марахимов А.Р., Кудайбергенов Ж.К.,
Худайбергенов К.К., Охундадаев У.Р. 725

Организация фазинг-тестирования многопоточных приложений на основе метода распараллеливания независимых переходов

Доронин О.В. 734

Метод защиты нейронных сетей от компьютерных бэкдор-атак на основе идентификации триггеров закладок

Менисов А.Б., Ломако А.Г., Дудкин А.С. 742

Программная инструментальная система создания адаптивных пользовательских интерфейсов

Тагирова Л.Ф., Субботин А.В., Зубкова Т.М. 751

Метод обнаружения инцидентов информационной безопасности по аномалиям в биометрических поведенческих чертах пользователя

Есипов Д.А., Асланова Н., Шабала Е.Е.,
Щетинин Д.С., Попов И.Ю. 760

Облегченная система рекомендаций для анализа социальных сетей с использованием гибридного алгоритма классификатора BERT-SVM (на англ. яз.)

Кирутика Н.С., Тайламбал Г. 769

МАТЕМАТИЧЕСКОЕ И КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ

Моделирование случайных процессов на основе преобразования Карунена–Лоэва

Ефимов А.С. 779

Управление численной диссипацией гибридного метода крупных частиц в задачах с вихревой неустойчивостью

Садин Д.В. 785

Численная модель импульсного подкритического стримерного сверхвысокочастотного разряда для задач плазменного поджига топливных смесей в газовой фазе

Буллат П.В., Волков К.Н., Мельникова А.И., Ренев М.Е. 792

Численное исследование прямой, винтовой и спиральной капиллярной трубок для хладагента CO₂ (на англ. яз.)

Правин Джадхав, Анджан Кумар Саху,
Сунита Баллал 804

КРАТКИЕ СООБЩЕНИЯ

Восстановление информации в зашумленном канале методом фантомной визуализации со спектральным мультиплексированием в оптическом диапазоне

Опарин Е.Н., Шумидай В.С., Исмаилов А.О.,
Цыпкин А.Н. 812

Главный редактор — В.О. Никифоров, д.т.н., профессор

Editor-in-Chief — Vladimir O. Nikiforov, D.Sc, Professor

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

Беззатеев С.В. — д.ф.-м.н., доцент (Россия)
Беликов А.В. — д.ф.-м.н., профессор (Россия)
Бессмертный И.А. — д.т.н., профессор (Россия)
Бобцов А.А. — д.т.н., профессор (Россия)
Бухановский А.В. — д.т.н. (Россия)
Волков К. Н. — д.ф.-м.н. (Великобритания)
Дукельский К.В. — к.т.н., доцент (Россия)
Ефимов Д.В. — PhD (Франция)
Заколдаев Д.А. — к.т.н., доцент (Россия)
Иорш И.В. — д.ф.-м.н. (Россия)
Кармановский Н.С. (заместитель главного редактора) — к.т.н. доцент (Россия)
Комаров И.И. — к.ф.-м.н., доцент (Россия)
Никоноров Н.В. — д.т.н., профессор (Россия)
Полищук Г.С. — к.т.н. (Россия)
Романов А.Е. — д.ф.-м.н., профессор (Россия)
Сидоркина И.Г. — д.т.н., профессор (Россия)
Степанов О.А. — д.т.н., профессор (Россия)
Татарникова Т.М. — д.т.н., доцент (Россия)
Тимофеев А. В. — д.т.н. (Казахстан)
Успенская М.В. — д.т.н., профессор (Россия)
Цыпкин А. Н. — д.ф.-м.н. (Россия)

Ответственный секретарь — И.В. Малькова

Издание зарегистрировано в Федеральной службе по надзору в сфере связи, информационных технологий и массовых коммуникаций (свидетельство ПИ №ФС77-67990 от 6 декабря 2016 г.).

Языки журнала: русский, английский.

ISSN 2226-1494 (print version), ISSN 2500-0373 (online version)

Англоязычное название: «Scientific and Technical Journal of Information Technologies, Mechanics and Optics». Транслитерированное название: «Nauchno-Tekhnicheskii Vestnik Informatsionnykh Tekhnologii, Mekhaniki i Optiki». Сокращенное название: «Sci.Tech. J. Inf. Technol. Mech. Opt.»

Журнал включен в каталог периодических изданий Ulrich's Periodical Directory.

Журнал входит в крупнейшие международные реферативные и наукометрические базы Scopus, EBSCO, ProQuest, EastView, ИВИС и другие.

Журнал входит в утвержденный Высшей аттестационной комиссией «Перечень ведущих рецензируемых научных журналов и изданий, выпускаемых в Российской Федерации, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени доктора и кандидата наук».

Периодичность издания — 6 выпусков в год.

Плата за публикации и редактирование не взимается.

Редакторы: Ю.В. Михайлова, Л.Н. Точилина, Т.В. Точилина
Перевод — Н.А. Корнев
Компьютерная верстка — Е.С. Егорова

Оф. 2136, Университет ИТМО, ул. Ломоносова., д. 9,
Санкт-Петербург, Российская Федерация, 191002

Телефон (812) 480 02 75
<http://ntv.ifmo.ru>
E-mail: ntvitmo@itmo.ru

Статьи Журнала доступны по лицензии Creative Commons
«Attribution-NonCommercial» 4.0 Всемирная.

EDITORIAL BOARD

Sergey V. Bezzateev — D.Sc, Associate Professor (Russia)
Andrey V. Belikov — D.Sc, Professor (Russia)
Igor A. Bessmertny — D.Sc, Professor (Russia)
Alexey A. Bobtsov — D.Sc, Professor (Russia)
Alexandr V. Boukhanovsky — D.Sc (Russia)
Konstantin N. Volkov — D.Sc (Great Britain)
Konstantin V. Dukelskii — PhD, Associate Professor (Russia)
Denis V. Efimov — PhD (France)
Danil A. Zakoldaev — PhD, Associate Professor (Russia)
Ivan V. Iorsh — D.Sc (Russia)
Nikolai S. Karmanovskiy (Deputy Chief Editor) — PhD, Associate Professor (Russia)
Igor I. Komarov — PhD, Associate Professor (Russia)
Nikolay V. Nikonorov — D.Sc, Professor (Russia)
Grigori S. Polishchuk — PhD (Russia)
Alexey E. Romanov — D.Sc, Professor (Russia)
Irina G. Sidorkina — D.Sc, Professor (Russia)
Oleg A. Stepanov — D.Sc, Professor (Russia)
Tatiana M. Tatarnikova — D.Sc, Associate Professor (Russia)
Andrey V. Timofeev — PhD (Kazakhstan)
Mayya V. Uspenskaya — D.Sc, Professor (Russia)
Anton N. Tsyppkin — D.Sc, Professor (Russia)

Executive secretary — Irina V. Malkova

Publication is registered by the Federal Inspectorate Service for Communication, Information Technologies and Communication Media, certificate PI FS77-67 990 dated December, 6th, 2016.

Languages of the journal: Russian, English

ISSN 2226-1494 (print version), ISSN 2500-0373 (online version)

The English title is "Scientific and Technical Journal of Information Technologies, Mechanics and Optics".

Transliterated title is "Nauchno-Tekhnicheskii Vestnik Informatsionnykh Tekhnologii, Mekhaniki i Optiki".

Abbreviated parallel title: "Sci.Tech. J. Inf. Technol. Mech. Opt."

The journal is catalogued in Ulrich's Periodical Directory.

The journal is included in the largest international abstract and scientometric databases Scopus, EBSCO, ProQuest, East View, EVIS and others.

The journal is included in the "List of leading scientific journals and periodicals under review in the Russian Federation, where the main scientific results of theses for the PhD and doctor of sciences degree must be published" approved by the Higher Attestation Commission.

Publication frequency is 6 times a year.

Publication and editing are free of charge.

Editors: Yulia V. Mikhailova, Lyubov N. Tochilina, Tatiana V. Tochilina
English Language Technical Editor — Nikolay A. Kornev
Computer layout — Ekaterina S. Egorova

of. 2136, ITMO University, Lomonosova Street, 9,
St.Petersburg, 191002, Russia

Phone (812) 480 02 75
<http://ntv.ifmo.ru/en>
E-mail: ntvitmo@itmo.ru



Подписано к печати 01.08.2022 Тираж 350 экз. Заказ № 88 (140)

Отпечатано в учреждении «Университетские телекоммуникации»
Адрес: 197101, Санкт-Петербург, Кронверкский пр., д. 49, литера А

ФОТОНИКА И ФОТОИНФОРМАТИКА PHOTONICS AND OPTOINFORMATICS

doi: 10.17586/2226-1494-2022-22-4-635-642

УДК 535-14

Исследование дисперсионных свойств конгруэнтного кристалла ниобата лития в терагерцовом диапазоне частот

Владимир Сергеевич Шумигой¹✉, Егор Николаевич Опарин²,
Александра Олеговна Набилкова³, Максим Владимирович Мельник⁴,
Антон Николаевич Цыпкин⁵, Сергей Аркадьевич Козлов⁶

^{1,2,3,4,5,6} Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

¹ vshumigay@itmo.ru✉, <https://orcid.org/0000-0001-8597-1196>

² en_oparin@itmo.ru, <https://orcid.org/0000-0002-4009-7594>

³ aonabilkova@itmo.ru, <https://orcid.org/0000-0001-5699-2211>

⁴ mmelnik@itmo.ru, <https://orcid.org/0000-0002-0796-0659>

⁵ tsyppkinan@itmo.ru, <https://orcid.org/0000-0002-9254-1116>

⁶ sakozlov@itmo.ru, <https://orcid.org/0000-0002-2692-627X>

Аннотация

Предмет исследования. В терагерцовом диапазоне частот рассмотрены дисперсионные кривые показателя конгруэнтного ниобата лития (cLN), срезанного перпендикулярно осям x и z . **Метод.** В исследовании применен метод терагерцовой спектроскопии с разрешением во времени, основанный на измерении эллиптичности пробного пучка, прошедшего через изначально изотропный детектирующий кристалл, который при воздействии терагерцового поля становится двулучепреломляющим. Величина наведенного двулучепреломления пропорциональна амплитуде терагерцового поля. С помощью фурье-анализа терагерцового импульса, прошедшего через кристалл cLN, и опорного импульса, не взаимодействующего с объектом, построены частотные зависимости показателя преломления и коэффициента поглощения исследуемого объекта. **Основные результаты.** Представлены дисперсионные кривые показателя преломления для действительной части показателя преломления кристалла cLN, срезанного вдоль плоскостей (100) и (001), в диапазоне частот 0,25–1,25 ТГц. Выполнено моделирование распространения полупериодного импульса в средах с дисперсиями на основании данных научных работ других авторов. В результате найдены временные формы выходных сигналов. Сделан вывод о неточности дисперсионных кривых из выбранных работ. Выявлены параметры, оптимизация которых позволила устранить неточности отображения дисперсионной зависимости для высокочастотной области терагерцового спектра. **Практическая значимость.** Полученные результаты о дисперсии очень важны для проектирования устройств, основанных на нелинейно-оптических эффектах. Полученные результаты полезны для генерации разностных частот, оптического выпрямления и генерации терагерцового излучения, а также для областей, где требуются точные данные о терагерцовых дисперсионных свойствах нелинейных кристаллов, в том числе cLN.

Ключевые слова

терагерцовое излучение, дисперсия, конгруэнтный кристалл, ниобат лития, терагерцовая спектроскопия с разрешением во времени, показатель преломления

Благодарности

Работа выполнена при финансовой поддержке Минобрнауки Российской Федерации в рамках Государственного задания (Паспорт № 2019-0903).

Ссылка для цитирования: Шумигой В.С., Опарин Е.Н., Набилкова А.О., Мельник М.В., Цыпкин А.Н., Козлов С.А. Исследование дисперсионных свойств конгруэнтного кристалла ниобата лития в терагерцовом диапазоне частот // Научно-технический вестник информационных технологий, механики и оптики. 2022. Т. 22, № 4. С. 635–642. doi: 10.17586/2226-1494-2022-22-4-635-642

Investigation of congruent lithium niobate crystal dispersion properties in the terahertz frequency range

Vladimir S. Shumigai¹, Egor N. Oparin², Aleksandra O. Nabilkova³, Maksim V. Melnik⁴, Anton N. Tsyarkin⁵, Sergei A. Kozlov⁶

^{1,2,3,4,5,6} ITMO University, Saint Petersburg, Russian Federation

¹ vshumigay@itmo.ru, <https://orcid.org/0000-0001-8597-1196>

² en_oparin@itmo.ru, <https://orcid.org/0000-0002-4009-7594>

³ aonabilkova@itmo.ru, <https://orcid.org/0000-0001-5699-2211>

⁴ mmelnik@itmo.ru, <https://orcid.org/0000-0002-0796-0659>

⁵ tsypkinan@itmo.ru, <https://orcid.org/0000-0002-9254-1116>

⁶ sakozlov@itmo.ru, <https://orcid.org/0000-0002-2692-627X>

Abstract

Dispersion curves of the refraction index of a congruent lithium niobate (cLN) crystal cut perpendicular to the x and z axes in the terahertz frequency range are considered. In the study, the method of terahertz time domain spectroscopy with time resolution is used passed through an initially isotropic detecting crystal which becomes birefringent when exposed to a terahertz field. The magnitude of the induced birefringence is proportional to the amplitude of the terahertz field. Using Fourier analysis of a terahertz pulse passing through a cLN crystal and a reference pulse that does not interact with the object, the frequency dependences of the refractive index and the absorption coefficient of the object under study are constructed. Dispersion curves are presented for the real part of the refractive index of a cLN crystal cut along the planes (100) and (001), in the frequency range 0.25–1.25 THz. Simulation of the propagation of a one-and-a-half-cycle pulse in media with dispersion is performed based on the data of scientific papers by other authors. As a result, the temporal forms of the output signals are found. Conclusion about the inaccuracy of the dispersion curves from the selected works is made. The parameters has been identified whose optimization made it possible to eliminate inaccuracies in the display of the dispersion dependence for the high-frequency region of the terahertz spectrum were identified. The results obtained are very important for the design of devices based on nonlinear optical effects. These data will be useful for the generation of difference frequencies, optical rectification and generation of terahertz radiation as well as for areas where accurate data on the terahertz dispersion properties of nonlinear crystals, including cLN, are required.

Keywords

terahertz radiation, dispersion, congruent crystal, lithium niobate, terahertz time-domain spectroscopy, refractive index

Acknowledgements

This work was supported by the Ministry of Science and Education of the Russian Federation (Passport No. 2019-0903).

For citation: Shumigai V.S., Oparin E.N., Nabilkova A.O., Melnik M.V., Tsyarkin A.N., Kozlov S.A. Investigation of congruent lithium niobate crystal dispersion properties in the terahertz frequency range. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2022, vol. 22, no. 4, pp. 635–642 (in Russian). doi: 10.17586/2226-1494-2022-22-4-635-642

Введение

Впервые терагерцовое (ТГц) излучение было экспериментально продемонстрировано около 100 лет тому назад [1], но его разнообразные возможности применения начали активно изучаться только в течение последних десятилетий [1, 2]. В настоящее время существуют доступные источники как импульсного, так и непрерывного излучений ТГц диапазона. Импульсное ТГц излучение применяется в спектроскопии [3], при исследовании фазовых переходов, вызванных сильным полем [4], в ТГц нелинейной оптике [5, 6], и др. Во многих исследованиях используются прозрачные в ТГц диапазоне частот материалы. Пример таких материалов — кристаллы, применяемые для генерации или детектирования импульсного ТГц поля. Один из самых эффективных генераторов ТГц излучения — одноосный конгруэнтный кристалл ниобата лития (cLN) [7–9]. Более того, кристалл cLN является материалом с сильными нелинейно-оптическими свойствами [10]. Области применения широкополосного ТГц излучения требуют наличие информации о линейных свойствах материалов, таких как дисперсия и поглощение.

Один из методов получения полной информации о действительной и мнимой частях показателя прелом-

ления — метод ТГц спектроскопии с разрешением во времени [11, 12]. Данная методика основана на измерении эллиптичности пробного пучка, прошедшего через изначально изотропный детектирующий кристалл, который при воздействии ТГц поля преобразуется в двулучепреломляющий. Величина наведенного двулучепреломления пропорциональна напряженности электрического поля в импульсе ТГц излучения. Для измерения временной формы ТГц импульса выполняют изменение временной задержки зондирующего импульса на некоторую величину [13]. Далее с помощью фурье-анализа ТГц импульса, прошедшего через материал, и опорного импульса, не взаимодействующего с объектом, получают частотные зависимости показателя преломления и коэффициента поглощения исследуемого объекта.

Отметим, что дисперсионные свойства кристалла cLN в ТГц диапазоне частот были измерены в работах [14, 15], где кристалл продемонстрировал сильную дисперсию показателя преломления в диапазоне частот 0,2–2 ТГц. Данные результаты оказывали большое влияние на временную структуру импульса [16].

В настоящей работе приведены зависимости показателя преломления cLN, срезанного перпендикулярно осям x и z , от частоты излучения в диапазоне 0,25–

1,25 ТГц. Зависимости представлены для значений толщин кристалла: 0,52 мм, 1 мм и 2,21 мм. Приведено сравнение дисперсионных кривых с дисперсиями кристалла cLN в ТГц диапазоне частот, полученными в научных работах других авторов. Дополнительно выполнена оценка сравнения дисперсионного уширения ТГц импульса во времени в процессе моделирования его распространения в среде с заданной дисперсией. Показано, что изначально полупериодный ТГц импульс уширяется в средах с дисперсиями, полученными в работах других авторов, что не соответствует экспериментальным данным из настоящей работы. В результате сделан вывод, что дисперсионные кривые cLN из рассмотренных работ не согласовываются со значениями показателя преломления в ТГц диапазоне частот, полученными в настоящей работе. Предложена методика оптимизации параметров в процессе пост-обработки.

Экспериментальные методы

Измерения проведены на стандартной установке ТГц спектроскопии с разрешением во времени. В качестве источника излучения использован фемтосекундный Yb-лазер с центральной длиной волны 1050 нм, длительностью импульсов 100 фс и их энергией 30 нДж.

Схема ТГц спектрометра представлена на рис. 1. Инфракрасный пучок разделен на два в соотношении мощностей 90:10 с помощью светоделителя. Более мощный пучок освещает кристалл арсенида индия (InAs), помещенный в магнитное поле со значением индукции 2,4 Т, в котором за счет фотоэффекта Дембера производилась генерация ТГц излучения. С помощью системы линз ТГц импульсы направлялись через образец кристалла cLN в систему электрооптического детектирования на основе кристалла теллурида кадмия (CdTe). В ходе эксперимента были зарегистрированы временные формы ТГц импульсов, прошедшие через свободное пространство (О) и три образца (L_{1-3}) — x-cLN толщиной 1 мм и 2,21 мм, и z-cLN толщиной 0,52 мм. Спектр детектируемого ТГц излучения обладал полушириной 0,5 ТГц и центральной частотой 0,4 ТГц.

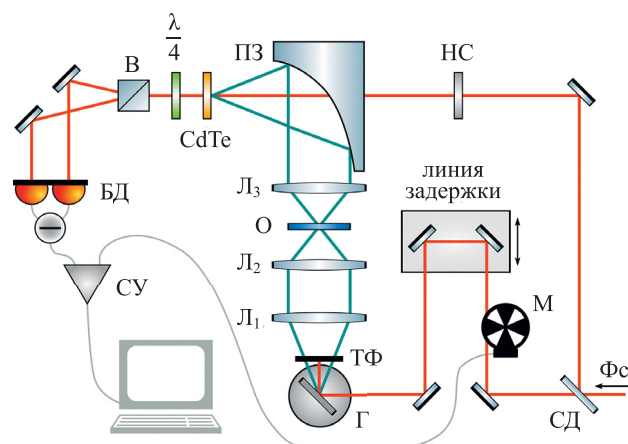


Рис. 1. Экспериментальная установка терагерцевой спектроскопии с разрешением во времени.

Фс — фемтосекундный инфракрасный импульс, СД — светоделитель 90:10, Г — генератор ТГц излучения, ТФ — фильтр, отрезающий инфракрасного излучения, L_{1-3} — линзы, О — образец, ПЗ — параболическое зеркало, НС — нейтральный светофильтр, В — призма Волластона, БД — балансный детектор, СУ и М — синхронный усилитель и оптико-механический модулятор

Fig. 1. Experimental setup of time-resolved terahertz spectroscopy.

Фс is femtosecond infrared pulse, СД is 90:10 beam splitter, Г is THz radiation generator, ТФ is filter that cuts off infrared radiation, L_{1-3} are lenses, О is sample, ПЗ is parabolic mirror, НС is neutral filter, В is Wollaston prism, БД is balanced detector, СУ and М are synchronous amplifier and opto-mechanical modulator

В результате экспериментальных данных о форме и спектре ТГц импульса, прошедшего через исследуемый кристалл cLN, построены дисперсионные кривые в диапазоне частот 0,25–1,25 ТГц. Действительная часть показателя преломления cLN в ТГц области частот найдена при помощи данных о спектре и форме ТГц импульса, прошедшего через кристалл. В качестве примера на рис. 2, а приведена типичная временная форма ТГц импульса, полученного при исследовании

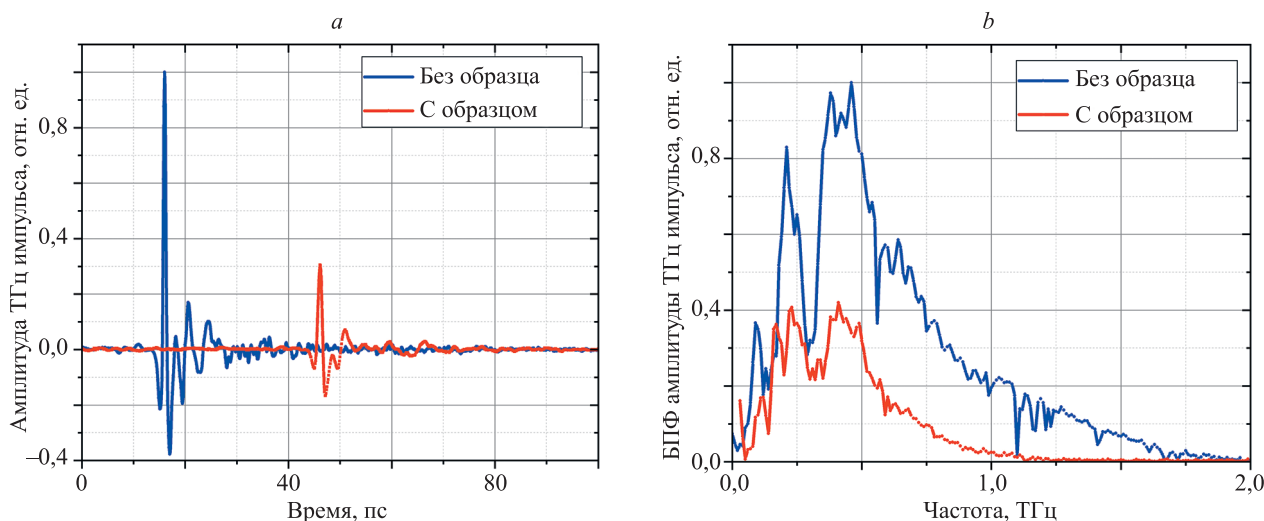


Рис. 2. Временная форма терагерцевого импульса (а), спектр терагерцевого импульса (b)

Fig. 2. Temporal form of the terahertz pulse (a), the spectrum of the terahertz pulse (b)

кристалла cLN, срезанного перпендикулярно оси x и обладающего толщиной 2,21 мм. На рис. 2, b изображен спектр детектируемого сигнала с центральной частотой 0,4 ТГц и полушириной 1 ТГц. Линии поглощения, возникающие на частотах выше 1 ТГц, вызваны содержанием водяных паров в воздухе.

Обработка экспериментальных данных

Для расчета комплексного показателя преломления кристаллов cLN для каждого образца зарегистрировано два импульса: $E_{sam}(t)$, прошедший через образец, и $E_{ref}(t)$, распространённый в отсутствии образца. Полученные сигналы преобразованы из временной области в спектральную при помощи быстрого преобразования Фурье и имеют вид:

$$E_{sam}(\omega) = \frac{4\hat{n}_s n_0}{(\hat{n}_s + n_0)^2} \exp\left[-i\hat{n}_s \frac{\omega l}{c}\right] E_0(\omega),$$

$$E_{ref}(\omega) = \exp\left[-in_0 \frac{\omega l}{c}\right] E_0(\omega),$$

где $\hat{n}_s = n_s - ik_s$ — комплексный показатель преломления образца; n_s — действительная часть \hat{n}_s ; k_s — мнимая часть \hat{n}_s ; i — мнимая единица; n_0 — показатель преломления воздуха; l — длина распространения излучения в образце, которая в рассматриваемом случае равна его толщине; ω — циклическая частота излучения; c — скорость света в вакууме; E_0 — амплитуда электрического поля.

Извлечение оптических характеристик образца требует расчета его передаточной функции:

$$H_0(\omega) = \frac{E_{sam}(\omega)}{E_{ref}(\omega)} = \frac{4\hat{n}_s n_0}{(\hat{n}_s + n_0)^2} \exp\left[-k_s \frac{\omega l}{c}\right] \exp\left[-i(n_s - n_0) \frac{\omega l}{c}\right].$$

В рассматриваемом спектральном диапазоне кристалл cLN можно считать оптически прозрачным, т. е. допускается комплексный показатель преломления $\hat{n}(\omega)$, входящий во френелевский коэффициент пропускания, аппроксимировать действительной частью n_s . Например, получим аргумент от упрощенной таким образом передаточной функции:

$$\angle H(\omega) = -i(n_s - n_0) \frac{\omega l}{c},$$

где $\angle H(\omega)$ — фаза комплекснозначной передаточной функции $H(\omega)$.

Отсюда может быть рассчитан показатель преломления образца:

$$n_s = n_0 - \frac{c}{\omega l} \angle H(\omega).$$

Описанный метод расчета крайне чувствителен к корректному определению фазового спектра ТГц излучения. Неточности может вносить ряд факторов. Например, множественное отражение ТГц импульса от границ образца приводит к возникновению в сиг-

нале повторных импульсов. Их наличие выражается в появлении в спектре осцилляций на частоте следования импульсов. По своей природе данный эффект близок к действию интерферометра Фабри–Перо. Однако интерференция в данном случае оказывается строго спектральной и может быть исключена путем наложения на сигнал временного окна.

Второй важный фактор — корректное определение рабочего диапазона спектрометра. Например, удовлетворительное соотношение сигнал-шум выполняется для задействованного в работе спектрометра в диапазоне 0,25–1,25 ТГц [17]. Неизбежное наличие шума в сигнале приводит к искажению фазовой информации на частотах ниже и выше указанных. Особенно важна низкочастотная часть спектра, которая в первую очередь влияет на развитие фазы во всем рассматриваемом диапазоне. При этом разрывы и скачки диапазона могут исказить кривую показателя преломления, создавая картину, похожую на область высокой дисперсии. Во избежание этого фаза на частотах ниже 0,25 ТГц была линейно экстраполирована до нуля, опираясь на поведение кривой в пределах рабочего диапазона.

С учетом перечисленных факторов построены дисперсионные кривые для cLN, срезанного перпендикулярно осям x и z . Одновременно построены дисперсионные кривые по данным научных работ [18–20], где кристалл cLN проявлял сильную дисперсию в ТГц диапазоне частот. Данные дисперсионные кривые представлены на рис. 3.

Сравнивая зависимости из работ [18–20] с кривыми, полученными из эксперимента в настоящей работе, сделаем вывод, что для x -срезанного cLN в диапазоне 0,5–1,25 ТГц наблюдается расхождение данных — экспериментальные кривые демонстрируют меньший рост. Для z -срезанного cLN значительное расхождение данных наблюдается на всем исследуемом диапазоне ТГц частот.

Выполним моделирование распространения ТГц импульсов в средах с вычисленными дисперсиями для сравнения экспериментальных данных о форме импульса с результатами численного моделирования.

Численное моделирование

Рассмотрим распространение фемтосекундных низкоинтенсивных импульсов ТГц диапазона частот в диэлектрической изотропной немагнитной среде. Математическая модель динамики импульсов в таких средах может быть описана уравнением [21]:

$$\frac{\partial E}{\partial z} - a \frac{\partial^3 E}{\partial \tau^3} + b \int_{-\infty}^{\tau} E d\tau' = 0,$$

где E — действительная часть электрического поля импульса; $\tau = t - N_0 \frac{z}{c}$ — время в бегущей системе координат, c — скорость света в вакууме, z — координата, вдоль которой распространяется импульс. N_0 , a и b — параметры, описывающие эффективный показатель преломления среды и его дисперсию:

$$n(\omega) = N_0 + ac\omega^2 - b \frac{c}{\omega^2}. \quad (1)$$

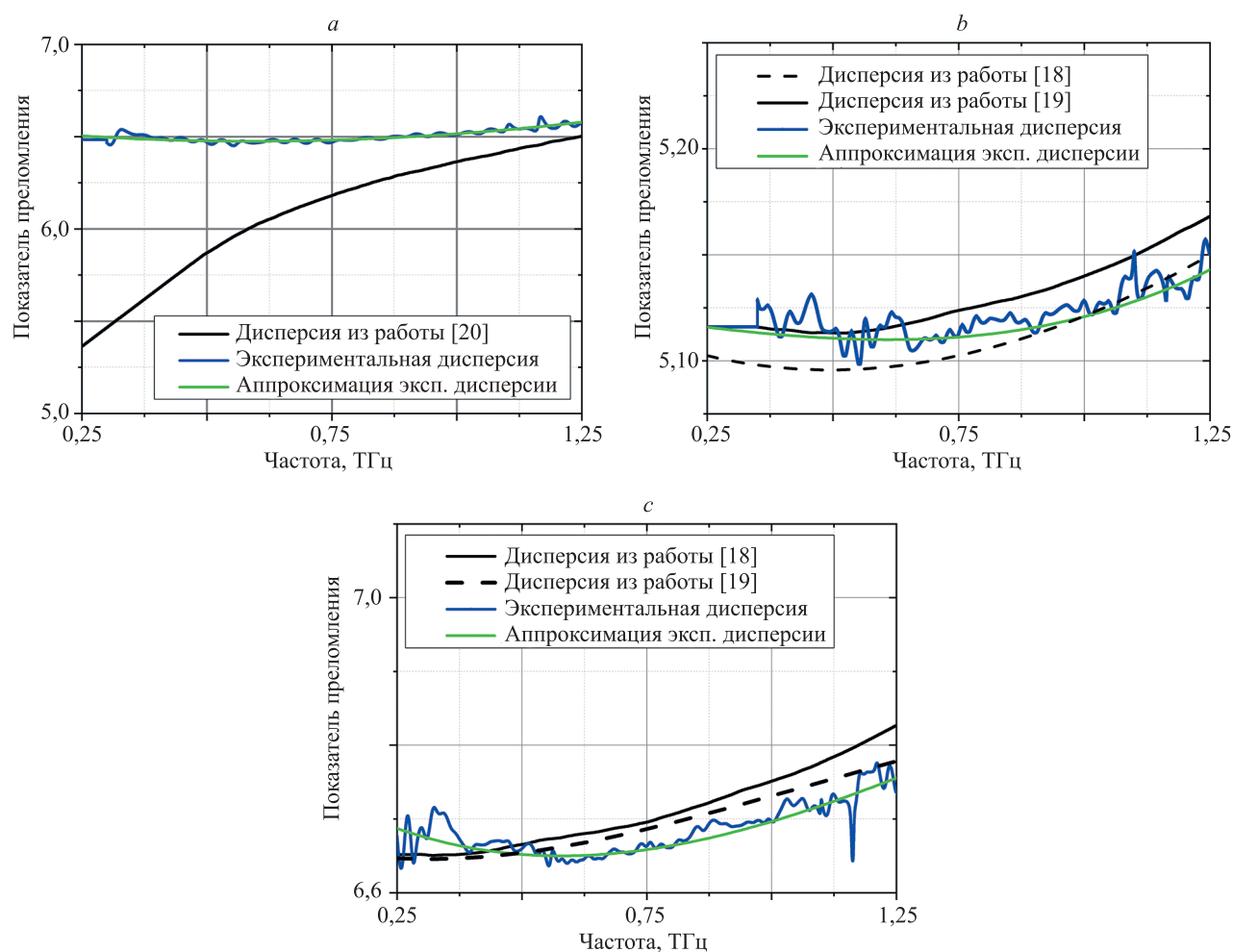


Рис. 3. Дисперсионные кривые из работ [18–20] и вычисленные дисперсии для кристалла cLN, срезанного перпендикулярно осям: z для случая, когда кристалл изотропен (а); x для необыкновенной (b) и обыкновенной (c) волн

Fig. 3. Dispersion curves from works [18–20] and calculated dispersions for the cLN crystal cut perpendicular to the axes: z for the case when the crystal is isotropic (a); x for extraordinary (b) and ordinary (c) waves

Рассчитанные дисперсии кристалла cLN, срезанного вдоль плоскостей (100) и (001), и начальная временная форма ТГц импульса обработаны программой, аппроксимирующей экспериментальные кривые с помощью уравнения (1). Выполним моделирование распространения полупериодного ТГц импульса в среде с заданной дисперсией. При правильных параметрах смоделированный импульс должен совпадать по форме и по времени с импульсом, детектируемым после прохождения кристалла cLN в эксперименте. Соответствующие импульсы представлены на рис. 4.

Результаты моделирования с большой точностью совпали с данными из эксперимента и показали, что изначально полупериодный импульс не уширяется после прохождения через образец cLN толщинами 0,52 мм, 1 мм и 2,21 мм. В результате сделан вывод, что используемая методика расчета действительной и мнимой частей показателя преломления дает корректные результаты.

Убедившись в том, что экспериментальные данные сходятся с результатами моделирования с высокой точностью, было произведено численное моделирование распространения импульса заданной формы в кристалле cLN со значениями дисперсий из работ [18–20].

По результатам численного моделирования распространения полупериодного импульса в средах с дисперсиями найдены временные формы выходных сигналов. Данные импульсы представлены на рис. 5 и рис. 6.

Анализ временных форм ТГц импульсов, прошедших через среды со значениями дисперсии из работ [18–20], показал, что полупериодный импульс испытывает дисперсионное уширение, тем самым увеличивая число колебаний, чего на эксперименте не наблюдалось. В соответствии с этим был сделан вывод о важности выбора параметров в процессе пост-обработки для дальнейшего определения показателя преломления исследуемой среды.

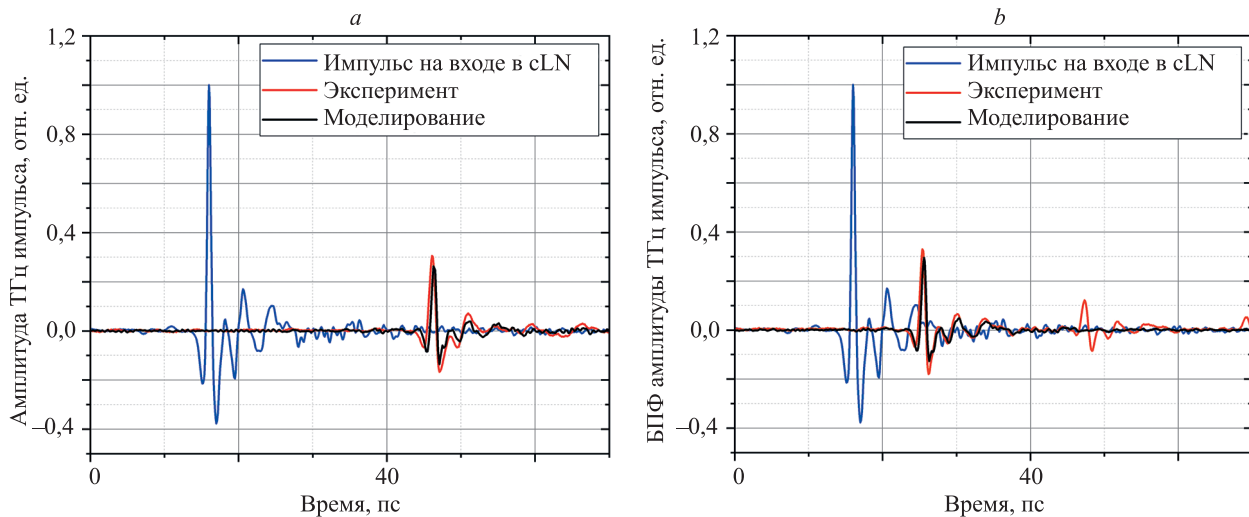


Рис. 4. Временные формы терагерцовых импульсов — до и после прохождения через кристалл cLN: x-срезанного толщиной 2,21 мм (a), z-срезанного толщиной 0,52 мм (b)

Fig. 4. Temporal forms of terahertz pulses — before and after passing through the cLN crystal: x-cut with a thickness of 2.21 mm (a), z-cut with a thickness of 0.52 mm (b)

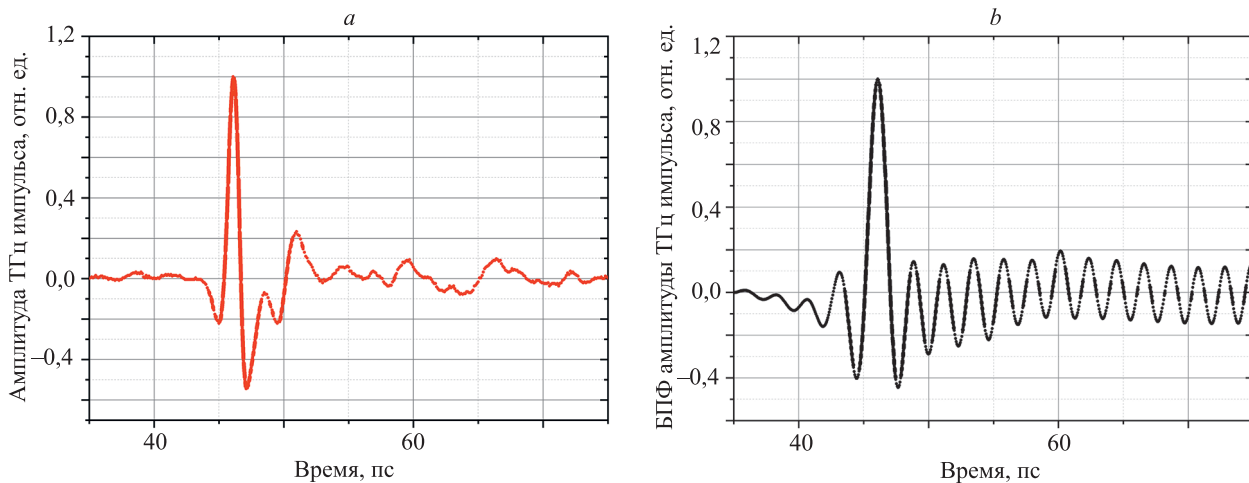


Рис. 5. Временные формы терагерцовых импульсов после распространения в x-cLN толщиной 2,21 мм с: экспериментальной дисперсией (a), дисперсией из работы [19] (b)

Fig. 5. Temporal forms of terahertz pulses after propagation in x-cLN with a thickness of 2.21 mm with: experimental dispersion (a), dispersion from work [19] (b)

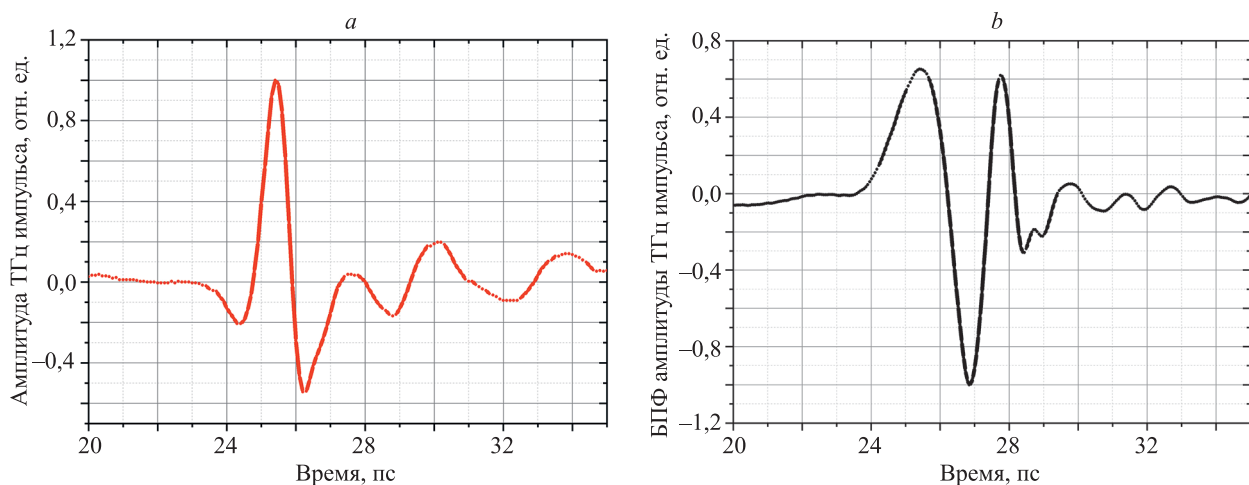


Рис. 6. Временные формы терагерцовых импульсов после распространения в z-cLN толщиной 0,52 мм с: экспериментальной дисперсией (a), дисперсией из работы [20] (b)

Fig. 6. Temporal forms of terahertz pulses after propagation in a 0.52 mm thick z-cLN with: experimental dispersion (a), dispersion from work [20] (b)

Заклучение

В результате проведения экспериментов с использованием установки терагерцовой спектроскопии с разрешением во времени были найдены дисперсионные кривые показателя преломления кристалла cLN, срезанного перпендикулярно осям x и z , в диапазоне частот 0,25–1,25 ТГц. Численное моделирование распространения полуторапериодного импульса в средах с найденными дисперсиями показало, что терагерцовый импульс не претерпевает дисперсионное уширения, что находится в полном соответствии с экспериментальными данными. Выполнен сравнительный анализ полученных дисперсионных кривых показателя преломления с данными работ [18–20]. Результаты, полученные в ходе численного моделирования распространения терагерцового излучения в кристалле cLN с такими дисперсиями, показали, что полуторапериодный импульс дисперсионно уширился в x -cLN толщиной 1 мм и 2,21 мм, и в z -cLN толщиной 0,52 мм, чего не наблюдалось при эксперименте.

Расхождение дисперсионных кривых показателя преломления кристалла cLN можно объяснить разным

выбором параметров в процессе пост-обработки данных из эксперимента. В связи с этим была предложена методика подбора параметров для пост-обработки, исключающая ошибки при определении дисперсии образцов. Согласно ей, необходимо накладывать на выходной сигнал временное окно, чтобы не учитывать повторные импульсы, обусловленные множественным отражением терагерцового импульса от границ образца. Также важным параметром является рабочий диапазон спектрометра. Его необходимо точно определить для получения корректных данных. Необходимо задействовать диапазон частот, для которого сохраняются большие значения соотношения сигнал-шум, так как наличие шума в сигнале приводит к искажению фазовой информации на участках спектра с низким соотношением сигнал-шум.

Рассмотренные примеры показали, что точные знания о дисперсионных характеристиках позволяют учитывать особенности материалов в терагерцовом диапазоне частот и для их использования в системах терагерцовой фотоники.

Литература

1. Glagolewa-Arkadiewa A. Short electromagnetic waves of wavelength up to 82 microns // *Nature*. 1924. V. 113. N 2844. P. 640–640. <https://doi.org/10.1038/113640a0>
2. Evenson K.M., Wells J.S., Matarrese L.M. Absolute Frequency Measurements of The CO₂ CW Laser at 28 THz (10.6μm) // *Applied Physics Letters*. 1970. V. 16. N 6. P. 251–253. <https://doi.org/10.1063/1.1653183>
3. Hangyo M., Tani M., Nagashima T. Terahertz time-domain spectroscopy of solids: A review // *International Journal of Infrared and Millimeter Waves*. 2005. V. 26. N 12. P. 1661–1690. <https://doi.org/10.1007/s10762-005-0288-1>
4. Liu M.K., Hwang H.Y., Tao H., Strikwerda A.C., Fan K.B., Keiser G.R., Sternbach A.J., West K.G., Kittiwatanakul S., Lu J.W., Wolf S.A., Omenetto F.G., Zhang X., Nelson K.A., Averitt R.D. Terahertz-field-induced insulator-to-metal transition in vanadium dioxide metamaterial // *Nature*. 2012. V. 487. N 7407. P. 345–348. <https://doi.org/10.1038/nature11231>
5. Тсупкин А.Н., Мельник М.В., Жуклова М.О., Воронцова И.О., Путилин С.Е., Козлов С.А., Zhang X.-C. High Kerr nonlinearity of water in THz spectral range // *Optics express*. 2019. V. 27. N 8. P. 10419–10425. <https://doi.org/10.1364/OE.27.010419>
6. Тсупкин А., Жуклова М., Воронцова И., Куля М., Путилин С., Козлов С., Choudhary S., Boyd R.W. Giant third-order nonlinear response of liquids at terahertz frequencies // *Physical Review Applied*. 2021. V. 15. N 5. P. 054009. <https://doi.org/10.1103/PhysRevApplied.15.054009>
7. Kawase K., Sato M., Taniuchi T., Ito H. Coherent tunable THz-wave generation from LiNbO₃ with monolithic grating coupler // *Applied Physics Letters*. 1996. V. 68. N 18. P. 2483–2485. <https://doi.org/10.1063/1.115828>
8. Bodrov S.B., Ilyakov I.E., Shishkin B.V., Stepanov A.N. Efficient terahertz generation by optical rectification in Si-LiNbO₃-air-metal sandwich structure with variable air gap // *Applied Physics Letters*. 2012. V. 100. N 20. P. 201114. <https://doi.org/10.1063/1.4719674>
9. Nagashima K., Kosuge A. Design of rectangular transmission gratings fabricated in LiNbO₃ for high-power terahertz-wave generation // *Japanese Journal of Applied Physics*. 2010. V. 49. N 12R. P. 122504. <https://doi.org/10.1143/JJAP.49.122504>
10. Zhukova M., Melnik M., Vorontsova I., Tsypkin A., Kozlov S. Estimations of low-inertia cubic nonlinearity featured by electro-optical crystals in the thz range // *Photonics*. 2020. V. 7. N 4. P. 98. <https://doi.org/10.3390/photonics7040098>
11. Neu J., Schmuttenmaer C.A. Tutorial: An introduction to terahertz time domain spectroscopy (THz-TDS) // *Journal of Applied Physics*. 2018. V. 124. N 23. P. 231101. <https://doi.org/10.1063/1.5047659>

References

1. Glagolewa-Arkadiewa A. Short electromagnetic waves of wavelength up to 82 microns. *Nature*, 1924, vol. 113, no. 2844, pp. 640–640. <https://doi.org/10.1038/113640a0>
2. Evenson K.M., Wells J.S., Matarrese L.M. Absolute Frequency Measurements of The CO₂ CW Laser at 28 THz (10.6μm). *Applied Physics Letters*, 1970, vol. 16, no. 6, pp. 251–253. <https://doi.org/10.1063/1.1653183>
3. Hangyo M., Tani M., Nagashima T. Terahertz time-domain spectroscopy of solids: A review. *International Journal of Infrared and Millimeter Waves*, 2005, vol. 26, no. 12, pp. 1661–1690. <https://doi.org/10.1007/s10762-005-0288-1>
4. Liu M.K., Hwang H.Y., Tao H., Strikwerda A.C., Fan K.B., Keiser G.R., Sternbach A.J., West K.G., Kittiwatanakul S., Lu J.W., Wolf S.A., Omenetto F.G., Zhang X., Nelson K.A., Averitt R.D. Terahertz-field-induced insulator-to-metal transition in vanadium dioxide metamaterial. *Nature*, 2012, vol. 487, no. 7407, pp. 345–348. <https://doi.org/10.1038/nature11231>
5. Tsypkin A.N., Melnik M.V., Zhukova M.O., Vorontsova I.O., Putilin S.E., Kozlov S.A., Zhang X.-C. High Kerr nonlinearity of water in THz spectral range. *Optics express*, 2019, vol. 27, no. 8, pp. 10419–10425. <https://doi.org/10.1364/OE.27.010419>
6. Tsypkin A., Zhukova M., Vorontsova I., Kulya M., Putilin S., Kozlov S., Choudhary S., Boyd R.W. Giant third-order nonlinear response of liquids at terahertz frequencies. *Physical Review Applied*, 2021, vol. 15, no. 5, pp. 054009. <https://doi.org/10.1103/PhysRevApplied.15.054009>
7. Kawase K., Sato M., Taniuchi T., Ito H. Coherent tunable THz-wave generation from LiNbO₃ with monolithic grating coupler. *Applied Physics Letters*, 1996, vol. 68, no. 18, pp. 2483–2485. <https://doi.org/10.1063/1.115828>
8. Bodrov S.B., Ilyakov I.E., Shishkin B.V., Stepanov A.N. Efficient terahertz generation by optical rectification in Si-LiNbO₃-air-metal sandwich structure with variable air gap. *Applied Physics Letters*, 2012, vol. 100, no. 20, pp. 201114. <https://doi.org/10.1063/1.4719674>
9. Nagashima K., Kosuge A. Design of rectangular transmission gratings fabricated in LiNbO₃ for high-power terahertz-wave generation. *Japanese Journal of Applied Physics*, 2010, vol. 49, no. 12R, pp. 122504. <https://doi.org/10.1143/JJAP.49.122504>
10. Zhukova M., Melnik M., Vorontsova I., Tsypkin A., Kozlov S. Estimations of low-inertia cubic nonlinearity featured by electro-optical crystals in the thz range. *Photonics*, 2020, vol. 7, no. 4, pp. 98. <https://doi.org/10.3390/photonics7040098>
11. Neu J., Schmuttenmaer C.A. Tutorial: An introduction to terahertz time domain spectroscopy (THz-TDS). *Journal of Applied Physics*, 2018, vol. 124, no. 23, pp. 231101. <https://doi.org/10.1063/1.5047659>

12. Peretti R., Mitryukovskiy S., Froberger K., Mebarki M.A., Eliet S., Vanwolleghem M., Lampin J.-F. THz-TDS time-trace analysis for the extraction of material and metamaterial parameters // *IEEE Transactions on Terahertz Science and Technology*. 2018. V. 9. N 2. P. 136–149. <https://doi.org/10.1109/TTHZ.2018.2889227>
13. Nahata A., Weling A.S., Heinz T.F. A wideband coherent terahertz spectroscopy system using optical rectification and electro-optic sampling // *Applied Physics Letters*. 1996. V. 69. N 16. P. 2321–2323. <https://doi.org/10.1063/1.117511>
14. Pálfalvi L., Hebling J., Kuhl J., Pter Á., Polgár K. Temperature dependence of the absorption and refraction of Mg-doped congruent and stoichiometric LiNbO₃ in the THz range // *Journal of Applied Physics*. 2005. V. 97. N 12. P. 123505. <https://doi.org/10.1063/1.1929859>
15. Wang T.D., Lin S.T., Lin Y.Y., Chiang A.C., Huang Y.C. Forward and backward terahertz-wave difference-frequency generations from periodically poled lithium niobate // *Optics Express*. 2008. V. 16. N 9. P. 6471–6478. <https://doi.org/10.1364/OE.16.006471>
16. Дроздов А.А., Козлов С.А. Фазовая самомодуляция одноперiodных оптических волн // *Научно-технический вестник Санкт-Петербургского государственного университета информационных технологий, механики и оптики*. 2011. № 2(72). С. 99–105.
17. Grachev Y.V., Osipova M.O., Kuz'mina A.V., Bepalov V.G. Determining the working band of frequencies of a pulsed terahertz spectrometer // *Journal of Optical Technology*. 2014. V. 81. N 8. P. 468–471. <https://doi.org/10.1364/JOT.81.000468>
18. Li D., Ma G., Ge J., Hu S., Dai N. Terahertz pulse shaping via birefringence in lithium niobate crystal // *Applied Physics B*. 2009. V. 94. N 4. P. 623–628. <https://doi.org/10.1007/s00340-008-3334-6>
19. Unferdorben M., Szaller Z., Hajdara I., Hebling J., Pálfalvi L. Measurement of refractive index and absorption coefficient of congruent and stoichiometric lithium niobate in the terahertz range // *Journal of Infrared, Millimeter, and Terahertz Waves*. 2015. V. 36. N 12. P. 1203–1209. <https://doi.org/10.1007/s10762-015-0165-5>
20. Mao Z.-L., Hou B.-H., Wang L., Sun Y.-M., Hao W. The study of the terahertz spectral of LiNbO₃ crystal // *Proc. of the Joint 31st International Conference on Infrared Millimeter Waves and 14th International Conference on Terahertz Electronics*. 2006. P. 465–465. <https://doi.org/10.1109/ICIMW.2006.368673>
21. Самарцев В., Козлов С. Основы фемтосекундной оптики. М.: Физматлит, 2009. 292 с.
22. Peretti R., Mitryukovskiy S., Froberger K., Mebarki M.A., Eliet S., Vanwolleghem M., Lampin J.-F. THz-TDS time-trace analysis for the extraction of material and metamaterial parameters. *IEEE Transactions on Terahertz Science and Technology*, 2018, vol. 9, no. 2, pp. 136–149. <https://doi.org/10.1109/TTHZ.2018.2889227>
23. Nahata A., Weling A.S., Heinz T.F. A wideband coherent terahertz spectroscopy system using optical rectification and electro-optic sampling. *Applied Physics Letters*, 1996, vol. 69, no. 16, pp. 2321–2323. <https://doi.org/10.1063/1.117511>
24. Pálfalvi L., Hebling J., Kuhl J., Pter Á., Polgár K. Temperature dependence of the absorption and refraction of Mg-doped congruent and stoichiometric LiNbO₃ in the THz range. *Journal of Applied Physics*, 2005, vol. 97, no. 12, pp. 123505. <https://doi.org/10.1063/1.1929859>
25. Wang T.D., Lin S.T., Lin Y.Y., Chiang A.C., Huang Y.C. Forward and backward terahertz-wave difference-frequency generations from periodically poled lithium niobate. *Optics Express*, 2008, vol. 16, no. 9, pp. 6471–6478. <https://doi.org/10.1364/OE.16.006471>
26. Drozdov A., Kozlov S. Phase self-modulation of single-cycle optical waves. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2011, no. 2(72), pp. 99–105. (in Russian)
27. Grachev Y.V., Osipova M.O., Kuz'mina A.V., Bepalov V.G. Determining the working band of frequencies of a pulsed terahertz spectrometer. *Journal of Optical Technology*, 2014, vol. 81, no. 8, pp. 468–471. <https://doi.org/10.1364/JOT.81.000468>
28. Li D., Ma G., Ge J., Hu S., Dai N. Terahertz pulse shaping via birefringence in lithium niobate crystal. *Applied Physics B*, 2009, vol. 94, no. 4, pp. 623–628. <https://doi.org/10.1007/s00340-008-3334-6>
29. Unferdorben M., Szaller Z., Hajdara I., Hebling J., Pálfalvi L. Measurement of refractive index and absorption coefficient of congruent and stoichiometric lithium niobate in the terahertz range. *Journal of Infrared, Millimeter, and Terahertz Waves*, 2015, vol. 36, no. 12, pp. 1203–1209. <https://doi.org/10.1007/s10762-015-0165-5>
30. Mao Z.-L., Hou B.-H., Wang L., Sun Y.-M., Hao W. The study of the terahertz spectral of LiNbO₃ crystal. *Proc. of the Joint 31st International Conference on Infrared Millimeter Waves and 14th International Conference on Terahertz Electronics*, 2006, pp. 465–465. <https://doi.org/10.1109/ICIMW.2006.368673>
31. Samartsev V., Kozlov S. *Basics of Femtosecond Optics*. Moscow, Fizmatlit Publ., 2009, 292 p. (in Russian)

Авторы

Шумигай Владимир Сергеевич — лаборант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, <https://orcid.org/0000-0001-8597-1196>, vshumigay@itmo.ru

Опарин Егор Николаевич — лаборант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, <https://orcid.org/0000-0002-4009-7594>, en_oparin@itmo.ru

Набилкова Александра Олеговна — лаборант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, <https://orcid.org/0000-0001-5699-2211>, aonabilkova@itmo.ru

Мельник Максим Владимирович — кандидат физико-математических наук, научный сотрудник, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, <https://orcid.org/0000-0002-0796-0659>, mmelnik@itmo.ru

Цыпкин Антон Николаевич — доктор физико-математических наук, научный сотрудник, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, <https://orcid.org/0000-0002-9254-1116>, tsyppkinan@itmo.ru

Козлов Сергей Аркадьевич — доктор физико-математических наук, профессор, руководитель международного научного центра оптической и квантовой информатики, биофотоники, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, <https://orcid.org/0000-0002-2692-627X>, sakozlov@itmo.ru

Authors

Vladimir S. Shumigai — Laboratory Assistant, ITMO University, Saint Petersburg, 197101, Russian Federation, <https://orcid.org/0000-0001-8597-1196>, vshumigay@itmo.ru

Egor N. Oparin — Laboratory Assistant, ITMO University, Saint Petersburg, 197101, Russian Federation, <https://orcid.org/0000-0002-4009-7594>, en_oparin@itmo.ru

Aleksandra O. Nabilkova — Laboratory Assistant, ITMO University, Saint Petersburg, 197101, Russian Federation, <https://orcid.org/0000-0001-5699-2211>, aonabilkova@itmo.ru

Maksim V. Melnik — PhD (Physics and Mathematics), Scientific Researcher, ITMO University, Saint Petersburg, 197101, Russian Federation, <https://orcid.org/0000-0002-0796-0659>, mmelnik@itmo.ru

Anton N. Tsyppkin — D. Sc. (Physics and Mathematics), Scientific Researcher, ITMO University, Saint Petersburg, 197101, Russian Federation, <https://orcid.org/0000-0002-9254-1116>, tsyppkinan@itmo.ru

Sergei A. Kozlov — D. Sc. (Physics and Mathematics), Professor, Head of International Research Center of Biophotonics, Optical and Quantum Information, ITMO University, Saint Petersburg, 197101, Russian Federation, <https://orcid.org/0000-0002-2692-627X>, sakozlov@itmo.ru

Статья поступила в редакцию 26.05.2022

Одобрена после рецензирования 06.06.2022

Принята к печати 26.07.2022

Received 26.05.2022

Approved after reviewing 06.06.2022

Accepted 26.07.2022



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

ОПТИЧЕСКИЕ СИСТЕМЫ И ТЕХНОЛОГИИ OPTICAL ENGINEERING

doi: 10.17586/2226-1494-2022-22-4-643-649

УДК 535-45

Исследование влияния наведенных механических напряжений, вызванных герметизацией стеклоприпоем, на значение коэффициента поляризационной экстинкции в двулучепреломляющем оптическом волокне

Евгений Эдуардович Калугин^{1✉}, Азамат Булатович Мухтубаев²,
Игорь Касьянович Мешковский³

^{1,2,3} Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

¹ znabyaklg@gmail.com✉, <https://orcid.org/0000-0002-8474-1791>

² mukhtubaev.ab@gmail.com, <https://orcid.org/0000-0002-4848-0483>

³ igorkm@itmo.ru, <https://orcid.org/0000-0003-3470-1000>

Аннотация

Предмет исследования. В работе рассмотрено влияние наведенных механических напряжений, вызванных герметизацией стеклоприпоем пары оптических двулучепреломляющих волокон с эллиптической напрягающей оболочкой на величину коэффициента поляризационной экстинкции. Предложен вариант с размещением в области герметизации нерабочих волокон для создания симметрии наведенных механических напряжений. Выполнено экспериментальное исследование вклада наведенных механических напряжений на величину коэффициента поляризационной экстинкции. Проведена оценка влияния температуры на коэффициент поляризационной экстинкции в области герметизации оптических двулучепреломляющих волокон. **Метод.** Герметизация волокон в металлической трубке произведена путем плавления преформы стеклоприпоя индукционным нагревателем. Оценка коэффициента поляризационной экстинкции получена методом широкополосной интерферометрии с использованием сканирующего поляризационного интерферометра Майкельсона. **Основные результаты.** Измерена величина коэффициента поляризационной экстинкции в четырех образцах с длиной рабочих волокон 4 м. Эксперимент показал, что создание изометрии структуры в области герметизации с помощью добавления нерабочих волокон позволило уменьшить степень изменения коэффициента поляризационной экстинкции с 0,082 дБ/К до 0,035 дБ/К в температурном диапазоне от минус 15 °С до плюс 70 °С. Полученный результат позволил произвести герметизацию нескольких волокон в одной трубке для уменьшения габаритов устройств. **Практическая значимость.** Выполненное исследование может быть полезно в разработках оптоэлектронных устройств, где требуется герметизация выводов двулучепреломляющего оптического волокна в металлическом корпусе.

Ключевые слова

оптические двулучепреломляющие волокна, коэффициент поляризационной экстинкции, стеклоприпой, связь поляризационных мод, герметизация

Благодарности

Работа выполнена при финансовой поддержке программы «Приоритет 2030».

Ссылка для цитирования: Калугин Е.Э., Мухтубаев А.Б., Мешковский И.К. Исследование влияния наведенных механических напряжений, вызванных герметизацией стеклоприпоем, на значение коэффициента поляризационной экстинкции в двулучепреломляющем оптическом волокне // Научно-технический вестник информационных технологий, механики и оптики. 2022. Т. 22, № 4. С. 643–649. doi: 10.17586/2226-1494-2022-22-4-643-649

Polarization extinction ratio in polarization maintaining fiber sealed with glass solder

Evgeniy E. Kalugin¹✉, Azamat B. Mukhtubaev², Igor K. Meshkovskiy³

^{1,2,3} ITMO University, Saint Petersburg, 197101, Russian Federation

¹ znabyaklg@gmail.com✉, <https://orcid.org/0000-0002-8474-1791>

² mukhtubaev.ab@gmail.com, <https://orcid.org/0000-0002-4848-0483>

³ igorkm@itmo.ru, <https://orcid.org/0000-0003-3470-1000>

Abstract

The paper considers the effect of sealing with glass solder a pair of optical polarization maintaining fibers with an elliptical stress cladding on the value of the polarization extinction ratio. A variant with the placement of non-working fibers in the sealing area to create the symmetry of induced mechanical stresses is proposed. An experimental study of the contribution of the induced mechanical stresses on the value of the polarization extinction ratio has been performed. The assessment of temperature effect on the coefficient of polarization extinction in the place of sealing of optical polarization maintaining fibers has been made. Sealing of fibers in a metal tube was performed using a glass solder preform and an induction heater. Evaluation of the polarization extinction ratio was obtained by white-light interferometry using the scanning Michelson interferometer. The value of the polarization extinction ratio was measured on 4 samples with a length of working fibers of 4 meters. This experiment shows that creation of isometry structure in the area of sealing by addition of non-working fibers allows decreasing the polarization extinction ratio from 0.082 dB/K to 0.035 dB/K in the temperature range from -15°C to $+70^{\circ}\text{C}$. This method allows the sealing of several fibers in one tube to reduce the size of the devices. The performed research can be useful in the development of optoelectronic devices where it is required to introduce optical birefringent fibers in a sealed housing.

Keywords

polarization maintaining (PM) fibers, polarization extinction ratio, glass solder, cross-polarization coupling, sealing

Acknowledgements

The work is financially supported by Priority 2030 program.

For citation: Kalugin E.E., Mukhtubaev A.B., Meshkovskiy I.K. Polarization extinction ratio in polarization maintaining fiber sealed with glass solder. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2022, vol. 22, no. 4, pp. 643–649 (in Russian). doi: 10.17586/2226-1494-2022-22-4-643-649

Введение

Оптические волокна с сохранением поляризации часто используются при разработке волоконно-оптических датчиков. В таких волокнах поляризация сохраняется за счет создания в их структуре асимметрии, что создает анизотропию и формирует двулучепреломление. Такие волокна называют двулучепреломляющими (ДЛП). В основном используется два вида ДЛП волокон: волокна с эллиптической сердцевинной и с наведенным механическим напряжением в структуре. В волокнах первого типа двулучепреломление создается за счет геометрии сердцевин: разное значение эффективных показателей преломлений для каждой из осей эллипса. Второй тип основан на эффекте фотоупругости [1, 2]. В таких ДЛП волокнах постоянные распространения оптического излучения двух ортогональных мод различаются. В ДЛП волокнах выделяют две основные поперечные оси с дифференцированными показателями преломления: одна пересекает область с наведенными напряжениями («медленная ось»), а другая перпендикулярна первой («быстрая ось») [1, 2].

Потребность формирования герметичного ввода оптического волокна в корпус оптоэлектронных устройств продолжает расти в связи с повышающимися требованиями индустрий, использующих оптоэлектронные устройства, к производительности и надежности этих устройств в целом. Ввод оптического излучения в волновод оптоэлектронных устройств, например многофункциональной интегральной оптической схемы (МИОС), требует сверхточного метода позиционирования и крепления оптического волокна

на весь срок службы устройства. Сдвиг или растяжение волокна может привести к потерям мощности оптического излучения в области стыковки волокна с волноводом МИОС. Оптические свойства электрооптических кристаллов, на которых формируют МИОС, меняются в зависимости от прикладываемого электрического напряжения, наличия водяных паров в окружающей среде, изменения температуры и газового состава окружающей атмосферы [3–6]. По этим причинам, требования надежности исключают использование эпоксидной смолы в качестве фиксирующего волокно адгезива из-за ползучести материала, выделения газов и предрасположенности к деградации со временем [7]. Метод с металлизацией голого волокна и последующей впайкой металлическим припоем в металлическую трубку не только дорог, но и образует другие проблемы с надежностью. Из-за различия коэффициента термического расширения кварца и металлов, в используемых припоях при изменении температуры возникают наведенные механические напряжения, которые влияют на оптические свойства оптических волокон [8].

Для таких устройств, как волоконно-оптический гироскоп, используются ДЛП волокна в качестве чувствительного элемента. При возникновении внешних механических воздействий происходит перераспределение напряжений в структуре волокна, что в свою очередь приводит к изменению двулучепреломления и к перекачке оптической мощности из одной поляризационной моды в ортогональную [9, 10]. Напряжения также возникают при изменении температуры, вследствие различия коэффициента термического расширения материалов. Точечные места перекачки оптической мощ-

ности в ДЛП волокнах описываются коэффициентом поляризационной экстинкции (Polarization Extinction Ratio, PER) [1]:

$$PER = \frac{P_{\perp}}{P_{\parallel}},$$

где P_{\perp} — оптическая мощность, перекачавшаяся в ортогональное состояние поляризации; P_{\parallel} — оптическая мощность основного состояния поляризации.

Локальные точки связи поляризационных мод приводят к ошибке сигнала волоконно-оптического гироскопа. Чем больше значение PER, тем больше сдвиг фазы Саньяка [11]. В некоторых волоконно-оптических гироскопах используется МИОС с X-разветвителем, соответственно с каждым из портов МИОС стыкуется ДЛП волокно. Отсюда возникает необходимость формирования гермовывода для пары ДЛП волокон с каждой из сторон, так как делать отдельные выводы часто не позволяют допустимые габариты устройства.

Цель работы — сравнение созданных симметричной и асимметричной структур герметизации оптических ДЛП волокон. Для этого использованы оптические ДЛП волокна с эллиптической напрягающей оболочкой, трубка из ковара, преформа стеклоприпоя в форме кольца и индукционный нагреватель.

Технология герметизации

Для тестирования метода герметизации оптического волокна в пропускной трубке из ковара (сплав 29НК) разработан индукционный нагреватель, разогревающий коваровую трубку, которая плавит преформу стеклоприпоя. Индукционный нагреватель работает в режиме коммутации при нулевом напряжении (Zero Voltage Switch, ZVS).

Мощность нагрева преформы регулируется входным напряжением схемы (максимум 12 В) и обеспечивает нагрев трубки до 400 °С. На рис. 1, а, d представлены фотография индукционного нагревателя в сборе и общая схема герметизации.

Для проведения исследования использованы трубки из ковара и преформы стеклоприпоя в форме кольца (рис. 1, с), подходящие для герметизации 2–4 волокон в металлической трубке. Температура плавления преформ стеклоприпоя 320–350 °С. Материалы данных компонентов выбраны как наиболее близкие по коэффициенту термического расширения к кварцу, из которого изготовлены оптические волокна: коэффициент термического расширения кварца $0,5 \cdot 10^{-6}$ 1/К, ковар $5 \cdot 10^{-6}$ 1/К, стеклоприпой $7,5 \cdot 10^{-6}$ 1/К. Длина коваровой трубки (рис. 1, b) составила 6,5 мм, внешний диаметр — 3 мм, внутренний диаметр 1,5 мм с сужением до 0,8 мм с одной стороны. Сужение предусмотрено для большей плотности заполнения расплавленного стеклоприпоя. Во внутренней части трубки сделан выступ, на который размещена преформа стеклоприпоя, чтобы при плавлении он затекал в узкую часть трубки. Чтобы исключить искажения векторов силовых линий магнитного поля индуктора при изготовлении образцов использована оснастка из керамики и органического стекла.

Трубка закреплена вертикально, волокна зачищены до кварцевого основания на длину около 10 мм и помещены в трубку таким образом, чтобы «голый» участок волокон находился в области герметизации. Преформа стеклоприпоя предварительно расположена во внутреннем выступе трубки.

Для проведения эксперимента были изготовлены четыре образца. В образце № 1 использовано два отрезка оптического ДЛП волокна. В образцах № 2–4 в герметизации участвовало четыре отрезка волокна. Два нерабочих волокна использованы для уплотнения заполнения расплавленного стеклоприпоя и создания изометрии механических напряжений для рабочих волокон. Нерабочие волокна имели длину около 20 см и были зачищены до кварцевой части. Внутри трубки волокна расположены в форме знака «+» таким образом, чтобы рабочие волокна были напротив друг друга (рис. 1, d).

Методика исследования

Для анализа величины коэффициента поляризационной экстинкции в исследуемых образцах использована методика широкополосной интерферометрии. С помощью данной методики можно добиться высокой точности в определении места локальной поляризационно-модовой связи в оптическом тракте, а по амплитуде дать оценку величине PER [12]. В настоящей работе использован сканирующий интерферометр Майкельсона с анализатором на входе для выделения обеих ортогональных поляризационных мод исследуемого образца.

С каждым из образцов выполнены исследования температурных зависимостей коэффициента поляризационной экстинкции в области герметизации оптического ДЛП волокна (рис. 2), запаянного при помощи стеклоприпоя в коваровой трубке. Коваровая трубка каждого из образцов располагалась на элементе Пельтье, а контакт между ними заполнен термопастой. Для контроля и измерения температуры использована термопара, которая также была в контакте с коваровой трубкой через термопасту. Электрическое питание элемента Пельтье контролировалось отдельно стоящим блоком питания с ручным контролем напряжения.

Исследование состояло из трех этапов: нагрев от комнатной температуры до плюс 70 °С, остывание до комнатной температуры и охлаждение до минус 15 °С. Измерения выполнялись каждые 10 °С при нагреве и остывании, и каждые 3 °С при охлаждении. Каждое измерение производилось трижды.

Результаты исследования

В результате проведенного исследования получены значения величины PER в области герметизации оптических ДЛП волокон образцов № 1–4 и зависимости PER в этих образцах от температуры коваровой трубки. Результаты представлены в таблице и на рис. 3.

Из полученных результатов видно, что в точке герметизации оптического ДЛП волокна в коваровой трубке при помощи стеклянного припоя возникает уве-

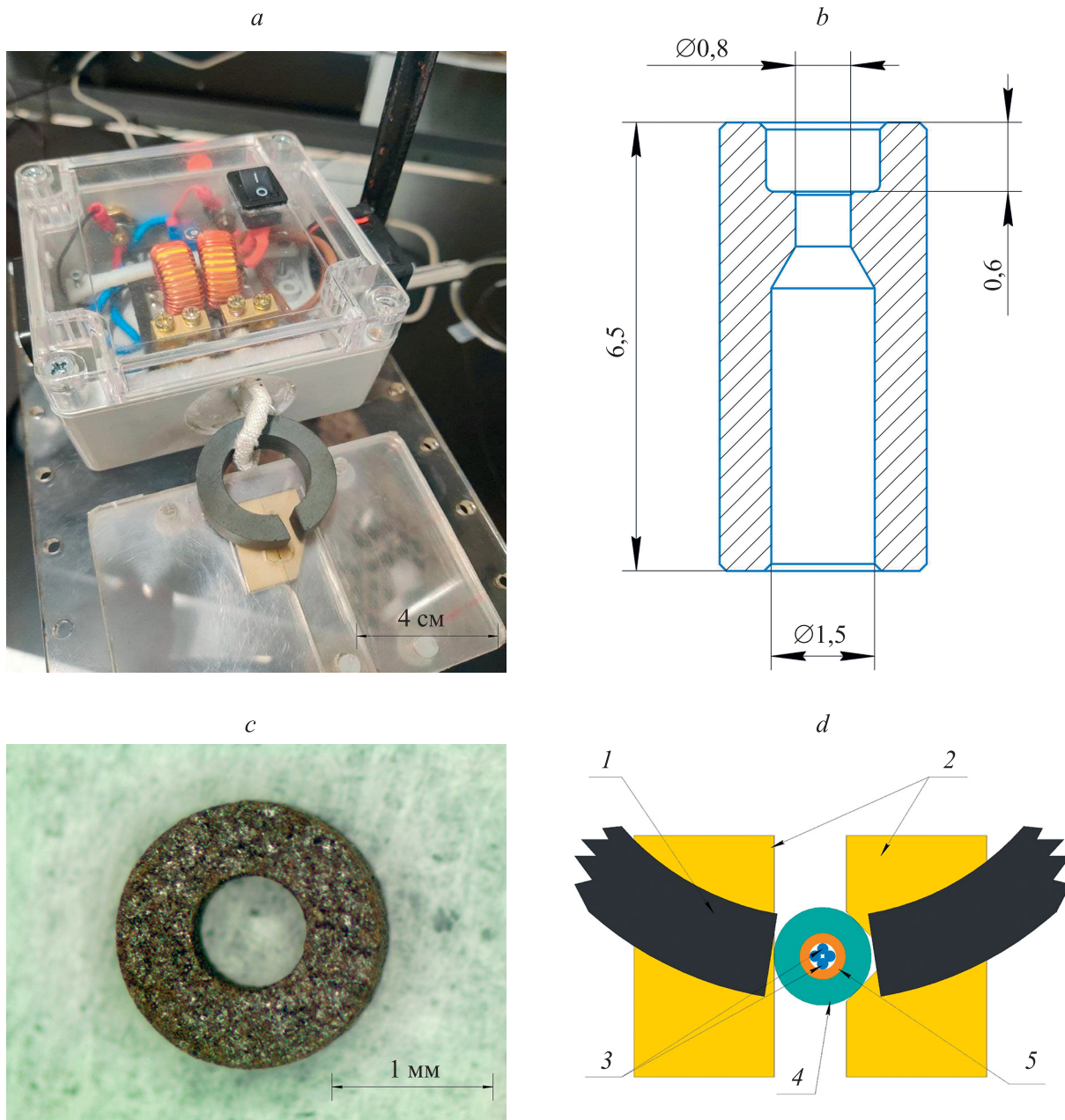


Рис. 1. Индукционный нагреватель в сборе (а); чертеж продольного сечения коваровой трубки (b); преформа стеклоприпоя (c); общая схема герметизации (d).

1 — ферритовое кольцо; 2 — керамические зажимы; 3 — нерабочие оптические волокна; 4 — коваровая трубка; 5 — преформа стеклоприпоя

Fig. 1. Induction heater assembly (a); drawing of a longitudinal section of a kovar tube (b); solder glass preform (c); general sealing scheme (d).

1 — ferrite ring, 2 — ceramic clamps, 3 — “dumb” fibers, 4 — kovar tube, 5 — glass solder preform

личение связи поляризационных мод, так коэффициент поляризационной экстинкции на образце № 1 увеличивается от -65 дБ (уровень чувствительности системы) до -38 дБ, а на образцах № 2–4 в среднем до -30 дБ.

Температурные испытания показали, что при нагреве коваровой трубки в запаянном оптическом волокне коэффициент поляризационной экстинкции понижается с -25 дБ при комнатной температуре до -31 дБ при плюс 70 °C на образце № 4, а при остывании возвращается к прежнему уровню. При охлаждении коэффи-

циент поляризационной экстинкции увеличивается с -39 дБ при комнатной температуре до -32 дБ при минус 10 °C на образце № 1.

Однако значения графиков образца № 4 отличаются от остальных, так как в образцах № 1–3 в одном волокне при нагреве коэффициент поляризационной экстинкции понижается, а в другом увеличивается. Обратная картина наблюдается при охлаждении.

В ходе проведенных экспериментов выявлено, что добавление в области герметизации нерабочих воло-

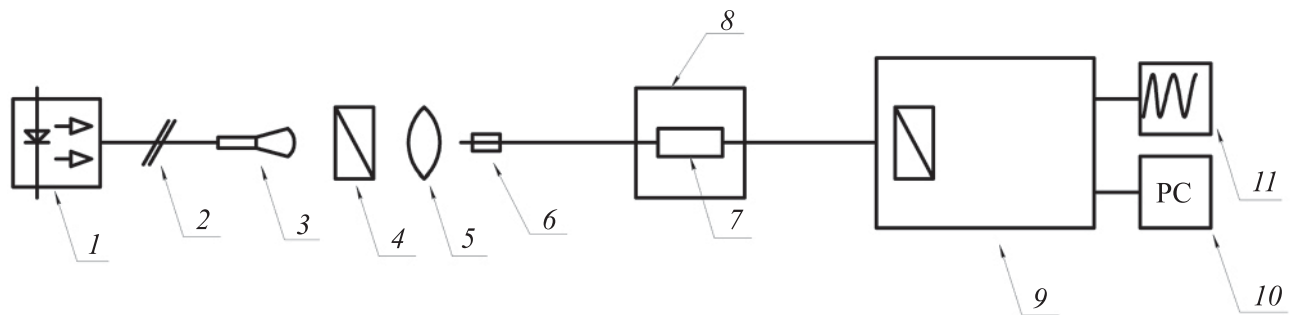


Рис. 2. Схема экспериментальной установки для исследования зависимости коэффициента поляризационной экстинкции полученных образцов от температуры:

1 — источник излучения; 2 — оптическая розетка; 3 — коллиматор; 4 — поляризатор; 5 — объектив; 6 — держатель; 7 — объект исследования; 8 — элемент Пельтье; 9 — поляризационный интерферометр Майкельсона; 10 — персональный компьютер (PC); 11 — осциллограф

Fig. 2. Scheme of the pilot plant for studying the dependence of PER of the obtained samples vs. temperature:

1 — light source, 2 — optical socket, 3 — collimator, 4 — polarizer, 5 — lens, 6 — clamp, 7 — object of study, 8 — thermoelectric cooler, 9 — Michelson polarization interferometer, 10 — PC, 11 — oscilloscope

Таблица. Результаты сканирования образцов № 1–4 после герметизации

Table. Results of scanning samples 1–4

Номер образца	Волокно	Значение коэффициента поляризационной экстинкции PER, дБ
1	1	–27,45
	2	–37,95
2	1	–42,32
	2	–38,60
3	1	–28,06
	2	–19,95
4	1	–25,13
	2	–26,48

кон увеличивает коэффициент поляризационной экстинкции, но при этом уменьшает его зависимость от температуры, так, например, в волокне 2 образца № 1 значения меняются от –39 дБ до –32 дБ, а в волокне 2 образца № 3 – от –21 дБ до –19 дБ.

Как известно из работ [13–15] величина коэффициента поляризационной экстинкции в оптических ДЛП волокнах зависит от величины механического давления и области его воздействия на волокно. Это наглядно подтверждено на рис. 3.

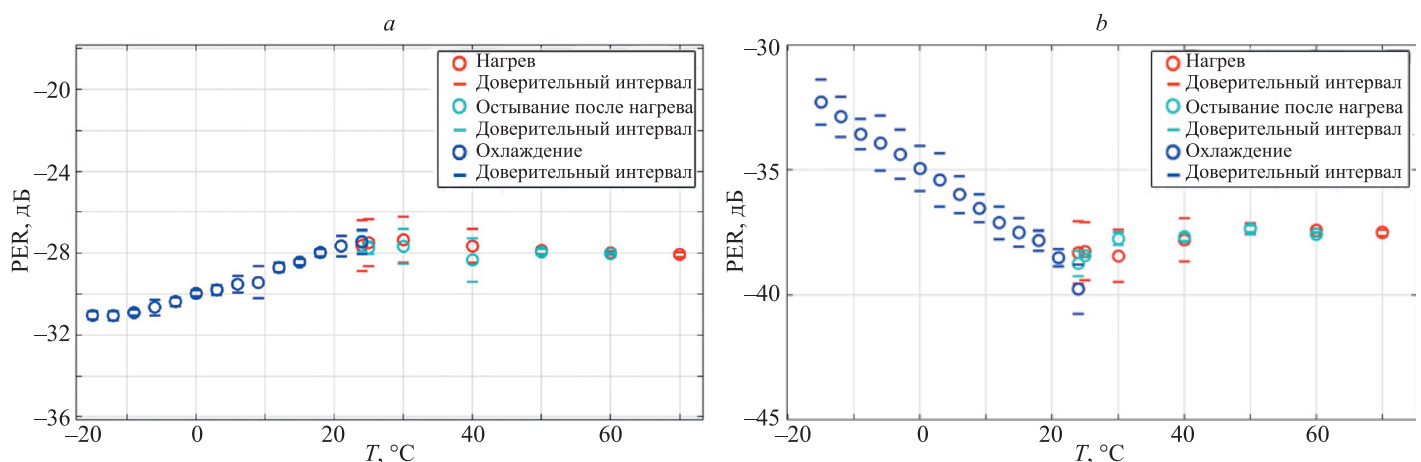


Рис. 3. Зависимость коэффициента поляризационной экстинкции от температуры для: образца № 1 — в волокнах 1 (a) и 2 (b); образца № 2 — в волокнах 1 (c) и 2 (d); образца № 3 — в волокнах 1 (e) и 2 (f); образца № 4 — в волокнах 1 (g) и 2 (h)
Fig. 3. Dependence of PER vs. temperature: fiber 1 sample 1 (a); fiber 2 sample 1 (b); fiber 1 sample 2 (c); fiber 2 sample 2 (d); fiber 1 sample 3 (e); fiber 2 sample 3 (f); fiber 1 sample 4 (g); fiber 2 sample 4 (h)

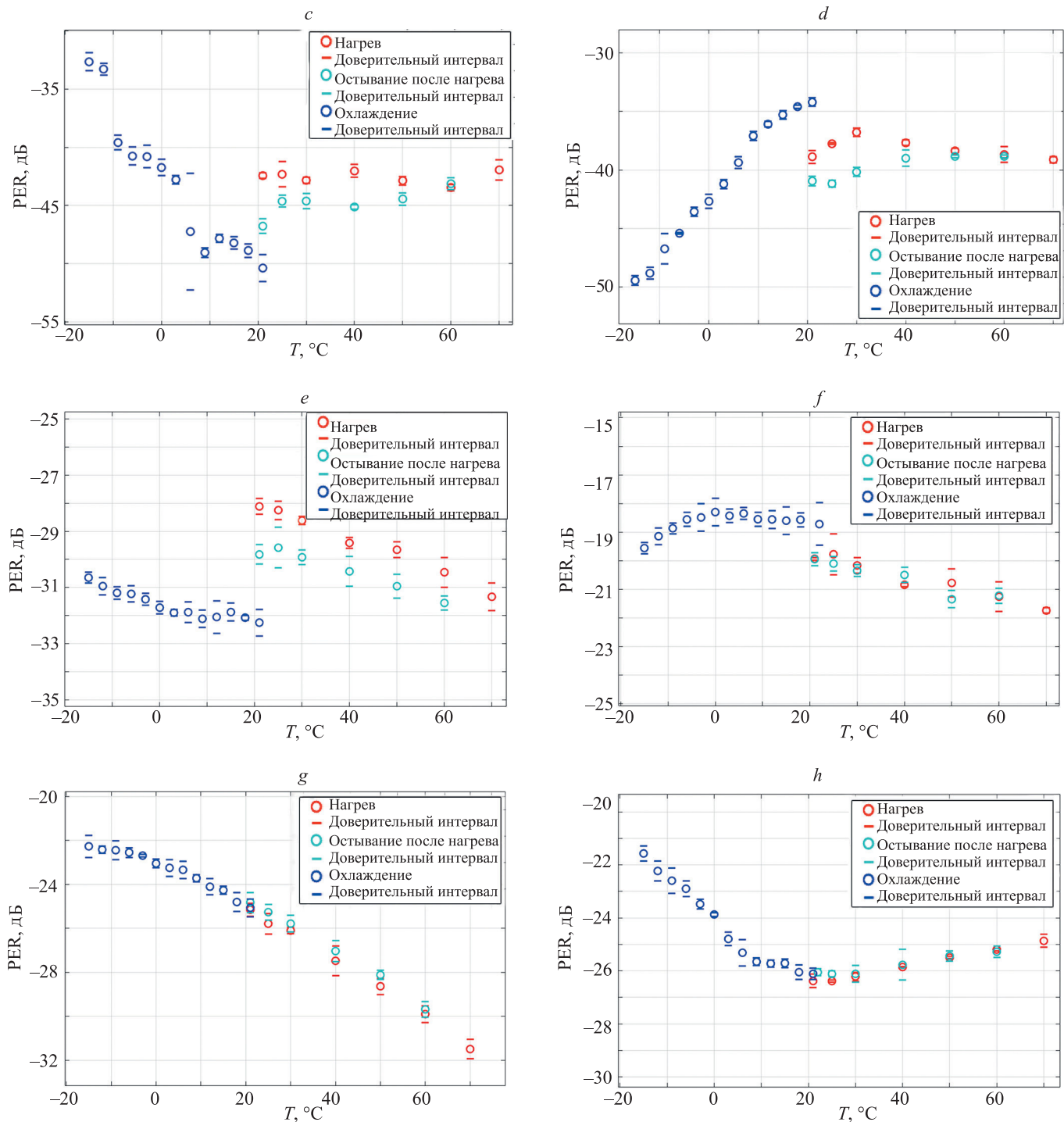


Рис. 3. Продолжение

Fig 3. Continued

Закключение

В работе изучено влияние герметизации стеклоприпоем пары оптических двулучепреломляющих волокон на величину коэффициента поляризационной экстинкции. В волокнах исследована зависимость коэффициента поляризационной экстинкции от напряжений, вызванных изменением температуры. Полученные результаты подтвердили, что механическое давление с разной величиной и в различных областях двулучепре-

ломляющего волокна по-разному влияет на изменение величины коэффициента поляризационной экстинкции. Эксперимент показал, что создание изометрии структуры в области герметизации с помощью добавления нерабочих волокон позволяет уменьшить степень изменения коэффициента поляризационной экстинкции с 0,082 дБ/К до 0,035 дБ/К в температурном диапазоне от минус 15 °С до плюс 70 °С. Данный способ позволяет производить герметизацию нескольких волокон в одной трубке для уменьшения габаритов устройств.

Литература

1. Lefèvre H. The Fiber-Optic Gyroscopes / 2nd ed. Artech House, 2014. P. 317–331.
2. Kaminow I.P. Polarization-maintaining fibers // *Applied Scientific Research*. 1984. V. 41. N 3–4. P. 257–270. <https://doi.org/10.1007/BF00382456>
3. Moody J., Morris R. Hermetically packaged LiNbO₃ optical circuit with oxidizing fill gas. Patent US8070368B1. 2010.
4. Iwamoto N.E., Sanders S.J., Gallagher T.J., Yates S., Taranta A., Grooms J.C., Warrensford K.L. Stabilizing gas in an optical gyro based on an inorganic waveguide. Patent EP3270107B1. 2018.
5. Mottet A., Jillard S., Hauden J., Grossard N., Porte H., Tchahame J., Veyrie D., Gilard O. Packaging Improvement of LiNbO₃ modulators and Space evaluation results // *Proceedings of SPIE*. 2018. V. 11180. P. 111803K. <https://doi.org/10.1117/12.2536047>
6. Donaldson A., Freer S.J. Lithium niobate devices. Patent GB2306690A. 1997.
7. Dietz R.L. Sealing optical fibers without metallization: design guidelines // *Proceedings of SPIE*. 2004. V. 5454. P. 111–120. <https://doi.org/10.1117/12.554264>
8. Shaw M., Galeotti R., Coppo G. Method of fixing an optical fibre in a laser package // *Proc. of the Electronic Components and Technology Conference*. 2001. P. 1441–1446. <https://doi.org/10.1109/ECTC.2001.928025>
9. Ликумович Л.Б. Поляризационные эффекты в волоконных интерферометрах на основе двулучепреломляющих световодов: диссертация на соискание ученой степени доктора физико-математических наук. СПб., 2008. 333 с.
10. Marcuse D. Coupled-mode theory for anisotropic optical waveguides // *Bell System Technical Journal*. 1975. V. 54. N 6. P. 985–995. <https://doi.org/10.1002/j.1538-7305.1975.tb02878.x>
11. Мухтубаев А.Б. Влияние поляризационных преобразований на сдвиг фазы Саньяка в волоконно-оптическом гироскопе: диссертация на соискание ученой степени кандидата технических наук. СПб., 2020. 199 с.
12. Аксарин С.М. Исследование поляризационных методов и технологий согласования волоконно-оптических и интегрально-оптических волноводов: диссертация на соискание ученой степени кандидата технических наук. СПб., 2014. 116 с.
13. Mukhtubayev A.B., Aksarin S.M., Strigalev V.E. Polarization extinction ratio of the polarization crosstalk caused by point pressure force in the polarization-maintaining fiber // *Optical Fiber Technology*. 2017. V. 38. P. 119–122. <https://doi.org/10.1016/j.yofte.2017.09.010>
14. Namihira Y., Kudo M., Mushiaka Y. Effect of mechanical stress on the transmission characteristics of optical fibers // *Trans. Inst. Electron. Commun. Eng. Japan*. 1977. V. 60. N 7. P. 107–115.
15. Suhir E. Mechanical behavior of optical fibers and interconnects: application of analytical modelling // *Encyclopedia of Continuum Mechanics*. Springer, 2020. P. 1528–1532. https://doi.org/10.1007/978-3-662-55771-6_371

Авторы

Калугин Евгений Эдуардович — инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, <https://orcid.org/0000-0002-8474-1791>, znabyaklg@gmail.com
Мухтубаев Азамат Булатович — кандидат технических наук, заведующий лабораторией, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, <https://orcid.org/0000-0002-4848-0483>, mukhtubayev.ab@gmail.com
Мешковский Игорь Касьянович — доктор технических наук, профессор, директор НИЦ СФ, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, <https://orcid.org/0000-0003-3470-1000>, igorkm@itmo.ru

Статья поступила в редакцию 11.04.2022
 Одобрена после рецензирования 20.05.2022
 Принята к печати 12.07.2022

References

1. Lefèvre H. *The Fiber-Optic Gyroscopes*. 2nd ed. Artech House, 2014. pp. 317–331.
2. Kaminow I.P. Polarization-maintaining fibers. *Applied Scientific Research*, 1984, vol. 41, no. 3–4, pp. 257–270. <https://doi.org/10.1007/BF00382456>
3. Moody J., Morris R. Hermetically packaged LiNbO₃ optical circuit with oxidizing fill gas. *Patent US8070368B1*, 2010.
4. Iwamoto N.E., Sanders S.J., Gallagher T.J., Yates S., Taranta A., Grooms J.C., Warrensford K.L. Stabilizing gas in an optical gyro based on an inorganic waveguide. *Patent EP3270107B1*, 2018.
5. Mottet A., Jillard S., Hauden J., Grossard N., Porte H., Tchahame J., Veyrie D., Gilard O. Packaging Improvement of LiNbO₃ modulators and Space evaluation results. *Proceedings of SPIE*, 2018, vol. 11180, pp. 111803K. <https://doi.org/10.1117/12.2536047>
6. Donaldson A., Freer S.J. Lithium niobate devices. *Patent GB2306690A*, 1997.
7. Dietz R.L. Sealing optical fibers without metallization: design guidelines. *Proceedings of SPIE*, 2004, vol. 5454, pp. 111–120. <https://doi.org/10.1117/12.554264>
8. Shaw M., Galeotti R., Coppo G. Method of fixing an optical fibre in a laser package. *Proc. of the 51st Electronic Components and Technology Conference*, 2001, pp. 1441–1446. <https://doi.org/10.1109/ECTC.2001.928025>
9. Liokumovich L.B. *Polarization Effects in Fiber Interferometers Based on Birefringent Optical Fiber Waveguides*. Dissertation for the degree of doctor of physical and mathematical sciences. St. Petersburg, 2008, 333 p. (in Russian)
10. Marcuse D. Coupled-mode theory for anisotropic optical waveguides. *Bell System Technical Journal*, 1975, vol. 54, no. 6, pp. 985–995. <https://doi.org/10.1002/j.1538-7305.1975.tb02878.x>
11. Mukhtubayev A.B. *The polarization cross-coupling influence on the Sagnac phase shift in a fiber-optic gyroscope*. Academic dissertation candidate of engineering. St. Petersburg, 2020, 199 p. (in Russian)
12. Aksarin S.M. *Investigation of polarization Methods and Technologies for Matching Fiber-Optic and Integrated-Optical Waveguides*. Dissertation for the degree of candidate of technical sciences. St. Petersburg, 2014, 116 p. (in Russian)
13. Mukhtubayev A.B., Aksarin S.M., Strigalev V.E. Polarization extinction ratio of the polarization crosstalk caused by point pressure force in the polarization-maintaining fiber. *Optical Fiber Technology*, 2017, vol. 38, pp. 119–122. <https://doi.org/10.1016/j.yofte.2017.09.010>
14. Namihira Y., Kudo M., Mushiaka Y. Effect of mechanical stress on the transmission characteristics of optical fibers. *Trans. Inst. Electron. Commun. Eng. Japan*, 1977, vol. 60, no. 7, pp. 107–115.
15. Suhir E. Mechanical behavior of optical fibers and interconnects: application of analytical modelling. *Encyclopedia of Continuum Mechanics*. Springer, 2020, pp. 1528–1532. https://doi.org/10.1007/978-3-662-55771-6_371

Authors

Evgeniy E. Kalugin — Engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, <https://orcid.org/0000-0002-8474-1791>, znabyaklg@gmail.com
Azamat B. Mukhtubayev — PhD, Head of Laboratory, ITMO University, Saint Petersburg, 197101, Russian Federation, <https://orcid.org/0000-0002-4848-0483>, mukhtubayev.ab@gmail.com
Igor K. Meshkovskiy — D. Sc., Professor, Director of the Research Center of Light-Guided Photonics, ITMO University, Saint Petersburg, 197101, Russian Federation, <https://orcid.org/0000-0003-3470-1000>, igorkm@itmo.ru

Received 11.04.2022
 Approved after reviewing 20.05.2022
 Accepted 12.07.2022



Работа доступна по лицензии
 Creative Commons
 «Attribution-NonCommercial»

doi: 10.17586/2226-1494-2022-22-4-650-658

УДК 543.426

Метод дистанционного контроля радиационных параметров космических аппаратов на основе рентгенофлуоресцентного анализа

Людмила Анатольевна Лукьянова¹, Игорь Владимирович Свитнев²✉,
Елена Александровна Харитоновна³, Илья Евгеньевич Гаврилов⁴

^{1,2,3,4} Санкт-Петербургский государственный университет, Санкт-Петербург, 199034, Российская Федерация

¹ lalukianova@mail.ru, <https://orcid.org/0000-0001-5458-3019>

² isvitnev@mail.ru✉, <https://orcid.org/0000-0002-8862-1797>

³ Haritonova_ea@mail.ru, <https://orcid.org/0000-0002-4851-2756>

⁴ Ilyusha-2145@mail.ru, <https://orcid.org/0000-0003-1124-0336>

Аннотация

Предмет исследования. Существующие международные юридические акты, инструменты и процедуры не гарантируют обеспечения паритетных условий для освоения и использования космического пространства. Необходимы средства объективного контроля космических аппаратов, на борту которых могут находиться изделия с делящимися материалами. Инспекция таких объектов может проводиться методами рентгенофлуоресценции. Однако в рассматриваемой предметной области применение подобных методов мало изучено. В данной работе предложен метод получения спектров рентгенофлуоресцентного излучения материалов, исследуемого объекта на основе расчета пространственно-энергетических характеристик рентгеновского излучения. **Метод.** Спектры рентгенофлуоресцентного исследования объектов получены на основе расчета пространственно-энергетических характеристик рентгеновского излучения по разработанным авторами оригинальным геометрической (математической) и имитационным моделям. При расчетах учтена сложная слоистая система объекта с учетом долей высокоэнергетического флуоресцентного излучения вышележащих слоев. Предложен численный эксперимент с использованием программы, которая позволяет выбрать проекции объекта, подвергаемого рентгеновскому облучению, длину волны и интенсивность параметров излучателя. С помощью полученного пространственно-энергетического распределения квантов излучения и физических свойств среды прохождения излучения решена задача поиска координат и углов пересечения треков пучков квантов в каждой области объекта. Результат программной обработки отображен в виде результирующего спектра. Полученный спектр позволил сделать вывод о химическом составе материалов инспектируемого объекта. Пучок рентгеновского излучения, достигающий объекта, смоделирован в виде пятна площадью, соизмеримой с сечением аппарата в виде выбранного геометрического примитива — квадрата. Площадь пятна падающих фотонов рассчитана по заранее задаваемому углу расхождения. **Основные результаты.** На основе открытых научных работ выбрана физическая модель объекта с ядерными делящимися материалами W88 (США). Приняты следующие характеристики рентгеновских пучков (подпучков): когерентный пучок фотонов с длиной волны 0,005 нм; угол рассеяния пучка 1°, число квантов от $1 \cdot 10^{15}$ (для дистанции облучения 10 м) до $1 \cdot 10^{21}$ (для дистанции 10 км); площадь детектора излучателя 4 м². Получен результирующий спектр рентгеновской флуоресценции, представляющий химический состав узлов и блоков аппарата и объекта, находящегося внутри. Наличие на борту объекта специфических изделий подтверждается характеристическими линиями с нормированными длинами волн, которые демонстрируют присутствие химических элементов радиоактивного ряда. **Практическая значимость.** Полученные результаты могут найти применение при разработке оборудования и программного обеспечения устройств космических аппаратов, при контроле наличия распадающихся материалов на борту инспектируемого аппарата.

Ключевые слова

делящиеся материалы, длина волны, рентгеновское излучение, флуоресценция, элементный состав, энергия кванта

Ссылка для цитирования: Лукьянова Л.А., Свитнев И.В., Харитоновна Е.А., Гаврилов И.Е. Метод дистанционного контроля радиационных параметров космических аппаратов на основе рентгенофлуоресцентного анализа // Научно-технический вестник информационных технологий, механики и оптики. 2022. Т. 22, № 4. С. 650–658. doi: 10.17586/2226-1494-2022-22-4-650-658

© Лукьянова Л.А., Свитнев И.В., Харитоновна Е.А., Гаврилов И.Е., 2022

Method for remote control of radiation parameters of spacecraft based on X-ray fluorescence analysis

Lyudmila A. Lukyanova¹, Igor V. Svitnev²✉, Elena A. Kharitonova³, Ilya E. Gavrilov⁴

^{1,2,3,4} St. Petersburg State University (SPbSU), Saint Petersburg, 199034, Russian Federation

¹ lalukyanova@mail.ru, <https://orcid.org/0000-0001-5458-3019>

² isvitnev@mail.ru✉, <https://orcid.org/0000-0002-8862-1797>

³ Xaritonova_ea@mail.ru, <https://orcid.org/0000-0002-4851-2756>

⁴ Ilyusha-2145@mail.ru, <https://orcid.org/0000-0003-1124-0336>

Abstract

Existing international legal acts, instruments and procedures do not guarantee equal conditions for the exploration and use of outer space. There is a need for means of objective control of spacecraft carrying products with fissile materials. Inspection of such objects can be carried out by X-ray fluorescence methods. However, in the subject area under consideration, the use of such methods has been little studied. In this paper, a method is proposed for obtaining the spectra of X-ray fluorescence radiation of materials, the object under study based on the calculation of the spatial and energy characteristics of X-ray radiation. The X-ray fluorescence spectra of objects are obtained on the basis of the calculation of the spatial and energy characteristics of X-ray radiation according to the original geometric (mathematical) and simulation models developed by the authors. The calculations take into account the complex layered system of the object, taking into account the proportions of high-energy fluorescent radiation of the overlying layers. An original numerical experiment is proposed using a program that allows one to choose the projections of an object subjected to X-ray irradiation, the wavelength and intensity of the emitter parameters. Using the obtained spatial-energy distribution of radiation quanta and the physical properties of the radiation transmission medium, the problem of finding the coordinates and angles of intersection of tracks of quanta beams in each area of the object is solved. The result of software processing is displayed as the resulting spectrum. The obtained spectrum makes it possible to draw a conclusion about the chemical composition of the materials of the inspected object. The X-ray beam reaching the object is modeled as a spot with an area commensurate with the cross section of the device in the form of a selected geometric primitive — a square. The spot area of the incident photons is calculated from a predetermined divergence angle. On the basis of open literary sources, a physical model of an object with nuclear fissile materials W88 (USA) was chosen. The following characteristics of X-ray beams (subbeams) are accepted: a coherent beam of photons with a wavelength of 0.005 nm; beam scattering angle of 1 degree; emitter detector area of 4 m². The resulting X-ray fluorescence spectrum was obtained which gives an idea of the chemical composition of the units and blocks of the apparatus and the object inside. The presence of specific products on the object board is confirmed by characteristic lines with normalized wavelengths indicating the presence of chemical elements belonging to the radioactive series. The results obtained can be used in the development of hardware and software for spacecraft devices that monitor the presence of fissile materials on board of the inspected vehicle.

Keywords

fissile materials, wavelength, X-rays, fluorescence, elemental composition, quantum energy

For citation: Lukyanova L.A., Svitnev I.V., Kharitonova E.A., Gavrilov I.E. Method for remote control of radiation parameters of spacecraft based on X-ray fluorescence analysis. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2022, vol. 22, no. 4, pp. 650–658 (in Russian). doi: 10.17586/2226-1494-2022-22-4-650-658

Введение

Современная международная деятельность в области исследования и освоения космического пространства характеризуется: расширением перечня государств, реализующих практические программы; прогрессирующим объемом межгосударственного сотрудничества; увеличением масштабов действий, предпринимаемых, в том числе неправительственными юридическими лицами и партнерскими компаниями; постоянным появлением новых технологий; возрастанием числа запусков космических аппаратов различного назначения. В этих условиях сложно обнаружить нарушение паритетных отношений при использовании космического пространства.

В соответствии с принципами, государства-участники Договора¹ обязуются не выводить на орбиту вокруг Земли какие бы то ни было объекты с ядерным ору-

жием или любыми другими видами оружия массового уничтожения, не устанавливать такое оружие на небесных телах и не размещать такое оружие в космическом пространстве каким-либо иным образом. Кроме того, требуется сводить к минимуму количество радиоактивного материала в космосе и связанные с этим риски использования ядерных источников энергии в космическом пространстве².

В интересах контроля выполнения взятых обязательств государствами-участниками действует система регистрации объектов, запускаемых в космическое пространство³, согласно которой осуществляется национальная регистрация запускающими государствами космических объектов, включающая, в том числе сведения

¹ Договор о принципах деятельности государств по исследованию и использованию космического пространства, включая Луну и другие небесные. Принят резолюцией 2222 (XXI) Генеральной Ассамблеи ООН от 19.12.1966 г. Статьи IV и V.

² Принципы, касающиеся использования ядерных источников энергии в космическом пространстве. Приняты Генеральной Ассамблеей ООН в резолюции 47/68 от 14.12.1992 г. Часть 2, Глава В.

³ Конвенция о регистрации объектов, запускаемых в космическое пространство. Принята 12.11.1974 резолюцией 3235 (XXIX) на 2280-ом пленарном заседании 29-ой сессии Генеральной Ассамблеи ООН. Часть 1, глава С, Статьи I–IV.

об общем назначении космического аппарата. Вместе с тем, в рамках принятого соглашения, предусматривается возможность обеспечения государств-участников дополнительными средствами и процедурами, способствующими идентификации космических объектов. Исходя из этого, одной из мер проверки соответствия заявленным задачам запущенных космических аппаратов может рассматриваться их инспекция.

Идея осмотра чужих спутников появилась еще в семидесятых годах, и в настоящее время ведущие космические державы разрабатывают, испытывают и вводят в строй свои орбитальные средства для этой цели^{1,2,3,4}.

Спутник-инспектор — аппарат, который может приблизиться к другому спутнику и осмотреть его. Основная задача спутников-инспекторов заключается в сборе данных о других объектах, находящихся на орбите. Относительно небольшие размеры аппарата обеспечивают его малозаметность.

Реализованные на сегодняшний день методы инспекции с использованием оптико-электронной аппаратуры [1] не могут дать полного понимания о фактическом назначении аппарата.

Постановка задачи

Наиболее близкий по физическому содержанию метод инспекции назначения космического аппарата гамма-спектрометрия [2]. Сущность данного метода состоит в регистрации гамма-квантов, испускаемых радиоактивными изотопами, накопленными за время работы космических ядерных энергетических установок. Соответственно, гамма-спектрометрия не дает представление о делящихся материалах, находящихся в подкритическом состоянии.

В качестве одного из возможных подходов к решению рассматриваемой проблемы предложен метод обнаружения изделий, размещенных на борту космического аппарата и содержащих радиоактивные элементы, основанный на рентгенофлуоресцентном анализе (рис. 1).

Рентгенофлуоресцентный анализ — современный физический метод элементного анализа, т. е. качественного и количественного определения содержания химических элементов в конструкционных узлах и блоках исследуемого объекта.

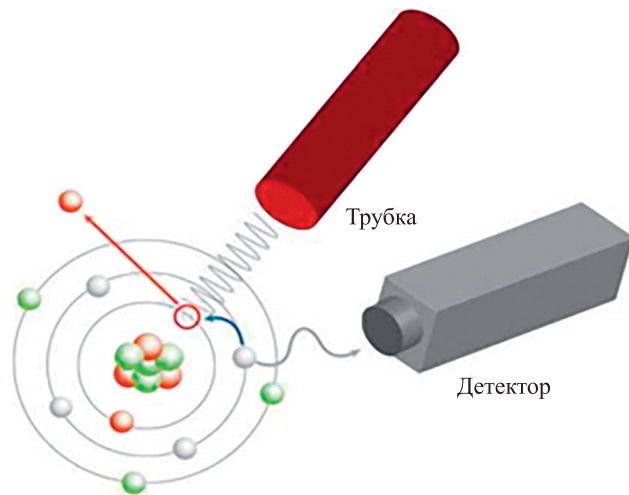


Рис. 1. Принципиальная схема применения метода рентгенофлуоресцентного анализа

Fig. 1. Schematic diagram of the application of the X-ray fluorescence method

Традиционно рентгенофлуоресцентный метод используется при определении главной компоненты в ходе анализа материалов в металлургической, строительной, топливной промышленности, а также в геологии. В последнее время наблюдается расширение сферы практического приложения метода в области анализа объектов окружающей среды, в медицине и научно-исследовательских целях.

Рентгеновские кванты, испускаемые источником рентгеновского излучения, проходят сквозь исследуемый материал, вызывая в нем рентгеновскую флуоресценцию. Кванты, испущенные материалом, покидают его, некоторая их часть попадает в детектор и подвергается анализу.

При прохождении пучка рентгеновского излучения через слой вещества интенсивность пучка ослабевает в направлении его распространения вследствие процессов: поглощения и рассеяния. В случае поглощения фотона первичного излучения из атома выбрасывается фотоэлектрон и образуется вакансия в одной из внутренних оболочек. Уменьшение энергии атома путем заполнения этой вакансии более удаленным от ядра электроном возможно переходами двух типов: радиационным (с испусканием фотона характеристического излучения) и безрадиационным (с выбрасыванием из атома еще одного электрона). В первом случае атом флуоресцирует, во втором случае — нет [3].

В зависимости от относительной вероятности переходов этих двух типов доля случаев, в которых испускаются характеристические фотоны, может быть больше или меньше.

В соответствии с положением в периодической системе, каждый химический элемент [4] обладает своей конфигурацией заполняемых электронных уровней, следствием чего является уникальный для каждого элемента набор переходов электронов с одного уровня на другой. Энергия кванта рентгеновской флуоресценции элемента зависит от разницы энергий уровней, между которыми происходит переход. Длина волны излуче-

¹ Спутник «Космос-2519». Инспектор на орбите. Военное обозрение [Электронный ресурс]. URL: <https://topwar.ru/128281-sputnik-kosmos-2519-inspektor-na-orbite.html> (дата обращения: 19.02.2022).

² США увеличивают группировку спутников-инспекторов GSSAP. Военное обозрение [Электронный ресурс]. URL: <https://topwar.ru/191611-ssha-uvlichivajut-gruppirovku-sputnikov-inspektorov-gssap.html> (дата обращения: 19.02.2022).

³ Инспектор и буксир. Экспериментальный космический аппарат «Шицзянь-21» (Китай). Военное обозрение: сайт. 2017 [Электронный ресурс]. URL: <https://topwar.ru/191898-inspektor-i-buksir-jeksperimentalnyj-kosmicheskij-apparat-shiczyan-21-kitaj.html> (дата обращения: 18.02.2022).

⁴ GSSAP (Geosynchronous Space Situational Awareness Program) [Электронный ресурс]. URL: <https://directory.eoportal.org/web/eoportal/satellite-missions/g/gssap> (дата обращения: 20.02.2022).

ния прямо пропорциональна энергии кванта. Таким образом, существует однозначная взаимосвязь между длинами волн излучения и порядковыми номерами элементов, это излучение испустивших.

Использование рентгенофлуоресцентного анализа в интересах инспекции космического аппарата предоставляет возможность быстрого получения информации о качественном составе исследуемого объекта по отдельным химическим элементам от бериллия (Be) до урана (U), а при применении аппаратов-эталонов и проведении серии измерений — приблизительноном количественном содержании простых веществ [5]. Это позволяет делать аргументированные предположения об устройстве и назначении аппарата и находящихся на нем объектов. При этом необходимо учитывать ряд существенных факторов, влияющих на эффективность применения предлагаемого метода. К ним относятся:

- большая энергозатратность;
- высокие требования к углу расхождения пучка квантов, а также к точности наведения излучателя на исследуемый аппарат;
- вероятность случайного повреждения радиоэлектронного оборудования космического аппарата, оказавшихся на линии пролета пучка частиц, либо вывод из строя исследуемого объекта при избыточной интенсивности излучения на небольшой дистанции.

Проверка целесообразности и безопасности применения метода рентгенофлуоресцентного анализа в задачах инспекции орбитальных средств осуществляется путем компьютерного моделирования процесса прохождения рентгеновского излучения через космический аппарат с последующим заключением о наличии в составе конструктивных блоков специфических химических элементов.

Моделирование прохождения пучка рентгеновского излучения через космический аппарат

Имитационная модель процесса прохождения квантов рентгеновского излучения через вещество базируется на решении дифференциальных уравнений, описывающих расчет интенсивности флуоресценции, и задачи вычислительной геометрии прохождения рентгеновского излучения через трехмерные объекты произвольной формы, и реализована в программной среде Microsoft Visual Studio Community 2017.

Исходные данные и допущения.

1. Исследуемый объект представлен в виде графической модели, оригинальной геометрической модели прохождения квантов рентгеновского излучения через среду и расчета значений альbedo, а также физической модели в виде набора уравнений, задающих определенные области пространства, соответствующие тем или иным функциональным элементам и узлам космического аппарата.
2. Вследствие больших расстояний до inspectируемого аппарата можно допустить, что лучи рентгеновского излучения движутся перпендикулярно моделируемому пятну на аппарате.

3. На площади пятна выбирается область, покрывающая габариты модели. Область разделяется на подобласти, задаваемые программно.
4. Траекторией всего излучения подпучка квантов, проходящего через подобласти объекта, является линия, перпендикулярная пятну.
5. Поскольку форма подобластей объекта может быть произвольной, то для простоты расчетов и организации памяти программы она выбрана квадратной. Форма подобластей не оказывает значительного влияния на точность вычислений.
6. Весь падающий поток квантов рентгеновского излучения разбивается на подпучки меньшего диаметра в форме цилиндров, оси симметрии которых принимаются как пути прохождения всех квантов рентгеновского излучения в этих пучках.
7. Частота флуоресценции для разных химических элементов задается коэффициентами выхода флуоресценции уровня в виде вероятности возникновения «скачка» поглощения квантов рентгеновского излучения — $P_{пр}$. Вероятность $P_{пр}$ тем меньше, чем легче элемент и выше энергетический уровень.
8. Химические элементы до алюминия, не включительно, не способны создавать рентгенофлуоресцентное излучение достаточной интенсивности; их наличием при создании модели можно пренебречь, так как их вклад в результирующий спектр незначителен.
9. Значения геометрического параметра модели прохождения рентгеновского излучения через среду β рассчитано авторами на основе геометрической модели для расчета альbedo.

Вербальное описание модели. С помощью разработанной авторским коллективом программы исследуемый объект воспроизводится в виде демонстрационной графической модели, а также физической модели в форме набора уравнений, задающих определенные подобласти пространства, соответствующие тем или иным функциональным элементам и узлам космического аппарата. К каждой из подобластей привязан набор физических коэффициентов, характеризующих проницаемость квантов первичного излучения и рентгеновскую флуоресценцию в зависимости от химического состава вещества, заполняющего подобласть [6].

Программа предлагает выбрать проекцию космического аппарата, подвергаемого облучению, длину волны и интенсивность излучения. Проекция космического аппарата разбивается на подобласти, число которых зависит от требуемой точности и скорости вычислений. Центр каждой подобласти считается точкой попадания всех квантов излучения в космический аппарат. Рассчитывается длина пути квантов в заданных математически подобластях пространства, занимаемых теми или иными узлами, и углы падения квантов рентгеновского излучения при переходе из одной среды в другую. С использованием данных о пробегах квантов и физических свойствах пройденных ими сред решаются задачи вычислительной геометрии — поиск координат и углов пересечений прямых линий и трехмерных фигур произвольной формы. На конечном этапе вычисляется рентгеновская флуоресценция каждой подобласти.

Полученные значения сводятся вместе и отображаются в виде результирующего спектра.

Графическая интерпретация моделируемых процессов представлена на рис. 2.

Математическое представление модели. Подобласти космического аппарата характеризуются вероятностью P проникновения квантов рентгеновского излучения без взаимодействия на глубину ξ , рассеяния в обратном направлении и выходе за пределы отражателя в направлении детектора. Ее значение может быть получено из выражения:

$$\frac{dP}{d\Omega d\xi} = \rho_i \frac{N_A}{A_i} \cdot \sec\Theta_0 \exp\left[\frac{\bar{\mu}_0(E_0)\xi}{\cos\Theta_0}\right] \cdot \left(\frac{E_1^s}{E_0}\right) \times \\ \times \left\{ \frac{d\sigma_k}{d\Omega}(E_0, \Theta_s) \cdot Z \cdot \exp\left[\frac{\bar{\mu}_1(E_1^s)\xi}{\cos\Theta_1}\right] + \frac{d\sigma_R}{d\Omega}(E_0, \Theta_s) \cdot \exp\left[\frac{\bar{\mu}_0(E_0)\xi}{\cos\Theta_0}\right] \right\},$$

где ξ — глубина рассеивающего материала, г·см⁻²; ρ_i — плотность i -го материала, г/см³; N_A — число Авогадро, 6,022 140 76·10²³ моль⁻¹; A_i — атомная масса i -го элемента, а.е.м.; $\Theta_0, \Theta_1, \Theta_s$ — углы падения, рассеяния и выхода квантов, град.; ($\bar{\mu}_0$ и $\bar{\mu}_1$ — линейные коэффициенты ослабления рентгеновского излучения оболочкой космического аппарата и материалами подобласти космического аппарата, см⁻¹ (E_0 и E_1^s — энергия рент-

геновского излучения падающего потока и отраженная в точке, кэВ; ($\frac{d\sigma_k}{d\Omega}$ и $\frac{d\sigma_R}{d\Omega}$ — дифференциальные сечения комптоновского и когерентного рассеяний, см²); Z — атомный номер; Ω — геометрическая постоянная сечения взаимодействия.

Подобласти разбиваются на геометрические примитивы, принимая во внимание возможность пересечения отдельных объектов космического аппарата и путей квантов в подобластях. Проводится перебор слоев материалов космического аппарата при условии, что сведения по обратному рассеянию излучения от тонких рассеивателей с $Z \leq 26$ и $E_0 = 50\text{--}300$ кэВ в справочном материале не приводятся.

Поскольку для исследуемого диапазона энергии (0,0001–14,4 нм) толщины оболочки космического аппарата составляют величину меньше или порядка двух длин свободного пробега, то необходимо получить выражения для расчета дифференциального токового энергетического (a_E) и дифференциального числового (a_N) альбеда. Для получения математических зависимостей a_E и a_N необходимо знать значение энергии падающего потока рентгеновского излучения на элемент подобласти.

Рентгеновское излучение с энергией E_0 падает на элемент поверхности, с учетом предлагаемой ориги-

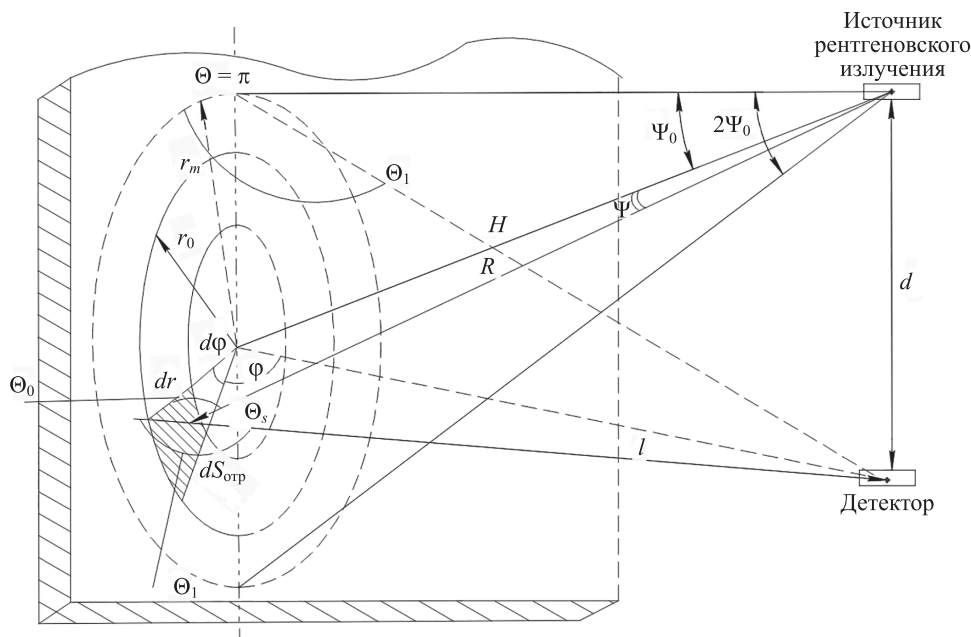


Рис. 2. Геометрическая схема прохождения квантов рентгеновского излучения через среду и расчета значений альбеда: $\Theta_0, \Theta_1, \Theta_s$ — углы падения, выхода и рассеяния квантов соответственно, рад; Ψ_0 — половина угла раскрытия диаграммы направленности генератора; Ψ — угол между нормалью к элементу поверхности и направлением квантов рентгеновского излучения; φ — угол рассеяния квантов рентгеновского излучения в подобластях; H — расстояние от источника до космического аппарата; R — длина пробега рентгеновского кванта; l — длина пробега отраженного рентгеновского кванта до детектора; d — величина базы «источник-детектор»; r — радиус отражаемой поверхности; r_0, r_m — характеристики отражающей поверхности для расчета альбеда

Fig. 2. The geometric scheme of the passage of X-ray quanta through matter and calculation of albedo values: $\Theta_0, \Theta_1, \Theta_s$ — the angles of incidence, output and scattering of quanta, respectively; Ψ_0 — half of the angle of opening of the radiation pattern of the generator; Ψ — the angle between the normal to the surface element and the direction of the X-ray quanta; φ — scattering angle of X-ray quanta in subdomains; H — the distance from the source to the spacecraft; R — the length of the path of the X-ray quantum; l — the length of the path of the reflected X-ray quantum to the detector; d — the value of the base “source-detector”; r — the radius of the reflected surface; r_0, r_m — characteristics of the reflecting surface for calculating albedo

нальной геометрической модели прохождения излучения (рис. 2), площадью $dS_{\text{отр}} = r dr d\varphi$ с энергией $dE_{\text{пад}}$:

$$dE_{\text{пад}} = \frac{E_0}{2\pi R^2} \cos\Psi \exp(-\mu_0 R) dS_{\text{отр}},$$

где $R = (H^2 + r^2)^{1/2}$; μ_0 — энергетический коэффициент ослабления рентгеновского излучения, падающего на отражатель.

Тогда с помощью предлагаемой геометрической модели прохождения излучения через среду и расчета альбеда (рис. 2) можно рассчитать значение падающей энергии $dE_{\text{пад}}$ на элемент отражаемой поверхности $dS_{\text{отр}}$.

Авторская модель (рис. 2) позволяет сформулировать математическую запись для расчетов параметров выражений.

Тогда энергия излучения, падающая на отражаемую поверхность радиусом $r = H \tan\Psi_0$, будет равна:

$$E_{\text{пад}} = \int_0^{H \tan\Psi_0} \int_0^{2\pi} \frac{E_0}{2\pi R^2} \exp(-\mu_0 R) \cos\Psi r dr d\varphi.$$

Величина отраженной энергии E_1^s в точке приема составляет:

$$E_1^s = E_{\text{пад}} \Omega A_E \beta,$$

где $\Omega = 4\pi \sin^2\left(\frac{\Psi_0}{2}\right)$; β — геометрический параметр системы:

$$\beta = \int_0^{H \tan\Psi_0} \int_0^{2\pi} \times \frac{\exp(-9\sqrt{H^2 + r^2}) \exp(-\mu_1 \sqrt{H^2 + d^2 + r^2 - 2dr \cos\varphi})}{8\pi^2 (H^2 + r^2)(H^2 + d^2 + r^2 - 2dr \cos\varphi)} r dr d\varphi.$$

Коэффициенты $\bar{\mu}_0$, μ_1 имеют непосредственную взаимосвязь с зарядовыми числами атомов вещества. Общий пучок излучения разбивается относительно проекции на космический аппарат по подобластям с учетом значений дифференциальных альбедо A_N и a_E . Определение A_N в зависимости от толщины конструкционного материала космического аппарата осуществляется по следующему выражению:

$$A_N = \rho_i \frac{N_A}{A_i} \frac{1}{\cos\Theta_1} \left\{ \frac{d\sigma_k}{d\Omega}(E_0, \Theta_0) Z \times \times \left(\frac{\cos\Theta_0 \cos\Theta_1}{\mu_0(E_0) \cos\Theta_1 + \bar{\mu}_1(E_1^s) \cos\Theta_0} \right) \times \times \left[1 - \exp\left(-\frac{\bar{\mu}_0(E_0)}{\cos\Theta_0} + \frac{\bar{\mu}_1(E_1^s)}{\cos\Theta_1} \cdot \xi\right) \right] \frac{d\sigma_R}{d\Omega}(E_0, \Theta_0) \times \times \left(\frac{\cos\Theta_0}{2\bar{\mu}_0(E_0)} \right) \left[1 - \exp\left(-\frac{2\bar{\mu}_0(E_0)}{\cos\Theta_0}\right) \right] \right\}.$$

Определим a_E по формуле:

$$a_E = \rho_i \frac{N_A}{A_i} \sec\Theta_0 \left\{ \frac{d\sigma_k}{d\Omega}(E_0, \Theta_0) Z \times$$

$$\times \left(\frac{\cos\Theta_0 \cos\Theta_1}{\mu_0(E_0) \cos\Theta_1 + \bar{\mu}_1(E_1^s) \cos\Theta_0} \right) \times \times \left[1 - \exp\left(-\frac{\bar{\mu}_0(E_0)}{\cos\Theta_0} + \frac{\bar{\mu}_1(E_1^s)}{\cos\Theta_1} \cdot \xi\right) \right] \left(\frac{E_1^s}{E_0} \right) + + \frac{d\sigma_R}{d\Omega}(E_0, \Theta_0) \left(\frac{\cos\Theta_0}{2\bar{\mu}_0(E_0)} \right) \left[1 - \exp\left(-\frac{2\bar{\mu}_0(E_0)}{\cos\Theta_0}\right) \right] \right\}.$$

В каждой подобласти считается, что все кванты прошли по одной прямой линии, перпендикулярной к центру подобласти. В зависимости от вероятности проникновения P квантов рентгеновского излучения в подобласти рассеяния в обратном направлении и выходу к детектору, выбираемого направления движения пучка квантов в программе расчета динамически вычисляются точки пересечений отдельных объектов космического аппарата и путей квантов в подобластях.

Определение максимального пробега квантов флуоресценции осуществляется по следующей функциональной зависимости:

$$R_{\text{макс}} = f(E_0, N_0, K_{\Sigma E^*}, \varepsilon_d, P_{\text{ск}}, A_N),$$

где N_0 — выход квантов, с^{-1} ; $K_{\Sigma E^*}$ — суммарный коэффициент ослабления энергии рентгеновского излучения корпусом космического аппарата; ε_d — эффективность регистрации детектором квантов рентгеновского излучения, %; $P_{\text{ск}}$ — вероятность поглощения квантов рентгеновского излучения с испусканием фотоэлектрона.

На основании полученных значений a_E , A_N и $P_{\text{пр}}$, а также учета толщин элементов конструкции космического аппарата выбирается оптимальная энергия излучателя рентгеновского излучения.

Вычисление длины λ_i пути квантов рентгеновского излучения в заданной подобласти аппарата произведена с учетом факторов, определяющих виды взаимодействия [7]:

$$\lambda_i = f(E_0, \Psi, \varphi, \bar{\mu}_0, \bar{\mu}_1, S_q, R, A_N),$$

где S_q — скачок поглощения (характеристика изменения коэффициента поглощения элементом на соответствующем крае поглощения).

Выполним расчет числа отраженных квантов рентгеновского излучения на поверхности детектора [8]

$$N_{\text{отр}} = f(N_0, A_N, \beta, S_{\text{отр}}, S_d, R, K_{\Sigma E^*}, d),$$

где S_d — площадь детектора, м^2 .

Решение задачи вычислительной геометрии позволяет вычислить рентгеновскую флуоресценцию каждой подобласти космического аппарата (X, Y, Z, φ, Θ) и доли интенсивности отдельных линий флуоресцентного излучения, а также суммарные интенсивности всех линий, серий, соответствующих скачку поглощения.

Расчет интенсивности флуоресценции (I_2) осуществляется по следующей зависимости:

$$I_2 = f(I_1, \omega_q, \bar{\mu}_0, \bar{\mu}_1, S_q, R, A_N, \lambda_i, \Psi, \varphi, \Theta, E_0),$$

где I_1 — интенсивность падающего потока, квант/(с·ст); ω_q — выход флуоресценции, %.

После проведения моделирования результаты представляются в графическом виде.

Если пришедшее на детектор излучение значительно превышает фоновое космическое излучение на длине волны, соответствующей характеристическому излучению того или иного элемента, то делается вывод о наличии на борту космического аппарата соответствующего химического элемента.

В итоге были рассмотрены физические, геометрические и математические основы, положенные в основу разработки программы, позволяющей получить результаты, поставленные в цели работы.

Алгоритм программы

Модель¹ реализована в программной среде MS Visio с применением дополнительных программных средств, чтобы более наглядно показать работу метода дистанционного контроля радиационных параметров космического аппарата на основе рентгенофлуоресцентного анализа.

Примем следующие начальные исходные данные: минимальное число фотонов рентгеновского излучения для обнаружения ядерных материалов на борту космического аппарата от 10^{15} до 10^{21} для дистанций от 10 до 10 000 м (длина волны излучения $\lambda = 0,005$ нм, угол расхождения пучка источника излучения инспектора $\varphi = 1^\circ$, площадь детектора $S = 4$ м²). Упрощенный вариант алгоритма программы при словесном способе его представления разработан с учетом исходных данных и допущений, рассмотренных ранее, и включает в себя следующие укрупненные блоки.

Блоки ввода-вывода: (1) обеспечивает ввод исходных данных по объекту инспекции и аппарату-инспектору и получение данных о моделируемой ситуации; (2) загружает выбранные параметры модели излучателя аппарата-инспектора, модели фонового излучения и трехмерную модель инспектируемого космического аппарата с данными об используемых в нем материалах; (7) выводит зависимость интенсивности излучения от длины волны в виде графика и трехмерную модель космического аппарата с траекториями пучков рентгеновского излучения, пересекших объект. Трехмерная визуализация осуществляется на основе расчетов положения пучка рентгеновского излучения и положения объекта инспекции на экране с учетом решенных задач вычислительной геометрии по поиску координат и углов пересечения прямых линий (пучков рентгеновского излучения) и трехмерных фигур произвольной формы в каждой подобласти аппарата.

Блоки обработки данных: (3) производит расчет ослабления пучка излучения при прохождении им космического пространства до поверхности космического аппарата; (4) производит расчет пересечения элементарными подпучками слоев материала космического аппарата; (5) циклически перебирает элементарные

подпучки; (5А) рассчитывает флуоресцентное излучение при прохождении элементарного подпучка от поверхностного слоя до центра космического аппарата с учетом доли высокоэнергетического флуоресцентного излучения вышележащих слоев, направленного вглубь космического аппарата. Реализует функции нахождения коэффициентов ослабления и поглощения, расчета доли сквозного непроореагировавшего излучения и флуоресцентного излучения. Данный блок представляет наиболее долгий и ресурсоемкий участок алгоритма, поскольку использует большое количество физических констант и вспомогательных математических алгоритмов; (5Б) рассчитывает ослабление флуоресцентного излучения при прохождении слоев космического аппарата в обратном направлении до поверхности космического аппарата; (6) производит расчет ослабления пучка излучения с поверхности космического аппарата при прохождении им космического пространства до детектора аппарата-инспектора.

Результаты исследования

С учетом ранее рассмотренных исходных данных и допущений, разработанной программы выполнен численный эксперимент с целью получения спектра рентгенофлуоресцентного анализа материалов исследуемого объекта.

На рис. 3 представлен результат моделирования рентгеновской флуоресценции, вызванной прохождением пучка рентгеновского излучения через космический аппарат, содержащий материалы, относящиеся к радиоактивному ряду. Яркими линиями обозначены длины волн, явно указывающие на присутствие таких химических элементов, как уран и плутоний.

Результаты расчетов, выполненных с помощью разработанной программы, могут стать основой для обоснования требований к излучательной аппаратуре спутника-инспектора.

Отдельное внимание необходимо уделить принятым допущениям. В разработанной программе не обрабатываются свойства материалов низкой плотности и

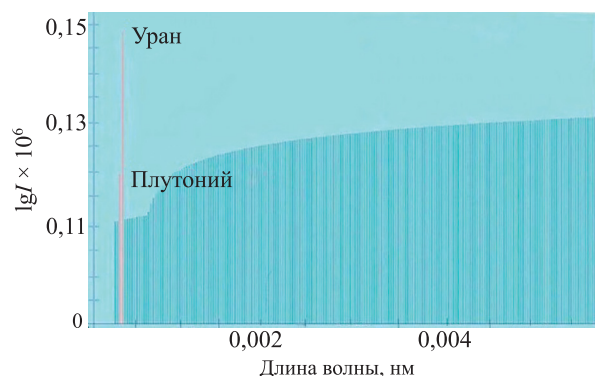


Рис. 3. Совмещенные спектры рентгенофлуоресцентного излучения материалов космических аппаратов (красный цвет) и фонового рентгеновского излучения космического пространства (синий цвет)

Fig. 3. Combined spectra of the X-ray fluorescence radiation of spacecraft materials (red) and the background X-ray radiation of outer space (blue)

¹ [Электронный ресурс]. <https://disk.yandex.ru/i/s7nufJ17OLQ6mA> (дата обращения: 19.02.2022).

элементным составом до алюминия, поскольку поглощение рентгеновского кванта такими материалами с последующим выходом флуоресценции маловероятно. Кроме того, не учитываются неровности поверхности аппарата, так как вызываемая ими разница путей, пройденных отдельными лучами, пренебрежимо мала по сравнению с общим расстоянием, преодолеваемым ими в космосе. Задаваемые углы ограничены двадцатью угловыми минутами. Все цифровые массивы в машинной памяти представлены в формате числовых данных float языка программирования C.

Указанные допущения не оказывают существенного влияния на адекватность моделирования физических процессов, изначально характеризующихся точностью значений физических коэффициентов, используемых в расчетах.

Вывод

Разработаны модель и программа имитационного моделирования прохождения пучка рентгеновского

излучения через материалы космических аппаратов. В зависимости от расстояния между спутником-инспектором и исследуемым объектом, длины волны и энергии излучения, а также угла падения квантов на поверхность функциональных узлов космических аппаратов определяется интенсивность флуоресцентного излучения и делается вывод об элементном составе материалов внутреннего наполнения объекта.

Заключение

Рассмотрена возможность применения метода рентгенофлуоресцентного анализа для инспектирования космических аппаратов с целью соблюдения требований международного законодательства в области использования космического пространства.

Полученные результаты могут найти применение в ходе обоснования технических проектов, предназначенных для решения задач инспекции орбитальных средств.

Литература

1. Клименко Н.Н., Назаров А.Е. Перспективная космическая система для наблюдения геостационарной орбиты // Вестник «НПО имени С.А. Лавочкина». 2015. № 4(30). С. 16–22.
2. Ляпко П.С., Олейников И.И., Улин С.Е. Методы инспекции космических аппаратов с ядерными энергетическими установками на борту // Краткие сообщения по физике ФИАН. 2016. Т. 43. № 11. С. 24–30.
3. Лосев Н.Ф. Количественный рентгеноспектральный флуоресцентный анализ. М.: Наука, 1969. 336 с.
4. Маркова Ю.Н., Анчутина Е.А., Кербер Е.В., Максимовская В.В., Золотарев К.В. Оценка пригодности методики измерений состава горных пород для анализа растительных материалов рентгенофлуоресцентным методом с использованием синхротронного излучения // Стандартные образцы. 2013. № 4. С. 19–24.
5. Симаков В.А. Рентгенофлуоресцентный анализ руд и продуктов их переработки (при разведке месторождений): автореферат диссертации на соискание ученой степени доктора технических наук: 02.00.02. Москва, 2000. 40 с.
6. Shand C.A., Wendler R. Portable X-ray fluorescence analysis of mineral and organic soils and the influence of organic matter // Journal of Geochemical Exploration. 2014. V. 143. P. 31–42. <https://doi.org/10.1016/j.gexplo.2014.03.005>
7. Мазурицкий М.И., Лерер А.М., Новакович А.А., Махно А.С., Махно П.В. Канализация рентгеновской флуоресценции, возбуждаемой внутри полых микрокапилляров // Инженерный вестник Дона. 2013. № 3(26). С. 65.
8. Симаков В.А., Кордюков С.В. Применение стандартных образцов состава при рентгеноспектральном флуоресцентном анализе твердых полезных ископаемых // Стандартные образцы. 2013. № 4. С. 11–15.

Авторы

Лукьянова Людмила Анатольевна — старший преподаватель, Санкт-Петербургский государственный университет, Санкт-Петербург, 199034, Российская Федерация, <https://orcid.org/0000-0001-5458-3019>, lalukianova@mail.ru

Свитнев Игорь Владимирович — кандидат военных наук, доцент, доцент, заведующий кафедрой, Санкт-Петербургский государственный университет, Санкт-Петербург, 199034, Российская Федерация, <https://orcid.org/0000-0002-8862-1797>, isvitnev@mail.ru

Харитонов Елена Александровна — кандидат медицинских наук, доцент, Санкт-Петербургский государственный университет, Санкт-

References


1. Klimenko N.N., Nazarov A.E. Advanced space-borne system for surveillance of geostationary orbit. *Vestnik "NPO im. S. A. Lavochkina"*, 2015, no. 4(30), pp. 16–22. (in Russian)
2. Lyanko P.S., Oleinikov I.I., Ulin S.E. Inspection methods for spacecrafts with nuclear power plants. *Bulletin of the Lebedev Physics Institute*, 2016, vol. 43, no. 11, pp. 332–335. <https://doi.org/10.3103/S106833561611004X>
3. Losev N.F. *Quantitative X-ray Spectrum Fluorescence Analysis*. Moscow, Nauka Publ., 1969, 336 p. (in Russian)
4. Markova Yu.N., Anchutina E.A., Kerber E.V., Maksimovskaya V.V., Zolotarev K.V. Applicability evaluation of procedure for measurement of rock composition for analysis of plant materials by x-ray fluorescence method with the use of synchrotron radiation. *Reference Materials*, 2013, no. 4, pp. 19–24. (in Russian)
5. Simakov V.A. *X-ray fluorescence analysis of ores and products of their processing (in the mineral exploration)*. Dissertation abstract for the degree of doctor of technical sciences: 02.00.02. Moscow, 2000, 40 p. (in Russian)
6. Shand C.A., Wendler R. Portable X-ray fluorescence analysis of mineral and organic soils and the influence of organic matter. *Journal of Geochemical Exploration*, 2014, vol. 143, pp. 31–42. <https://doi.org/10.1016/j.gexplo.2014.03.005>
7. Mazuritskiy M.I., Lerer A.M., Novakovich A.A., Makhno A.S., Makhno P.V. Channeling of X-ray fluorescence excited in hollow microcapillaries. *Engineering Journal of Don*, 2013, no. 3(26), pp. 65. (in Russian)
8. Simakov V.A., Kordyukov S.V. The use of reference materials of composition with x-ray fluorescent analysis of solid minerals. *Reference Materials*, 2013, no. 4, pp. 11–15. (in Russian)

Authors

Lyudmila A. Lukyanova — Senior Lecturer, St. Petersburg State University (SPbSU), Saint Petersburg, 199034, Russian Federation, <https://orcid.org/0000-0001-5458-3019>, lalukianova@mail.ru

Igor V. Svitnev — PhD (Military Science), Associate Professor, Associate Professor, Head of Department, St. Petersburg State University (SPbSU), Saint Petersburg, 199034, Russian Federation, <https://orcid.org/0000-0002-8862-1797>, isvitnev@mail.ru

Elena A. Kharitonova — PhD (Medicine), Associate Professor, St. Petersburg State University (SPbSU), Saint Petersburg, 199034, Russian

Петербург, 199034, Российская Федерация,  57217586798, <https://orcid.org/0000-0002-4851-2756>, Haritonova_ea@mail.ru

Гаврилов Илья Евгеньевич — студент, Санкт-Петербургский государственный университет, Санкт-Петербург, 199034, Российская Федерация, <https://orcid.org/0000-0003-1124-0336>, Ilyusha-2145@mail.ru

Federation,  57217586798, <https://orcid.org/0000-0002-4851-2756>, Haritonova_ea@mail.ru

Ilya E. Gavrilov — Student, St. Petersburg State University (SPbSU), Saint Petersburg, 199034, Russian Federation, <https://orcid.org/0000-0003-1124-0336>, Ilyusha-2145@mail.ru

Статья поступила в редакцию 07.04.2022
Одобрена после рецензирования 30.05.2022
Принята к печати 14.07.2022

Received 07.04.2022
Approved after reviewing 30.05.2022
Accepted 14.07.2022



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

doi: 10.17586/2226-1494-2022-22-4-659-665

УДК 535.08

Волоконно-оптический амплитудный датчик направления и величины изгиба

Андрей Анатольевич Дмитриев¹✉, Кирилл Вадимович Гребнев²,
Даниил Сергеевич Смирнов³, Сергей Владимирович Варжель⁴

^{1,2,3,4} Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

¹ Dmitrieva1994@yandex.ru✉, <https://orcid.org/0000-0002-0674-8381>

² greb.k1r@yandex.ru, <https://orcid.org/0000-0002-8461-301X>

³ dsmirnov@itmo.ru, <https://orcid.org/0000-0003-4324-9327>

⁴ Vsv187@gmail.ru, <https://orcid.org/0000-0002-3120-8109>

Аннотация

Предмет исследования. Предложен вариант реализации волоконно-оптического датчика направления и величины изгиба. В отличие от известных спектральных измерительных комплексов, рассматриваемое решение предполагает использование амплитудной методики опроса, позволяющей увеличить быстродействие сенсора, применив более простые и доступные компоненты. **Метод.** Исследован чувствительный элемент на основе специальных дифракционных структур, который состоит из пар чирпированных волоконных брэгговских решеток. Чувствительные элементы закреплены на оснастке в виде стального стержня, подвергнутого изгибному воздействию. **Основные результаты.** Продемонстрирована способность датчика определять величину и направление изгиба в диапазоне отклонения от 0 до 30 мм со среднеквадратичным отклонением измеренных значений от реальных в 0,536 мм. Данный результат измерений достигнут с помощью обработки данных, полученных с трех измерительных устройств, нейронной сетью, скрытый слой которой составил 10 нейронов, при этом функцией активации установлена сигмоида. **Практическая значимость.** Результаты исследований имеют существенное значение для современных систем мониторинга. Реализация датчика направления и величины изгиба в формате волоконно-оптического устройства способна преодолеть ограничения пьезоэлектрических сенсоров, благодаря высокой помехозащищенности и стойкости к воздействию внешних сред. Предложенное технологическое решение позволяет отойти от спектральной методики проведения измерений, получившей широкое распространение в волоконно-оптических сенсорных системах. Применение амплитудного датчика величины и направления изгиба дает возможность его использования в составе устройств, где существует необходимость в точном позиционировании управляющих элементов или структурных компонент, подвергаемых изгибному воздействию. Благодаря измерению воздействия при помощи оценки оптической мощности сигнала, конструкция датчика не предполагает наличие сложного измерительного устройства. Работоспособность сенсора может быть обеспечена при помощи каскада недорогих, но при этом высокоскоростных и долговечных фотоприемных устройств.

Ключевые слова

волоконная брэгговская решетка, волоконно-оптический датчик, датчик направления и величины изгиба

Благодарности

Работа выполнена при финансовой поддержке программы «Приоритет 2030».

Ссылка для цитирования: Дмитриев А.А., Гребнев К.В., Смирнов Д.С., Варжель С.В. Волоконно-оптический амплитудный датчик направления и величины изгиба // Научно-технический вестник информационных технологий, механики и оптики. 2022. Т. 22, № 4. С. 659–665. doi: 10.17586/2226-1494-2022-22-4-659-665

Fiber-optic amplitude bend direction and magnitude sensor

Andrei A. Dmitriev¹✉, Kirill V. Grebnev², Daniil S. Smirnov³, Sergei V. Varzhel⁴

^{1,2,3,4} ITMO University, Saint Petersburg, 197101, Russian Federation

¹ Dmitrievaa1994@yandex.ru✉, <https://orcid.org/0000-0002-0674-8381>

² greb.k1r@yandex.ru, <https://orcid.org/0000-0002-8461-301X>

³ dsmirnov@itmo.ru, <https://orcid.org/0000-0003-4324-9327>

⁴ Vsv187@gmail.ru, <https://orcid.org/0000-0002-3120-8109>

Abstract

A variant of the implementation of a fiber-optic sensor for the direction and magnitude of the bend is proposed. Unlike existing spectral measuring systems, the solution under consideration involves the use of an amplitude polling technique which makes it possible to increase the speed of the sensor when using simpler and more affordable components. A sensitive element based on special diffraction structures consisting of pairs of chirped fiber Bragg gratings has been studied. The sensing elements are mounted on a tooling — a steel rod subjected to bending. The ability of the sensor to determine the magnitude and direction of bending in the deviation range from 0 to 30 mm was demonstrated with a standard deviation of the measured values from the real values of 0.536 mm. This measurement result is achieved by processing data obtained from three measuring devices and by the neural network with a hidden layer of 10 neurons and the sigmoid as the activation function. The research results are essential for modern monitoring systems. The implementation of the direction and magnitude of the bend sensor in the format of a fiber-optic device allows you to overcome the limitations of piezoelectric sensors, due to high noise immunity and resistance to environmental influences. The proposed technological solution makes it possible to avoid the spectral measurement technique that has become widely used in fiber-optic sensor systems. The use of an amplitude sensor for the magnitude and direction of bending will allow its use in devices where there is a need for precise positioning of control elements or structural components subjected to bending. Also, due to the measurement of the desired bending effect by estimating the optical power of the signal, the design of the sensor does not require the presence of a complex measuring device, and the sensor's performance can be ensured using a cascade of inexpensive, but at the same time high-speed and durable photodetectors.

Keywords

fiber Bragg grating, fiber optic sensor, bend direction and magnitude sensor

Acknowledgements

The work is financially supported by Priority 2030 Program.

For citation: Dmitriev A.A., Grebnev K.V., Smirnov D.S., Varzhel S.V. Fiber-optic amplitude bend direction and magnitude sensor. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2022, vol. 22, no. 4, pp. 659–665 (in Russian). doi: 10.17586/2226-1494-2022-22-4-659-665

Введение

Технологический прогресс в отрасли телекоммуникаций дал стимул разработки волоконно-оптических датчиков, которые обладают рядом преимуществ по сравнению с другими существующими измерительными технологиями [1]. Такими преимуществами являются: повышенная чувствительность, невосприимчивость к электромагнитному воздействию, универсальность форм-фактора, высокая скорость отклика, а также пригодность для работы в суровых условиях окружающей среды и возможность создания массивов датчиков с помощью мультиплексирования с разделением по времени и длине волны [2, 3]. Функциональные особенности волоконно-оптических датчиков использованы для замены традиционных пьезоэлектрических решений в широком диапазоне применений, включая деформационные, вибрационные, электрические, акустические измерения, датчики ускорения, вращения, давления, температуры, влажности, вязкости и др.

Измерительные устройства на основе волоконной брэгговской решетки (ВБР) — один из типов волоконно-оптических датчиков, которые совместимы с оптическими системами связи и способны выполнять сканирование на большом удалении от блока обработки без подведения электропитания к непосредственному месту расположения чувствительного элемента.

ВБР — периодическая структура модуляции показателя преломления, индуцированная в сердцевине оптического волокна. В последние годы особое внимание уделяется изготовлению и применению чирпированных ВБР, которые характеризуются непостоянством периода по длине. Чирпированные ВБР ведут себя как каскад ВБР, каждая из которых отражает узкий спектр оптического излучения. Такие решетки применяются в основном для оптической связи, так как обеспечивают компенсацию дисперсии путем введения дифференциальной групповой задержки, зависящей от длины волны [4, 5]. В последнее время чирпированные ВБР приобрели значительный интерес со стороны построения волоконно-оптических датчиков, благодаря своим ключевым характеристикам [6–8]. В настоящей работе исследована конструкция датчика направления и величины изгиба с применением трех пар чирпированных ВБР в качестве чувствительного элемента.

Оптическое волокно и запись чирпированных волоконных брэгговских решеток

Для записи чирпированных ВБР использовано стандартное телекоммуникационное оптическое волокно SMF-28, предварительно подвергнутое водородной обработке в течение двух недель при температуре 50 °C и давлении 10 МПа для повышения фоторефрактивно-

сти [9, 10]. При таких параметрах оптического волокна появляется возможность записи chirпированных ВБР с коэффициентом отражения близким к 100 % [11, 12].

Запись chirпированных ВБР произведена с помощью оптической схемы интерферометра Тальбота [13]. Во время записи использованы следующие параметры:

- фазовая маска с линейно изменяющимся периодом в 1 нм/см;
- источник ультрафиолетового излучения — KrF эксимерная лазерная система Optosystems MOPA CL-7550;
- плотность энергии лазерного импульса — 220 мДж/см²;
- время экспозиции 120 с при частоте следования импульсов 10 Гц;
- для фокусировки лазерного излучения в области записи применена цилиндрическая линза с фокусным расстоянием 500 мм.

В результате записи получены chirпированные ВБР с коэффициентом отражения примерно 100 %.

Отметим, что благодаря универсальности данной системы, при ее использовании возможно изготовление таких специальных дифракционных структур как: наклонные ВБР [14], ВБР с фазовым сдвигом [15], массивы [12] и суперпозиции ВБР [16].

Детали эксперимента

Оснастка датчика величины и направления изгиба представляет собой металлический стержень, на котором закреплены три оптических волокна, расположенных по периметру окружности стержня с шагом в 120° относительно друг друга (рис. 1). Внутри каждого волновода индуцирована структура из двух chirпированных ВБР со смещенной на значение полуширины на полувисоте длины волны брэгговского резонанса внутри пары, а также коэффициентом отражения около 100 %.

Одна из структур в паре chirпированных решеток (чувствительная) жестко закреплена на стержне, и испытывает на себе механическое натяжение-сжатие при наведении изгибного воздействия. Вторая (опорная) остается в первоначальном положении.

При приложении механического воздействия чувствительные ВБР претерпевают изменение геометрии, которое приводит к смещению периода данной структуры и центральной длины волны отражения (λ) согласно условию Брэгга (рис. 2).

Смещение центральной длины волны брэгговского резонанса чувствительного элемента приводит к изме-

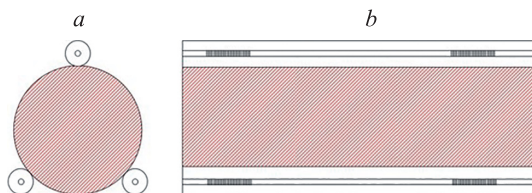


Рис. 1. Схематическое изображение чувствительного элемента измерения величины и направления изгиба: разрез металлического стержня (а); структура волновода (б)

Fig. 1. Scheme of the sensing element for measuring the magnitude and direction of bending

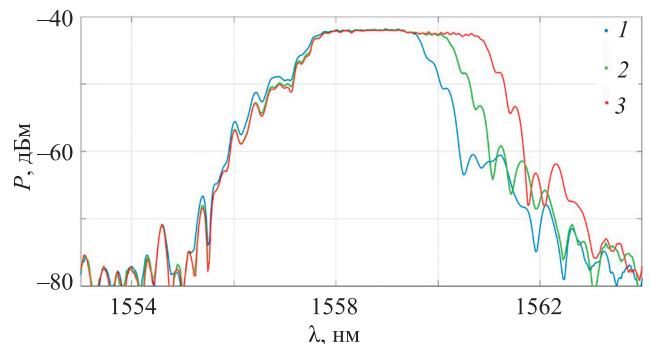


Рис. 2. Спектр пропускания чувствительного элемента при трех различных положениях наведения изгибного воздействия: без приложения изгибного воздействия (кривая 1); с отклонениями конца оснастки на 10 мм (кривая 2) и 20 мм (кривая 3)

Fig. 2. The transmission spectrum of the sensing element at three different exposure positions: without application of bending action (curve 1); with a bending of the tooling end by 10 mm (curve 2) and 20 mm (curve 3)

нению ширины общего спектра отражения (P) структуры чувствительной опорной chirпированных ВБР. Данные изменения приводят к изменению отраженной от структуры оптической мощности, которая может быть измерена при помощи фотоприемного устройства.

Оснастка датчика закреплена на измерительном стенде (рис. 3). Красными линиями обозначено наводимое изгибное воздействие на оснастку. Источником излучения для опроса структур в рамках эксперимента выступает подключенный через ответвитель 1×3 СЛД.

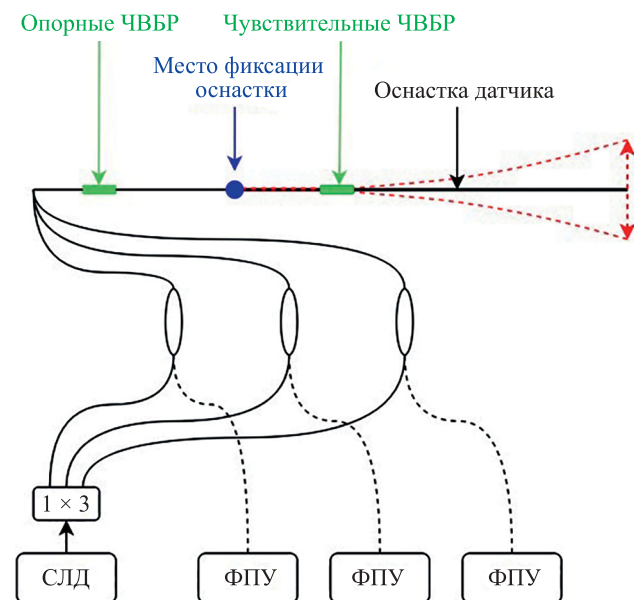


Рис. 3. Схематическое изображение испытательного стенда измерения величины и направления изгиба.

ЧВБР — chirпированная волоконная брэгговская решетка; СЛД — суперлюминесцентный диод; ФПУ — фотоприемное устройство

Fig. 3. Scheme of the test bench for measuring the magnitude and direction of bending

ЧВБР — chirped fiber Bragg grating; СЛД — superluminescent diode; ФПУ — photodetector

Излучение, проходя через оптическое волокно, попадает на опрашиваемые структуры, где, отражаясь от пары чирпированных ВБР, распространяется в обратном направлении и через Y-ответвитель попадает на ФПУ. При приложении изгибающего воздействия три пары специальных структур изменяют приходящую на ФПУ отраженную оптическую мощность. Для контролируемой установки направления и величины изгиба изготовлен шаблон с отверстиями, расположенными с шагом 5 мм при вариации угла 30° (рис. 4, а).

Изгибное воздействие на оснастку датчика выполнено на изготовленном испытательном стенде с помощью установки свободного конца стержня в соответствующие отверстия шаблона (рис. 4, б). Отметим, что температурное воздействие окружающей среды на чувствительный элемент оказывает влияние как на чувствительные, так и на опорные структуры в составе пар чирпированных ВБР. В этом случае спектр отражения испытывает совместное эквивалентное смещение в частотной области. Данное свойство приводит к отсутствию необходимости в дополнительном механизме температурной компенсации. Эксперимент по исследованию температурной стабильности подобного чувствительного элемента в диапазоне от 24°C до 100°C

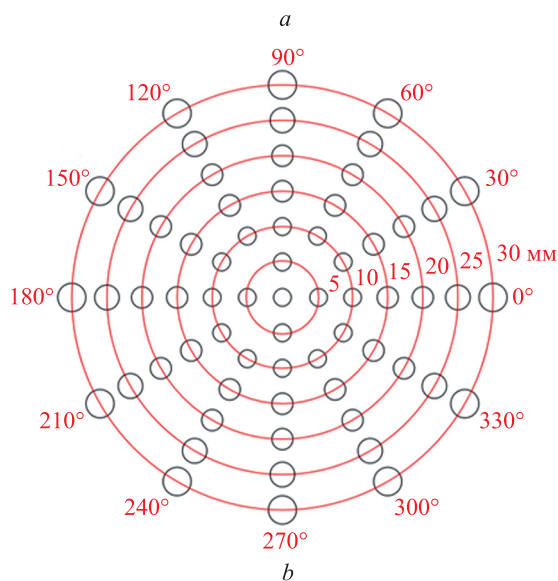


Рис. 4. Схематическое (а) и фотографическое (б) изображения шаблона установки величины и направления изгиба

Fig. 4. Schematic (a) and photographic (b) images of the bending magnitude and direction setting pattern

в составе амплитудного тензометрического датчика представлен в работе [17].

Результаты эксперимента

В ходе эксперимента закрепленная на измерительном стенде оснастка установлена на углы отклонения (α), в диапазоне от 0° до 330° с шагом 30° при значениях отклонения от 10 мм до 30 мм с шагом 5 мм. Отклонение стержня от исходного положения соответствует углом в диапазоне от 5° до 13° с шагом в 2° . При проведении эксперимента зарегистрирована комбина-

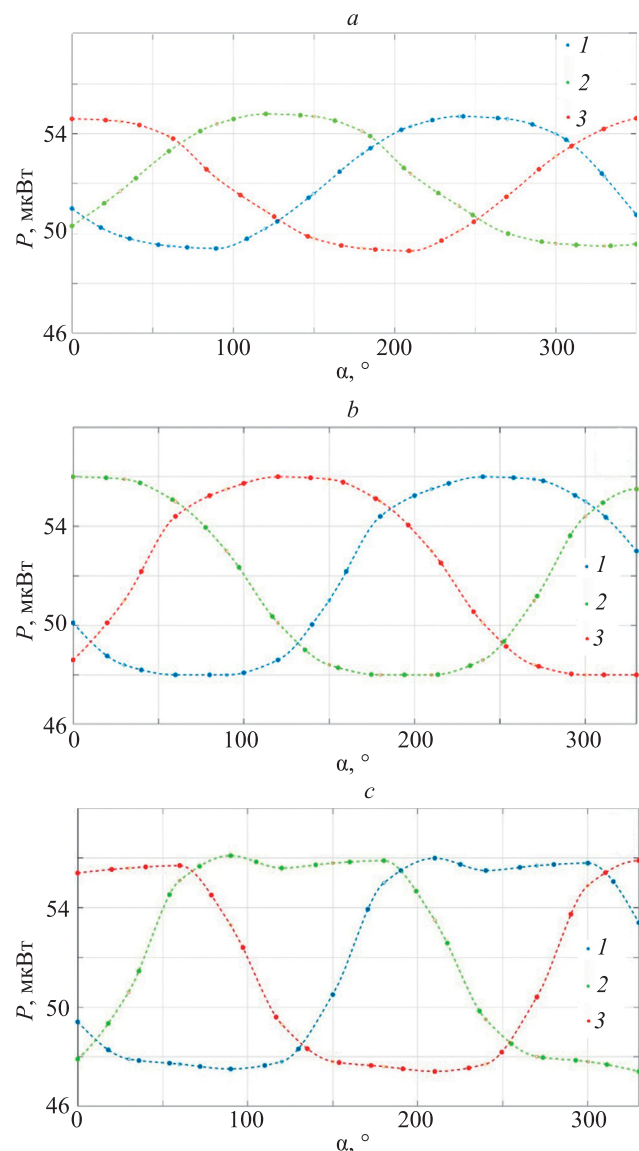


Рис. 5. Изменение отраженной оптической мощности (P) от угла направления изгиба (α) для трех чувствительных элементов при величинах изгиба: 10 мм (а), 20 мм (б), 30 мм (с).

Зависимости, характеризующие: первый (кривая 1), второй (кривая 2) и третий (кривая 3) чувствительные элементы

Fig. 5. Combination of modulation characteristics of the reflected optical power vs. the angle of the bend direction at the amount of bend: 10 mm (a), 20 mm (b), 30 mm (c).

Dependencies characterize the first (curve 1), second (curve 2), and third (curve 3) sensitive elements

ция значений отраженной оптической мощности, состоящая из показаний трех опрашивающих устройств в каждой точке измерения. В ходе исследования получены характеристики изменения отраженной оптической мощности для трех структур при изменении величины и направления изгиба (рис. 5).

Для определения величины и направления изгиба полученные наборы данных использованы для обучения искусственной нейронной сети с применением инструмента MATLAB 2020b Neural Network Toolbox.

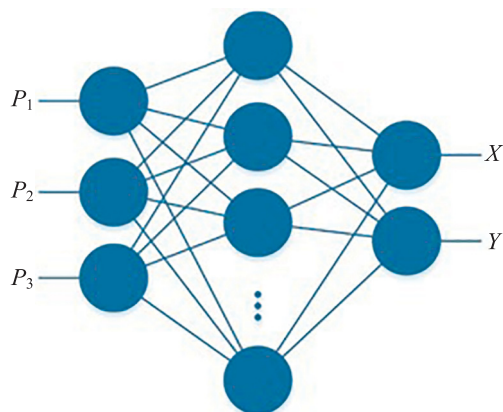


Рис. 6. Архитектура нейронной сети

Fig. 6. Neural network architecture

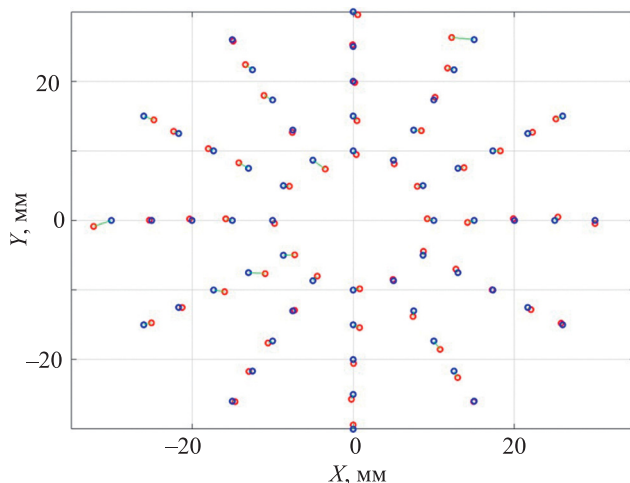


Рис. 7. Распределение измеренных (красные метки) и реальных (синие метки) значений величины и направления изгиба

Fig. 7. Distribution of measured (red marks) and real (blue marks) values of the magnitude and direction of the bend

В рамках данной задачи входными данными нейронной сети являются три значения оптической мощности, отраженной в каждом волокне — (P_1 , P_2 и P_3), а выходными значениями — направление (X) и величина изгиба (Y), интерпретируемые как координаты отклонения иглы (рис. 6). Скрытый слой нейронной сети составил 10 нейронов, а функцией активации установлена сигмоида. Обучение нейронной сети произведено с применением метода градиентного спуска.

В результате обучения нейронной сети на базе набора данных из эксперимента получены весовые коэффициенты, которые были применены для проверки работы алгоритма оценки величины и направления отклонения. Результаты проверки приведены на рис. 7, среднее квадратичное отклонение измеренного расположения относительно реального составило 0,536 мм.

Заключение

В работе представлен датчик величины и направления изгиба, в основе чувствительного элемента которого использованы пары чирпированных волоконных брэгговских решеток. Смещение спектрального отклика одной структуры относительно другой внутри пары произвело изменение отраженной оптической мощности, которая детектируется устройствами опроса. Такой принцип проведения измерений определил представленный датчик как амплитудное устройство измерения. В отличие от классических сенсоров на основе волоконной брэгговской решетки с использованием спектральной методики опроса, амплитудные сенсоры обладают более высокой скоростью работы, а также сравнительно низкой себестоимостью устройств опроса. Заметим, что такой чувствительный элемент не требует дополнительного механизма температурной компенсации вследствие близкого расположения компонентов и соответственно синхронного смещения характеристик чувствительной и опорной структур в спектральной области.

Исследованный прототип измерительного устройства позволяет определять направление и величину изгиба с высокой точностью и скоростью опроса. Среднее квадратичное отклонение измеренных значений от реальных значений составило 0,536 мм в диапазоне отклонения свободного конца оснастки от 0 до 30 мм. Датчики величины и направления изгиба, использующие метод, исследованный в рамках проведенной работы, могут найти применение в приложениях для мониторинга состояния конструктивных элементов зданий и сооружений, медицинских исследованиях, а также в других областях науки и техники.

Литература

References

1. Annamdas V.G.M. Review on developments in fiber optical sensors and applications // *International Journal of Materials Engineering*. 2011. V. 1. N 1. P. 1–16. <https://doi.org/10.5923/j.ijme.20110101.01>
2. Mihailov S.J. Fiber Bragg grating sensors for harsh environments // *Sensors*. 2012. V. 12. N 2. P. 1898–1918. <https://doi.org/10.3390/s120201898>
3. Kersey A.D., Davis M.A., Patrick H.J., LeBlanc M., Koo K.P., Askins C.G., Putnam M.A., Friebele E.J. Fiber grating sensors // *Journal of Lightwave Technology*. 1997. V. 15. N 8. P. 1442–1463. <https://doi.org/10.1109/50.618377>
4. Byron K.C., Sugden K., Bricheno T., Bennion I. Fabrication of chirped Bragg gratings in photosensitive fibre // *Electronics Letters*. 1993. V. 29. N 18. P. 1659–1660. <https://doi.org/10.1049/el:19931104>
5. Hill K.O., Bilodeau F., Malo B., Kitagawa T., Thériault S., Johnson D.C., Albert J., Takiguchi K. Chirped in-fiber Bragg gratings for compensation of optical-fiber dispersion // *Optics Letters*. 1994. V. 19. N 17. P. 1314–1316. <https://doi.org/10.1364/OL.19.001314>
6. Feng K.-M., Chai J.-X., Grubsky V., Starodubov D.S., Hayee M.I., Lee S., Jiang X., Willner A.E., Feinberg J. Dynamic dispersion compensation in a 10-Gb/s optical system using a novel voltage tuned nonlinearly chirped fiber Bragg grating // *IEEE Photonics Technology Letters*. 1999. V. 11. N 3. P. 373–375. <https://doi.org/10.1109/68.748240>
7. Markowski K., Jedrzejewski K., Marzecki M., Osuch T. Linearly chirped tapered fiber-Bragg-grating-based Fabry–Perot cavity and its application in simultaneous strain and temperature measurement // *Optics Letters*. 2017. V. 42. N 7. P. 1464–1467. <https://doi.org/10.1364/OL.42.001464>
8. Wei P., Lang H., Liu T., Xia D. Detonation velocity measurement with chirped fiber Bragg grating // *Sensors*. 2017. V. 17. N 11. P. 2552. <https://doi.org/10.3390/s17112552>
9. Lemaire P.J., Atkins R.M., Mizrahi V., Reed W.A. High pressure H₂ loading as a technique for achieving ultrahigh UV photosensitivity and thermal sensitivity in GeO₂ doped optical fibres // *Electronics Letters*. 1993. V. 29. N 13. P. 1191–1193. <https://doi.org/10.1049/el:19930796>
10. Varzhel' S.V., Mun'ko A.S., Konnov K.A., Gribaev A.I., Kulikov A.V. Recording Bragg gratings in hydrogenated birefringent optical fiber with elliptical stress cladding // *Journal of Optical Technology*. 2016. V. 83. N 10. P. 638–641. <https://doi.org/10.1364/JOT.83.000638>
11. Mikhneva A.A., Gribaev A.I., Varzhel' S.V., Frolov E.A., Novikova V.A., Konnov K.A., Zaleskaya Y.K. Inscription and investigation of the spectral characteristics of chirped fiber Bragg gratings // *Journal of Optical Technology*. 2018. V. 85. N 9. P. 531–534. <https://doi.org/10.1364/JOT.85.000531>
12. Dmitriev A.A., Gribaev A.I., Varzhel S.V., Konnov K.A., Motorin E.A. High-performance fiber Bragg gratings arrays inscription method // *Optical Fiber Technology*. 2021. V. 63. P. 102508. <https://doi.org/10.1016/j.yofte.2021.102508>
13. Gribaev A.I., Pavlishin I.V., Stam A.M., Idrisov R.F., Varzhel S.V., Konnov K.A. Laboratory setup for fiber Bragg gratings inscription based on Talbot interferometer // *Optical and Quantum Electronics*. 2016. V. 48. N 12. P. 540. <https://doi.org/10.1007/s11082-016-0816-3>
14. Дмитриев А.А., Гребнев К.В., Варжель С.В., Плотноков М.Ю. Волоконно-оптический датчик вибрации на основе SMF-MMF-SMF перехода и наклонной решетки Брэгга // *Научно-технический вестник информационных технологий, механики и оптики*. 2021. Т. 21. № 6. С. 801–807. <https://doi.org/10.17586/2226-1494-2021-21-6-801-807>
15. Новикова В.А., Варжель С.В., Лосева Е.А., Дмитриев А.А. Экспериментальное исследование и моделирование волоконных брэгговских решёток с фазовым сдвигом // *Оптический журнал*. 2021. Т. 88. № 6. С. 36–44. <https://doi.org/10.17586/1023-5086-2021-88-06-36-44>
16. Idrisov R.F., Gribaev A.I., Stam A.M., Varzhel' S.V., Slozhenikina Yu.I., Konnov K.A. Inscription of superimposed fiber Bragg gratings using a Talbot interferometer // *Journal of Optical Technology*. 2017. V. 84. N 10. P. 694–697. <https://doi.org/10.1364/JOT.84.000694>
17. Dmitriev A.A., Varzhel S.V., Grebnev K.V., Anokhina E.V. Strain gauge based on n-pairs of chirped fiber Bragg gratings // *Optical Fiber Technology*. 2022. V. 70. P. 102893. <https://doi.org/10.1016/j.yofte.2022.102893>
1. Annamdas V.G.M. Review on developments in fiber optical sensors and applications. *International Journal of Materials Engineering*, 2011, vol. 1, no. 1, pp. 1–16. <https://doi.org/10.5923/j.ijme.20110101.01>
2. Mihailov S.J. Fiber Bragg grating sensors for harsh environments. *Sensors*, 2012, vol. 12, no. 2, pp. 1898–1918. <https://doi.org/10.3390/s120201898>
3. Kersey A.D., Davis M.A., Patrick H.J., LeBlanc M., Koo K.P., Askins C.G., Putnam M.A., Friebele E.J. Fiber grating sensors. *Journal of Lightwave Technology*, 1997, vol. 15, no. 8, pp. 1442–1463. <https://doi.org/10.1109/50.618377>
4. Byron K.C., Sugden K., Bricheno T., Bennion I. Fabrication of chirped Bragg gratings in photosensitive fibre. *Electronics Letters*, 1993, vol. 29, no. 18, pp. 1659–1660. <https://doi.org/10.1049/el:19931104>
5. Hill K.O., Bilodeau F., Malo B., Kitagawa T., Thériault S., Johnson D.C., Albert J., Takiguchi K. Chirped in-fiber Bragg gratings for compensation of optical-fiber dispersion. *Optics Letters*, 1994, vol. 19, no. 17, pp. 1314–1316. <https://doi.org/10.1364/OL.19.001314>
6. Feng K.-M., Chai J.-X., Grubsky V., Starodubov D.S., Hayee M.I., Lee S., Jiang X., Willner A.E., Feinberg J. Dynamic dispersion compensation in a 10-Gb/s optical system using a novel voltage tuned nonlinearly chirped fiber Bragg grating. *IEEE Photonics Technology Letters*, 1999, vol. 11, no. 3, pp. 373–375. <https://doi.org/10.1109/68.748240>
7. Markowski K., Jedrzejewski K., Marzecki M., Osuch T. Linearly chirped tapered fiber-Bragg-grating-based Fabry–Perot cavity and its application in simultaneous strain and temperature measurement. *Optics Letters*, 2017, vol. 42, no. 7, pp. 1464–1467. <https://doi.org/10.1364/OL.42.001464>
8. Wei P., Lang H., Liu T., Xia D. Detonation velocity measurement with chirped fiber Bragg grating. *Sensors*, 2017, vol. 17, no. 11, pp. 2552. <https://doi.org/10.3390/s17112552>
9. Lemaire P.J., Atkins R.M., Mizrahi V., Reed W.A. High pressure H₂ loading as a technique for achieving ultrahigh UV photosensitivity and thermal sensitivity in GeO₂ doped optical fibres. *Electronics Letters*, 1993, vol. 29, no. 13, pp. 1191–1193. <https://doi.org/10.1049/el:19930796>
10. Varzhel' S.V., Mun'ko A.S., Konnov K.A., Gribaev A.I., Kulikov A.V. Recording Bragg gratings in hydrogenated birefringent optical fiber with elliptical stress cladding. *Journal of Optical Technology*, 2016, vol. 83, no. 10, pp. 638–641. <https://doi.org/10.1364/JOT.83.000638>
11. Mikhneva A.A., Gribaev A.I., Varzhel' S.V., Frolov E.A., Novikova V.A., Konnov K.A., Zaleskaya Y.K. Inscription and investigation of the spectral characteristics of chirped fiber Bragg gratings. *Journal of Optical Technology*, 2018, vol. 85, no. 9, pp. 531–534. <https://doi.org/10.1364/JOT.85.000531>
12. Dmitriev A.A., Gribaev A.I., Varzhel S.V., Konnov K.A., Motorin E.A. High-performance fiber Bragg gratings arrays inscription method. *Optical Fiber Technology*, 2021, vol. 63, pp. 102508. <https://doi.org/10.1016/j.yofte.2021.102508>
13. Gribaev A.I., Pavlishin I.V., Stam A.M., Idrisov R.F., Varzhel S.V., Konnov K.A. Laboratory setup for fiber Bragg gratings inscription based on Talbot interferometer. *Optical and Quantum Electronics*, 2016, vol. 48, no. 12, pp. 540. <https://doi.org/10.1007/s11082-016-0816-3>
14. Dmitriev A.A., Grebnev K.V., Varzhel S.V., Plotnikov M.Yu. A fiber optic vibration sensor based on SMF-MMF-SMF Structure And A Tilted Fiber Bragg grating. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2021, vol. 21, no. 6, pp. 801–807. (in Russian). <https://doi.org/10.17586/2226-1494-2021-21-6-801-807>
15. Novikova V.A., Varzhel' S.V., Loseva E.A., Dmitriev A.A. Experimental investigation and simulation of phase-shifted fiber Bragg gratings. *Journal of Optical Technology*, 2021, vol. 88, no. 6, pp. 315–320. <https://doi.org/10.1364/JOT.88.000315>
16. Idrisov R.F., Gribaev A.I., Stam A.M., Varzhel' S.V., Slozhenikina Yu.I., Konnov K.A. Inscription of superimposed fiber Bragg gratings using a Talbot interferometer. *Journal of Optical Technology*, 2017, vol. 84, no. 10, pp. 694–697. <https://doi.org/10.1364/JOT.84.000694>
17. Dmitriev A.A., Varzhel S.V., Grebnev K.V., Anokhina E.V. Strain gauge based on n-pairs of chirped fiber Bragg gratings. *Optical Fiber Technology*, 2022, vol. 70, pp. 102893. <https://doi.org/10.1016/j.yofte.2022.102893>

Авторы

Дмитриев Андрей Анатольевич — инженер-исследователь, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 57215003358](https://orcid.org/0000-0002-0674-8381), <https://orcid.org/0000-0002-0674-8381>, Dmitrievaa1994@yandex.ru

Гребнев Кирилл Вадимович — инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 57390827800](https://orcid.org/0000-0002-8461-301X), <https://orcid.org/0000-0002-8461-301X>, greb.k1r@yandex.ru

Смирнов Даниил Сергеевич — младший научный сотрудник, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 57204933153](https://orcid.org/0000-0003-4324-9327), <https://orcid.org/0000-0003-4324-9327>, dsmirnov@itmo.ru

Варжель Сергей Владимирович — кандидат физико-математических наук, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 55247304200](https://orcid.org/0000-0002-3120-8109), <https://orcid.org/0000-0002-3120-8109>, Vsv187@gmail.ru

Authors

Andrei A. Dmitriev — Research Engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 57215003358](https://orcid.org/0000-0002-0674-8381), <https://orcid.org/0000-0002-0674-8381>, Dmitrievaa1994@yandex.ru

Kirill V. Grebnev — Engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 57390827800](https://orcid.org/0000-0002-8461-301X), <https://orcid.org/0000-0002-8461-301X>, greb.k1r@yandex.ru

Daniil S. Smirnov — Junior Researcher, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 57204933153](https://orcid.org/0000-0003-4324-9327), <https://orcid.org/0000-0003-4324-9327>, dsmirnov@itmo.ru

Sergei V. Varzhel — PhD (Physics & Mathematics), Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 55247304200](https://orcid.org/0000-0002-3120-8109), <https://orcid.org/0000-0002-3120-8109>, Vsv187@gmail.ru

Статья поступила в редакцию 14.04.2022
Одобрена после рецензирования 02.06.2022
Принята к печати 14.07.2022

Received 14.04.2022
Approved after reviewing 02.06.2022
Accepted 14.07.2022



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

АВТОМАТИЧЕСКОЕ УПРАВЛЕНИЕ И РОБОТОТЕХНИКА AUTOMATIC CONTROL AND ROBOTICS

doi: 10.17586/2226-1494-2022-22-4-666-673

УДК 681.51

Компенсация внешних возмущений для многоканальных систем с запаздыванием в управлении

Хак Тунг Нгуен^{1✉}, Сергей Михайлович Власов², Антон Александрович Пыркин³,
Александра Викторовна Скобелева⁴

^{1,2,3,4} Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

¹ nguyenkhactunghvhq1994@gmail.com✉, <https://orcid.org/0000-0001-6430-1927>

² smvlasov@itmo.ru, <https://orcid.org/0000-0002-8345-7553>

³ a.pyrkin@gmail.com, <https://orcid.org/0000-0001-8806-4057>

⁴ skala.job@gmail.com, <https://orcid.org/0000-0002-1504-5856>

Аннотация

Предмет исследования. Проблема компенсации внешних возмущающих воздействий для линейных многоканальных объектов с запаздыванием на входе является важной и актуальной. Предложено решение данной проблемы в задачах управления динамическими объектами и в ряде других задач. **Метод.** Предлагаемый метод основан на принципе внутренней модели, где необходима идентификация параметров возмущения. На первом этапе описана схема выделения возмущения, которое представлено как синусоидальный сигнал с неизвестной частотой, амплитудой и фазой. На втором этапе решена задача идентификации частот синусоидального и мультисинусоидального сигналов. На последнем этапе разработан алгоритм стабилизации и приведения состояния объекта к нулю, осуществленный с помощью обратной связи. **Основные результаты.** Представлена новая схема компенсации внешних возмущений для линейных многоканальных объектов с запаздыванием на входе. Предложен новый алгоритм идентификации частот мультисинусоидального сигнала. Выполнен анализ возможностей предложенного метода оценивания с использованием компьютерного моделирования в среде MATLAB Simulink. **Практическая значимость.** Разработанный метод может быть эффективно применен для широкого класса прикладных задач, связанных с управлением роботами, манипуляторами различного назначения.

Ключевые слова

адаптивное управление, многоканальные системы, идентификация, внутренняя модель, запаздывание

Ссылка для цитирования: Нгуен Х.Т., Власов С.М., Пыркин А.А., Скобелева А.В. Компенсация внешних возмущений для многоканальных систем с запаздыванием в управлении // Научно-технический вестник информационных технологий, механики и оптики. 2022. Т. 22, № 4. С. 666–673. doi: 10.17586/2226-1494-2022-22-4-666-673

Compensation of external disturbances for MIMO systems with control delay

Khac T. Nguyen^{1✉}, Sergey M. Vlasov², Anton A. Pyrkin³, Aleksandra V. Skobeleva⁴

^{1,2,3,4} ITMO University, Saint Petersburg, 197101, Russian Federation

¹ nguyenkhactunghvhq1994@gmail.com✉, <https://orcid.org/0000-0001-6430-1927>

² smvlasov@itmo.ru, <https://orcid.org/0000-0002-8345-7553>

³ a.pyrkin@gmail.com, <https://orcid.org/0000-0001-8806-4057>

⁴ skala.job@gmail.com, <https://orcid.org/0000-0002-1504-5856>

Abstract

The problem of compensation of external disturbing influences for MIMO system with input delay is important and relevant. A solution to this problem is proposed in the problems of dynamic objects control and in a number of others. The proposed method is based on the principle of an internal model and requires the identification of perturbation parameters. At the first stage, a scheme for extracting a disturbance is presented which is represented as a sinusoidal signal with an unknown frequency, amplitude, and phase. At the second stage, the problem of identifying the frequencies

of a sinusoidal and multisinusoidal signal is solved. In the last stage, an algorithm for stabilizing the state of the object to zero is developed using feedback. A new scheme for compensating external disturbances for a MIMO system with an input delay is proposed. A new algorithm for identifying the frequencies of a multisinusoidal signal is proposed. The analysis of the possibilities of the proposed estimation method using computer simulation in the MATLAB Simulink environment is carried out. The developed method can be effectively applied to a wide class of applied tasks related to the control of robots and robotic manipulators for various purposes.

Keywords

adaptive control, MIMO systems, identification, internal model, delay

For citation: Nguyen K.T., Vlasov S.M., Pyrkin A.A., Skobeleva A.V. Compensation of external disturbances for MIMO systems with control delay. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2022, vol. 22, no. 4, pp. 666–673 (in Russian). doi: 10.17586/2226-1494-2022-22-4-666-673

Введение

В данной работе предложен новый алгоритм компенсации параметрически неопределенных гармонических возмущений, действующих в многоканальных системах с запаздываниями в каналах управления.

Компенсация внешнего неизвестного возмущения — одна из основных и актуальных проблем теории автоматического управления. Исследование возмущенных систем с запаздыванием в канале управления очень важно для широкого практического применения и реализации таких систем в различных областях. Большое количество работ посвящено управлению под неизвестным возмущающим воздействием [1–6]. Однако, несмотря на разнообразие методов решения и моделей объектов, проблема компенсации гармонических возмущений в нелинейном объекте с запаздыванием в канале управления не рассматривалась.

На реальную техническую установку воздействуют различные физические факторы окружающей среды. Один из таких факторов — временная задержка, которая неизбежно появляется из-за процессов разной природы в большинстве практических приложений. Например, если заводы управляются удаленно через Интернет, то данные передаются не мгновенно, а с задержкой, которая зависит от качества связи. При этом некоторые вопросы, связанные с задержкой информации, остаются открытыми и требуют дальнейших исследований. Пренебрежение этими явлениями может привести к снижению качества поведения системы или даже вызвать автоколебания с последующей потерей устойчивости. Именно поэтому проектирование методов с учетом задержек является актуальной и востребованной задачей.

В теории автоматического управления задача контроля системы с запаздыванием также важна и актуальна, и давно привлекает внимание исследователей [7–11]. Выделение объектов с запаздыванием в отдельный класс обусловлено сложностью их изучения по сравнению с объектами, которые не содержат временную задержку. Характерные особенности систем управления объектами с запаздыванием: зависимость состояния управляемого процесса от его истории, пренебрежение влиянием запаздывания, что приводит к ухудшению качества системы.

В настоящее время разработано большое количество алгоритмов компенсации внешнего возмущающего воздействия с запаздыванием. Использование идентификационного подхода для компенсации мультигар-

монических возмущений реализовано для линейных [12, 13] и нелинейных [14] систем. В работе [15] представлен алгоритм управления линейным многоканальным объектом по выходу. В компенсации внешнего возмущающего воздействия с запаздыванием может применяться метод прямого адаптивного управления [16–18], основанный на принципе внутренней модели.

В настоящей работе рассмотрена задача стабилизации линейных многоканальных объектов в условиях неизмеряемых векторов состояния, внешних синусоидальных возмущений и известного постоянного запаздывания. Параметры объекта и нижние границы частоты возмущения считаются известными, а все параметры синусоидальных возмущений (амплитуда, частота и фаза) неизвестны. Решение задачи состоит из следующих этапов: предлагается наблюдатель возмущений; выполняется идентификация частот возмущений каждого канала; происходит стабилизация состояния объекта к нулю с помощью обратной связи.

Предложена новая схема компенсации возмущений для многоканальных систем, которая улучшает качество стабилизации объединенного выходного сигнала замкнутой системы. Описан метод улучшения качества оценивания частот мультисинусоидального сигнала, и обеспечена экспоненциальная сходимости к нулю ошибок оценивания. Метод компенсации возмущения основан на методе внутренней модели и требует идентификации параметров возмущения. Отличие нового метода от работ [12–14] заключается в том, что требуется идентификация только частот мультисинусоидального возмущения.

Постановка задачи

Рассмотрим класс линейных устойчивых возмущенных объектов управления вида:

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{U}_\tau(t) + \mathbf{B}\delta(t), \quad (1)$$

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{Q}\delta(t), \quad (2)$$

где $\mathbf{x} \in \mathbb{R}^n$ — неизмеряемый вектор состояния; $\mathbf{U}_\tau = [u_1(t - \tau) \dots u_q(t - \tau)]^T \in \mathbb{R}^q$ — вектор управления с запаздываниями; $\mathbf{y}(t) = [y_1(t) \dots y_q(t)]^T \in \mathbb{R}^q$ — измеряемый вектор выхода; $\delta(t) = [\delta_1(t) \dots \delta_q(t)]^T \in \mathbb{R}^q$ — вектор неизмеряемого ограниченного внешнего возмущения; $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\mathbf{B} \in \mathbb{R}^{n \times q}$, $\mathbf{C} \in \mathbb{R}^{q \times n}$, $\mathbf{Q} \in \mathbb{R}^q$ — матрицы соответствующих размерностей; τ — известное постоянное запаздывание; q — количество системных выходов.

В качестве возмущений $\delta_1(t), \dots, \delta_q(t)$ рассмотрим сигналы

$$\begin{cases} \delta_1(t) = A_1 \sin(\omega_1 t + \varphi_1), \\ \vdots \\ \delta_q(t) = A_q \sin(\omega_q t + \varphi_q), \end{cases}$$

представленные соответственно в виде синусоид с неизвестными амплитудами A_1, \dots, A_q , частотами $\omega_1, \dots, \omega_q$ и фазовыми сдвигами $\varphi_1, \dots, \varphi_q$.

Требуется синтезировать закон управления $U(t)$, обеспечивающий стабилизацию объединенного выходного сигнала замкнутой системы

$$\lim_{t \rightarrow \infty} \|y(t)\| = 0$$

с учетом следующих допущений.

Допущение 1: Параметры $\mathbf{A}, \mathbf{B}, \mathbf{C}$ известны.

Допущение 2: Тройка матриц $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ полностью управляемая и наблюдаемая.

Допущение 3: Нижняя граница частот ω_0 известна $\omega_i \geq \omega_0, i = 1, q$.

Выделение сигнала возмущения

Так как вектор состояния \mathbf{x} недоступен для прямых измерений, опишем объект управления (1), (2) в виде соотношения вход-выход:

$$\mathbf{y} = \mathbf{W}(s)[\mathbf{U}_\tau + \boldsymbol{\delta}], \quad (3)$$

где $\mathbf{W}(s) = \mathbf{C}(\mathbf{sI} - \mathbf{A})^{-1}\mathbf{B} = \frac{\beta_{ij}(s)}{\alpha(s)}$ — передаточная функция системы; \mathbf{I} — единичная матрица; $\alpha(s)$ и $\beta_{ij}(s)$ полиномы порядка n и γ соответственно с известными постоянными коэффициентами, такие что

$$\beta_{ij}(s) = b_{ij}s^\gamma + b_{\gamma-1ij}s^{\gamma-1} + \dots + b_{1ij}s + b_{0ij},$$

$$i = \overline{1, q}, j = \overline{1, q},$$

$$\alpha(s) = s^n + a_{n-1}s^{n-1} + \dots + a_1s + a_0,$$

где γ — степень числителя передаточной функции системы.

$$\frac{\beta_{ij}(s)}{\alpha(s)} = \begin{bmatrix} \frac{\beta_{11}(s)}{\alpha(s)} & \dots & \frac{\beta_{1q}(s)}{\alpha(s)} \\ \vdots & \ddots & \vdots \\ \frac{\beta_{q1}(s)}{\alpha(s)} & \dots & \frac{\beta_{qq}(s)}{\alpha(s)} \end{bmatrix}.$$

Перепишем систему (3) в виде:

$$\mathbf{y} = \left[\frac{\beta_{11}(s)}{\alpha(s)} [u_1(t - \tau) + \delta_1] + \dots + \frac{\beta_{iq}(s)}{\alpha(s)} [u_q(t - \tau) + \delta_q] \right].$$

Рассматриваемый объект управления является устойчивым. Применим фильтры следующего вида

$$\bar{\mathbf{y}} = \left[\frac{\beta_{11}(s)}{\alpha(s)} [u_1(t - \tau)] + \dots + \frac{\beta_{iq}(s)}{\alpha(s)} [u_q(t - \tau)] \right].$$

Рассмотрим расхождение по выходу

$$\tilde{\mathbf{y}} = \bar{\boldsymbol{\delta}} = \mathbf{y} - \bar{\mathbf{y}}$$

или

$$\bar{\boldsymbol{\delta}} = \mathbf{W}(s)[\boldsymbol{\delta}], \quad (4)$$

где $i = \overline{1, q}$.

Перепишем уравнение (4) в виде:

$$\bar{\boldsymbol{\delta}}(s) = \begin{bmatrix} \frac{\beta_{11}(s)}{\alpha(s)} \delta_1 + \dots + \frac{\beta_{1q}(s)}{\alpha(s)} \delta_q \\ \vdots \\ \frac{\beta_{q1}(s)}{\alpha(s)} \delta_1 + \dots + \frac{\beta_{qq}(s)}{\alpha(s)} \delta_q \end{bmatrix}.$$

В некоторых случаях для удобства используем эквивалентное представление возмущающего сигнала

$$\begin{cases} \bar{\delta}_1(t) = \bar{A}_{11} \sin(\omega_1 t + \bar{\varphi}_{11}) + \dots + \bar{A}_{1q} \sin(\omega_q t + \bar{\varphi}_{1q}) \\ \vdots \\ \bar{\delta}_q(t) = \bar{A}_{q1} \sin(\omega_1 t + \bar{\varphi}_{q1}) + \dots + \bar{A}_{qq} \sin(\omega_q t + \bar{\varphi}_{qq}). \end{cases}$$

Синтез наблюдателей возмущения

Каждая составляющая $\bar{\delta}_i$ возмущения может быть смоделирована как выход линейных генераторов [19]

$$\begin{aligned} \bar{\delta}_i &= \mathbf{h}_i^T \boldsymbol{\xi}_i, \\ \dot{\boldsymbol{\xi}}_i &= \mathbf{\Gamma}_i \boldsymbol{\xi}_i, \end{aligned}$$

где $\boldsymbol{\xi}_i \in \mathbb{R}^{q_i}$ — матрицы состояния генераторов с начальными значениями $\boldsymbol{\xi}_i(0)$; $\mathbf{\Gamma}_i$ — матрица постоянных коэффициентов; \mathbf{h}_i — векторы соответствующей размерности. Матрицы $\mathbf{\Gamma}_i$ неизвестны, но известны размеры генераторов $q_i, i = 1, q$.

Рассмотрим наблюдатели внешних возмущений [20]

$$\dot{\boldsymbol{\psi}}_i(t) = \mathbf{G}\boldsymbol{\psi}_i(t) + \mathbf{I}\bar{\boldsymbol{\delta}}_i(t), \quad i = \overline{1, q}, \quad (5)$$

где $\boldsymbol{\psi}_i \in \mathbb{R}^m$, $\mathbf{G} \in \mathbb{R}^{m \times m}$ — произвольная гурвицева матрица, образующая с $\mathbf{I} \in \mathbb{R}^m$ постоянный вектор, выбираемый по условию, что пара (\mathbf{G}, \mathbf{I}) — полностью управляема.

Сигналы возмущения $\bar{\delta}_i$ могут быть представлены в виде линейной регрессионной модели [20, 21]

$$\bar{\delta}_i = \boldsymbol{\psi}_i^T \boldsymbol{\zeta}_i, \quad (6)$$

где $\boldsymbol{\zeta}_i$ — векторы неизвестных постоянных коэффициентов, зависящих от параметров матрицы от матриц $\mathbf{\Gamma}_i$ (в конечном итоге — от частот мультисинусоидальных сигналов), векторы $\boldsymbol{\psi}_i$ формируются наблюдателями (5), $i = 1, q$.

Идентификация частот возмущений

Рассмотрим метод оценки частот несмещенных гармонических и мультигармонических сигналов с постоянными параметрами.

Заметим, что в матрице $\bar{\boldsymbol{\delta}}_i(t)$ каждым сигналом возмущения является сумма всех синусоидальных сиг-

налов от каждого канала системы с одинаковыми частотами. Тогда задача заключается в идентификации частот мультисинусоидального сигнала. Наблюдатели для возмущающих сигналов $\bar{\delta}_i$ могут быть синтезированы для каждой из подсистем отдельно и имеют одинаковый вид. В этой связи рассмотрим только первую подсистему $\bar{\delta}_1$.

Сигнал возмущения $\bar{\delta}_1$ перепишем в виде:

$$\bar{\delta}_1 = \sum_{i=1}^q \bar{A}_i \sin(\omega_i t + \bar{\varphi}_i). \quad (7)$$

Генератор сигналов вида (7) опишем дифференциальным уравнением

$$(p^2 + \omega_1^2)(p^2 + \omega_2^2) \dots (p^2 + \omega_q^2) \bar{\delta}_1(t) = 0, \quad (8)$$

где $p = \frac{d}{dt}$ — оператор дифференцирования, $i = \overline{1, q}$.

С учетом выражения (8) получим

$$p^{2q} \bar{\delta}_1(t) + \theta_1 p^{2q-1} \bar{\delta}_1(t) + \dots + \theta_{q-1} p^2 \bar{\delta}_1(t) + \theta_q \bar{\delta}_1(t) = 0,$$

где

$$\begin{cases} \theta_1 = \omega_1^2 + \omega_2^2 + \dots + \omega_q^2, \\ \theta_2 = \omega_1^2 \omega_2^2 + \omega_1^2 \omega_3^2 + \dots + \omega_{q-1}^2 \omega_q^2, \\ \dots \\ \theta_q = \omega_1^2 \omega_2^2 \dots \omega_q^2. \end{cases}$$

Сигнал $\bar{\delta}_1(t)$ представим как выход линейного генератора конечной размерности [19]

$$\begin{aligned} \bar{\delta}_1(t) &= \mathbf{h}_1^T \xi_1(t), \\ \dot{\xi}_1(t) &= \mathbf{\Gamma}_1 \xi_1(t), \\ \mathbf{\Gamma}_1 &= \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ -\theta_q & 0 & -\theta_{q-1} & \dots & 0 \end{bmatrix}, \\ \mathbf{h}_1 &= [1 \ 0 \ 0 \ \dots \ 0]. \end{aligned}$$

Рассмотрим сигнал (7) при $q=1$:

$$\bar{\delta}_1(t) = \bar{A}_1 \sin(\omega_1 t + \bar{\varphi}_1). \quad (9)$$

Сигнал (9) представим в виде [22]

$$\bar{\delta}_1(t-2d) + \bar{\delta}_1(t) = 2\cos(\omega_1 d) \bar{\delta}_1(t-d), \quad (10)$$

где $d > 0$ — оператор запаздывания.

Рассмотрим сигнал (7) при $q \in \mathbb{N}$ гармоник

$$\bar{\delta}_1(t) = \sum_{i=1}^q \bar{A}_i \sin(\omega_i t + \bar{\varphi}_i) \quad (11)$$

и дискретный аналог сигнала (11)

$$\bar{\delta}_1[k] = \sum_{i=1}^q \bar{A}_i \sin(\omega_i kT + \bar{\varphi}_i),$$

где $T > 0$ — период дискретизации сигнала, $t = kT$, $k = 0, 1, \dots$

Получим аналогичный результат для непрерывного сигнала (11). Для этого применим оператор запаздывания d как и в (10), к измеряемому сигналу (11). Сигналы с кратным d запаздыванием, в силу ассоциативного закона, могут быть записаны с использованием оператора запаздывания (11) как

$$\begin{cases} \bar{\delta}_1(t-d) = \Omega \bar{\delta}_1(t), \\ \bar{\delta}_1(t-2d) = \Omega^2 \bar{\delta}_1(t), \\ \vdots \\ \bar{\delta}_1(t-qd) = \Omega^q \bar{\delta}_1(t), \end{cases} \quad (12)$$

где Ω — оператор запаздывания:

$$[\Omega(\cdot)] = \begin{cases} 0, & t < d, \\ (\cdot), & (t-d), t \geq d. \end{cases}$$

С учетом (12) перепишем уравнение (10) в виде

$$(\Omega^2 - 2\cos\omega_1 d \Omega + 1) \bar{\delta}_1(t) = 0. \quad (13)$$

Утверждение. Для сигнала, представленного как сумма q гармоник, справедливо выражение:

$$(\Omega^2 - 2\Omega c_1 + 1) \cdot \dots \cdot (\Omega^2 - 2\Omega c_q + 1) \bar{\delta}_1(t) = 0, \quad (14)$$

где $c_i = \cos\omega_i d$, $i = \overline{1, q}$.

Доказательство. Для доказательства Утверждения используем метод математической индукции.

При $q = 1$ выражение (14) выполняется согласно (13).

Предположим, что уравнение (14) решается при $q = k$, тогда

$$(\Omega^2 - 2\Omega c_1 + 1) \cdot \dots \cdot (\Omega^2 - 2\Omega c_k + 1) \bar{\delta}_1^k(t) = 0. \quad (15)$$

Следовательно, необходимо доказать, что уравнение (14) может быть решено при $q = k+1$, тогда получим:

$$\begin{aligned} (\Omega^2 - 2\Omega c_1 + 1) \cdot \dots \cdot (\Omega^2 - 2\Omega c_{k+1} + 1) \bar{\delta}_1^{k+1}(t) &= \\ = (\Omega^2 - 2\Omega c_1 + 1) \cdot \dots \cdot (\Omega^2 - 2\Omega c_k + 1) \bar{\delta}_1^k(t) + & (16) \\ + (\Omega^2 - 2\Omega c_1 + 1) \cdot \dots \cdot (\Omega^2 - 2\Omega c_k + 1) \beta_{k+1}(t). \end{aligned}$$

Из уравнений (15) и (16) имеем

$$\begin{aligned} (\Omega^2 - 2\Omega c_1 + 1) \cdot \dots \cdot (\Omega^2 - 2\Omega c_{k+1} + 1) \bar{\delta}_1^{k+1}(t) &= \\ = (\Omega^2 - 2\Omega c_1 + 1) \cdot \dots \cdot (\Omega^2 - 2\Omega c_k + 1) \beta_{k+1}(t), \end{aligned}$$

где $\beta_{k+1} = \bar{A}_{k+1} \sin(\omega_{k+1} t + \bar{\varphi}_{k+1})$.

Заметим, что сигнал β_{k+1} можно рассматривать в соответствии с одной синусоидой (9), тогда

$$(\Omega^2 - 2\Omega c_{k+1} + 1) \beta_{k+1} = 0. \quad (17)$$

Применим оператор $(\Omega^2 - 2\Omega c_{k+1} + 1) \beta_{k+1}$ к (16) и получим

$$\begin{aligned} (\Omega^2 - 2\Omega c_1 + 1) \cdot \dots \cdot (\Omega^2 - 2\Omega c_{k+1} + 1) \bar{\delta}_1^{k+1}(t) &= \\ = (\Omega^2 - 2\Omega c_1 + 1) \cdot \dots \cdot (\Omega^2 - 2\Omega c_k + 1) \times & (18) \\ \times (\Omega^2 - 2\Omega c_{k+1} + 1) \beta_{k+1}. \end{aligned}$$

С учетом выражений (17) и (18) получим

$$(\Omega^2 - 2\Omega c_1 + 1) \cdot \dots \cdot (\Omega^2 - 2\Omega c_{k+1} + 1) \bar{\delta}_1^{k+1}(t) = 0.$$

Утверждение доказано.

На основе уравнения (14) построим регрессионную модель для сигнала с q количеством гармоник

$$\Xi(t) = \chi^T(t) \varpi, \quad (19)$$

где $\Xi \in \mathbb{R}^1$ — зависимая функция, $\chi = [\chi_1 \ \chi_2 \ \dots \ \chi_q]^T \in \mathbb{R}^q$ — регрессор, $\varpi = [\varpi_1 \ \varpi_2 \ \dots \ \varpi_q]^T \in \mathbb{R}^q$ — вектор неизвестных параметров.

Или, более конкретно,

$$(\Omega^2 + 1)^q \bar{\delta}_1(t) = \chi_1 \varpi_1(t) + \chi_2 \varpi_2(t) + \dots + \chi_q \varpi_q(t).$$

Компоненты уравнения (19):

— $\Xi(t)$ с использованием биннома Ньютона:

$$\Xi(t) = (\Omega^2 + 1)^q \bar{\delta}_1(t).$$

— Вектор неизвестных параметров ϖ_i связаны с c_i формулами Виета:

$$\begin{cases} \varpi_1 = c_1 + c_2 + \dots + c_q, \\ \varpi_2 = -c_1 c_2 - c_1 c_3 - \dots - c_{q-1} c_q, \\ \vdots \\ \varpi_q = (-1)^{q+1} c_1 c_2 \cdot \dots \cdot c_q. \end{cases}$$

— Регрессор $\chi_i(t)$:

$$\begin{cases} \chi_1(t) = 2\Omega(\Omega^2 + 1)^{q-1} \bar{\delta}_1(t), \\ \chi_2(t) = 2^2 \Omega^2 (\Omega^2 + 1)^{q-2} \bar{\delta}_1(t), \\ \vdots \\ \chi_q(t) = 2^q \Omega^q \bar{\delta}_1(t). \end{cases}$$

Алгоритм оценивания параметров модели (19)

Для оценивания параметров вектора ϖ модели (19) используется метод динамического расширения регрессора (DREM) [23].

Применим блок запаздывания v_i , $i = \overline{1, q-1}$ для известных элементов регрессионной модели (19):

$$\Xi(t - v_i) = \chi^T(t - v_i) \varpi.$$

Обозначим

$$\vartheta_e = \eta_e \varpi, \quad (20)$$

где

$$\begin{aligned} \vartheta_e &= [\Xi(t) \ \Xi(t - v_1) \ \dots \ \Xi(t - v_{q-1})]^T, \\ \eta_e &= [\chi_1^T(t) \ \chi_1^T(t - v_1) \ \dots \ \chi_1^T(t - v_{q-1})]. \end{aligned}$$

Умножив (20) на $\text{adj}(\eta_e(t))$ получим

$$\vartheta(t) = \Delta(t) \varpi, \quad (21)$$

где $\Delta(t) = \det(\eta_e(t)) \in \mathbb{R}^1$, $\vartheta(t) = \text{adj}(\eta_e) \vartheta_e(t) \in \mathbb{R}^q$.

Запишем уравнение (21) покомпонентно:

$$\vartheta_i(t) = \Delta(t) \varpi_i, \quad i = \overline{1, q}.$$

Представим алгоритм оценивания параметров ϖ_i в виде:

$$\hat{\varpi}_i(t) = -\kappa_i \Delta(t) (\vartheta_i(t) - \Delta(t) \hat{\varpi}_i(t)),$$

где κ_i — настраиваемый параметр, $i = \overline{1, q}$.

Чтобы получить оценку за конечное время, заменяем ошибку оценки $\hat{\varpi}_i(t)$ по определению на $\varpi_i - \hat{\varpi}_i(t)$:

$$\varpi_i - \hat{\varpi}_i(t) = \varpi_i E(t) - \hat{\varpi}_i(0) E(t), \quad (22)$$

где $\dot{E}(t) = -\kappa_i \Delta^2(t) E(t)$, $E(0) = 1$ или $E(t) = e^{-\kappa_i \int_0^t \Delta^2(s) ds}$.

Выразим явно значение параметра ϖ_i из соотношения (22)

$$\hat{\varpi}_i^{fi}(t) = \frac{\hat{\varpi}_i(t) - E(t) \hat{\varpi}_i(0)}{1 - E(t)}.$$

Алгоритм оценивания частот

Для оценки частоты применим функцию $\arccos(\cdot)$, основанную на параметре $\hat{c}_i^{fi}(t)$:

$$\hat{\omega}_i^{fi}(t) = \frac{1}{d} \arccos(\hat{c}_i^{fi}(t)), \quad i = \overline{1, q}.$$

Заметим, что матрица Γ_1 зависит от частот мультигармонического сигнала $\bar{\delta}_1(t)$.

Тогда оценка матрицы $\hat{\Gamma}_1$ имеет вид:

$$\hat{\Gamma}_1 = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ -\hat{\theta}_q & 0 & -\hat{\theta}_{q-1} & \dots & 0 \end{bmatrix},$$

где

$$\begin{cases} \hat{\theta}_1 = \hat{\omega}_1^2 + \hat{\omega}_2^2 + \dots + \hat{\omega}_q^2, \\ \hat{\theta}_2 = \hat{\omega}_1^2 \hat{\omega}_2^2 + \hat{\omega}_1^2 \hat{\omega}_3^2 + \dots + \hat{\omega}_{q-1}^2 \hat{\omega}_q^2, \\ \vdots \\ \hat{\theta}_q = \hat{\omega}_1^2 \hat{\omega}_2^2 \cdot \dots \cdot \hat{\omega}_q^2. \end{cases}$$

Аналогично найдем $\hat{\Gamma}_i$, $i = \overline{1, q}$ для остальных каналов.

Синтез закона управления

С учетом выражения (6) наблюдатель (5) может быть представлен как автономная модель

$$\dot{\Psi}_i = (G + I \zeta_i^T) \Psi_i. \quad (23)$$

Однако векторы состояния Ψ_i неизмеримы, потому что матрицы ζ_i неизвестны. Выполним оценку векторов состояния возмущения $\hat{\Psi}_i$ с помощью оценки векторов $\hat{\zeta}_i$, при этом применим следующее уравнение

$$\hat{\zeta}_i = \mathbf{h}_i^T \mathbf{M}_i^{-1},$$

где \mathbf{M}_i — решение уравнения

$$\mathbf{M}_i \hat{\mathbf{\Gamma}}_i - \mathbf{G} \mathbf{M}_i = \mathbf{l} \mathbf{h}_i^T,$$

где $\hat{\mathbf{\Gamma}}_i$ — оценка матриц $\mathbf{\Gamma}_i$, $i = \overline{1, q}$.

Так как объект работает в условиях запаздывания, на основе фундаментального решения автономной модели (23) определим будущие значения вектора состояния наблюдателя

$$\boldsymbol{\psi}_i(t + \tau) = \exp\{(\mathbf{G} + \mathbf{l} \hat{\zeta}_i^T) \tau\} \boldsymbol{\psi}_i(t). \quad (24)$$

На основе уравнений (6) и (24) получим:

$$\hat{\delta}_i(t + \tau) = \hat{\rho}_i^T \boldsymbol{\psi}_i(t),$$

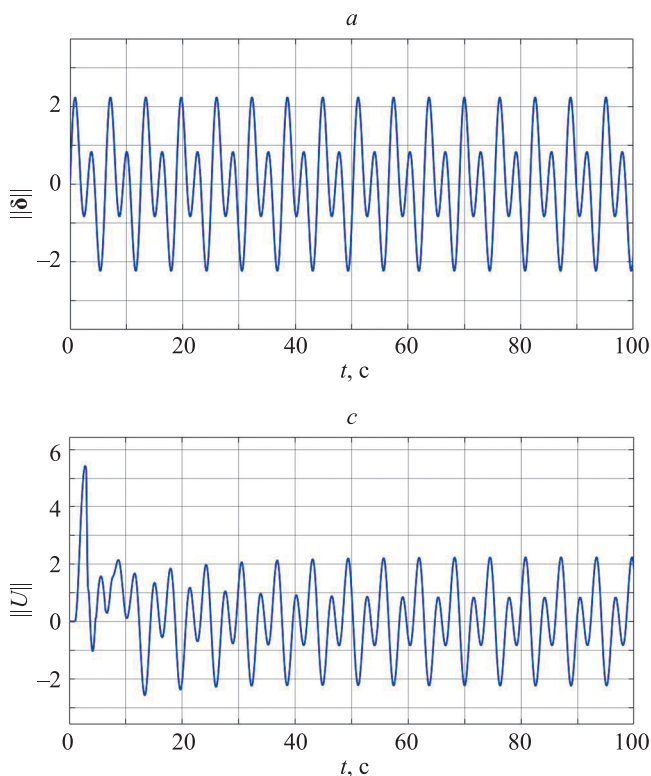
где $\hat{\rho}_i^T = \hat{\zeta}_i^T \exp\{(\mathbf{G} + \mathbf{l} \hat{\zeta}_i^T) \tau\}$.

Структуру настраиваемого регулятора определим в соответствии с принципом непосредственной компенсации:

$$\mathbf{U}(t) = -\hat{\delta}(t + \tau) = \frac{-1}{\mathbf{W}(s)} \hat{\delta}(t). \quad (25)$$

С учетом выражений (4) и (24) закон управления (25) можно записать

$$\mathbf{U}(t) = -\frac{1}{\mathbf{W}(s)} \hat{\delta}(t) = -\frac{1}{\mathbf{W}(s)} \begin{bmatrix} \hat{\delta}_1(t + \tau) \\ \vdots \\ \hat{\delta}_q(t + \tau) \end{bmatrix} = -\frac{1}{\mathbf{W}(s)} \begin{bmatrix} \hat{\rho}_1^T \boldsymbol{\psi}_1(t) \\ \vdots \\ \hat{\rho}_q^T \boldsymbol{\psi}_q(t) \end{bmatrix}.$$



Матричная передаточная функция $\boldsymbol{\beta}^{-1}(s)$ является инверсией функции $\boldsymbol{\beta}(s)$ уравнения объекта управления в виде соотношения вход-выход (3).

Математическое моделирование

Приведем результаты численного моделирования, иллюстрирующие эффективность предложенного алгоритма оценивания частоты несмещенного гармонического сигнала с постоянными параметрами. Моделирование выполнено с помощью программной среды MATLAB Simulink.

Рассмотрим объект управления второго порядка с двумя входными и двумя выходными каналами

$$\mathbf{y} = \mathbf{W}(s)[\mathbf{U}_\tau + \boldsymbol{\delta}],$$

где

$$\mathbf{W}(s) = \begin{bmatrix} \frac{3s^2 + 5s + 10}{s^2 + 2s + 1} & \frac{s^2 + 3s + 7}{s^2 + 2s + 1} \\ \frac{2s^2 + 2s + 1}{s^2 + 2s + 1} & \frac{s^2 + 5s + 3}{s^2 + 2s + 1} \end{bmatrix}, \tau = 1 \text{ c.}$$

Пусть возмущающее воздействие имеет вид:

$$\boldsymbol{\delta} = [2\sin(t) \quad 3\sin(2t)]^T.$$

Векторы $\boldsymbol{\psi}_1$ и $\boldsymbol{\psi}_2$ являются состояниями наблюдателей (5) со следующими параметрами:

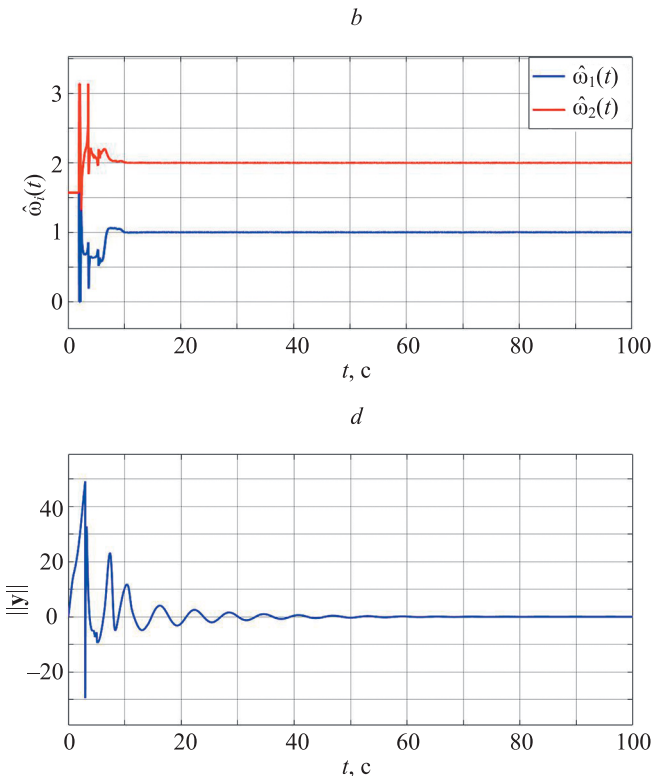


Рисунок. Временные диаграммы: возмущений $\boldsymbol{\delta}(t)$ (a); оценки частот $\hat{\omega}_1(t)$, $\hat{\omega}_2(t)$ (b); сигнала управления $\mathbf{U}(t)$ (c); переходные процессы для выхода $\mathbf{y}(t)$ (d)

Figure. Time diagrams of disturbances $\boldsymbol{\delta}(t)$ (a), frequency estimates $\hat{\omega}_1(t)$, $\hat{\omega}_2(t)$ (b), control signal $\mathbf{U}(t)$ (c), transients for output $\mathbf{y}(t)$ (d)

$$\mathbf{G} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -5 & -9 & -7 & -8 \end{bmatrix}, \mathbf{I} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Параметр запаздывания для параметризации сигнала возмущения: $d = 0,1$ с.

Параметры алгоритма DREM: $\nu = 0,05$, $\kappa_j = 1$.

В результате моделирования сформирована временная диаграмма функции возмущения $\delta(t)$ (рисунок, *a*), переходные процессы для оценки частот $\hat{\omega}_1(t) = 1$, $\hat{\omega}_2(t) = 2$ (рисунок, *b*) и для сигнала управления $U(t)$ (рисунок, *c*). На рисунке, *d* продемонстрировано, что все переменные выходных системы асимптотически стремятся к нулю.

Графики переходных процессов демонстрируют ограниченность всех сигналов в системе (рисунок, *a-d*). Алгоритм управления (25) позволяет сохранить устой-

чивость в многоканальных системах с входным запаздыванием в условиях внешних возмущений.

Заключение

Предложена новая схема компенсации внешних возмущающих воздействий в многоканальных системах с входным запаздыванием в условиях внешних возмущений. Схема обеспечивает ограниченность всех сигналов в системе и сходимости регулируемых переменных к нулю. Предложен новый подход к идентификации частот мультиусинусоидального сигнала. Показана экспоненциальная сходимость оценивания частот мультиусинусоидального сигнала к истинными значениям. Подход может быть расширен в случае выбора объекта с неизвестными параметрами и запаздыванием, что возможно в качестве направления дальнейших исследований.

Литература

- Bobtsov A., Kremlev A. Adaptive compensation of biased sinusoidal disturbances with unknown frequency // *IFAC Proceedings Volumes*. 2005. V. 38. N 1. P. 131–136. <https://doi.org/10.3182/20050703-6-CZ-1902.00022>
- Nikiforov V.O. Adaptive nonlinear servocompensation of unknown external disturbances // *IFAC Proceedings Volumes*. 1999. V. 32. N 2. P. 4506–4511. [https://doi.org/10.1016/S1474-6670\(17\)56768-3](https://doi.org/10.1016/S1474-6670(17)56768-3)
- Bobtsov A., Pyrkin A., Titov A., Kremlev A. Compensation of harmonic disturbance for nonlinear plant with parametric and functional uncertainty // *IFAC Proceedings Volumes*. 2011. V. 44. N 1. P. 1528–1533. <https://doi.org/10.3182/20110828-6-IT-1002.02324>
- Nikiforov V. Adaptive non-linear tracking with complete compensation of unknown disturbances // *European Journal of Control*. 1998. V. 4. N 2. P. 132–139. [https://doi.org/10.1016/S0947-3580\(98\)70107-4](https://doi.org/10.1016/S0947-3580(98)70107-4)
- Nikiforov V.O. Adaptive servocompensation of input disturbances // *IFAC Proceedings Volumes*. 1996. V. 29. N 1. P. 5114–5119. [https://doi.org/10.1016/S1474-6670\(17\)58492-X](https://doi.org/10.1016/S1474-6670(17)58492-X)
- Wang J., Vedyakov A., Vediakova A., Pyrkin A., Bobtsov A., Shavetov S. Output adaptive controller for a class of MIMO systems with input delay and multisinusoidal disturbance // *IFAC-PapersOnLine*. 2015. V. 48. N 11. P. 892–899. <https://doi.org/10.1016/j.ifacol.2015.09.304>
- Olbro A. Stabilizability, detectability, and spectrum assignment for linear autonomous systems with general time delays // *IEEE Transactions on Automatic Control*. 1978. V. 23. N 5. P. 887–890. <https://doi.org/10.1109/TAC.1978.1101879>
- Manitius A., Olbro A. Finite spectrum assignment problem for systems with delays // *IEEE Transactions on Automatic Control*. 1979. V. 24. N 4. P. 541–552. <https://doi.org/10.1109/TAC.1979.1102124>
- Anderson R., Spong M. Bilateral control of teleoperators with time delay // *IEEE Transactions on Automatic Control*. 1989. V. 34. N 5. P. 494–501. <https://doi.org/10.1109/9.24201>
- Krstic M., Smyshlyaev A. Backstepping boundary control for first-order hyperbolic PDEs and application to systems with actuator and sensor delays // *Systems and Control Letters*. 2008. V. 57. N 9. P. 750–758. <https://doi.org/10.1016/j.sysconle.2008.02.005>
- Bresch-Pietri D., Krstic M. Adaptive trajectory tracking despite unknown input delay and plant parameters // *Automatica*. 2009. V. 45. N 9. P. 2074–2081. <https://doi.org/10.1016/j.automatica.2009.04.027>
- Pyrkin A.A., Bobtsov A.A., Nikiforov V.O., Vedyakov A.A., Kolyubin S.A., Borisov O.I. Output control approach for delayed linear systems with adaptive rejection of multiharmonic disturbance // *IFAC Proceedings Volumes*. 2014. V. 43. N 3. P. 12110–12115. <https://doi.org/10.3182/20140824-6-ZA-1003.01787>
- Пыркин А.А., Бобцов А.А., Никифоров В.О., Колубин С.А., Ведяков А.А., Борисов О.И., Громов В.С. Компенсация полигар-

References

- Bobtsov A., Kremlev A. Adaptive compensation of biased sinusoidal disturbances with unknown frequency. *IFAC Proceedings Volumes*, 2005, vol. 38, no. 1, pp. 131–136. <https://doi.org/10.3182/20050703-6-CZ-1902.00022>
- Nikiforov V.O. Adaptive nonlinear servocompensation of unknown external disturbances. *IFAC Proceedings Volumes*, 1999, vol. 32, no. 2, pp. 4506–4511. [https://doi.org/10.1016/S1474-6670\(17\)56768-3](https://doi.org/10.1016/S1474-6670(17)56768-3)
- Bobtsov A., Pyrkin A., Titov A., Kremlev A. Compensation of harmonic disturbance for nonlinear plant with parametric and functional uncertainty. *IFAC Proceedings Volumes*, 2011, vol. 44, no. 1, pp. 1528–1533. <https://doi.org/10.3182/20110828-6-IT-1002.02324>
- Nikiforov V. Adaptive non-linear tracking with complete compensation of unknown disturbances. *European Journal of Control*, 1998, vol. 4, no. 2, pp. 132–139. [https://doi.org/10.1016/S0947-3580\(98\)70107-4](https://doi.org/10.1016/S0947-3580(98)70107-4)
- Nikiforov V.O. Adaptive servocompensation of input disturbances. *IFAC Proceedings Volumes*, 1996, vol. 29, no. 1, pp. 5114–5119. [https://doi.org/10.1016/S1474-6670\(17\)58492-X](https://doi.org/10.1016/S1474-6670(17)58492-X)
- Wang J., Vedyakov A., Vediakova A., Pyrkin A., Bobtsov A., Shavetov S. Output adaptive controller for a class of MIMO systems with input delay and multisinusoidal disturbance. *IFAC-PapersOnLine*, 2015, vol. 48, no. 11, pp. 892–899. <https://doi.org/10.1016/j.ifacol.2015.09.304>
- Olbro A. Stabilizability, detectability, and spectrum assignment for linear autonomous systems with general time delays. *IEEE Transactions on Automatic Control*, 1978, vol. 23, no. 5, pp. 887–890. <https://doi.org/10.1109/TAC.1978.1101879>
- Manitius A., Olbro A. Finite spectrum assignment problem for systems with delays. *IEEE Transactions on Automatic Control*, 1979, vol. 24, no. 4, pp. 541–552. <https://doi.org/10.1109/TAC.1979.1102124>
- Anderson R., Spong M. Bilateral control of teleoperators with time delay. *IEEE Transactions on Automatic Control*, 1989, vol. 34, no. 5, pp. 494–501. <https://doi.org/10.1109/9.24201>
- Krstic M., Smyshlyaev A. Backstepping boundary control for first-order hyperbolic PDEs and application to systems with actuator and sensor delays. *Systems and Control Letters*, 2008, vol. 57, no. 9, pp. 750–758. <https://doi.org/10.1016/j.sysconle.2008.02.005>
- Bresch-Pietri D., Krstic M. Adaptive trajectory tracking despite unknown input delay and plant parameters. *Automatica*, 2009, vol. 45, no. 9, pp. 2074–2081. <https://doi.org/10.1016/j.automatica.2009.04.027>
- Pyrkin A.A., Bobtsov A.A., Nikiforov V.O., Vedyakov A.A., Kolyubin S.A., Borisov O.I. Output control approach for delayed linear systems with adaptive rejection of multiharmonic disturbance. *IFAC Proceedings Volumes*, 2014, vol. 43, no. 3, pp. 12110–12115. <https://doi.org/10.3182/20140824-6-ZA-1003.01787>

- монического возмущения, действующего на состояние и выход линейного объекта с запаздыванием в канале управления // Автоматика и телемеханика. 2015. № 12. С. 43–64.
14. Pyrkina A.A., Bobtsov A.A., Chepinskiy S.A., Kapitanyuk Y.A. Compensation of unknown multiharmonic disturbance for nonlinear plant with delay in control // IFAC Proceedings Volumes. 2010. V. 43. N 14. P. 481–486. <https://doi.org/10.3182/20100901-3-IT-2016.00065>
 15. Wang J., Vedyakov A.A., Vediakova A.O., Pyrkina A.A., Bobtsov A.A., Shavetov S.V. Output adaptive controller for a class of MIMO Systems with input delay and multisinusoidal disturbance // IFAC-PapersOnLine. 2015. V. 48. N 11. P. 892–899. <https://doi.org/10.1016/j.ifacol.2015.09.304>
 16. Nikiforov V.O. Adaptive compensation of external disturbances // Control of Complex Systems: Preprint 125 / ed. by A.L. Fradkov, A.A. Stotsky. St. Petersburg: IPME, 1995. P. 34–40.
 17. Johnson C.D. Accommodation of external disturbances in linear regulator and servomechanism problems // IEEE Transactions on Automatic Control. 1971. V. 16. N 6. P. 635–644. <https://doi.org/10.1109/TAC.1971.1099830>
 18. Francis D.A., Wonham W.M. The internal model principle for linear multivariable regulators // Applied Mathematics and Optimization. 1975. V. 2. N 2. P. 170–194. <https://doi.org/10.1007/BF01447855>
 19. Nikiforov V.O. Adaptive servomechanism controller with an implicit reference model // International Journal of Control. 1997. V. 68. N 2. P. 277–286. <https://doi.org/10.1080/002071797223604>
 20. Никифоров В.О. Адаптивное и робастное управление с компенсацией внешних возмущений. СПб.: Наука, 2003. 282 с.
 21. Никифоров В.О. Наблюдатели внешних детерминированных возмущений I. Объекты с известными параметрами // Автоматика и телемеханика. 2004. № 10. С. 13–24.
 22. Нгуен Х.Т., Власов С.М., Пыркин А.А., Попков И.В. Новый алгоритм идентификации частоты синусоидального сигнала с постоянными параметрами // Научно-технический вестник информационных технологий, механики и оптики. 2022. Т. 22. № 1. С. 18–24. <https://doi.org/10.17586/2226-1494-2022-22-1-18-24>
 23. Aranovskiy S., Bobtsov A., Ortega R., Pyrkina A. Performance enhancement of parameter estimators via dynamic regressor extension and mixing // IEEE Transactions on Automatic Control. 2017. V. 62. N 7. P. 3546–3550. <https://doi.org/10.1109/TAC.2016.2614889>
 13. Pyrkina A.A., Bobtsov A.A., Nikiforov V.O., Kolyubin S.A., Vedyakov A.A., Borisov O.I., Gromov V.S. Compensation of polyharmonic disturbance of state and output of a linear plant with delay in the control channel. *Automation and Remote Control*, 2015, vol. 76, no. 12, pp. 2124–2142. <https://doi.org/10.1109/TAC.2016.2614889>
 14. Pyrkina A.A., Bobtsov A.A., Chepinskiy S.A., Kapitanyuk Y.A. Compensation of unknown multiharmonic disturbance for nonlinear plant with delay in control. *IFAC Proceedings Volumes*, 2010, vol. 43, no. 14, pp. 481–486. <https://doi.org/10.3182/20100901-3-IT-2016.00065>
 15. Wang J., Vedyakov A.A., Vediakova A.O., Pyrkina A.A., Bobtsov A.A., Shavetov S.V. Output adaptive controller for a class of MIMO Systems with input delay and multisinusoidal disturbance. *IFAC-PapersOnLine*, 2015, vol. 48, no. 11, pp. 892–899. <https://doi.org/10.1016/j.ifacol.2015.09.304>
 16. Nikiforov V.O. Adaptive compensation of external disturbances. *Control of Complex Systems: Preprint 125*. Ed. by A.L. Fradkov, A.A. Stotsky. St. Petersburg, IPME, 1995, pp. 34–40.
 17. Johnson C.D. Accommodation of external disturbances in linear regulator and servomechanism problems. *IEEE Transactions on Automatic Control*, 1971, vol. 16, no. 6, pp. 635–644. <https://doi.org/10.1109/TAC.1971.1099830>
 18. Francis D.A., Wonham W.M. The internal model principle for linear multivariable regulators. *Applied Mathematics and Optimization*, 1975, vol. 2, no. 2, pp. 170–194. <https://doi.org/10.1007/BF01447855>
 19. Nikiforov V.O. Adaptive servomechanism controller with an implicit reference model. *International Journal of Control*, 1997, vol. 68, no. 2, pp. 277–286. <https://doi.org/10.1080/002071797223604>
 20. Nikiforov V.O. *Adaptive and Robust Control with Compensation of the External Disturbances*. St. Petersburg, Nauka Publ., 2003, 282 p. (in Russian)
 21. Nikiforov V.O. Observers of external deterministic disturbances. I. Objects with known parameters. *Automation and Remote Control*, 2004, vol. 65, no. 10, pp. 1531–1541. <https://doi.org/10.1023/B:AURC.0000044264.74470.48>
 22. Nguyen Kh.T., Vlasov S.M., Pyrkina A.A., Popkov I.V. A new algorithm for the identification of sinusoidal signal frequency with constant parameters. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2022, vol. 22, no. 1, pp. 18–24. (in Russian). <https://doi.org/10.17586/2226-1494-2022-22-1-18-24>
 23. Aranovskiy S., Bobtsov A., Ortega R., Pyrkina A. Performance enhancement of parameter estimators via dynamic regressor extension and mixing. *IEEE Transactions on Automatic Control*, 2017, vol. 62, no. 7, pp. 3546–3550. <https://doi.org/10.1109/TAC.2016.2614889>

Авторы

Нгуен Хак Тунг — аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, <https://orcid.org/0000-0001-6430-1927>, nguyenkhactungvhq1994@gmail.com

Власов Сергей Михайлович — кандидат технических наук, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, <https://orcid.org/0000-0002-8345-7553>, smvlasov@itmo.ru

Пыркин Антон Александрович — доктор технических наук, профессор, профессор, декан факультета, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, <https://orcid.org/0000-0001-8806-4057>, a.pyrkina@gmail.com

Скобелева Александра Викторовна — студент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, <https://orcid.org/0000-0002-1504-5856>, skala.job@gmail.com

Authors

Khac T. Nguyen — PhD Student, ITMO University, Saint Petersburg, 197101, Russian Federation, <https://orcid.org/0000-0001-6430-1927>, nguyenkhactungvhq1994@gmail.com

Sergey M. Vlasov — PhD, Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, <https://orcid.org/0000-0002-8345-7553>, smvlasov@itmo.ru

Anton A. Pyrkina — D. Sc., Full Professor, Dean of Faculty, ITMO University, Saint Petersburg, 197101, Russian Federation, <https://orcid.org/0000-0001-8806-4057>, a.pyrkina@gmail.com

Aleksandra V. Skobeleva — Student, ITMO University, Saint Petersburg, 197101, Russian Federation, <https://orcid.org/0000-0002-1504-5856>, skala.job@gmail.com

Статья поступила в редакцию 18.03.2022
Одобрена после рецензирования 20.05.2022
Принята к печати 12.07.2022

Received 18.03.2022
Approved after reviewing 20.05.2022
Accepted 12.07.2022



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

КОМПЬЮТЕРНЫЕ СИСТЕМЫ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ COMPUTER SCIENCE

doi: 10.17586/2226-1494-2022-22-4-674-680

УДК 004.021

Построение криптографических схем, основанных на эллиптических кривых над рациональными числами

Вадим Валерьевич Давыдов¹✉, Жан-Мишель Никодэмович Дакуо²,
Иван Дмитриевич Иогансон³, Алтана Феликсовна Хуцаева⁴

^{1,2,3,4} Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

¹ vvdavydov@itmo.ru✉, <https://orcid.org/0000-0002-5544-2434>

² jeandakuo@itmo.ru, <https://orcid.org/0000-0002-4084-8829>

³ ivan.ioganson@itmo.ru, <https://orcid.org/0000-0002-0856-2249>

⁴ afkhutsaeva@itmo.ru, <https://orcid.org/0000-0001-5494-7142>

Аннотация

Предмет исследования. Исследована возможность использования в криптографических схемах эллиптических кривых над полем рациональных чисел ненулевого ранга. **Метод.** Впервые предложено построение криптосистем, безопасность которых основана на сложности решения математической задачи о рюкзаке на эллиптических кривых над рациональными числами ненулевых рангов. **Основные результаты.** Описан новый подход использования эллиптических кривых для криптографических схем. Выполнен ряд экспериментов для оценки поведения высот точек эллиптических кривых бесконечного порядка. Представлена модель криптосистемы, стойкой к вычислениям на квантовом компьютере и основанной на использовании рациональных точек кривой бесконечного порядка. Проведено исследование криптографической стойкости и эффективности предлагаемой схемы. Реализована атака на поиск секрета в криптосистеме, показано, что сложность атаки экспоненциальна. **Практическая значимость.** Рассмотренное решение может быть применено при построении реальных криптографических схем и протоколов.

Ключевые слова

эллиптические кривые, рациональные числа, ранг кривых, асимметричное шифрование, задача о рюкзаке

Благодарности

Работа выполнена при поддержке программы «Приоритет 2030».

Ссылка для цитирования: Давыдов В.В., Дакуо Ж.-М.Н., Иогансон И.Д., Хуцаева А.Ф. Построение криптографических схем, основанных на эллиптических кривых над рациональными числами // Научно-технический вестник информационных технологий, механики и оптики. 2022. Т. 22, № 4. С. 674–680. doi: 10.17586/2226-1494-2022-22-4-674-680

Building cryptographic schemes based on elliptic curves over rational numbers

Vadim V. Davydov¹✉, Jean-Michelle N. Dakuo², Ivan D. Ioganson³,
Altana F. Khutsaeva⁴

^{1,2,3,4} ITMO University, Saint Petersburg, 197101, Russian Federation

¹ vvdavydov@itmo.ru✉, <https://orcid.org/0000-0002-5544-2434>

² jeandakuo@itmo.ru, <https://orcid.org/0000-0002-4084-8829>

³ ivan.ioganson@itmo.ru, <https://orcid.org/0000-0002-0856-2249>

⁴ afkhutsaeva@itmo.ru, <https://orcid.org/0000-0001-5494-7142>

Abstract

The possibility of using elliptic curves over the rational field of non-zero ranks in cryptographic schemes is studied. For the first time, the construction of cryptosystems is proposed the security of which is based on the complexity of solving the knapsack problem on elliptic curves over rational numbers of non-zero ranks. A new approach to the use

© Давыдов В.В., Дакуо Ж.-М.Н., Иогансон И.Д., Хуцаева А.Ф., 2022

of elliptic curves for cryptographic schemes is proposed. A few experiments have been carried out to estimate the heights characteristic of points on elliptic curves of infinite order. A model of a cryptosystem resistant to computations on a quantum computer and based on rational points of an infinite order curve is proposed. A study of the security and effectiveness of the proposed scheme has been carried out. An attack on the secret search in such a cryptosystem is implemented and it is shown that the complexity of the attack is exponential. The proposed solution can be applied in the construction of real cryptographic schemes as well as cryptographic protocols.

Keywords

elliptic curves, rational numbers, curve rank, asymmetric encryption, knapsack problem

Acknowledgements

This research was supported by Priority 2030 Federal Academic Leadership Program.

For citation: Davydov V.V., Dakuo J.-M.N., Ioganson I.D., Khutsaeva A.F. Building cryptographic schemes based on elliptic curves over rational numbers. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2022, vol. 22, no. 4, pp. 674–680 (in Russian). doi: 10.17586/2226-1494-2022-22-4-674-680

Введение

В настоящее время многие популярные крипто-системы с открытым ключом основаны на задачах факторизации больших целых чисел и дискретного логарифмирования в конечных группах. Например, в мультипликативной группе конечного поля и группе точек на эллиптической кривой над конечным полем. Эллиптические кривые представляют особый интерес, так как в построенных криптосистемах используются ключи меньшего размера при том же уровне безопасности по сравнению с аналогами из других областей криптографии [1].

На данный момент не существует эффективных алгоритмов, решающих задачу дискретного логарифмирования в группе точек эллиптической кривой за полиномиальное время. Это позволяет работать в полях меньшего размера, следовательно и генерировать ключи меньшего размера при том же уровне безопасности по сравнению с аналогами. Выполнение группового закона на эллиптических кривых позволяет реализовывать и применять уже известные алгоритмы в криптографии.

Один из классических протоколов на эллиптических кривых — протокол Диффи–Хеллмана [2], позволяющий осуществлять выработку общего секретного ключа. В зависимости от изначальных условий, например, если пользователи не установили общий ключ, существуют различные варианты решения таких задач. Так, в криптосистеме Мессе–Омуры [3] благодаря свойствам эллиптических кривых осуществляется коммутативное шифрование: пользователи передают друг другу сообщения, не создавая при этом общего ключа. На эллиптических кривых также может быть реализована схема Эль–Гамала [4], в которой осуществляются более эффективные вычисления, по сравнению со схемой на модульной арифметике. Это связано с тем, что операция возведения в степень сводится к сложению точек.

Кроме конечных полей, кривые можно рассматривать и над полем рациональных чисел. Эллиптические кривые над рациональным полем обладают особой групповой структурой. Луи Морделл (1922 год) выдвинул гипотезу о конечности числа рациональных точек для кривых первого рода [5], а Андре Вейль (1928 год) ввел обобщение гипотезы [6]. Впоследствии была

сформулирована теорема Морделла–Вейля о том, что кривые над рациональным полем образуют конечнопорожденную абелеву группу, изоморфную группам точек кручения и точек бесконечного порядка. В 1935 году Найгел и Лутц [7] показали, каким образом можно найти точки подгруппы кручения. В 1978 году Барри Мазуром было получено, что подгруппа точек кручения конечна и мала. Отметим, что подгруппа точек бесконечного порядка исследована недостаточно, количество точек генераторов данной подгруппы задает ранг кривой. В настоящее время известна кривая с максимальным рангом 28 [8], а большинство кривых имеют ранг 0 или 1. Задача подсчета точного ранга кривой является сложной, а поиск кривых высокого ранга остается открытой проблемой.

В настоящей работе предложено построение модели криптосистемы с использованием задачи о рюкзаке [9] на эллиптических кривых над рациональными числами.

Задача о рюкзаке является NP-полной и находит применение во многих областях. Первое упоминание о ней можно найти в трудах Данцига [10] и Мэтьюза [11], в общем случае она заключается в поиске таких m_i из множества $M = \{m_1, m_2, \dots, m_n\}$, которые при умножении на известные коэффициенты из множества $K = \{k_1, k_2, \dots, k_m\}$ в сумме дают целое число C :

$$C = \sum_{i=1}^m m_i \cdot k_i.$$

Впервые криптосистема, основанная на задаче о рюкзаке, была предложена Ральфом Мерклом и Мартином Хеллманом [12], где использована сверхвозрастающая последовательность секретного ключа. С помощью ключа осуществлена расшифровка текста, а в качестве открытого ключа использована не-сверхвозрастающая последовательность, по которой расшифровать число C без знания секретного ключа невозможно. Существует эффективная аппроксимирующая LLL-атака на данную криптосистему, где с помощью редуцированного базиса решетки появляется возможность решить задачу упаковки рюкзака с определенной вероятностью. Однако данная атака не применима к эллиптическим кривым, так как, в отличие от классической криптосистемы рюкзака, вместо чисел используются точки кривой [13].

Формулировка математической задачи

Пусть $E(\mathbb{Q})$ — эллиптическая кривая над полем рациональных чисел (\mathbb{Q}), заданная уравнением в обобщенной форме Вейерштрасса:

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6,$$

где коэффициент $a_i \in \mathbb{Q}$ для $i \in \{1, 2, 3, 4, 6\}$, переменные $x, y \in \mathbb{Q}$.

По теореме Морделла–Вейля [13] группа точек на такой кривой — конечно сгенерированная абелева группа:

$$E(\mathbb{Q}) \cong E(\mathbb{Q})_{tors} \oplus \mathbb{Z}^r,$$

где $E(\mathbb{Q})_{tors}$ и \mathbb{Z}^r — группы точек кручения кривой и бесконечного порядка соответственно; r — ранг эллиптической кривой.

Группа точек кручения кривой может быть представлена следующим образом:

$$E(\mathbb{Q})_{tors} = \{ \text{точка } P \in E(\mathbb{Q}) \text{ и число } n \in \mathbb{N} \text{ такое, что } nP = \mathcal{O}, \}$$

где \mathcal{O} — точка на бесконечности. По теореме Мазура [14] порядок такой группы не превышает 16, также известно, каким группам изоморфна $E(\mathbb{Q})_{tors}$. Наибольший интерес при построении криптографических схем представляет группа \mathbb{Z}^r . Если ранг кривой равен нулю, то такая схема неприменима в криптографии, так как количество рациональных точек на кривой конечно и не превышает 16, что позволяет осуществить элементарный перебор точек.

Однако при $r > 0$ такие кривые обладают следующей особенностью. Пусть на кривой $E(\mathbb{Q}) \{T_1, T_2, \dots, T_n\}$ — множество точек кручения конечного порядка, а $\{P_1, P_2, \dots, P_r\}$ — множество точек бесконечного порядка. Тогда любая точка $E(\mathbb{Q})$ имеет вид:

$$Q = a_1T_1 + a_2T_2 + \dots + a_nT_n + b_1P_1 + b_2P_2 + \dots + b_rP_r, \quad (1)$$

где $\{a_1, \dots, a_n\}, \{b_1, \dots, b_r\} \in \mathbb{Z}$, и сложение точек осуществляется по групповому закону.

Основная математическая проблема, на сложности которой можно строить криптосистемы — поиск коэффициентов $\{a_1, \dots, a_n\}, \{b_1, \dots, b_r\}$, при известных точке Q и группе $E(\mathbb{Q})/2E(\mathbb{Q})$. Если не существует алгоритма, с помощью которого данная задача может быть решена за полиномиальное время, тогда ее использование допустимо. Для оценки необходимо учитывать специфические свойства эллиптических кривых над рациональными числами. Одним из таких свойств является высота точки.

Пусть высота дроби $\frac{m}{n} \in \mathbb{Q}$, $\text{НОД}(m, n) = 1$, где НОД — наибольший общий делитель, задается как [15]:

$$h\left(\frac{m}{n}\right) = \log(\max\{|m|, |n|\}).$$

Высотой точки эллиптической кривой называют высоту ее значения по координате x . Спецификой процес-

са последовательного сложения точек кривой является предсказуемый рост высоты координат точки.

Также существует понятие о канонической высоте \hat{h} точки P на кривой $E(\mathbb{Q})$ [15]:

$$\hat{h}(P) = \lim_{n \rightarrow \infty} \frac{h(2^n P)}{4^n}.$$

Для канонической высоты верно, что для всех $P \in E(\mathbb{Q})$ и $m \in \mathbb{Z}$:

$$\hat{h}(mP) = m^2 \hat{h}(P).$$

Таким образом, значение высоты точки можно аппроксимировать квадратичной функцией.

Высота позволяет получить некоторую информацию о коэффициентах, но не об их значениях. Следовательно, чем больше ранг эллиптической кривой, тем сложнее поиск коэффициентов в уравнении (1) из-за большего числа возможных вариантов.

Проведем оценку необходимого количество точек бесконечного порядка для достижения уровня безопасности, при котором время поиска таких коэффициентов будет больше полиномиального.

В уравнении (1) злоумышленнику известны координаты точки Q , соответственно, ему известно и значение высоты.

В случае, когда $r = 0$, число рациональных точек на кривой конечно. Как следует из теоремы Мазура, порядок подгруппы точек кручения не может быть больше 16, а порядки самих точек не превышают 12. Тогда необходимо перебрать один или два коэффициента у суммы точек подгруппы $E(\mathbb{Q})_{tors}$, что выполняется за линейное время $O(1)$.

Если $r = 1$, уравнение (1) приобретает вид:

$$Q = a_1T_1 + a_2T_2 + \dots + a_nT_n + b_1P_1.$$

В данном случае по значению $h(Q)$ возможно определить небольшой диапазон для значения b_1 и найти коэффициент за полиномиальное время.

Если $r > 1$, зная высоту точки Q , нет возможности определить, какие коэффициенты $\{a_1, \dots, a_n\}, \{b_1, \dots, b_r\}$ использовались в уравнении (1).

Выберем кривую с $r = 2$ и тривиальной подгруппой точек кручения. Тогда уравнение (1) можно переписать в виде:

$$Q = b_1P_1 + b_2P_2.$$

Задача злоумышленника — поиск коэффициентов b_1 и b_2 . Очевидно, что с увеличением количества точек бесконечного порядка (или ранга кривой), а следовательно, и коэффициентов, увеличивается сложность поиска.

Оценка вычислительной сложности поиска коэффициентов и точек в уравнении

Рассмотрим подробнее уравнение (1). Пусть злоумышленнику известны точки кручения, точки бесконечного порядка и итоговая точка (или ее высота).

Тогда задача злоумышленника — найти коэффициенты $\{a_1, \dots, a_n\}, \{b_1, \dots, b_r\}$.

Очевидно, что можно оптимизировать полный перебор коэффициентов, так как знание итоговой высоты дает дополнительную информацию о коэффициентах. Для простоты положим, что группа точек кручения тривиальна. Для поиска коэффициентов предлагается использовать следующий алгоритм поиска (Алгоритм).

Алгоритм. Алгоритм поиска коэффициентов уравнения (1) при условии, что подгруппа точек кручения T тривиальна

ФУНКЦИЯ Перебор_по_точкам($Q, r, \{P_1, \dots, P_r\}$)

Входные данные: Q — целевая точка,
 r — количество точек
 бесконечного
 порядка,
 $\{P_1, P_2, \dots, P_r\}$ — набор точек
 бесконечного порядка.

НАЧАЛО

ЕСЛИ $r = 2$

$i_1 := 1$

ПОКА $h(i_1 \times P_1) < h(Q)$

$i_1 := i_1 + 1$

$i_2 := 0$

ПОКА $h(i_2 \times P_2) < h(Q)$

ПОКА $h(i_1 \times P_1 + i_2 \times P_2) > h(Q)$

$i_1 := i_1 - 1$

ПОКА $h(i_1 \times P_1 + i_2 \times P_2) > h(Q)$

$i_2 := i_2 + 1$

ЕСЛИ $Q = i_1 \times P_1 + i_2 \times P_2$

ВЕРНУТЬ $\{\{i_1, P_1\}, \{i_2, P_2\}\}$

ВЕРНУТЬ 0

ИНАЧЕ

$i := 0$

ПОКА $h(i \times P_r) < h(Q)$

$res := \text{Перебор_по_точкам}(Q - i \times P_r, r - 1, \{P_1, \dots, P_{r-1}\})$

ЕСЛИ $res \neq 0$

ВЕРНУТЬ $res \cup \{\{i, P_r\}\}$

$i := i + 1$

ВЕРНУТЬ 0

КОНЕЦ

Заметим, что алгоритм поиска рекурсивно перебирает все возможные коэффициенты для точек бесконечного порядка. Тогда, в случае двух точек, данный алгоритм будет перебирать $O(\sqrt{h(Q)})$ значений. Каждая дополнительная точка будет увеличивать сложность перебора еще в $O(\sqrt{h(Q)})$. Таким образом, можно сделать вывод, что вычислительная сложность алгоритма поиска будет равна $O((h(Q))^{\frac{r-1}{2}})$.

Модель криптографической схемы, построенной на эллиптических кривых над рациональными числами

Построим криптографические схемы и протоколы. Приведем модель криптосистемы, основанной на зада-

че о рюкзаке. Выберем эллиптическую кривую $E(\mathbb{Q})$ с $r > 1$, наборами точек кручения $\{T_1, \dots, T_n\}$ и бесконечного порядка $\{I_1, \dots, I_r\}$.

Зададим секретный ключ как сверхвозрастающую последовательность точек P_1, P_2, \dots, P_n , при условии, что такие точки задаются как набор линейных комбинаций:

$$P_i = q_1 T_1 + \dots + q_n T_n + k_1 I_1 + \dots + k_r I_r; q, k \in \{0, 1\}.$$

Под сверхвозрастающей последовательностью точек P_1, P_2, \dots, P_n будем понимать:

$$\forall i \in \{2 \dots n\}, h(P_i) > h\left(\sum_{j=1}^{i-1} P_j\right).$$

Зададим секретное преобразование ϕ , удовлетворяющее следующим свойствам:

- 1) сохраняется операция сложения над точками эллиптической кривой, т. е. $\phi(P + Q) = \phi(P) + \phi(Q)$;
- 2) сверхвозрастающая последовательность точек P_1, P_2, \dots, P_n преобразуется в несверхвозрастающую $\phi(P_1), \phi(P_2), \dots, \phi(P_n)$;
- 3) преобразований подобных ϕ должно существовать достаточное количество, чтобы избежать атаки перебором.

Открытым ключом является последовательность $\phi(P_1), \phi(P_2), \dots, \phi(P_n)$.

Для шифрования сообщения $m = (b_1, b_2, \dots, b_n)$, где $b_i \in \{0, 1\}$, вычислим точку: $C = \sum_{i=1}^n b_i \cdot \phi(P_i)$.

Для расшифрования сообщения вычислим $Q = \phi^{-1}(C)$. При условии, что i от n до 1, получим:

- 1) если $h(Q) \geq h(P_i)$, то $b_i = 1$, иначе $b_i = 0$;
- 2) $Q = Q - P_i$.

В результате расшифрования получим сообщение $m = (b_1, b_2, \dots, b_n)$.

Основная проблема такой модели — большие размеры ключей, так как необходимо хранить координаты точек, размеры значений которых достаточно велики при больших значениях коэффициентов. Можно утверждать, что такая криптосистема обладает стойкостью к квантовому компьютеру, а также задача о рюкзаке не может быть вскрыта с помощью алгоритма LLL (Ленстра–Ленстра–Ловаса), так как работа ведется с точками эллиптической кривой.

Результаты экспериментов

Для проверки полученных результатов и оценки применения кривых различных рангов в криптографии выполнен ряд экспериментов. Основные эксперименты проведены с целью исследования поведения высот, а также уровня безопасности используемой задачи упаковки рюкзака. Для моделирования разработаны программы на языке Python с использованием программного обеспечения SAGE Math на персональном компьютере со следующими характеристиками: процессор AMD Ryzen 7 5800H, NVIDIA GeForce RTX 3070 Laptop GPU, оперативная память 32 ГБ, SSD 2024 ГБ.

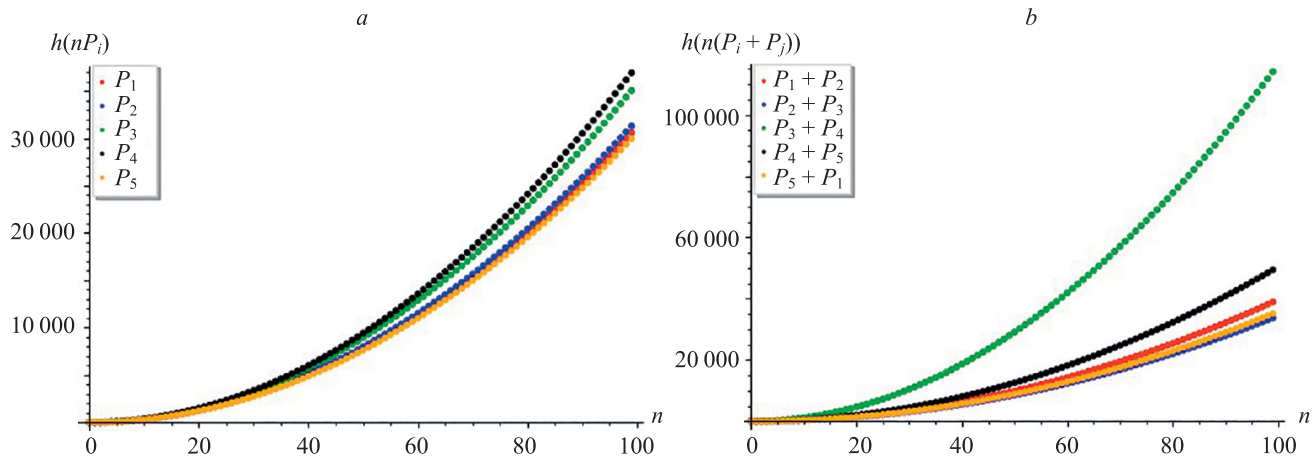


Рис. 1. Зависимости высот отдельных точек $\{P_1, \dots, P_5\}$ (a) и суммы двух из них (b) на эллиптической кривой $E(\mathbb{Q})$ от умножающего коэффициента n

Fig. 1. The dependences of the individual points $\{P_1, \dots, P_5\}$ heights (a) and the sum of two of them (b) on the elliptic curve $E(\mathbb{Q})$ on the multiplying factor n

Для определения аппроксимирующей функции для высот точек в зависимости от помножающего коэффициента осуществлен следующий эксперимент. Выбрана кривая $E(\mathbb{Q})$: $y^2 + y = x^3 - 79x + 342$ с рангом $r = 5$ и точками бесконечного порядка $P_1 = (3, 11)$, $P_2 = (7, 11)$, $P_3 = (10, 23)$, $P_4 = (-6, 24)$, $P_5 = (4, 9)$. На рис. 1 показан рост высот каждой точки и суммы двух точек при умножении на различные коэффициенты от 1 до 100.

Как видно из полученных результатов, наблюдается квадратичный рост высот, в то же время для каждой точки рост незначительно отличается. Отсюда следует, что при поиске коэффициентов в уравнении (1) достаточно сложно определить, у какой точки был наибольший коэффициент.

Рассмотрим следующую задачу. Пусть на кривой $E(\mathbb{Q})$ точка Q задана в виде:

$$Q = b_1 P_i + b_2 P_j.$$

Зафиксируем сумму коэффициентов $b_1 + b_2 = 100$, последовательно увеличивая коэффициент b_1 и уменьшая коэффициент b_2 . На рис. 2 показана зависимость высоты суммы при изменении соотношения умножающих коэффициентов.

Заметим, что минимальное значение итоговой высоты рис. 2 наблюдается при примерном равенстве коэффициентов $b_1 = b_2$, в то время как максимальные значения получаются, когда коэффициенты сильно различаются. Однако примерно одна и та же высота точки может получаться большим числом различных вариантов, например, когда коэффициенты дают меньшую или большую суммы. Необходимо иметь в виду, что при получении большого значения высоты точки кривой с помощью расчетов трудно определить, с помощью каких именно коэффициентов она была получена.

Проведем сравнение алгоритма полного перебора всех значений и предложенного алгоритма поиска, для

оценки сложности решения задачи. В таблице показано время, полученное в процессе моделирования для разных кривых и с различным ограничением сверху на коэффициенты уравнения. За относительную единицу принято время решения задачи поиска коэффициентов с помощью алгоритма поиска для эллиптической кривой $E(\mathbb{Q})$: $y^2 + xy = x^3 + 1$ с $r = 2$ и максимальным значением коэффициентов 10.

Как видно из таблицы, предложенный алгоритм поиска работает гораздо быстрее алгоритма полного перебора при использовании кривых более высоких рангов. Отметим, что предложенный алгоритм имеет экспоненциальную сложность, что не позволяет эффективно решать задачу поиска умножающих коэффициентов.

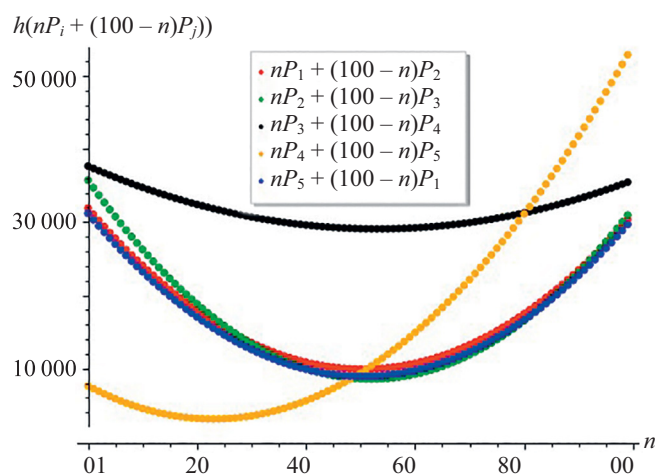


Рис. 2. Зависимость высоты от изменения соотношения умножающих коэффициентов b_1, b_2 для суммы двух точек при сохранении значения суммы

Fig. 2. The dependence of the height on the change in the ratio of multiplying coefficients b_1, b_2 for the sum of two points while maintaining the sum value

Таблица. Оценка времени решения задачи поиска коэффициентов
Table. Solution time estimation of the coefficient search problem

Эллиптическая кривая	Ранг	Максимальное значение коэффициентов	Время решения задачи поиска коэффициентов с помощью	
			алгоритма полного перебора, отн. ед.	предложенного алгоритма поиска, отн. ед.
$y^2 + xy = x^3 + 1$	2	10	2,1459	1
		15	5,6280	1,7683
$y^2 = x^3 - 52x + 100$	3	10	34,3741	30,1693
		15	136,7492	81,2948
$y^2 + y = x^3 - 79x + 342$	5	10	12508,3143	5724,5763
		15	98604,8883	33741,7067

Заклучение

В работе проведено исследование применимости эллиптических кривых над рациональными числами в криптографии. Получены важные теоретические результаты – математическая задача, которую можно использовать при построении асимметричных криптографических схем, а также модель криптосистемы, использующей рассматриваемые кривые. Основная отличительная особенность таких схем — наличие на кривых над рациональным полем точек бесконечного порядка, позволяющих генерировать уникальные точки

конечнопорожденной группы. Данная особенность открывает новые возможности для задач криптографии, учитывая ряд нерешенных проблем в области эллиптических кривых над рациональными числами. Отметим, что основной недостаток предложенных криптографических схем — размер секретного ключа достаточно велик по сравнению с современными криптосистемами, построенными на эллиптических кривых над конечными полями. В рамках дальнейших исследований планируется уменьшение размеров ключей, используя свойства эллиптических кривых и их рациональных точек.

Литература

1. Koblitz N., Menezes A., Vanstone S. The state of elliptic curve cryptography // *Designs, Codes and Cryptography*. 2000. V. 19. N 2-3. P. 173–193. <https://doi.org/10.1023/A:1008354106356>
2. Koblitz N. Elliptic curve cryptosystems // *Mathematics of Computation*. 1987. V. 48. N 177. P. 203–209. <https://doi.org/10.1090/S0025-5718-1987-0866109-5>
3. Washington L.C. *Elliptic Curves: Number Theory and Cryptography*. Chapman and Hall/CRC, 2008. 536 p.
4. Silverman J.H. *The Arithmetic of Elliptic Curves*. New York: Springer, 2009. 513 p. (Graduate Texts in Mathematics; V. 106).
5. Mordell L. On the rational solutions of the indeterminate equation of the third and fourth degree // *Proceedings Cambridge Philosophical Society*. 1922. V. 21. P. 179–192.
6. Weil A. L'arithmétique sur les courbes algébriques // *Acta Mathematica*. 1929. V. 52. N 1. P. 281–315. <https://doi.org/10.1007/BF02592688>
7. Silverman J.H. Heights and elliptic curves // *Arithmetic Geometry*. New York, NY: Springer, 1986. P. 253–265. https://doi.org/10.1007/978-1-4613-8655-1_10
8. Klagsbrun Z., Sherman T., Weigandt J. The Elkies curve has rank 28 subject only to GRH // *Mathematics of Computation*. 2019. V. 88. N 316. P. 837–846. <https://doi.org/10.1090/mcom/3348>
9. Menezes A.J., Van Oorschot P.C., Vanstone S.A. *Handbook of Applied Cryptography*. CRC press, 2018. 810 p.
10. Danzig T. *Numbers: The Language of Science*. Revised. Now York, Macmillan, 1933.
11. Mathews G.B. On the partition of numbers // *Proceedings of the London Mathematical Society*. 1896. V. 1. N 1. P. 486–490. <https://doi.org/10.1112/plms/s1-28.1.486>
12. Diffie W., Hellman M. New directions in cryptography // *IEEE Transactions on Information Theory*. 1976. V. 22. N 6. P. 644–654. <https://doi.org/10.1109/TIT.1976.1055638>
13. Noro K., Kobayashi K. Knapsack cryptosystem on elliptic curves // *Cryptology ePrint Archive*. 2009. P. 2009/091.

References

1. Koblitz N., Menezes A., Vanstone S. The state of elliptic curve cryptography. *Designs, Codes and Cryptography*, 2000, vol. 19, no. 2-3, pp. 173–193. <https://doi.org/10.1023/A:1008354106356>
2. Koblitz N. Elliptic curve cryptosystems. *Mathematics of Computation*, 1987, vol. 48, no. 177, pp. 203–209. <https://doi.org/10.1090/S0025-5718-1987-0866109-5>
3. Washington L.C. *Elliptic Curves: Number Theory and Cryptography*. Chapman and Hall/CRC, 2008, 536 p.
4. Silverman J.H. *The Arithmetic of Elliptic Curves*. New York, Springer, 2009, 513 p. Graduate Texts in Mathematics, vol. 106.
5. Mordell L. On the rational solutions of the indeterminate equation of the third and fourth degree. *Proceedings Cambridge Philosophical Society*, 1922, vol. 21, pp. 179–192.
6. Weil A. L'arithmétique sur les courbes algébriques. *Acta Mathematica*, 1929, vol. 52, no. 1, pp. 281–315. <https://doi.org/10.1007/BF02592688>
7. Silverman J.H. Heights and elliptic curves. *Arithmetic Geometry*. New York, NY, Springer, 1986, pp. 253–265. https://doi.org/10.1007/978-1-4613-8655-1_10
8. Klagsbrun Z., Sherman T., Weigandt J. The Elkies curve has rank 28 subject only to GRH. *Mathematics of Computation*, 2019, vol. 88, no. 316, pp. 837–846. <https://doi.org/10.1090/mcom/3348>
9. Menezes A.J., Van Oorschot P.C., Vanstone S.A. *Handbook of Applied Cryptography*. CRC press, 2018, 810 p.
10. Danzig T. *Numbers: The Language of Science*. Revised. Now York, Macmillan, 1933.
11. Mathews G.B. On the partition of numbers. *Proceedings of the London Mathematical Society*, 1896, vol. 1, no. 1, pp. 486–490. <https://doi.org/10.1112/plms/s1-28.1.486>
12. Diffie W., Hellman M. New directions in cryptography. *IEEE Transactions on Information Theory*, 1976, vol. 22, no. 6, pp. 644–654. <https://doi.org/10.1109/TIT.1976.1055638>
13. Noro K., Kobayashi K. Knapsack cryptosystem on elliptic curves. *Cryptology ePrint Archive*, 2009, pp. 2009/091.

14. Mazur B., Goldfeld D. Rational isogenies of prime degree // *Inventiones Mathematicae*. 1978. V. 44. N 2. P. 129–162. <https://doi.org/10.1007/BF01390348>
15. Lozano-Robledo Á. Elliptic curves, modular forms, and their L-functions // *The Student Mathematical Library*. 2011. V. 58. <http://dx.doi.org/10.1090/stml/058>

14. Mazur B., Goldfeld D. Rational isogenies of prime degree. *Inventiones Mathematicae*, 1978, vol. 44, no. 2, pp. 129–162. <https://doi.org/10.1007/BF01390348>
15. Lozano-Robledo Á. Elliptic curves, modular forms, and their L-functions. *The Student Mathematical Library*, 2011, vol. 58. <http://dx.doi.org/10.1090/stml/058>

Авторы

Давыдов Вадим Валерьевич — преподаватель, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 57203909696](https://orcid.org/0000-0002-5544-2434), [vvdavydov@itmo.ru](https://orcid.org/0000-0002-5544-2434)
Дакуо Жан-Мишель Никодэмович — студент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, <https://orcid.org/0000-0002-4084-8829>, jeandakuo@itmo.ru
Иогансон Иван Дмитриевич — инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, <https://orcid.org/0000-0002-0856-2249>, ivan.ioganson@itmo.ru
Хуцаева Алтана Феликсовна — инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, <https://orcid.org/0000-0001-5494-7142>, afkhutsaeva@itmo.ru

Authors

Vadim V. Davydov — Lecturer, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 57203909696](https://orcid.org/0000-0002-5544-2434), [vvdavydov@itmo.ru](https://orcid.org/0000-0002-5544-2434)
Jean-Michelle N. Dakuo — Student, ITMO University, Saint Petersburg, 197101, Russian Federation, <https://orcid.org/0000-0002-4084-8829>, jeandakuo@itmo.ru
Ivan D. Ioganson — Engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, <https://orcid.org/0000-0002-0856-2249>, ivan.ioganson@itmo.ru
Altana F. Khutsaeva — Engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, <https://orcid.org/0000-0001-5494-7142>, afkhutsaeva@itmo.ru

Статья поступила в редакцию 14.04.2022
 Одобрена после рецензирования 20.05.2022
 Принята к печати 12.07.2022

Received 14.04.2022
 Approved after reviewing 20.05.2022
 Accepted 12.07.2022



Работа доступна по лицензии
 Creative Commons
 «Attribution-NonCommercial»

doi: 10.17586/2226-1494-2022-22-4-681-690

УДК 004:658.011.56

Алгоритм генерации проектных решений для управления данными и проектно-производственными процедурами на этапах жизненного цикла электронного изделия

Юлия Валерьевна Донецкая✉

Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация
donetskaya_julia@mail.ru✉, <https://orcid.org/0000-0001-5293-5025>

Аннотация

Предмет исследования. Интеграция автоматизированных систем на предприятиях обеспечивает информационную поддержку этапов жизненного цикла изделия и электронное взаимодействие между сотрудниками в процессе выполнения работ. При выполнении проектно-производственных процедур сотрудники предприятий решают различные проектные задачи. Задачи связаны с анализом большого объема информации об изделии, представляемой в виде онтологии. Для решения таких задач требуется разработка алгоритма для извлечения информации из онтологии на основе заданных требований. **Метод.** Разработанный алгоритм состоит из пяти этапов. На первом этапе формируется пространство поиска проектных решений. Далее для каждого варианта проектного решения рассчитаны значения целевой функции и выбрано лучшее проектное решение (выполняется условие минимизации значения целевой функции). Третий этап связан с выбором проектных решений, близких к найденному лучшему решению (определяется по вычисляемому расстоянию Хэмминга). Следующие этапы характеризуются анализом элементов множества вариантов проектных решений и формированием искомого проектного решения. **Основные результаты.** Предложены последовательности действий, выполняемые на этапах алгоритма генерации проектных решений. **Практическая значимость.** Алгоритм может быть реализован на предприятиях для обеспечения процедуры решения проектных задач, что позволит разработать сигнатуры и семантики унифицированных сервисов для применения цифрового паспорта.

Ключевые слова

алгоритм, цифровой паспорт, проектное решение, онтология

Ссылка для цитирования: Донецкая Ю.В. Алгоритм генерации проектных решений для управления данными и проектно-производственными процедурами на этапах жизненного цикла электронного изделия // Научно-технический вестник информационных технологий, механики и оптики. 2022. Т. 22, № 4. С. 681–690. doi: 10.17586/2226-1494-2022-22-4-681-690

An algorithm for generating design solutions for data and design-production procedures management at the stages of the lifecycle of an electronic product

Julia V. Donetskaya✉

ITMO University, Saint Petersburg, 197101, Russian Federation
donetskaya_julia@mail.ru✉, <https://orcid.org/0000-0001-5293-5025>

Abstract

The integration of automated systems at enterprises provides information support for the stages of the product life cycle and electronic interaction between employees in the process of performing work. This means that performing design and production procedures employees of enterprises solve various design problems. The tasks are related to the analysis of a large amount of information about the product presented in the form of ontology. This requires the development of an algorithm to extract information from the ontology based on given requirements. The developed algorithm consists of several stages. At the first stage, a search space for design solutions is formed. At the second stage, for each variant of the design solution the values of the objective function are calculated, and the best design solution is selected. The best solution is the one for which the condition of minimizing the value of the objective function is satisfied. The third stage

© Донецкая Ю.В., 2022

is associated with the choice of design solutions that are close to the found best solution. The best solution is determined by the computed Hamming distance. The fourth and fifth stages are characterized by the analysis of the elements of the set of options for design solutions and the formation of the desired design solution Sequences of actions performed at the stages of the algorithm for generating design solutions are proposed. The proposed algorithm can be implemented at enterprises to provide a procedure for solving design problems. The presented algorithm allows the development of signatures and semantics of unified services for the use of a digital passport.

Keywords

algorithm, digital passport, design solution, ontology

For citation: Donetskaya Ju.V. An algorithm for generating design solutions for data and design-production procedures management at the stages of the lifecycle of an electronic product. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2022, vol. 22, no. 4, pp. 681–690 (in Russian). doi: 10.17586/2226-1494-2022-22-4-681-690

Введение

Задача создания и использования цифрового паспорта электронного изделия на предприятиях обусловлена двумя причинами. Первая причина заключается в организации электронного взаимодействия исполнителей договоров с заказчиками вне зависимости от того, в каких организациях они работают и какие системы используют. Вторая связана с автоматизацией генерации проектных решений для управления данными об изделии и проектно-производственными процедурами [1, 2].

Обе причины означают, что цифровые паспорта должны содержать данные и знания об изделиях и формирующих их проектно-производственных процедурах, выполняемых на этапах жизненного цикла. Для этого на предприятиях: внедряют системы, реализующие управление ресурсами предприятия, данными и знаниями об изделии, производстве и качеством продукции, или создают специализированные базы данных и знаний; разрабатывают механизмы интеграции и способы автоматизации формирования данных, определяющие содержание технической документации.

Рассмотрим научные работы, в которых нашли подтверждение данные процессы.

В работе [3] рассмотрена процедура автоматизации формирования инструкций по использованию устройств сигнализации, централизации и блокировки. Для процесса автоматизации разработаны одноименный программный модуль и база данных. Программный модуль позволяет обрабатывать перечни шаблонов документа, заполняемых обязательных и не обязательных полей, а также общий перечень разрабатываемых разделов инструкций. База данных содержит планы станций железной дороги, включающие информацию о: светофорах, стрелках, переездах, рельсовых цепях, роде тока и используемой частоте, а также расположении и принципе действия устройств.

Благодаря тому, что шаблоны регламентированы требованиями отраслевого формата технической документации, местоположение полей и разделов однозначно известны. Это позволяет разработчику в интерактивном режиме формировать или выбирать требуемый шаблон, после чего приступить к его заполнению. Причем вариант шаблона соответствует виду эксплуатируемого устройства и содержит только те перечни разделов инструкции, которые должны быть сформированы. Отметим, что заполнение содержа-

ния разделов инструкции частично автоматизировано. Например, информация, хранящаяся в базе данных, вводится в текст инструкции автоматически, а разработчик может выбирать требуемые ему данные из списков. Готовые документы проверяются на корректность заполнения всех полей и сохраняются одновременно в двух форматах — текстовом MS Word и специализированном — на дисках.

Значимость решения задачи автоматизации разработки технической документации доказана в [4]. Показано, что ошибки, допущенные при разработке, оказывают влияние на безопасность управления движением поездов, а именно, на изменение технического состояния систем железнодорожной автоматики и телемеханики, что может создавать условия опасных отказов объектов управления.

В работах [5–7] рассмотрены способы решения задач автоматизации формирования конструкторской документации табличного вида. Формирование выполнено на основе анализа значений столбцов базы данных, содержащей сведения о составе и структуре изделий. Поиск данных, определяющих содержание заданного вида документации, выполняется по запросу обозначения проектируемого изделия и дополнительных параметров согласно виду документа. Результат поиска — информация для разработки конструкторского документа, отсортированная и представленная согласно требованиям ГОСТ. Оформленный и сформированный документ в формате «.doc» или «.pdf» хранится на диске совместно со специализированным форматом, содержащим структуру шаблона.

Постановка задачи генерации проектных решений

В результате анализа научных работ можно сделать вывод, что автоматизация создания технической документации связана с несколькими задачами. Необходима разработка хранилища данных об изделии, получаемых, как на этапах разработки и подготовки производства, так и при производстве, в процессе испытаний изделий [8–10] и исследовании их отказов [11–13]. В качестве хранилища предложено использовать цифровой паспорт изделия, содержащий информацию этапов жизненного цикла изделия. Паспорт формируется на основе онтологического подхода.

Требуется разработка процедур анализа и синтеза проектных решений и описание предъявляемых требо-

ваний. Требования — четкие значения, такие как обозначение изделия, для которого формируется документ. Тогда поиск информации в хранилище заключается в анализе значений строк в конкретных столбцах таблиц. Результат поиска — совокупность информации, образующая содержание документации, по полному или частичному соответствию значения строки значению требования.

Заметим, что задачи автоматизации не ограничиваются только формированием данных, определяющих содержание некоторых видов конструкторских документов. Работы по цифровизации процессов предприятий определяют необходимость постановки задач автоматизации в более широком смысле — требуется разработка алгоритмов для управления не только данными об изделии, но и для управления проектно-производственными процедурами на предприятиях. Проектные решения образуются совокупностью данных, используемых для разработки и корректировки технической — конструкторской, программной или технологической — документации, для управления проектными и производственными процедурами. Причем требования, предъявляемые к таким проектным решениям, могут быть заданы нечеткими значениями.

Цель работы — разработка алгоритма автоматизации генерации проектных решений для управления данными и проектно-производственными процедурами.

Пусть исходными данными для алгоритма являются:

RS = (rs_1, rs_2, \dots, rs_s) — множество s требований, предъявляемых к содержанию цифрового паспорта,

RP = (rp_1, rp_2, \dots, rp_p) — множество p требований, предъявляемых к проектным решениям [14],

а также модель цифрового паспорта электронного изделия, содержащего знания о данных об изделии (в виде информационных объектов систем) и формирующих их проектно-производственных процедурах, выполняемых на этапах жизненного цикла.

В настоящей работе модель цифрового паспорта — метаописание онтологии цифрового паспорта. Поскольку онтология — кортеж, состоящий из множеств классов (**T**) и атрибутов (**A**), матриц сопоставлений (**P**) и смежности (**R**) [15], то ее метаописание можно определить в виде [16]:

D = (d_1, d_2, \dots, d_k), **D** \subseteq **T** — множество k информационных объектов, формируемых на этапах жизненного цикла средствами систем управления ресурсами предприятия, данными об изделии, производственной деятельностью или качеством продукции;

Pr = (pr_1, pr_2, \dots, pr_n), **Pr** \subseteq **T** — множество n проектно-производственных процедур, выполняемых на этапах жизненного цикла;

$$\mathbf{C} = \begin{bmatrix} c_{11} & \dots & c_{1n} \\ \dots & c_{ij} & \dots \\ c_{k1} & \dots & c_{kn} \end{bmatrix} = \begin{bmatrix} C_1 \\ C_2 \\ \dots \\ C_k \end{bmatrix}, i = \overline{1, k}; j = \overline{1, n} — модель$$

компонентов цифрового паспорта, где $c_{ij} \in C_i$, $c_{ij} = \langle d_i, pr_j \rangle$, если информационный объект $d_i \in \mathbf{D}$ связан отношением «формируется» с проектно-производственной процедурой $pr_j \in \mathbf{Pr}$, $c_{ij} = \langle 0, 0 \rangle$ в ином случае (в отличие от указанного в [16]);

U = (u_1, u_2, \dots, u_p), **U** \subseteq **A** — модель параметров компонентов цифрового паспорта;

$$\mathbf{G} = \begin{bmatrix} g_{11}(C_1) & \dots & g_{1p}(C_1) \\ \dots & g_{ij}(C_i) & \dots \\ g_{k1}(C_k) & \dots & g_{kp}(C_k) \end{bmatrix} = \begin{bmatrix} G_1(C_1) \\ G_2(C_2) \\ \dots \\ G_k(C_k) \end{bmatrix}, i = \overline{1, k}; j = \overline{1, p},$$

G \subseteq **P** — модель описаний компонентов цифрового паспорта, где $g_{ij}(C_i) \in G_i(C_i)$, $C_i \in \mathbf{C}$, $g_{ij}(C_i) = \langle C_i, u_j \rangle$, если компонент цифрового паспорта $C_i \in \mathbf{C}$ связан отношением «описывается» с параметром $u_j \in \mathbf{U}$, $g_{ij}(C_i) = \langle 0, 0 \rangle$, в ином случае (в отличие от указанного в [16]);

Z(C) = $|Z_{ij}(C_i, C_j)|$ — модель замещения описаний компонентов цифрового паспорта, где $Z_{ij}(C_i, C_j) = \begin{cases} 0, & G_i(C_i) \not\subseteq G_j(C_j) \\ 1, & G_i(C_i) \subseteq G_j(C_j) \end{cases}, i, j = \overline{1, k}$;

X(C) = $|X_{ij}(C_i, C_j)|$, **X(C)** \subseteq **R**, $i, j = \overline{1, k}$ — модель взаимодействия описаний компонентов цифрового паспорта, где $X_{ij}(C_i, C_j) = 1$, если описание компонента цифрового паспорта $G_i(C_i) \in \mathbf{G}$ связано отношениями (согласно [15]) с описанием $G_j(C_j) \in \mathbf{G}$, $X_{ij}(C_i, C_j) = 0$ в ином случае.

Тогда для любого предприятия отрасли приборостроения задача автоматизации генерации проектных решений для управления данными и проектно-производственными процедурами сводится к анализу описаний компонентов цифрового паспорта множества **G** согласно требованиям **RP** и **RS**.

Анализ элементов множества **G** на основе требований **RS** позволяет определить те из них, что образуют содержание цифрового паспорта на конкретном предприятии. Поскольку требования **RS** заданы вербально, то каждому элементу $G_i(C_i) \in \mathbf{G}$, $i = \overline{1, k}$ соответствует значение функции принадлежности нечеткой переменной, которая согласно [17] имеет вид:

$$\langle \overline{C}_i, \overline{\mathbf{U}}, \mu_{C_i}(\overline{\mathbf{U}}) \rangle,$$

где $\overline{C}_i \in \mathbf{C}$ — элементы матрицы **C**, удовлетворяющие требованиям **RS**, которые определяются на основе экспертных оценок; $\overline{\mathbf{U}} \subseteq \mathbf{U}$ — подмножество множества параметров компонентов цифрового паспорта, удовлетворяющих требованиям **RS**; $\mu_{C_i}(\overline{\mathbf{U}})$ — функция принадлежности, характеризующая вероятность того, что параметры из множества $\overline{\mathbf{U}}$ описывают компонент цифрового паспорта $\overline{C}_i \in \mathbf{C}$.

Вычисление значений функций принадлежности описано в работе [17], результат множества — ($\omega_1^{C_1}, \omega_2^{C_1}, \dots, \omega_p^{C_1}$), ($\omega_1^{C_2}, \omega_2^{C_2}, \dots, \omega_p^{C_2}$), ..., ($\omega_1^{C_k}, \omega_2^{C_k}, \dots, \omega_p^{C_k}$). Для решения задачи поиска проектного решения полученные значения приведены к четким — для каждого элемента $G_i(C_i) \in \mathbf{G}$, $i = \overline{1, k}$ вычислены значения $\Omega = (\Omega^{C_1}, \Omega^{C_2}, \dots, \Omega^{C_k})$ методом центроида

$$\Omega^{C_i} = \frac{\max(\omega_j^{C_i})}{\sum_{j=1}^p \max(\omega_j^{C_i})}, i = \overline{1, k}, j = \overline{1, p}, \quad (1)$$

где $\omega_j^{C_i}$ — значение функции принадлежности μ_{C_i} , соответствующее значению $u_j \in \mathbf{U}$; $\Omega^{C_i} \in \Omega$ — значение центроида для элемента $G_i(C_i) \in \mathbf{G}$, содержащее четкое значение нечеткой переменной.

Анализ элементов множества \mathbf{G} , согласно требованиям \mathbf{RP} , позволяет определить те из них, которые образуют проектное решение на основе содержания цифрового паспорта на конкретном предприятии. Используя алгоритм Такаги–Сугено–Канга, в работе [18] описана процедура синтеза проектных решений, содержащая нечеткие правила логического вывода. Получим, на основании выражения (1), для вычисления значений $\bar{\Omega}^{C_i}$ ($i = 1, \bar{k}$):

$$\begin{aligned}\bar{\Omega}^{C_1} &= w_1' \cdot \left(b_{10} + \sum_{j=1}^p b_{1j} \cdot r p_j \right), \\ \bar{\Omega}^{C_2} &= w_2' \cdot \left(b_{20} + \sum_{j=1}^p b_{2j} \cdot r p_j \right), \\ &\dots \\ \bar{\Omega}^{C_k} &= w_k' \cdot \left(b_{k0} + \sum_{j=1}^p b_{kj} \cdot r p_j \right),\end{aligned}$$

где $w_i' \in \mathbf{W}$ ($i = \overline{1, \bar{k}}$) — результат активизации подзаключений нечетких правил согласно работе [19]; b_{i0} , b_{ij} — коэффициенты, определяющие веса, вычисляемые на этапах нечеткого логического вывода.

Совокупность выражений $\bar{\Omega}^{C_1}, \bar{\Omega}^{C_2}, \dots, \bar{\Omega}^{C_k}$ обобщена и записана в матричной форме

$$\bar{\Omega} = \mathbf{M} \cdot \beta + \varepsilon, \quad (2)$$

где ε — ошибка; \mathbf{M} — детерминированная матрица размерности $k \times p$, согласно [19]:

$$\mathbf{M} = \mathbf{W} \cdot \mathbf{R} \mathbf{P} = \begin{bmatrix} w_1' \cdot r p_1 & \dots & w_1' \cdot r p_p \\ \dots & w_i' \cdot r p_j & \dots \\ w_k' \cdot r p_1 & \dots & w_k' \cdot r p_p \end{bmatrix},$$

где β — вектор-столбец вида $\beta = (b_{10}, \dots, b_{1p}, b_{20}, \dots, b_{2p}, \dots, b_{k0}, \dots, b_{kp})$; \mathbf{W} — множество результатов активизации подзаключений нечетких правил.

Тогда искомым проектным решением являются элементы матрицы $\mathbf{X}(\mathbf{C})$, которые удовлетворяют критерию подобия вида:

$$\Delta(\beta) = \sum_{i=1}^k (\Omega - \bar{\Omega}) = (\Omega - \mathbf{M} \cdot \beta)^T (\Omega - \mathbf{M} \cdot \beta) \rightarrow \min. \quad (3)$$

Преобразуя выражение (3), согласно работам [18, 19], получено выражение для вычисления элементов вектора-столбца β

$$\beta = (\mathbf{M}^T \mathbf{M})^{-1} \cdot (\mathbf{M}^T \Omega). \quad (4)$$

Рассмотрим варианты генерации проектных решений на основе биоинспирированных алгоритмов.

Варианты алгоритмов генерации проектных решений

Согласно работам [20–28], анализ метаописаний онтологий соответствует процедурам отображения или кластеризации, для автоматизации которых используются биоинспирированные алгоритмы.

Процедура отображения заключается в поиске семантических связей между вершинами графов мета-

онтологий. Основная задача поиска — расчет меры семантической близости. В [20] мера близости — комплексная мера, для вычисления которой предложен алгоритм многокритериальной оптимизации методом «роя частиц». Комплексная мера близости — среднее арифметическое значение пяти видов мер близости с учетом весового коэффициента. Значение вида меры близости вычислено на основе анализа значений вершин графов метаописаний онтологий, заданных вербально, а весового коэффициента — на основе экспертной оценки или автоматически с помощью обучаемой нейронной сети или алгоритма. В [21] мера семантической близости — эквивалентный тип семантической близости, вычисляемый в соответствии с методами из работ [22–26] и предложенным гибридным оптимизационным алгоритмом бактериального и кукушкиного поиска. Критерий — пороговое значение, превышение которого свидетельствует о значимости эквивалентной семантической близости.

В свою очередь процедура кластеризации означает объединение в группы схожих элементов метаописания онтологии. Согласно работам [27, 28], процедура выполнена с использованием алгоритмов на основе поведения пчелиной колонии, различаемых вычислением критериев подобия. В работе [27] критерий учитывает вес вершины и вес ребра графа метаописания онтологии, а в [28] предложено вычислить расстояние Хэмминга между элементами.

При этом алгоритм на основе поведения пчелиной колонии состоит из четырех фаз: инициализации, рабочих пчел, пчел-наблюдателей и пчел-разведчиков, обеспечения его точность и быстроедействие по сравнению с другими алгоритмами [29–31].

В фазе инициализации задается начальное пространство поиска проектных решений и параметры работы алгоритма. Пространством поиска определено множество проектных альтернатив, элементы которого анализируют при принятии проектного решения. Анализ выполнен на основе критерия подобия, формулируемого согласно требованиям или мере близости элементов друг другу. Параметры работы алгоритма: количество пчел-агентов, используемых для поиска проектного решения в фазах рабочих пчел; количество пчел-наблюдателей; условие завершения процедуры поиска.

Фаза рабочих пчел образована совокупностью нескольких действий. Прежде всего выполнен расчет значений целевых функций для каждой проектной альтернативы и выбор наилучшего проектного решения среди возможных. Любая проектная альтернатива может быть определена двумя способами. Первый способ — представление проектных альтернатив или вершин графа метаописаний онтологии в вербальной форме, а второй — в числовой форме. Причем если вершины заданы вербально, то каждой из них поставлено в соответствие числовое значение, используемое для расчета значения целевой функции.

Далее проведен поиск и расчет значения целевой функции соседнего проектного решения для выбранного наилучшего решения, по результатам которого реализован выбор лучшего проектного решения сравне-

нием значений целевых функций. Для поиска соседнего решения используется любой способ, например, один из предложенных в работах [27, 28, 32–34]. Расчет значения целевой функции и выбор наилучшего проектного решения осуществлен на основе критерия подобия.

Фаза пчел-наблюдателей включает действия, аналогичные действиям в фазе рабочих пчел. Из всех лучших решений, найденных рабочими пчелами, выбирается одно наилучшее с некоторой вероятностью. Для этого может быть использована одна из схем, предложенных, например в работе [35]. Поиск и расчет значения целевой функции соседнего решения для выбранного решения, по результатам которого реализуется выбор лучшего проектного решения, происходит сравнением значений целевых функций, что соответствует ранее изложенному подходу.

Фаза пчел-разведчиков заключается в формировании нового пространства поиска проектных решений на основе анализа проектных решений, отвергнутых на предыдущих фазах. Если не выполнено условие завершения поиска, то осуществляется возврат к фазе рабочих пчел; иначе полученное проектное решение сохраняется и алгоритм завершается.

Алгоритм генерации проектных решений

В настоящей работе на перечисленных фазах выполнены следующие этапы.

Формирование элементов множества $\mathbf{rsG} \subset \mathbf{G}$ по модели \mathbf{G} , а также вычисление значений функций принадлежности параметров \mathbf{U} и требований \mathbf{RP} выполняется на основе работ [15] и [17] соответственно.

- Этап формирования начального пространства поиска проектных решений означает вычисление значений $\mathbf{\Omega} = (\Omega^{C_1}, \Omega^{C_2}, \dots, \Omega^{C_k})$, $\mathbf{W} = (w_1', w_2', \dots, w_k')$ и вектора-столбца $\mathbf{\beta}$ в соответствии с выражениями (1) и (4). В итоге формируются два множества: $\mathbf{rsGSearch} \subseteq \mathbf{rsG}$, состоящего из незаменяемых элементов матрицы $\mathbf{Z}(\mathbf{C})$, и $\mathbf{rsGStorage} \subseteq \mathbf{rsG}$, состоящее из всех остальных элементов.
- Этап выбора и оценки проектных решений заключается в вычислении значений $\mathbf{\Omega}$ и $\Delta(\mathbf{\beta})$ для элементов множества $\mathbf{rsGSearch}$ на основе (2) и (3).
- На этапе поиска новых проектных решений вычислим и проанализируем значения расстояния Хэмминга между элементами анализируемого множества.
- Этап поиска лучших проектных решений содержит действия для фаз рабочих пчел и пчел-наблюдателей. Фазе рабочих пчел соответствует расчет вероятностей выбора каждого лучшего проектного решения. В фазе пчел-наблюдателей формируется множество лучших проектных решений, образующих искомое проектное решение.
- Этап формирования проектных решений соответствует фазе пчел-разведчиков. В отличие от традиционного подхода в настоящей работе не выполнен анализ проектных решений, которые были отброшены на предыдущих фазах. Тогда рассматриваемый этап заключается только в анализе элементов множества $\mathbf{rsGStorage}$.

Опишем вариант реализации каждого из этапов подробнее.

Формирование начального пространства поиска проектных решений. Рассматриваемый этап представлен следующей последовательностью:

- 1) расчет значений элементов множеств $\mathbf{\Omega}$ и \mathbf{W} для каждого $\mathbf{rsG}_i \in \mathbf{rsG}$ ($i = 1, k$) согласно выражению (1);
- 2) расчет значений элементов вектора-столбца $\mathbf{\beta}$ по выражению (4);
- 3) формирование элементов множеств $\mathbf{rsGSearch}$ и $\mathbf{rsGStorage}$ путем разбиения множества \mathbf{rsG} согласно анализу значений элементов модели $\mathbf{Z}(\mathbf{C})$.

Элементы сформированного множества $\mathbf{rsGSearch}$ используются при выполнении последующих этапов выбора и оценки проектных решений, поиска новых и лучших проектных решений, а элементы множества $\mathbf{rsGStorage}$ — только на последнем этапе формирования проектного решения.

Выбор и оценка проектных решений. Рассматриваемый этап состоит из:

- 1) расчета значений целевой функции $\Delta(\mathbf{\beta})$ согласно (3) и значений $\mathbf{\Omega}^{C_i} \in \mathbf{\Omega}$ ($i = 1, k$) согласно (2) для каждого элемента множества $\mathbf{rsGSearch}$;
- 2) выбора проектных решений, удовлетворяющих критерию подобия (3), для чего проведено сравнение элементов множества $\mathbf{rsGSearch}_i \in \mathbf{rsGSearch}$ ($i = 1, k$) между собой.

Сформированное в результате множество проектных решений, удовлетворяющих критерию подобия, используется для поиска новых проектных решений.

Поиск новых проектных решений. Рассматриваемый этап выполняется в двух фазах: рабочих пчел и пчел-наблюдателей. Этап определяет анализируемое множество проектных решений $\mathbf{analyze_set}$ в фазах: рабочих пчел $\mathbf{analyze_set} \subset \mathbf{rsGSearch}$ и пчел-наблюдателей $\mathbf{analyze_set} \subset \mathbf{rsGBest}$.

Для каждого элемента множества $\mathbf{analyze_set}$ и множества проектных решений, полученного на предыдущем этапе, выполняется расчет расстояния Хэмминга. Элементы, значения расстояния Хэмминга которых минимально, включены в состав множества $\mathbf{result_set}$ в фазах: рабочих пчел $\mathbf{result_set} \subset \mathbf{bestGSearch}$ и пчел-наблюдателей $\mathbf{result_set} \subset \mathbf{rsGBest}$.

Сформированное в результате множество используется для поиска лучших проектных решений.

Поиск лучших проектных решений. Рассматриваемый этап в фазе рабочих пчел означает расчет вероятностей выбора лучших решений из множества $\mathbf{bestGSearch}$, выполняемый любым известным способом, например, описанным в работе [35].

Отметим, что в фазе пчел-наблюдателей формируется множество лучших проектных решений, образующих проектное решение, удовлетворяющее требованиям \mathbf{RP} . Для каждого элемента $\mathbf{rsGBest}_i \in \mathbf{rsGBest}$ ($i = 1, k$) выполняется сравнение значений критерия подобия (3). Лучшие проектные решения — с минимальным значением критерия подобия — включаются в состав искомого решения $\mathbf{resX} \subset \mathbf{X}(\mathbf{C})$.

Полученное на этом этапе проектное решение передается на следующий этап для его дополнения.

Формирование проектных решений. Этап означает поиск элементов множества $\mathbf{rsGStorage}$ для формирования искомого проектного решения $\mathbf{resX} \subset \mathbf{X(C)}$. Для каждого элемента $\mathbf{rsGStorage}_j \in \mathbf{rsGStorage}$ ($j = |\mathbf{rsGStorage}|$) сравнивается значение критерия подобия со значением, соответствующим элементу $\mathbf{resX}_i \in \mathbf{resX}$ ($i = 1, k$).

Поскольку элементами множества $\mathbf{rsGStorage}$ являются элементы, аналогичные элементам множества $\mathbf{rsGSearch}$ в силу модели замещения, то логично утверждать, что значения критериев подобия двух аналогичных элементов будут равны. Тогда элемент $\mathbf{rsGStorage}_j \in \mathbf{rsGStorage}$ включается в искомое проектное решение.

Получаемый тем самым результат — искомое проектное решение, удовлетворяющее предъявленным к нему требованиям.

Пример реализации алгоритма

Для примера работы предложенного алгоритма рассмотрим решение задач поиска данных для управления проектно-производственными процедурами на примере цифрового паспорта электронного изделия из работы [15]. Пусть элементы моделей \mathbf{C} , $\bar{\mathbf{U}}$, \mathbf{rsG} , $\mathbf{X(C)}$ и $\mathbf{Z(C)}$ соответствуют требованиям \mathbf{RS} конкретного предприятия. Для краткости представим часть элементов и их значений, достаточных для демонстрации работы алгоритма.

Поскольку по [15] предприятие специализируется на выпуске блоков и модулей на основе печатных плат, приобретаемых у контрагентов, то элементы модели \mathbf{rsG} имеют вид:

$\mathbf{rsG}_1 = \langle \text{управление актуализацией данных о покупных изделиях, Объект «Изделие» (обозначение; наименование; версия; состояние; признак изготовителя (покупка); вид изделия; ограничительный признак)} \rangle$;

$\mathbf{rsG}_2 = \langle \text{разработка и изменение данных печатной платы, Объект «Изделие» вида «Сборочная единица» (обозначение (АБВГ.ХХХХХХ.001); наименование (Плата управления); представление; вид изделия; версия; состояние; признак изготовителя (собственное производство); признак изделия; габариты изделия; масса изделия)} \rangle$;

$\mathbf{rsG}_3 = \langle \text{разработка и изменение данных печатной платы, Объект «CAD-Документ» вида «Модель детали» (обозначение (АБВГ.ХХХХХХ.001); наименование (Плата управления); версия (0); состояние (сдано в архив); первичная применяемость (АБВГ.ХХХХХХ.001); вид модели)} \rangle$;

$\mathbf{rsG}_4 = \langle \text{разработка и изменение данных печатной платы, Объект «CAD-Документ» вида «Чертеж детали» (обозначение (АБВГ.ХХХХХХ.001); наименование (Плата управления); вид документа (конструкторский документ); версия (0); состояние (сдано в архив); первичная применяемость (АБВГ.ХХХХХХ.001); формат (А3); количество листов документа (2)} \rangle$;

$\mathbf{rsG}_5 = \langle \text{разработка и изменение данных печатной платы, Объект «Документ» вида «Спецификация» (обозначение (АБВГ.ХХХХХХ.001); наименование (Плата управления); вид документа (конструкторский доку-$

мент); версия (2); состояние (на проверке); первичная применяемость (АБВГ.ХХХХХХ.002); формат (А4); количество листов документа (5)} \rangle;

$\mathbf{rsG}_6 = \langle \text{разработка и изменение конструкции прибора, Объект «Изделие» (обозначение (АБВГ.ХХХХХХ.002); наименование (Прибор управления); представление; вид изделия; версия; состояние; признак изготовителя (собственное производство); признак изделия; габариты изделия; масса изделия)} \rangle$;

$\mathbf{rsG}_7 = \langle \text{разработка и изменение конструкции прибора, Объект «CAD-Документ» вида «Модель сборочной единицы» (обозначение (АБВГ.ХХХХХХ.002); наименование (Прибор управления); версия (0); состояние (сдано в архив); первичная применяемость (АБВГ.ХХХХХХ.002); вид модели)} \rangle$;

$\mathbf{rsG}_8 = \langle \text{разработка и изменение конструкции прибора, Объект «CAD-Документ» вида «Сборочный чертеж» (обозначение (АБВГ.ХХХХХХ.002СБ); наименование (Прибор управления Сборочный чертеж); вид документа (конструкторский документ); версия (0); состояние (сдано в архив); первичная применяемость (АБВГ.ХХХХХХ.002); формат (А3); количество листов документа (2)} \rangle$;

$\mathbf{rsG}_9 = \langle \text{разработка и изменение конструкции прибора, Объект «Документ» вида «Спецификация» (обозначение (АБВГ.ХХХХХХ.002); наименование (Прибор управления); вид документа (конструкторский документ); версия (0); состояние (Сдано в архив); первичная применяемость; формат (А4); количество листов документа (7)} \rangle$;

$\mathbf{rsG}_{10} = \langle \text{разработка и изменение программного обеспечения прибора, Объект «Изделие» (обозначение; наименование; представление; вид изделия; версия; состояние; признак изготовителя (собственное производство); признак изделия)} \rangle$;

$\mathbf{rsG}_{11} = \langle \text{разработка и изменение программного обеспечения прибора, Объект «Документ» (обозначение; наименование; вид документа; версия; состояние; первичная применяемость)} \rangle$;

$\mathbf{rsG}_{12} = \langle \text{разработка и изменение эксплуатационной документации, Объект «Документ» вида «Руководство по эксплуатации» (обозначение (АБВГ.ХХХХХХ.002РЭ); наименование (Прибор управления Руководство по эксплуатации); вид документа (эксплуатационный документ); версия (0); состояние (На проверке); первичная применяемость (АБВГ.ХХХХХХ.002); формат (А4); количество листов документа(40)} \rangle$.

При этом количество элементов в модели компонентов цифрового паспорта равно 12, а элементов в модели параметров — 16.

Элементами модели $\mathbf{X(C)}$ являются:

$X_{2,1}(C_2, C_1) = 1, X_{2,3}(C_2, C_3) = 1, X_{2,4}(C_2, C_4) = 1, X_{2,5}(C_2, C_5) = 1, X_{6,1}(C_6, C_1) = 1, X_{6,2}(C_6, C_2) = 1, X_{6,7}(C_6, C_7) = 1, X_{6,8}(C_6, C_8) = 1, X_{6,9}(C_6, C_9) = 1, X_{6,10}(C_6, C_{10}) = 1, X_{6,12}(C_6, C_{12}) = 1, X_{10,11}(C_{10}, C_{11}) = 1$. Остальные элементы равны нулю.

Элементами модели $\mathbf{Z(C)}$ являются:

$Z_{2,6}(C_2, C_6) = 1, Z_{2,10}(C_2, C_{10}) = 1, Z_{3,4}(C_3, C_4) = 1, Z_{3,7}(C_3, C_7) = 1, Z_{3,8}(C_3, C_8) = 1, Z_{4,3}(C_4, C_3) = 1, Z_{4,5}(C_4, C_5) = 1, Z_{4,8}(C_4, C_8) = 1, Z_{4,9}(C_4, C_9) = 1,$

Таблица. Результаты выполнения этапов разработанного алгоритма

Table. Results of the developed algorithm stages

Этапы реализации алгоритма		Задача поиска данных	
		о поставляемых заказчику изделиях и их составных частях	о составных частях изделий, подлежащих производству
Требования RP	вид изделия и его составной части	«сборочная единица»	
	признак изготовителя составной части изделия	«собственное производство»	
	вид документа	«эксплуатационный документ»	«конструкторский документ»
	контролируемое состояние	«сдано в архив»	«запущено в производство»
Расчет значений функций принадлежности		$(\omega_1^{C_1}, \omega_2^{C_1}, \dots, \omega_{16}^{C_1}), (\omega_1^{C_2}, \omega_2^{C_2}, \dots, \omega_{16}^{C_2}), \dots, (\omega_1^{C_{12}}, \omega_2^{C_{12}}, \dots, \omega_{16}^{C_{12}}),$ в том числе $(w_1', w_2', \dots, w_{12}')$.	
Этап формирования начального пространства поиска проектных решений		rsGSearch = (rsG_2, rsG_{12}) ; rsGStorage = (rsG_6) .	rsGSearch = (rsG_2, rsG_4) ; rsGStorage = $(rsG_6, rsG_5, rsG_8, rsG_9)$.
Последующие этапы		Выполнение циклов этапов выбора и оценки проектных решений, поиска новых проектных решений, поиска лучших проектных решений, а также формирования проектных решений.	
Результат работы		resX = (rsG_6, rsG_{12}) . Содержит обозначение и наименование объектов «Изделие», обозначение, наименование, версия и состояние эксплуатационного документа (объект «Документ» вида «Руководство по эксплуатации»).	resX = $(rsG_6, rsG_8, rsG_9, rsG_2, rsG_4)$. Содержит обозначение и наименование объектов «Изделие», обозначение, наименование, версия и состояние конструкторских документов (объект «CAD-Документ» вида «Модель ...» и вида «Чертеж ...», а также объект «Документ» вида «Спецификация»).

$Z_{4,11}(C_4, C_{11}) = 1, Z_{4,12}(C_4, C_{12}) = 1, Z_{5,4}(C_5, C_4) = 1,$
 $Z_{5,8}(C_5, C_8) = 1, Z_{5,9}(C_5, C_9) = 1, Z_{5,11}(C_5, C_{11}) = 1,$
 $Z_{5,12}(C_5, C_{12}) = 1, Z_{6,2}(C_6, C_2) = 1, Z_{6,10}(C_6, C_{10}) = 1,$
 $Z_{7,3}(C_7, C_3) = 1, Z_{8,3}(C_8, C_3) = 1, Z_{8,4}(C_8, C_4) = 1,$
 $Z_{8,5}(C_8, C_5) = 1, Z_{8,9}(C_8, C_9) = 1, Z_{8,11}(C_8, C_{11}) = 1,$
 $Z_{8,12}(C_8, C_{12}) = 1, Z_{9,4}(C_9, C_4) = 1, Z_{9,5}(C_9, C_5) = 1,$
 $Z_{9,8}(C_9, C_8) = 1, Z_{9,11}(C_9, C_{11}) = 1, Z_{9,12}(C_9, C_{12}) = 1,$
 $Z_{10,2}(C_{10}, C_2) = 1, Z_{10,6}(C_{10}, C_6) = 1, Z_{11,4}(C_{11}, C_4) = 1,$
 $Z_{11,5}(C_{11}, C_5) = 1, Z_{11,8}(C_{11}, C_8) = 1, Z_{11,9}(C_{11}, C_9) = 1,$
 $Z_{11,12}(C_{11}, C_{12}) = 1, Z_{12,4}(C_{12}, C_4) = 1, Z_{12,5}(C_{12}, C_5) = 1,$
 $Z_{12,8}(C_{12}, C_8) = 1, Z_{12,9}(C_{12}, C_9) = 1, Z_{12,11}(C_{12}, C_{11}) = 1.$
Остальные элементы равны нулю.

Пусть на предприятии сформулированы задачи поиска данных об изделиях и их составных частях, решаемые при управлении проектно-производственными процедурами. Управление проектными процедурами связано с задачей поиска данных о поставляемых заказчику изделиях и их составных частях. Управление производственной деятельностью — с поиском данных об изделиях, подлежащих производству. Пример основных требований **RP**, сформулированных для каждого из них, а также результаты этапов разработанного алгоритма приведены в таблице.

На первом этапе полученные результаты могут использоваться для принятия решений при управлении проектными и производственными процедурами, на втором — для обеспечения функционирования цифровых двойников приборов, выпускаемых предприятием.

Результат применения алгоритма — получение элементов онтологии с атрибутами: обозначение и наименование продукции контрагентов; наименование предприятия-изготовителя.

Заключение

В работе представлен алгоритм генерации проектных решений для управления данными об изделии и проектно-производственными процедурами, основанный на поведении пчелиной колонии. Предложены реализации нескольких этапов алгоритма, выполняемых в фазах инициализации, рабочих пчел, пчел-наблюдателей и пчел-разведчиков.

Этап формирования начального пространства поиска проектных решений означает формирование элементов двух множеств: **rsGSearch**, содержащего анализируемые описания компонентов цифрового паспорта, и **rsGStorage**, состоящего из всех остальных компонентов. Причем в состав множеств включаются только те компоненты цифрового паспорта, которые удовлетворяют требованиям **RS** конкретного предприятия.

Этап выбора и оценки проектных решений заключается в вычислении значений критерия подобия для элементов множества **rsGSearch**. На этапе поиска новых проектных решений вычисляют и сравнивают значения расстояния Хэмминга между вариантами проектных решений. Тем самым формируется множество вариантов

лучших проектных решений, удовлетворяющих требованиям **RP**.

Этап поиска лучших проектных решений в фазе рабочих пчел соответствует расчету вероятностей выбора каждого лучшего проектного решения, а в фазе пчел-наблюдателей — формированию множества лучших проектных решений, образующих искомое проектное решение.

Литература

1. Петров Р.Г. Технология разработки интегрированных систем управления техническими средствами на базе типовых приборов, типовой документации и программного обеспечения с использованием средств автоматизированного проектирования // Системы управления и обработки информации. 2016. № 1(32). С. 80–87.
2. Донецкая Ю.В. Цели и задачи разработки интегрированных систем управления данными // Информационные технологии в профессиональной деятельности и научной работе: сборник материалов Всероссийской научно-практической конференции с международным участием: в 2-х частях. Ч. 2. 2012. С. 8–11.
3. Седых Д.В., Белоусов С.В., Василенко М.Н. Автоматизация составления инструкций по пользованию устройствами сигнализации, централизации и блокировки // Известия Петербургского университета путей сообщения. 2017. Т. 14. № 2. С. 320–332.
4. Василенко М.Н., Бубнов В.П., Булавский П.Е., Василенко П.А. Ошибки в технической документации железнодорожной автоматики и телемеханики и их влияние на безопасность движения поездов // Автоматика на транспорте. 2019. Т. 5. № 1. С. 94–112. <https://doi.org/10.20295/2412-9186-2019-1-94-112>
5. Гурьянов А.В., Шукалов А.В., Жаринов И.О., Костишин М.О., Леоновец С.А. Программные шаблоны для автоматизированного оформления текстовых конструкторских документов на изделия авиационной промышленности // Вестник Череповецкого государственного университета. 2017. № 2. С. 15–22. <https://doi.org/10.23859/1994-0637-2017-2-77-2>
6. Гурьянов А.В., Коновалов П.В., Шукалов А.В., Жаринов И.О., Леоновец С.А. Автоматизация процесса формирования учетной документации с использованием базы данных радиоэлектронных компонентов // Программные продукты и системы. 2017. № 3(30). С. 517–523. <https://doi.org/10.15827/0236-235X.030.3.517-523>
7. Гурьянов А.В., Шукалов А.В., Жаринов И.О., Костишин М.О., Леоновец С.А. Автоматизация процессов подготовки программной документации на изделия радиоэлектронной промышленности // Программные продукты и системы. 2017. № 3(30). С. 504–509. <https://doi.org/10.15827/0236-235X.030.3.504-509>
8. Охтилев М.Ю., Гамов В.Ю., Черников А.Д. Создание единого виртуального электронного паспорта космической ракеты-носителя «Союз-2»: этапы, концепция и принципы построения, модель электронной структуры изделия // I-methods. 2018. Т. 10. № 4. С. 11–23.
9. Эртман Л.В., Рудаков В.Б., Бурцев А.С., Бакланов В.И., Филоненко П.А. Технология статистического контроля свойств наноматериалов и покрытий в условиях воздействия ионизирующих излучений и оценки надежности изделий космической техники, созданных на их основе // Инженерный журнал: наука и инновации. 2020. № 9. С. 8. <https://doi.org/10.18698/2308-6033-2020-9-2018>
10. Гурьянов А.В., Шукалов А.В., Заколдаев Д.А., Жаринов И.О., Нечаев В.А. Организация электронного документооборота между проектным, производственным и эксплуатирующим предприятиями в условиях цифровой экономики Индустрии 4.0 // Научно-технический вестник информационных технологий, механики и оптики. 2018. Т. 18. № 1. С. 106–112. <https://doi.org/10.17586/2226-1494-2018-18-1-106-112>
11. Автамонов П.Н., Бахмут А.Д., Крылов А.В., Охтилев М.Ю., Охтилев П.А., Соколов Б.В. Применение технологии поддержки принятия решений на различных этапах жизненного цикла космических средств в составе системы информации о техническом состоянии и надёжности // Вестник Самарского университета. Аэрокосмическая техника, технологии и машиностроение. 2017. Т. 16. № 3. С. 173–184. <https://doi.org/10.18287/2541-7533-2017-16-3-173-184>

Этап формирования проектных решений соответствует анализу элементов множества **rsGStorage**, что дополняет искомое проектное решение описаниями компонентов цифрового паспорта.

На основе полученных результатов предполагается разработать сигнатуры и семантики унифицированных сервисов для применения цифрового паспорта на предприятиях приборостроительной отрасли.

References

1. Petrov R.G. Technology of development of integrated control systems of technical facilities based on standard devices, standard documentation and software with use of computer assisted design facilities. *Control and Data Processing Systems*, 2016, no. 1(32), pp. 80–87. (in Russian)
2. Donetskaya Yu.V. Goals and objectives of the development of integrated data management systems. *Information technologies in professional activity and scientific work: collection of materials of the All-Russian scientific and practical conference with international participation*. In 2 parts. Part 2, 2012, pp. 8–11. (in Russian)
3. Sedykh D.V., Belousov S.V., Vasilenko M.N. Automation of instruction sheets on attention devices and signaling circuits compilation. *Proceedings of Petersburg Transport University*, 2017, vol. 14, no. 2, pp. 320–332. (in Russian)
4. Vasilenko M.N., Bubnov V.P., Bulavsky P.E. Errors in the technical documentation of railway automation and remote control and their impact on the safety of train traffic. *Transport automation research*, 2019, vol. 5, no. 1, pp. 94–112. (in Russian). <https://doi.org/10.20295/2412-9186-2019-1-94-112>
5. Guryanov A.V., Shukalov A.V., Zharinov I.O., Kostishin M.O., Leonovets S.A. Program templates for automated designs of text design documents on products of the aviation industry. *Cherepovets State University Bulletin*, 2017, no. 2, pp. 15–22. (in Russian). <https://doi.org/10.23859/1994-0637-2017-2-77-2>
6. Guryanov A.V., Konovalov P.V., Shukalov A.V., Zharinov I.O., Leonovets S.A. Automated generation of accounting documents using radioelectronic components database. *Software & Systems*, 2017, no. 3(30), pp. 517–523. (in Russian). <https://doi.org/10.15827/0236-235X.030.3.517-523>
7. Guryanov A.V., Shukalov A.V., Zharinov I.O., Kostishin M.O., Leonovets S.A. Automatization of software documentation preparation processes for radio-electronics industry devices. *Software & Systems*, 2017, no. 3(30), pp. 504–509. (in Russian). <https://doi.org/10.15827/0236-235X.030.3.504-509>
8. Okhtilev M.Yu., Gamov V.Yu., Chernikov A.D. Establishing a single virtual electronic passport of a space-socket-mounter “Soyuz-2”: stages, concept and principles of construction, model of the electronic structure of the product. *I-methods*, 2018, vol. 10, no. 4, pp. 11–23. (in Russian)
9. Ertman L.V., Rudakov V.B., Burtsev A.S., Baklanov V.I., Filonenko P.A. Technology of statistical control of the properties of nanomaterials and coatings exposed to ionizing radiation and assessment of the reliability of nanomaterial space products. *Engineering Journal: Science and Innovation*, 2020, no. 9, pp. 8. (in Russian). <https://doi.org/10.18698/2308-6033-2020-9-2018>
10. Gurjanov A.V., Shukalov A.V., Zakoldaev D.A., Zharinov I.O., Nechaev V.A. Electronic document flow between project, production and operating enterprises in the context of industry 4.0 digital economy. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2018, vol. 18, no. 1, pp. 106–112. (in Russian). <https://doi.org/10.17586/2226-1494-2018-18-1-106-112>
11. Avtamonov P.N., Bakhmut A.D., Krylov A.V., Okhtilev M.Yu., Okhtilev P.A., Sokolov B.V. Application of decision support technology at various stages of the life cycle of space facilities in assembly with the information system of technical condition and reliability. *Vestnik of Samara University. Aerospace and Mechanical Engineering*, 2017, vol. 16, no. 3, pp. 173–184. (in Russian). <https://doi.org/10.18287/2541-7533-2017-16-3-173-184>
12. Okhtilev M.YU., Klyucharyov A.A., Okhtilev P.A., Zyanchurin A.E. Technology of automated information and analytical support of the product life cycle on the example of unified virtual electronic passport of space facilities. *Journal of Instrument Engineering*, 2020, vol. 63,

12. Охтилев М.Ю., Ключарев А.А., Охтилев П.А., Зянчури А.Э. Технология автоматизированной информационно-аналитической поддержки жизненного цикла изделий на примере единого виртуального электронного паспорта космических средств // Известия ВУЗов. Приборостроение. 2020. Т. 63. № 11. С. 1012–1019. <https://doi.org/10.17586/0021-3454-2020-63-11-1012-1019>
13. Афанасьев В.Б. Онтологическое проектирование автоматизированной информационной системы поддержки качества продукции предприятия // Известия Тульского государственного университета. Технические науки. 2020. № 10. С. 12–22.
14. Donetskaya Ju.V., Gatchin Yu.A. Development of requirements for the content of a digital passport and design solutions // *Journal of Physics: Conference Series*. 2021. V. 1828. N 1. P. 012102. <https://doi.org/10.1088/1742-6596/1828/1/012102>
15. Донецкая Ю.В. Метод формирования и использования цифрового паспорта электронного изделия на предприятиях приборостроительной отрасли // Научно-технический вестник информационных технологий, механики и оптики. 2021. Т. 21. № 6. С. 969–976. <https://doi.org/10.17586/2226-1494-2021-21-6-969-976>
16. Donetskaya Ju.V. Creation of a digital passport for an electronic product and generation of design solutions based on it // *Frontiers in Artificial Intelligence and Applications*. 2021. V. 340. P. 379–385. <https://doi.org/10.3233/FAIA210210>
17. Donetskaya Ju.V. Design procedures for the analysis of the components and parameters of a digital passport // *Journal of Physics: Conference Series*. 2021. V. 1828. N 1. P. 012103. <https://doi.org/10.1088/1742-6596/1828/1/012103>
18. Donetskaya Ju.V., Gatchin Yu.A. Development of design procedures for the synthesis of design solutions for data management, design and production procedures at the stages of the life cycle of an electronic product // *Proc. of the 2020 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF)*. 2020. P. 9131470. <https://doi.org/10.1109/WECONF48837.2020.9131470>
19. Хижняков Ю.Н. Алгоритмы нечеткого, нейронного и нейро-нечеткого управления в системах реального времени: учебное пособие. Пермь: Изд-во ПНИПУ, 2013. 160 с.
20. Семенова А.В., Курейчик В.М. Оптимизация отображения онтологий методом роя частиц // *Онтология проектирования*. 2018. Т. 8. № 2(28). С. 285–295. <https://doi.org/10.18287/2223-9537-2018-8-2-285-295>
21. Кравченко Д.Ю., Кравченко А.Ю., Марков В.В. Гибридный биоинспирированный алгоритм отображения онтологий в задачах извлечения и управления данными // *Известия ЮФУ. Технические науки*. 2020. № 2(212). С. 16–28. <https://doi.org/10.18522/2311-3103-2020-2-16-28>
22. Andreasen T., Knappe R., Bulskov H. Domain specific similarity and retrieval // *Proc. of the 11th International Fuzzy Systems Association World Congress. V. 1*. 2005. P. 496–502.
23. Castano S., Ferrara A., Montanelli S., Racca G. Semantic information interoperability in open networked systems // *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2004. V. 3226. P. 215–230. https://doi.org/10.1007/978-3-540-30145-5_13
24. Кравченко Ю.А., Курситыс И.О. Биоинспирированный алгоритм приобретения новых знаний на основе классификации информационных ресурсов // *Известия ЮФУ. Технические науки*. 2019. № 2. С. 15–26. <https://doi.org/10.23683/2311-3103-2019-2-15-26>
25. Haase P., Siebes R., Van Harmelen F. Peer selection in peer-to-peer networks with semantic topologies // *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2004. V. 3226. P. 108–125. https://doi.org/10.1007/978-3-540-30145-5_7
26. Maedche A., Zacharias V. Clustering ontology-based metadata in the Semantic Web // *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2002. V. 2431. P. 348–360. https://doi.org/10.1007/3-540-45681-3_29
27. Бова В.В., Лещанов Д.В. Семантический поиск знаний в среде функционирования междисциплинарных информационных систем на основе онтологического подхода // *Известия ЮФУ. Технические науки*. 2017. № 7(192). С. 79–90. <https://doi.org/10.23683/2311-3103-2017-7-79-90>
28. Марков В.В., Кравченко Ю.А., Кузьмина М.А. Развитие методов семантической фильтрации на основе решения задачи кластеризации биоинспирированными алгоритмами // *Известия ЮФУ*. no. 11, pp. 1012–1019. (in Russian). <https://doi.org/10.17586/0021-3454-2020-63-11-1012-1019>
13. Afanasiev V.B. Ontological design of an automated information system for supporting the quality of products of the enterprise. *Izvestiya Tula State University. Technical Sciences*, 2020, no. 10, pp. 12–21. (in Russian)
14. Donetskaya Ju.V., Gatchin Yu.A. Development of requirements for the content of a digital passport and design solutions. *Journal of Physics: Conference Series*, 2021, vol. 1828, no. 1, pp. 012102. <https://doi.org/10.1088/1742-6596/1828/1/012102>
15. Donetskaya Ju. V. The method of forming and using a digital passport for an electronic product at enterprises of the instrument-making industry. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2021, vol. 21, no. 6, pp. 969–976. (in Russian). <https://doi.org/10.17586/2226-1494-2021-21-6-969-976>
16. Donetskaya Ju.V. Creation of a digital passport for an electronic product and generation of design solutions based on it. *Frontiers in Artificial Intelligence and Applications*, 2021, vol. 340, pp. 379–385. <https://doi.org/10.3233/FAIA210210>
17. Donetskaya Ju.V. Design procedures for the analysis of the components and parameters of a digital passport. *Journal of Physics: Conference Series*, 2021, vol. 1828, no. 1, pp. 012103. <https://doi.org/10.1088/1742-6596/1828/1/012103>
18. Donetskaya Ju.V., Gatchin Yu.A. Development of design procedures for the synthesis of design solutions for data management, design and production procedures at the stages of the life cycle of an electronic product. *Proc. of the 2020 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF)*, 2020, pp. 9131470. <https://doi.org/10.1109/WECONF48837.2020.9131470>
19. Khizhnyakov Yu.N. Fuzzy, Neural and Neuro-Fuzzy Control Algorithms in Real-Time Systems. Tutorial. Perm, Publishing House of PNIPU, 2013, 160 p. (in Russian)
20. Semenova A.V., Kureichik V.M. Ontology mapping using the method of particle swarm optimization. *Ontology of Designing*, 2018, vol. 8, no. 2(28), pp. 285–295. (in Russian). <https://doi.org/10.18287/2223-9537-2018-8-2-285-295>
21. Kravchenko D.Yu., Kravchenko A.Yu., Markov V.V. Hybrid bioinspired algorithm for ontologies mapping in the tasks of extraction and knowledge management. *Izvestiya SFedU. Engineering Sciences*, 2020, no. 2(212), pp. 16–28. (in Russian). <https://doi.org/10.18522/2311-3103-2020-2-16-28>
22. Andreasen T., Knappe R., Bulskov H. Domain specific similarity and retrieval. *Proc. of the 11th International Fuzzy Systems Association World Congress. V. 1*, 2005, pp. 496–502.
23. Castano S., Ferrara A., Montanelli S., Racca G. Semantic information interoperability in open networked systems. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2004, vol. 3226, pp. 215–230. https://doi.org/10.1007/978-3-540-30145-5_13
24. Kravchenko Yu.A., Kursitys I.O. Bioinspired algorithm for acquiring new knowledge on the basis of the information resources classification. *Izvestiya SFedU. Engineering Sciences*, 2019, no. 2, pp. 15–26. (in Russian). <https://doi.org/10.23683/2311-3103-2019-2-15-26>
25. Haase P., Siebes R., Van Harmelen F. Peer selection in peer-to-peer networks with semantic topologies. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2004, vol. 3226, pp. 108–125. https://doi.org/10.1007/978-3-540-30145-5_7
26. Maedche A., Zacharias V. Clustering ontology-based metadata in the Semantic Web. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2002, vol. 2431, pp. 348–360. https://doi.org/10.1007/3-540-45681-3_29
27. Bova V.V., Leshchanov D.V. The semantic search of knowledge in the environment of operation of interdisciplinary information systems based on ontological approach. *Izvestiya SFedU. Engineering Sciences*, 2017, no. 7(192), pp. 79–90. (in Russian). <https://doi.org/10.23683/2311-3103-2017-7-79-90>
28. Markov V.V., Kravchenko Yu.A., Kuzmina M.A. Development of semantic filtering methods based on solving the task of clustering by bioinspired algorithms. *Izvestiya SFedU. Engineering Sciences*, 2018, no. 4(198), pp. 175–185. (in Russian). <https://doi.org/10.23683/2311-3103-2018-4-175-185>

- Технические науки. 2018. № 4(198). С. 175–185. <https://doi.org/10.23683/2311-3103-2018-4-175-185>
29. Karaboga D., Gorkemli B., Ozturk C., Karaboga N. A comprehensive survey: artificial bee colony (ABC) algorithm and applications // *Artificial Intelligence Review*. 2014. V. 42. N 1. P. 21–57. <https://doi.org/10.1007/s10462-012-9328-0>
 30. Karaboga D., Akay B. A survey: algorithms simulating bee swarm intelligence // *Artificial Intelligence Review*. 2009. V. 31. N 1-4. P. 61–85. <https://doi.org/10.1007/s10462-009-9127-4>
 31. Bose A., Mali K. Fuzzy-based artificial bee colony optimization for gray image segmentation // *Signal, Image and Video Processing*. 2016. V. 10. N 6. P. 1089–1096. <https://doi.org/10.1007/s11760-016-0863-z>
 32. Wang P., Shi H., Yang X., Mi J. Three-way k -means: integrating k -means and three-way decision // *International Journal of Machine Learning and Cybernetics*. 2019. V. 10. N 10. P. 2767–2777. <https://doi.org/10.1007/s13042-018-0901-y>
 33. Kumar Yu., Shao G. A two-step artificial bee colony algorithm for clustering // *Neural Computing and Applications*. 2017. V. 28. N 3. P. 537–551. <https://doi.org/10.1007/s00521-015-2095-5>
 34. Su Z., Zhou H., Hao Y. Evidential evolving C-means clustering method based on artificial bee colony algorithm with variable strings and interactive evaluation mode // *Fuzzy Optimization and Decision Making*. 2021. V. 20. N 3. P. 293–313. <https://doi.org/10.1007/s10700-020-09344-7>
 35. Awadallah M.A., Al-Betar M.A., Bolaji A.L., Alsukni E.M., Al-Zoubi H. Natural selection methods for artificial bee colony with new versions of onlooker bee // *Soft Computing*. 2019. V. 23. N 15. P. 6455–6494. <https://doi.org/10.1007/s00500-018-3299-2>
 29. Karaboga D., Gorkemli B., Ozturk C., Karaboga N. A comprehensive survey: artificial bee colony (ABC) algorithm and applications. *Artificial Intelligence Review*, 2014, vol. 42, no. 1, pp. 21–57. <https://doi.org/10.1007/s10462-012-9328-0>
 30. Karaboga D., Akay B. A survey: algorithms simulating bee swarm intelligence. *Artificial Intelligence Review*, 2009, vol. 31, no. 1-4, pp. 61–85. <https://doi.org/10.1007/s10462-009-9127-4>
 31. Bose A., Mali K. Fuzzy-based artificial bee colony optimization for gray image segmentation. *Signal, Image and Video Processing*, 2016, vol. 10, no. 6, pp. 1089–1096. <https://doi.org/10.1007/s11760-016-0863-z>
 32. Wang P., Shi H., Yang X., Mi J. Three-way k -means: integrating k -means and three-way decision. *International Journal of Machine Learning and Cybernetics*, 2019, vol. 10, no. 10, pp. 2767–2777. <https://doi.org/10.1007/s13042-018-0901-y>
 33. Kumar Yu., Shao G. A two-step artificial bee colony algorithm for clustering. *Neural Computing and Applications*, 2017, vol. 28, no. 3, pp. 537–551. <https://doi.org/10.1007/s00521-015-2095-5>
 34. Su Z., Zhou H., Hao Y. Evidential evolving C-means clustering method based on artificial bee colony algorithm with variable strings and interactive evaluation mode. *Fuzzy Optimization and Decision Making*, 2021, vol. 20, no. 3, pp. 293–313. <https://doi.org/10.1007/s10700-020-09344-7>
 35. Awadallah M.A., Al-Betar M.A., Bolaji A.L., Alsukni E.M., Al-Zoubi H. Natural selection methods for artificial bee colony with new versions of onlooker bee. *Soft Computing*, 2019, vol. 23, no. 15, pp. 6455–6494. <https://doi.org/10.1007/s00500-018-3299-2>

Автор

Донецкая Юлия Валерьевна — кандидат технических наук, доцент, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, sc57200259228@mail.ru, <https://orcid.org/0000-0001-5293-5025>, donetskaya_julia@mail.ru

Author

Julia V. Donetskaya — PhD, Associate Professor, Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, sc57200259228@mail.ru, <https://orcid.org/0000-0001-5293-5025>, donetskaya_julia@mail.ru

Статья поступила в редакцию 15.02.2022
Одобрена после рецензирования 17.06.2022
Принята к печати 29.07.2022

Received 15.02.2022
Approved after reviewing 17.06.2022
Accepted 29.07.2022



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

doi: 10.17586/2226-1494-2022-22-4-691-698

УДК 004.942

Способ повышения эффективности комплексной обработки данных дистанционного зондирования Земли при решении задач мониторинга пространственных объектов

Сергей Александрович Карин^{1✉}, Александр Иванович Карин²

^{1,2} Военно-космическая академия имени А.Ф. Можайского, Санкт-Петербург, 197198, Российская Федерация

¹ Sergey.karin@gmail.com✉, <https://orcid.org/0000-0001-7139-1959>

² nil5-ksa@yandex.ru, <https://orcid.org/0000-0001-5746-5245>

Аннотация

Предмет исследования. Предложен способ повышения эффективности функционирования системы комплексной обработки данных дистанционного зондирования Земли в условиях ресурсной ограниченности при решении задач мониторинга пространственных объектов. **Методы.** Эффективность функционирования системы увеличена с помощью рационального распределения ее ресурсов по решаемым задачам. Сделан акцент на обеспечении приоритетного решения задач, которые имеют большее значение коэффициента относительной важности. Отмечена тенденция увеличения оперативности получения решения каждой из задач за счет исключения из рабочих планов решений тех ресурсов, которые перегружены другими задачами, но при этом вносят незначительный вклад в формирование интегрального результата. **Основные результаты.** Определены пути повышения эффективности функционирования системы комплексной обработки данных дистанционного зондирования Земли в условиях ресурсной ограниченности при решении задач мониторинга пространственных объектов. Рассмотрены варианты управления системой в указанных условиях и приведены их краткие характеристики. Представлены результаты имитационного моделирования процессов комплексной обработки данных на основе предложенных вариантов управления. **Практическая значимость.** Результаты имитационного моделирования показали, что использование описанного способа повышения эффективности зондирования Земли в условиях ресурсной ограниченности позволяет обеспечить требуемое качество принимаемых управленческих решений, особенно в условиях высокой динамичности изменения характеристик объектов мониторинга.

Ключевые слова

комплексная обработка, данные дистанционного зондирования Земли, ресурсы комплексной обработки, единый технологический цикл обработки данных ДЗЗ

Ссылка для цитирования: Карин С.А., Карин А.И. Способ повышения эффективности комплексной обработки данных дистанционного зондирования Земли при решении задач мониторинга пространственных объектов // Научно-технический вестник информационных технологий, механики и оптики. 2022. Т. 22, № 4. С. 691–698. doi: 10.17586/2226-1494-2022-22-4-691-698

A method for improving the efficiency of integrated processing of Earth remote sensing data in solving problems of spatial objects monitoring

Sergey A. Karin^{1✉}, Alexandr I. Karin²

^{1,2} Mozhaisky Military Aerospace Academy, Saint Petersburg, 197198, Russian Federation

¹ Sergey.karin@gmail.com✉, <https://orcid.org/0000-0001-7139-1959>

² nil5-ksa@yandex.ru, <https://orcid.org/0000-0001-5746-5245>

Abstract

A method is proposed for improving the efficiency of the system for complex processing of data obtained by remote sensing of the Earth in conditions of limited resources in solving problems of spatial objects monitoring. The efficiency of

the system functioning is increased through the rational allocating its resources according to the tasks to be solved, taking into account the priority solution of those that have a higher value of relative importance coefficient. It is also proposed to improve the efficiency of solving each of the tasks on the basis of exclusion from the work plans for solving those resources that are overloaded with other tasks, but at the same time make an insignificant contribution to the formation of an integral result. The simulation results show that the use of the proposed method for improving the efficiency of the system functioning for complex processing of Earth remote sensing data in conditions of limited resources, when solving problems of monitoring spatial objects, makes it possible to ensure the required quality of managerial decisions, especially in conditions of high dynamism of changing the monitoring objects characteristics.

Keywords

complex processing, earth remote sensing data, complex processing resources, single technological cycle of remote sensing data processing

For citation: Karin S.A., Karin A.I. A method for improving the efficiency of integrated processing of Earth remote sensing data in solving problems of spatial objects monitoring. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2022, vol. 22, no. 4, pp. 691–698 (in Russian). doi: 10.17586/2226-1494-2022-22-4-691-698

Введение

Отметим, что современные технические средства дистанционного зондирования Земли (ДЗЗ) воздушно-го и космического базирования — один из основных источников данных для решения информационно-расчетных задач. Решение задач выполняется в интересах государственных и муниципальных органов управления, которые занимаются мониторингом пространственных объектов. Технические средства ДЗЗ предоставляют более двух третей всех используемых при решении информационно-расчетных задач данных. При этом их потенциальные возможности возрастают опережающими темпами [1–4].

Обработка данных ДЗЗ осуществляется специальными программно-техническими комплексами обработки и описывается последовательностью связанных друг с другом операций в рамках единого технологического цикла.

В настоящей работе специальные программно-технические комплексы обработки названы кратко — ресурсами комплексной обработки или ресурсами. При этом каждый ресурс может выполнять одну или несколько операций в рамках единого технологического цикла.

Примеры ресурсов компьютерной обработки: различные комплексы автоматизации; тематические базы и банки данных, в которых накапливаются «сырые» данные и результаты их обработки; подразделения и организации, осуществляющие отдельные сложные этапы специальной обработки в рамках единого технологического цикла. Совокупность ресурсов, предназначенных для решения поступающих задач в рамках единого технологического цикла назовем системой обработки данных ДЗЗ (СОД ДЗЗ). В качестве примеров отдельных операций в рамках единого технологического цикла могут выступать: операции по приему сырых данных ДЗЗ с борта космических аппаратов или воздушных (пилотируемых и беспилотных) средств; операции первичной обработки данных и тематической обработки; разработки готовых отчетных информационных документов.

Основные задачи единого технологического цикла комплексной обработки данных ДЗЗ — вскрытие и анализ мероприятий (событий), которые осуществляются на объектах мониторинга с целью формирования

и доклада лицам, принимающим решения, выводов и аналитических справок по ним. Результаты работы СОД ДЗЗ — формирование отчетных информационных документов, которые содержат указанные сведения.

Следовательно, что для решения задачи по вскрытию одного мероприятия (события) на одном объекте мониторинга требуется разработка одного отчетного информационного документа.

СОД ДЗЗ предназначена для поддержки принятия решений на применение его разнородных и территориально-распределенных ресурсов (в том числе в условиях дефицита), представляемых в качестве совокупности взаимосвязанных сил (отделов, служб) и средств (программно-технических и информационных комплексов, стоящих на балансе государственных и муниципальных органов управления). При этом основные характеристики системы [5–8]: возможность адаптироваться к текущим внешним условиям функционирования; обеспечение автоматизированной поддержки принятия решений в цикле управления, особенно в условиях возрастания дефицита ресурсов, в процессе решения стоящих перед ним задач.

Сочетание данных характеристик способствует повышению качества принимаемых управленческих решений, особенно в условиях высокой динамичности изменения характеристик объектов мониторинга.

Постановка задачи рационального распределения ресурсов комплексной обработки данных ДЗЗ

Реализация единого технологического цикла в СОД ДЗЗ включает несколько этапов, от формирования заявок на разработку отчетных информационных документов и до выдачи потребителям результатов. Каждый ресурс компьютерной обработки при этом обеспечивает выполнение одной операции (подзадачи) в рамках единого технологического цикла, затрачивая при этом необходимое время. Предположим, что в один момент времени отдельный ресурс может осуществлять решение только одной подзадачи, при этом подзадачи могут выполняться параллельно. Исходя из того, как, когда и какую информацию ресурс предоставляет в качестве результата решения своей подзадачи, возможно определить степень его вклада в получение интегрального результата решения информационно-расчетных задач в целом.

Одна из важнейших задач, возникающая при реализации единого технологического цикла, — управление распределением ресурсов СОД ДЗЗ с целью поддержания устойчивого функционирования системы в заданных внешних и внутренних условиях (далее — условия обстановки).

Разнообразие ситуаций, возникающих при управлении ресурсами СОД ДЗЗ, вызвано: высокой динамичностью процессов, протекающих на объектах мониторинга; возрастанием противоречия между возможностями ресурсов СОД ДЗЗ по обработке и интенсивностью потоков решаемых информационно-расчетных задач (т. е. возрастанием ресурсных ограничений) [9, 10]. В связи с этим возникает необходимость оперативной выработки решений на формирование таких планов решения поступающих информационно-расчетных задач, которые удовлетворяли бы разнородным критериям, основным из которых является критерий обеспечения предельно возможной в текущих условиях эффективности функционирования СОД ДЗЗ.

Как показано в работах [11, 12], эффективность функционирования СОД ДЗЗ оценивается на заданном временном интервале $[0, t]$ показателем эффективности. Показатель эффективности определяется как отношение числа решенных задач к числу поступивших, с учетом коэффициентов их относительных значимостей и полноты решения, и формально имеет вид:

$$\Theta(t) = \frac{\sum_{n=1}^{N(t)} \xi_n K_{f,n} K_{a,n}}{\sum_{n=1}^{N(t)} \xi_n}, \quad (1)$$

где $N(t)$ — количество поступивших задач к моменту времени t ; ξ_n — коэффициент относительной важности задачи, $\xi_n \in [0, 1]$, при этом его исходное значение задается заранее потребителем результатов решения задачи; $K_{f,n}$ и $K_{a,n}$ — специальные параметры каждой решенной задачи к моменту времени t , обозначающие уровень завершенности (полноту результатов) и уровень актуальности результатов решения задачи, т. е. $K_{f,n} \in [0, 1]$ и $K_{a,n} \in [0, 1]$ соответственно.

Далее каждую n -ю задачу, которая поступила в момент времени t_n обозначим: $z_n = (t_n, K_{f,n}, K_{a,n}, \xi_n)$.

Таким образом, в соответствии с формулой (1), для повышения эффективности функционирования СКО ДЗЗ требуется обеспечить распределение имеющихся ресурсов по решаемым задачам с двумя условиями. Обоснованно увеличить число вовремя решенных задач с большим значением коэффициента относительной важности (приоритета) и обеспечить максимально возможную полноту результатов решения таких задач.

Исходные данные для решения рассматриваемой задачи рационального распределения ресурсов:

- 1) наличие характеристик имеющихся ресурсов комплексной обработки данных в СОД ДЗЗ;
- 2) наличие характеристик информационно-расчетных задач, решаемых в рамках единого технологического цикла обработки данных в СОД ДЗЗ, которые определяют состав привлекаемых ресурсов;
- 3) интенсивность потоков информационно-расчетных задач различных типов.

Характеристика способа повышения эффективности комплексной обработки данных ДЗЗ в условиях дефицита ресурсов

Модель функционирования СОД ДЗЗ представлена в работе [12] и относится к классу многоканальных сетей массового обслуживания с очередями, реализующими обслуживание с приоритетом.

Рассмотрим некоторые аспекты реализации единого технологического цикла обработки данных при решении информационно-расчетных задач.

Для отработки каждой задачи z_n подсистемой управления функционированием СОД ДЗЗ сформируем рабочий план ее решения $W_n = f(\Omega, z_n)$, где $\Omega = \{\omega_{m,k}\}$, m — тип ресурса, k — номер ресурса m -го типа. Типы ресурсов определены перечнем подзадач, входящих в состав задачи.

Содержательно рабочий план W_n представляет собой перечень элементарных операций обработки исходных данных в виде:

$$W_n = \langle w_1, w_2, w_3, \dots, w_j \rangle,$$

где $w_j = \langle \{\omega, \varepsilon\}_{j-1}, \tau_j, \chi_j, \{\omega\}_{j+1} \rangle$ — j -й пункт плана решения n -й информационно-расчетных задач, $j \in [1, J]$, J — количество элементарных операций в плане W_n .

При этом каждый j -й пункт плана содержит: $\{\omega, \varepsilon\}_{j-1}$ — перечень ресурсов (элементарных операций), необходимых для формирования исходных данных для текущего пункта плана (элементарной операции) и временных затрат на их передачу; τ_j — время выполнения текущего пункта плана (элементарной операции); χ_j — вклад текущего пункта плана (элементарной операции) в интегральный результат решения задачи, причем $\sum_{j=1}^J \chi_j = 1$; $\{\omega\}_{j+1}$ — список ресурсов (элементарных операций), для выполнения которых необходимы результаты выполнения текущего пункта плана (элементарной операции).

Отметим, что ресурсы могут отрабатывать свои подзадачи как параллельно, так и последовательно. Результаты работы одного ресурса могут являться исходными данными для других.

При условии, когда все операции рабочего плана выполнены за время не более допустимого, считается, что задача решена с требуемым качеством и вовремя.

Рабочий план решения информационно-расчетных задач может быть описан графами функциональных связей элементарных операций и временных затрат. Для удобства восприятия эти графы могут быть представлены совместно (рис. 1).

В условиях отсутствия ограничений на доступность ресурсов рабочий план гарантированно обеспечит выполнение всех операций за интервал времени, равный максимальному весу пути на графе временных затрат.

При наличии условий ограничений на доступность ресурсов выполнение рабочего плана затруднено возникновением конфликтов между планами решения различных задач при доступе к очередному ресурсу. Указанное обстоятельство приводит к возникновению

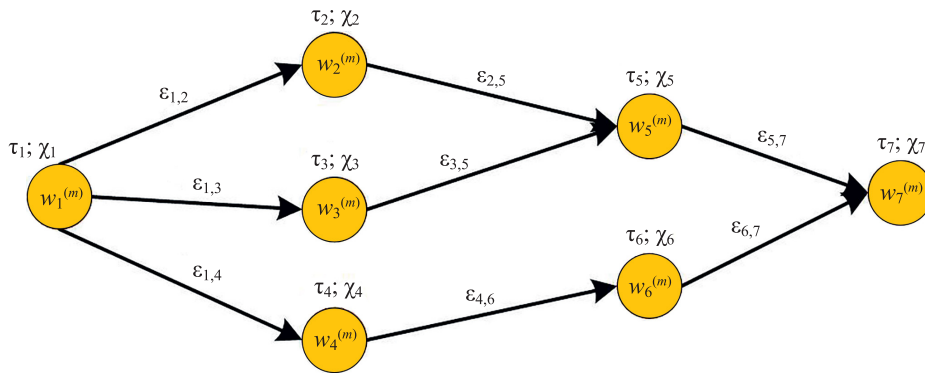


Рис. 1. Пример рабочего плана решения информационно-расчетных задач в виде сетевого графа

Fig. 1. An example of a work plan for solving information and calculation problems in the form of a network graph

задержек δ_j , обусловленных ожиданием освобождения ресурса, что иллюстрируется графом временных затрат (рис. 2).

В условиях функционирования СОД ДЗЗ, когда для решения той или иной поступившей информационно-расчетной задачи доступны все необходимые ресурсы, наилучшим вариантом является включение в рабочий план каждой задачи с требуемым уровнем полноты результатов и за нормативное время (значения коэффициентов $K_{a,n}$ и $K_{f,n}$ для этой задачи в выражении (1) будут максимальными и, как следствие, значение показателя эффективности на заданном временном интервале $[0, t]$ будет стремиться к единице).

По результатам имитационного моделирования [11–14], в условиях возрастания противоречия между производительностью системы комплексов обработки (ресурсов СОД ДЗЗ) и потоком решаемых задач, информационно-расчетные задачи, которым присвоено относительно небольшое значение коэффициента относительной важности, будут не решены, или результаты их решения не будут представлены заказчику в нормативные (требуемые) сроки.

Тогда используются следующие варианты управления, способные повлиять на данную ситуацию:

- 1) ранжирование информационно-расчетных задач по значению коэффициента их относительной важности и обеспечение приоритетного решения тех из них, которые находятся ближе к началу списка;

- 2) увеличение оперативности решения каждой из информационно-расчетных задач путем исключения из рабочих планов их решения тех ресурсов, которые перегружены другими задачами, но при этом вносят незначительный вклад в формирование интегрального результата. В этом случае в качестве важнейших параметров управления функционированием СОД ДЗЗ должны использоваться: порог исключения ресурсов обработки из рабочего плана решения информационно-расчетных задач, а также минимальное значение требуемого вклада ресурса. Данные варианты имеют существенные отличия, заключающиеся в следующем.

В случае использования первого варианта управления, когда на вход СОД ДЗЗ слишком часто будут поступать задачи с более высоким уровнем приоритета, в условиях ресурсной ограниченности системы часть задач с более низким значением коэффициента относительной важности решена не будет, или результаты их решения не будут представлены заказчику вовремя (значение коэффициента $K_{a,n}$ для таких задач будет равным нулю). Однако при этом будет обеспечена максимальная полнота результатов решенных задач (значение коэффициента $K_{f,n}$ для каждой информационно-расчетной задачи будет равным единице).

При втором варианте управления все информационно-расчетные задачи будут выполняться одна за другой последовательно. Ситуация, при которой какая-то из задач не будет решена, не возникнет. При этом каждая

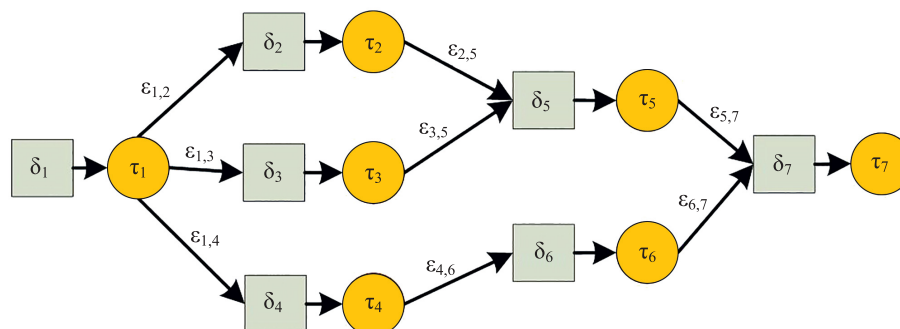


Рис. 2. Граф временных затрат решения информационно-расчетных задач с учетом ожидания доступа к ресурсам

Fig. 2. Graph of the time costs for solving information and calculation problems taking into account the expectation of access to resources

информационно-расчетная задача будет выполняться более оперативно (за счет возможного исключения некоторых ресурсов), и тем самым общая оперативность решения задач будет повышаться (значение коэффициента $K_{a,n}$ для большего числа задач будет равным единице или приближаться к нему). Однако полнота полученных результатов решения будет ниже, чем при первом варианте, т. е. значение коэффициента $K_{f,n}$ для таких задач будет заметно меньше единицы.

Тем самым первый вариант управления в условиях ресурсной ограниченности СОД ДЗЗ позволяет повысить значение показателя эффективности функционирования СОД ДЗЗ $\Theta(t)$ за счет увеличения числа нерешенных низкоприоритетных задач, в то время как второй вариант управления позволяет повысить значение данного показателя путем снижения полноты результатов решения информационно-расчетных задач.

Результаты имитационного моделирования процессов комплексной обработки данных на основе предложенных вариантов управления функционированием СОД ДЗЗ

Выполним оценку влияния исходных данных модели функционирования СОД ДЗЗ [9] (интенсивности потоков задач, текущих возможностей системы ресурсов обработки, степени вклада ресурсов в их решение и др.), а также параметров управления (порога исключения ресурсов обработки из рабочего плана, минимальное значение требуемого вклада ресурса и др.) на значения показателя эффективности функционирования СОД ДЗЗ. С этой целью разработан специальный программный комплекс имитационного моделирования [15, 16], который позволяет обеспечить: настройку и генерацию потоков задач разных типов (в соответствии с заданным законом распределения); постановку сгенерированных задач в очереди к ресурсам; расчет времени ожидания

задачи в очереди (δ_j); расчет значений показателя эффективности в соответствии с формулой (1).

Характеристика набора исходных данных, использованного для имитационного моделирования, представлена в табл. 1.

Решение поступающих информационно-расчетных задач осуществлено различными ресурсами (функциональными группами), характеристики которых представлены в табл. 2, при этом функциональные группы ω_3 и ω_4 могли выполнять свои подзадачи параллельно.

Расчеты выполнены при условии, что все задачи поступали круглосуточно, каждая функциональная группа отрабатывала свои задачи в течение 8 рабочих часов, после чего происходила смена сотрудников на новых.

Вклады ресурсов разных типов в получение интегрального результата решения информационно-расчетных задач разных типов представлены в табл. 3.

При имитационном моделировании выполнены испытания трех вариантов наборов значений настраиваемых параметров управления [10], при которых управление функционированием СОД ДЗЗ осуществлялось по следующим условиям.

Вариант 1. Исключение недоступных ресурсов из рабочего плана. Особенность варианта — при дефиците ресурсов акцент сделан на исключении из рабочего плана ресурсов, которые вносят минимальный вклад (не более 20 %) при условии, что расчетное время ожидания соответствующего ресурса превысило 3 часа.

Вариант 2. Ранжирование задач по приоритету. Особенность варианта — при дефиците ресурсов акцент сделан на перемещении в начало очереди задач, которые имели приоритет не менее 0,05.

Вариант 3. Смешанный вариант. При дефиците ресурсов осуществлено перемещение в начало очереди

Таблица 1. Исходные данные для имитационного моделирования

Table 1. Input data for simulation modeling

Номер типа задачи	Тип задачи	Значение коэффициента относительной важности ξ_n	Интенсивность потока информационно-расчетных задач, ед./сутки	Разнотипные ресурсы и нормативное время, затрачиваемое на выполнение задач данного типа, час			
				ω_1	ω_2	ω_3	ω_4
1	Разработка информационной справки по текущему состоянию объектов мониторинга и прилегающих территорий	0,15	7	1	1	1	1,5
2	Определение (расчет) изменений техногенного и природного характера на территориях расположения объектов мониторинга	0,25	3	1	1	1	1,5
3	Разработка (ведение) электронной карточки объекта мониторинга	0,2	4	1,5	1,5	1,5	2,5
4	Разработка 3D-модели на заданный объект мониторинга (район)	0,1	1	1	1,5	2	3
5	Решение различных аналитических задач	0,01–0,1	21	—	1,5	1	1,5

Таблица 2. Характеристики ресурсов
Table 2. Resource characteristics

Обозначение ресурса	Наименование ресурса (функциональная группа)	Количество единиц ресурса, чел.
ω_1	Сбор данных	1
ω_2	Первичная обработка данных	2
ω_3	Тематическая обработка данных	2
ω_4	Анализ данных	3

Таблица 3. Вклады ресурсов в получение интегрального результата решения информационно-расчетных задач, %
Table 3. Contributions of resources to obtaining the integral result of the solution of information and calculation problems, %

Обозначение ресурса	Типы задач и вклад ресурса в получение результата				
	1	2	3	4	5
ω_1	15	25	10	20	0
ω_2	25	25	10	20	10
ω_3	10	10	25	25	15
ω_4	40	40	35	35	75

ди тех задач, которые имели приоритет не менее 0,1. Выполнено исключение из рабочего плана тех ресурсов, которые вносят минимальный вклад (не более 20 %) при условии, что расчетное время ожидания соответствующего ресурса превысило 4 часа.

Результаты имитационного моделирования процессов функционирования СОД ДЗЗ представлены на рис. 3.

Из рис. 3 видно, что применение трех вариантов управления в разных условиях функционирования СОД ДЗЗ дает различные результаты.

Например, при низкой интенсивности потока поступающих задач рациональным (т. е. позволяющим получить максимальное значение показателя эффективности функционирования СОД ДЗЗ) является нулевой вариант управления (Вариант 1), при котором все задачи выполняются последовательно и в полном объеме. Этот вариант соответствует ситуации, когда все требуемые

ресурсы доступны. С увеличением потока информационно-расчетных задач наибольший выигрыш начинает приносить Вариант 2 с ранжированием задач. Тем не менее, в некоторый момент времени, когда ожидание доступности ресурсов начинает оказывать более существенное влияние на оперативность решения информационно-расчетных задач, приходится применять вариант управления (Вариант 3), учитывающий как ранжирование задач, так и исключение недоступных ресурсов.

Таким образом, в качестве основного подхода к повышению эффективности функционирования СОД ДЗЗ в заданных условиях (условиях текущей доступности ресурсов, а также интенсивности потока поступающих задач) должна использоваться технология обоснованного выбора вариантов управления на основе постоянного мониторинга текущего значения показателя $\Theta(t)$.

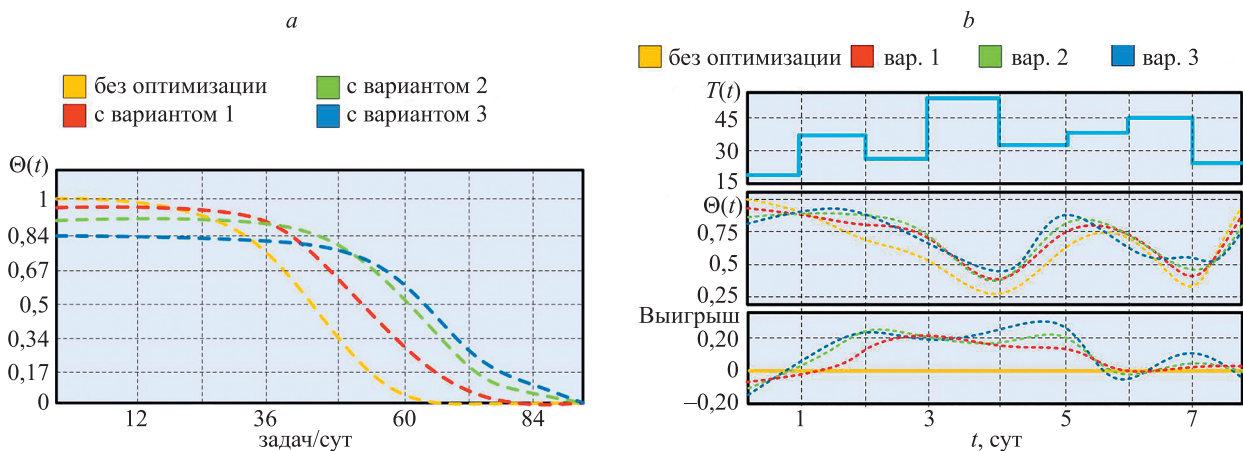


Рис. 3. Зависимости значений показателей эффективности для различных вариантов управления: от интенсивности потока задач (а) и от времени при изменении потока информационно-расчетных задач (б)

Fig. 3. The dependences of the values of performance indicators for various control options: vs. the intensity of the flow of tasks (a) and vs. the time when the flow of information and calculation tasks changes (b)

Заключение

В работе предложен способ повышения эффективности функционирования системы обработки данных дистанционного зондирования Земли при решении задач мониторинга пространственных объектов.

Способ позволяет осуществить адаптацию процессов комплексной обработки данных в системе обработки данных дистанционного зондирования Земли к текущим условиям обстановки и ресурсным ограничениям.

При этом обеспечена адаптация системы к текущим ресурсным возможностям на основе постоянного мониторинга параметров управления ее функционирования, которые определяют значения минимального вклада

ресурса в интегральный результат решения той или иной информационно-расчетной задачи. Определено максимально допустимое время ожидания доступности требуемого ресурса. Также обеспечен учет процессов, протекающих на объектах мониторинга, т. е. возможность учета коэффициентов относительной важности (приоритетов) той или иной поступившей информационно-расчетной задачи в зависимости от складывающихся условий обстановки.

В результате обоснованно выбран наиболее подходящий в текущих условиях вариант управления функционированием системы обработки данных дистанционного зондирования Земли.

Литература

1. Копытко В.К., Шептура В.Н. Проблемы построения единого информационного пространства Вооруженных Сил Российской Федерации и возможные пути их решения // Военная мысль. 2011. № 1. С. 16–26.
2. Севастьянов Н.Н., Бранец В.Н., Панченко В.А., Казинский Н.В., Кондранин Т.В., Негодяев С.С. Анализ современных возможностей создания малых космических аппаратов для дистанционного зондирования Земли // Труды МФТИ. 2009. Т. 1. № 3. С. 14–22.
3. Макриденко Л.А., Волков С.Н., Ходненко В.П., Золотой С.А. Концептуальные вопросы создания и применения малых космических аппаратов // Вопросы электромеханики. Труды ВНИИЭМ. 2010. Т. 114. № 1. С. 15–26.
4. Карин С.А. Построение предметно-ориентированной онтологии в системах обработки пространственных данных // Информационно-управляющие системы. 2014. № 4. С. 78–84.
5. Карин С.А., Дудин Е.А. Подходы к созданию распределенной системы сбора, хранения и поиска геопро пространственных данных // Информация и космос. 2014. № 3. С. 48–53.
6. Карин С.А. Интеграция в едином информационном пространстве разнородных геопро пространственных данных // Информационно-управляющие системы. 2012. № 2. С. 89–94.
7. Талалаев А.А. Организация конвейерно-параллельных вычислений для обработки потоков данных // Информационные технологии и вычислительные системы. 2011. № 1. С. 8–13.
8. Фраленко В.П., Агроник А.Ю. Средства, методы и алгоритмы эффективного распараллеливания вычислительной нагрузки в гетерогенных средах // Программные системы: теория и приложения. 2015. Т. 6. № 3(26). С. 73–92. <https://doi.org/10.25209/2079-3316-2015-6-3-73-92>
9. Карин С.А., Бережной И.В. Подходы к созданию перспективной системы комплексного мониторинга разнородных информационных ресурсов, имеющих геопро пространственную компоненту // Естественные и технические науки. 2016. № 6. С. 138–140.
10. Карин С.А., Бережной И.В. Технология обработки данных в сетцентрических системах сбора, обработки и анализа разнородной геопро пространственной информации // Естественные и технические науки. 2016. № 6. С. 141–143.
11. Карин С.А. Операционно-временная модель функционирования систем комплексной обработки геопро пространственных данных в условиях дефицита их ресурсов // Информационно-управляющие системы. 2017. № 2. С. 51–57. <https://doi.org/10.15217/issn1684-8853.2017.2.51>
12. Карин А.И., Карин С.А., Октябрьский В.В. Модели адаптивного управления функционированием систем комплексной обработки геопро пространственных данных при решении задач мониторинга территориально-распределенных объектов // Труды Военно-космической академии имени А.Ф. Можайского. 2019. № 671. С. 314–325.
13. Белоусов С.М. Математическая модель многопоточной системы массового обслуживания, управляемой планировщиком ресурсов // Вестник НГУ. Серия: Информационные технологии. 2006. Т. 4. № 1. С. 14–26.
14. Алферов А.В., Карин А.И., Карин С.А., Октябрьский В.В. Метод адаптивного определения приоритетов информационно-расчет-

References

1. Kopytko V.K., Sheptura V.N. Problems of building a unified information space of the Armed Forces of the Russian Federation and possible ways to solve them. *Military Thought*, 2011, no. 1, pp. 16–26. (in Russian)
2. Sevastyanov N.N., Branets V.N., Panchenko V.A., Kazinskiy N.V., Kondranin T.V., Negodyayev S.S. Advanced approaches to Earth observation small satellite development. *Proceedings of MIPT*, 2009, vol. 1, no. 3, pp. 14–22. (in Russian)
3. Makridenko L.A., Volkov S.N., Khodnenko V.P., Zolotoy S.A. Conceptual problems on creation and application of small spacecraft. *Electromechanical matters. VNIEM studies*, 2010, vol. 114, no. 1, pp. 15–26. (in Russian)
4. Karin S.A. Developing a domain-specific ontology in spatial data processing systems. *Information and Control Systems*, 2014, no. 4, pp. 78–84. (in Russian)
5. Karin S., Dudin E. Methods for creating distributed systems of geospatial data collection, storage, and search. *Information and Space*, 2014, no. 3, pp. 48–53. (in Russian)
6. Karin S.A. Integration in the single information space of heterogeneous geospatial data. *Information and Control Systems*, 2012, no. 2, pp. 89–94. (in Russian)
7. Talalaev A.A. Parallel-pipeline computations for data flows processing. *Journal of Information Technologies and Computing Systems*, 2011, no. 1, pp. 8–13. (in Russian)
8. Fralenko V., Agronik A. Tools, methods and algorithms for the efficient parallelization of computational loading in heterogeneous environments. *Program Systems: Theory and Applications*, 2015, vol. 6, no. 3(26), pp. 73–92. (in Russian). <https://doi.org/10.25209/2079-3316-2015-6-3-73-92>
9. Karin S.A., Berezhnoi I.V. Approaches to creating a promising system for comprehensive monitoring of heterogeneous information resources with a geospatial component. *Natural and Technical Sciences*, 2016, no. 6, pp. 138–140. (in Russian)
10. Karin S.A., Berezhnoi I.V. Data processing technology in network-centric systems for collecting, processing and analyzing heterogeneous geospatial information. *Natural and Technical Sciences*, 2016, no. 6, pp. 141–143. (in Russian)
11. Karin S.A. Operational and temporal model of complex geospatial data processing systems with insufficient resources. *Information and Control Systems*, 2017, no. 2, pp. 51–57. (in Russian). <https://doi.org/10.15217/issn1684-8853.2017.2.51>
12. Karin A.I., Karin S.A., Oktyabrskiy V.V. Models of adaptive control systems of complex geospatial data processing in solving problems of monitoring geographically distributed objects. *Proceedings of the Military Space academy named after A.F. Mozhaysky*, 2019, no. 671, pp. 314–325. (in Russian)
13. Belousov S.M. Mathematical model of a multi-threaded queuing system managed by a resource scheduler. *Vestnik NSU. Series: Information Technologies*, 2006, vol. 4, no. 1, pp. 14–26. (in Russian)
14. Alferov A.V., Karin A.I., Karin S.A., Oktyabrskiy V.V. Method for adaptive prioritization of information and calculation tasks in monitoring systems for potentially dangerous natural and man-made processes under resource constraints. *Proceedings of the Military*

- ных задач в системах мониторинга потенциально-опасных процессов природного и техногенного характера в условиях ресурсной ограниченности // Труды Военно-космической академии имени А.Ф. Можайского, 2021. № 676. С. 95–104.
15. Карин С.А. Программный комплекс обработки геопространственных данных в типовых территориально-распределенных защищенных автоматизированных системах сбора и обработки разнородной информации (DynamicGIS). Свидетельство о государственной регистрации программы для ЭВМ RU2014618310. Дата публикации 20.09.2014.
 16. Карин С.А. Программная библиотека формирования динамических очередей обработки разнородных данных в типовых территориально-распределенных защищенных автоматизированных системах сбора и обработки разнородной информации (DynamicQueue). Свидетельство о государственной регистрации программы для ЭВМ RU2014618196. Дата публикации 20.09.2014.
 - Space academy named after A.F. Mozhaisky*, 2021, no. 676, pp. 95–104. (in Russian)
 15. Karin S.A. Software package for processing geospatial data in typical geographically distributed secure automated systems for collecting and processing heterogeneous information (DynamicGIS). *Certificate of the computer program official registration RU2014618310*, 2014. (in Russian)
 16. Karin S.A. Software library for forming dynamic queues for processing heterogeneous data in typical geographically distributed secure automated systems for collecting and processing heterogeneous information (DynamicQueue). *Certificate of the computer program official registration RU2014618196*, 2014. (in Russian)

Авторы

Карин Сергей Александрович — кандидат военных наук, доцент, доцент, Военно-космическая академия имени А.Ф.Можайского, Санкт-Петербург, 197198, Российская Федерация, <https://orcid.org/0000-0001-7139-1959>, Sergey.karin@gmail.com

Карин Александр Иванович — научный сотрудник, Военно-космическая академия имени А.Ф.Можайского, Санкт-Петербург, 197198, Российская Федерация, <https://orcid.org/0000-0001-5746-5245>, nil5-ksa@yandex.ru

Authors

Sergey A. Karin — PhD (Military Science), Associate Professor, Associate Professor, Mozhaisky Military Aerospace Academy, Saint Petersburg, 197198, Russian Federation, <https://orcid.org/0000-0001-7139-1959>, Sergey.karin@gmail.com

Alexandr I. Karin — Scientific Researcher, Mozhaisky Military Aerospace Academy, Saint Petersburg, 197198, Russian Federation, <https://orcid.org/0000-0001-5746-5245>, nil5-ksa@yandex.ru

Статья поступила в редакцию 17.03.2022
Одобрена после рецензирования 08.06.2022
Принята к печати 31.07.2022

Received 17.03.2022
Approved after reviewing 08.06.2022
Accepted 31.07.2022



Работа доступна по лицензии
 Creative Commons
 «Attribution-NonCommercial»

doi: 10.17586/2226-1494-2022-22-4-699-707

УДК 004.72

Разработка модели обнаружения сетевых аномалий трафика в беспроводных распределенных самоорганизующихся сетях

Леонид Вячеславович Легашев¹✉, Любовь Сергеевна Гришина²,
 Денис Игоревич Парфенов³, Артур Юрьевич Жигалов⁴

^{1,2,3,4} Оренбургский государственный университет, Оренбург, 460018, Российская Федерация

¹ silentgir@gmail.com✉, <https://orcid.org/0000-0001-6351-404X>

² zabrodina97@inbox.ru, <https://orcid.org/0000-0003-2752-7198>

³ parfenovdi@mail.ru, <https://orcid.org/0000-0002-1146-1270>

⁴ leroy137.artur@gmail.com, <https://orcid.org/0000-0003-3208-1629>

Аннотация

Предмет исследования. Мобильные самоорганизующиеся сети — одно из перспективных направлений технологий граничных вычислений. Такие сети применяются в различных областях деятельности, в частности при разработке интеллектуальных транспортных систем. Особенность мобильных самоорганизующихся сетей заключается в постоянно изменяющейся их динамической топологии. В результате таких изменений необходимо использовать реактивные протоколы маршрутизации при передаче пакетов между узлами. Данные сети уязвимы к кибератакам, поэтому возникает необходимость разработки мер по идентификации сетевых угроз и разработке правил реагирования на них на основе моделей машинного обучения. Цель работы — разработка динамической модели обнаружения сетевых аномалий трафика в беспроводных распределенных самоорганизующихся сетях. **Метод.** Применены методы и алгоритмы интеллектуального анализа данных и машинного обучения. Предлагаемый подход к мониторингу трафика в беспроводных распределенных самоорганизующихся сетях состоит в реализации двух этапов: первоначального анализа трафика для выявления аномальных событий и последующего глубокого изучения инцидентов кибербезопасности для классификации типа атакующего воздействия. В рамках подхода построены модели на основе ансамблевых методов машинного обучения. Выполнен сравнительный анализ и выбор наиболее эффективных алгоритмов машинного обучения и их оптимальных гиперпараметров. **Основные результаты.** Проведена формализация модели обнаружения аномалий трафика в беспроводных распределенных самоорганизующихся сетях, и выделены основные количественные метрики производительности сети. Представлен обобщенный алгоритм обнаружения аномалий трафика в мобильных самоорганизующихся сетях. Выполнено экспериментальное исследование симуляции сегмента сети с позиции снижения производительности в условиях реализации различных сценариев возникновения сетевых атак. Показано, что сетевая распределенная атака вида «отказ в обслуживании» и кооперативная атака вида «Blackhole» оказывают наибольшее негативное влияние на производительность сегмента мобильной самоорганизующейся сети. Результаты моделирования сети применены для построения модели машинного обучения выявления аномалий и классификации типов атак. Результаты сравнительного анализа алгоритмов машинного обучения показали, что метод LightGBM наиболее эффективен для выявления аномалий сетевого трафика, при использовании которого доля правильных ответов составила 91 %. Тип проводимой атаки определен с долей правильных ответов 90 %. **Практическая значимость.** Предложенный подход к обнаружению сетевых аномалий за счет применения обученных моделей анализа трафика позволяет своевременно идентифицировать рассмотренные типы атак. Будущее направление развития данного исследования — рассмотрение новых сценариев возникновения сетевых атак и дополнительное онлайн-обучение построенных моделей идентификации. Разработанное программное средство обнаружения сетевых аномалий трафика в распределенных мобильных самоорганизующихся сетях может найти применение для любых типов беспроводных самоорганизующихся сетей.

Ключевые слова

мобильные самоорганизующиеся сети, метрики производительности, система обнаружения вторжений

Благодарности

Исследование выполнено при финансовой поддержке гранта Президента Российской Федерации для государственной поддержки молодых российских ученых — кандидатов наук (МК-2959.2021.1.6).

Ссылка для цитирования: Легашев Л.В., Гришина Л.С., Парфенов Д.И., Жигалов А.Ю. Разработка модели обнаружения сетевых аномалий трафика в беспроводных распределенных самоорганизующихся сетях // Научно-технический вестник информационных технологий, механики и оптики. 2022. Т. 22, № 4. С. 699–707. doi: 10.17586/2226-1494-2022-22-4-699-707

Development of a model for detecting network traffic anomalies in distributed wireless ad hoc networks

Leonid V. Legashev¹✉, Lubov S. Grishina², Denis I. Parfenov³, Arthur Yu. Zhigalov⁴

^{1,2,3,4} Orenburg State University, Orenburg, 460018, Russian Federation

¹ silentgir@gmail.com✉, <https://orcid.org/0000-0001-6351-404X>

² zabrodina97@inbox.ru, <https://orcid.org/0000-0003-2752-7198>

³ parfenovdi@mail.ru, <https://orcid.org/0000-0002-1146-1270>

⁴ leroy137.artur@gmail.com, <https://orcid.org/0000-0003-3208-1629>

Abstract

Mobile ad hoc networks are one of the promising directions of the edge computing technology and they are used in various applications, in particular, in the development of intelligent transport systems. A feature of mobile ad hoc networks lies in the constantly changing dynamic network topology, as a result of which it is necessary to use reactive routing protocols when transmitting packets between nodes. Mobile ad hoc networks are vulnerable to cyber-attacks, so there is a need to develop measures to identify network threats and develop rules for responding to them based on machine learning models. The subject of this study is the development of a dynamic model for detecting network traffic anomalies in wireless distributed ad hoc networks. Within the framework of this study, methods and algorithms of data mining and machine learning were applied. The proposed approach to traffic monitoring in wireless distributed ad hoc networks consists in the implementation of two stages: initial traffic analysis to identify anomalous events and subsequent in-depth study of cybersecurity incidents to classify the type of attack. Within the framework of this approach, the corresponding models are constructed based on ensemble methods of machine learning. A comparative analysis and selection of the most efficient machine learning algorithms and their optimal hyperparameters has been carried out. In this paper, a formalization of the traffic anomaly detection model in distributed wireless ad hoc networks is carried out, the main quantitative metrics of network performance are identified, a generalized algorithm for detecting traffic anomalies in mobile ad hoc networks is presented, and an experimental study of the network segment simulation is carried out from the point of view of performance degradation during the implementation of various network attack scenarios. Network distributed denial of service attacks and cooperative blackhole attacks have the greatest negative impact on the performance of the mobile ad hoc network segment. In addition, the network simulation results were used to build a machine learning model to detect anomalies and classify types of attacks. The results of a comparative analysis of machine learning algorithms showed that the use of the LightGBM method is the most effective for detecting network traffic anomalies with an accuracy of 91 %, and for determining directly the type of attack being carried out with an accuracy of 90 %. The proposed approach for network anomalies detection through the use of trained traffic analysis models makes it possible to identify the considered types of attacks in due time. The future development direction of this research is the consideration of new scenarios for the emergence of network attacks and online additional training of the constructed identification models. The developed software tool for detecting network traffic anomalies in distributed mobile ad hoc networks can be used for any type of wireless ad hoc networks.

Keywords

mobile ad hoc networks, performance metrics, intrusion detection system

Acknowledgements

The research was funded by the grant from President of the Russian Federation for state support of young Russian scientists (МК-2959.2021.1.6).

For citation: Legashev L.V., Grishina L.S., Parfenov D.I., Zhigalov A.Yu. Development of a model for detecting network traffic anomalies in distributed wireless ad hoc networks. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2022, vol. 22, no. 4, pp. 699–707 (in Russian). doi: 10.17586/2226-1494-2022-22-4-699-707

Введение

Особенность мобильных самоорганизующихся сетей (Mobile Ad-hoc Networks, MANETs) заключается в отсутствии заранее определенной сетевой инфраструктуры при передаче информации между двумя беспроводными устройствами. Каждый узел в мобильной беспроводной сети может служить как роутером, так и хостом, и осуществляет пересылку пакетов по

запросу. Для MANET характерна динамическая топология, повышенная мобильность сетевых узлов, а также многопереходная маршрутизация, когда связность двух элементов обеспечивается пересылкой данных через промежуточные элементы сети. Наиболее популярные протоколы сетевой маршрутизации в MANET — семейства реактивных и проактивных протоколов: AODV (Ad hoc On-Demand Distance Vector), DSDV (Destination-Sequenced Distance-Vector), OLSR (Optimized Link State

Routing), DSR (Dynamic Source Routing) и др. Как правило, узлы в MANET представляют собой мобильные устройства с небольшими объемами дисковой и оперативной памяти и процессорной мощностью, при этом такие сетевые устройства могут выполнять определенную вычислительную нагрузку. В частности, бинарная классификация сетевых угроз может производиться на конечных узлах MANET при наличии заранее обученного классификатора. Развитие архитектуры программно-конфигурируемых сетей (ПКС) и технологии граничных вычислений (edge computing) позволяют оптимальным образом увеличить эффективность функционирования MANETs, сделать их более гибкими с точки зрения обнаружения сетевых угроз и выработки правил реагирования на них. На рис. 1 представлена общая схема функционирования сегмента MANET с поддержкой ПКС-контроллера.

Особенность анализа функционирования MANETs — необходимость фиксирования взаимосвязей объектов сети и их характеристик не только в некоторый момент времени, но и в соответствующем пространственном положении. Объекты сети, кроме обычных координат долготы и широты, имеют высоту расположения (например, придорожные устройства,

такие как светофор, и антенны на автомобилях могут располагаться на разных высотах). В связи с этим расстояние между беспроводными роутерами такого типа необходимо рассчитывать с учетом высоты расположения. Зоны покрытия сигналов могут не перекрываться, и обеспечение стабильной связи между объектами нельзя гарантировать.

Реализация ПКС-контроллера в качестве элемента MANET отделяет плоскость данных (Data plane) от плоскости контроля (Control plane), позволяя проводить эффективный мониторинг сетевых потоков трафика практически в режиме реального времени, собирая статистические данные потоков и осуществляя маршрутизацию пакетов по всей сети. Коммутаторы с поддержкой OpenFlow осуществляют поиск по таблицам коммутации и обновляют информацию, если правило пересылки пакета не найдено. При возникновении сетевой угрозы ПКС-контроллер формирует сообщение с политикой реагирования (как правило, происходит изоляция злонамеренного узла и удаление его из всех таблиц коммутации) и рассылает его коммутаторам OpenFlow. Базовые станции используются для покрытия групп мобильных устройств. На каждой базовой станции работает механизм обнаружения сер-

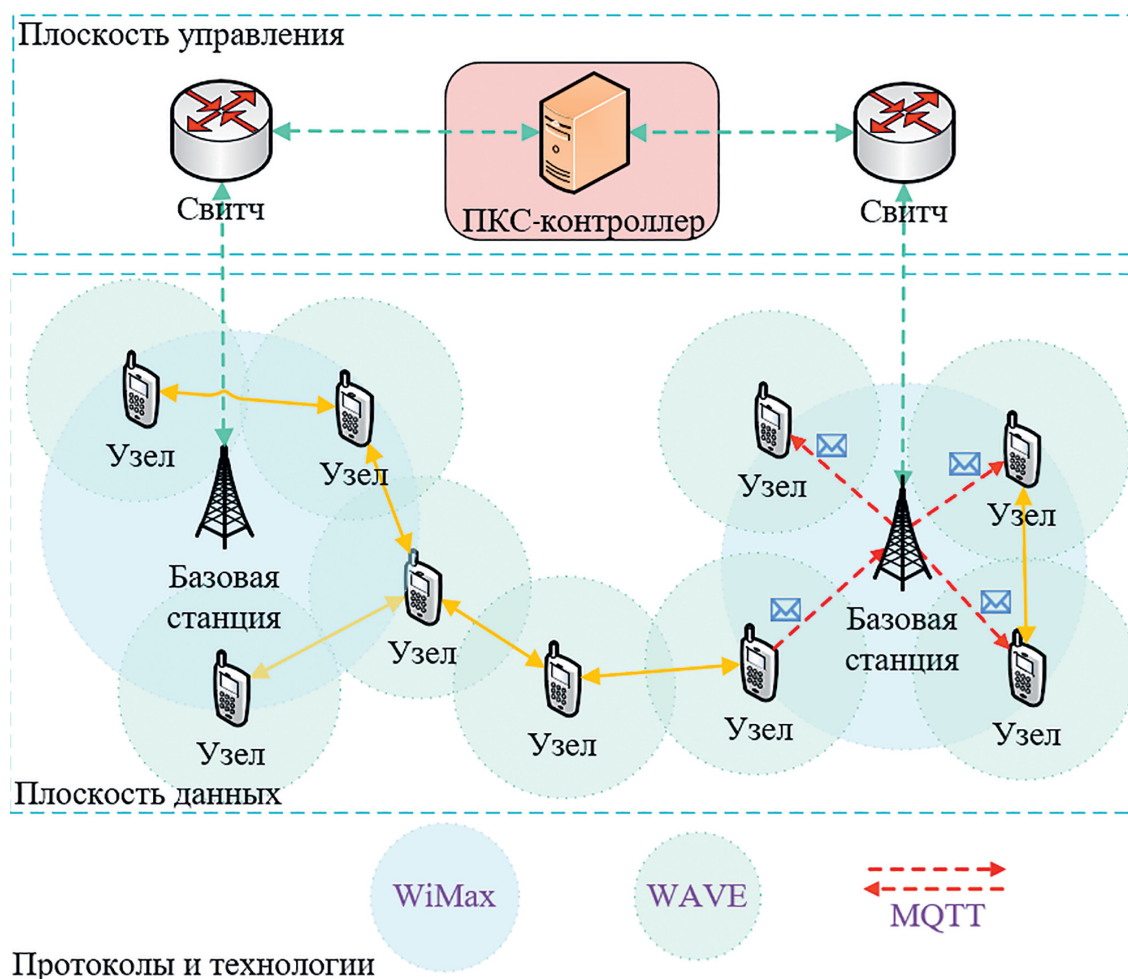


Рис. 1. Общая схема функционирования сегмента мобильной самоорганизующейся сети с поддержкой программно-конфигурируемых сетей

Fig. 1. General scheme of the mobile ad hoc network segment functioning with software-defined networking support

висов (service discovering) с целью регистрации мобильного узла и рассылки определенного тематического контента. Рассылка информации с датчиков конечных устройств может осуществляться с помощью протокола работы с телеметрией MQTT (Message Queuing Telemetry Transport) по модели подписки на контент. Конечный узел сети может осуществлять определенную рассылку (например, сервис информирования о погодных условиях, сервис информирования о дорожной ситуации и др.), предварительно зарегистрировавшись на базовой станции и указав способ доведения информации до остальных узлов сети.

Современное состояние исследований

Вопросы разработки архитектуры и моделей обеспечения безопасности MANETs рассмотрены во многих научных работах. В работе [1] описана архитектура безопасности MANET в виде пяти уровней: надежной инфраструктуры, безопасности связи и маршрутизации, сетевой безопасности и безопасности приложений. В [2] исследованы вопросы обеспечения безопасности протоколов маршрутизации интернета вещей в MANETs, с помощью описания классов используемых протоколов и расширения безопасности. В работе [3] представлена концептуальная облачная модель MANET для умных устройств на базе технологии 5G. В [4] разработана модель безопасности на основе метода шифрования Advanced Encryption Standard (AES) для генерации секретных ключей в кластерных MANETs. В работе [5] предложена модель байесовской игры с неполной информацией для анализа обычных и злонамеренных узлов в MANETs. В [6] описана энергоэффективная многомерная модель доверительного управления для достижения улучшенных параметров качества обслуживания в MANETs. В [7] представлены двухуровневая модель клиент-сервера для комплексной безопасности в MANETs, которая объединяет внутренние и внешние системы обнаружения вторжений в единое устройство.

Наиболее распространенными в MANETs являются атаки вида «отказ в обслуживании» DoS (Denial-of-Service) и DDoS (Distributed Denial-of-Service), а также Blackhole [8]. В случае реализации DoS- и DDoS-атак тот или иной узел сети временно выходит из строя, получая на обработку множество однотипных запросов, сгенерированных множеством злонамеренных узлов. При реализации Blackhole-атак, злонамеренный узел сбрасывает все входящие пакеты, затрудняя маршрутизацию сетевых потоков в MANETs.

Для повышения безопасности протокола маршрутизации AODV в MANET в работе [9] предложен подход изоляции вредоносных узлов на основе модели доверия. В [10] рассмотрено использование нечеткой логики для методики расчета доверия узлов сети, с целью обнаружения атак вида Blackhole, Grayhole и DDoS. В [11] описана вариация безопасного протокола маршрутизации AODV для защиты от кооперативной атаки вида Blackhole на базе хаотических отображений. Для обнаружения атак вида DDoS и Blackhole в работе [12] использован механизм аутентификации,

основанный на упрощенном алгоритме шифрования и MAC-аутентификации, который доказал эффективность предложенного решения на базе сетевого симулятора NS-2. В [13] использован сетевой симулятор NS-3 для изучения влияния атаки вида Blackhole на параметры производительности MANET, такие как пропускная способность, сквозная задержка и коэффициент потери пакетов. В работе [14] использована оптимизация муравьиной колонии для предотвращения атаки вида Blackhole в MANET. В [15] проведены экспериментальные исследования обнаружения DDoS-атак, с помощью метода опорных векторов и оптимизации метода роя частиц. В [16] представлен метод на основе бэггинга классификаторов для выявления аномалий сетевого трафика.

В [17] разработана модель распространения информации в автомобильных самоорганизующихся сетях с использованием протокола передачи телеметрии MQTT. Работа [18] направлена на повышение безопасности MQTT-протокола на уровне приложений с целью противодействия DoS-атакам в беспроводных сетях. В [19] представлен механизм управления потоком на базе протокола MQTT, снижающий показатели метрик отбрасывания пакетов и сквозной задержки по сравнению со стандартной реализацией MQTT.

Таким образом, обзор актуальных научных исследований в области обеспечения безопасности сетей MANET показал, что в условиях отсутствия заранее определенной сетевой инфраструктуры при передаче информации между беспроводными устройствами методы интеллектуального анализа данных достаточно эффективно могут идентифицировать атаки. При этом на данный момент отсутствует общая формализованная структура MANET, а также обобщенный алгоритм обнаружения аномалий трафика. В рамках настоящей работы рассмотрены более подробно сформулированные проблемы обеспечения безопасности сети MANET.

Модель обнаружения аномалий трафика в сегменте мобильной самоорганизующейся сети

Представим сегмент MANET в момент времени t_j в виде случайного геометрического неориентированного графа $G^j = (V, E^j)$, $E^j \subseteq V \times V$, $|V| = n$, $|E^j| = m$ с фиксированным множеством узлов $V = (v_1, v_2, \dots, v_n)$ размера n и множеством дуг $E^j = (e_1, e_2, \dots, e_m)$ размера m . Граф G^j представляет собой пространственную сеть, построенную посредством случайного размещения n узлов в трехмерной плоскости A размера $s_1 \times s_2 \times s_3$, при этом два узла соединяются дугой e_k только в том случае, если их расстояние в момент времени t_j находится в заданных зонах покрытия. В связи с динамикой объектов MANET и изменением расстояния между ними, в каждый следующий момент времени t_{j+1} изменяется множество дуг E^{j+1} и, следовательно, сам неориентированный граф G^{j+1} . Множество дуг E^{j+1} сохраняет ребра множества E^j , если к моменту времени t_{j+1} связь между соответствующими объектами не потеряна, и добавляет новые ребра, которые характеризуют новые образовавшиеся связи за момент времени $(t_{j+1} - t_j)$. Таким образом, два $G^j = (V, E^j)$ и $G^i = (V, E^i)$ описывают одну и ту

же MANET с одними и теми же участниками движения (вершины графа полностью совпадают), но связь между ними через момент времени Δt будет характеризоваться различными дугами E^j , E^i (множества частично совпадают, если сохранилась связь между некоторыми объектами).

С учетом принципов функционирования MANETs, введем пространственно-временные характеристики для каждого узла динамического графа G^j .

Каждая вершина графа G^j представляет собой беспроводной роутер в виде кортежа значений

$$v_i^j = \{r_i, mob_i^j, spd_i^j, (x_i^j, y_i^j, z_i^j)\}, i=1, |V|,$$

где r_i — зона покрытия i -го беспроводного роутера; mob_i^j — вид мобильности j -го узла в момент времени t_j ; spd_i^j — скорость движения i -го узла в момент времени t_j ; (x_i^j, y_i^j, z_i^j) — координаты местоположения i -го узла в момент времени t_j . По умолчанию все узлы графа G^j имеют линейную мобильность $mob_i^j = Linear$, однако если узел является стационарным, тогда его мобильность имеет вид $mob_i^j = Fixed$, а скорость $spd_i^j = 0$. Множество дуг E^j графа G^j представляет собой множество сетевых связей между узлами сети, размер которого динамически изменяется во времени в зависимости от текущей топологии MANET.

Определим расстояние между двумя беспроводными роутерами v_a^j и v_b^j в момент времени t_j как евклидово расстояние:

$$d(v_a^j, v_b^j) = \|v_a^j - v_b^j\|_2 = \sqrt{(x_a^j - x_b^j)^2 + (y_a^j - y_b^j)^2 + (z_a^j - z_b^j)^2}.$$

Дуга $e(v_a^j, v_b^j)$ строится только в том случае, когда $d(v_a^j, v_b^j) \leq r_a + r_b$, что означает пересечение зон покрытия двух узлов.

При установлении соединения между двумя произвольными узлами в сети графа G осуществляется пересылка сетевых пакетов согласно выбранному протоколу маршрутизации. Множество всех потоков сетевого трафика в сегменте MANET обозначим в виде кортежа значений

$$Z = \{flowID, bR, fB, fA\}, |Z| = fNUM,$$

где $flowID$ — уникальный идентификатор потока; bR — битрейт потока; fB и fA — множество базовых и приобретенных признаков потока. Введем понятие ПКС-контроллера в виде OpenFlow контроллера C и обозначим множество коммутаторов S_M , содержащих таблицы маршрутизации пакетов в сегменте MANET.

Модель обнаружения аномалий сетевого трафика

Построим модель машинного обучения на основе ансамблевых алгоритмов для первичного выявления аномалий внутри беспроводной распределенной самоорганизующейся сети, генерирующую потоки сетевого трафика в сегменте в формате множества Z . Задача состоит в распределении трафика и отсеивании «подоз-

рительного». Для решения данной задачи построим модель бинарного классификатора сетевых угроз и обозначим ее в виде функции $h(Z): Z \rightarrow \{0, 1\}$ которая присваивает каждому потоку трафика tz_i из множества всех сетевых потоков $Z = \{tz_1, tz_2, \dots, tz_n\}$ метку 0 в случае отсутствия сетевой атаки и метку 1 в случае ее наличия. Обозначим через $bCLF$ выбранный для бинарной классификации наиболее эффективный метод машинного обучения.

Модель идентификации типа сетевой атаки

Следующий этап анализа «подозрительного» сетевого трафика — идентификация конкретного типа атаки для выработки последующей стратегии противодействия. Построим модель мультиклассового классификатора сетевых угроз и обозначим ее в виде функции $f(Z): Z \rightarrow K$, которая присваивает каждому потоку трафика tz_i метку $k_j \in K$, $|K| > 2$, соответствующую нормальному трафику (benign traffic), либо конкретному виду сетевой атаки. Обозначим через $mCLF$ выбранный для мультиклассовой классификации метод машинного обучения. Результаты сравнения работы классификаторов представлены в разделе «Моделирование сценариев сетевых атак в рамках сегмента мобильной самоорганизующейся сети».

Оценим производительность сегмента MANET, а также эффективность и точность системы обнаружения аномалий, для этого введем в рассмотрение количественные метрики производительности.

1. Коэффициент доставки пакетов (Packet Delivery Ratio) — соотношение количества полученных пакетов к количеству отправленных, для исследуемых узлов сегмента сети.
2. Пропускная способность сети (Throughput) — соотношение размера успешно переданных по сети пакетов к общему времени симуляции сети $simT$.
3. Время передачи пакета от источника к получателю и обратно (round-trip-time) — разница времени получения ответа от получателя и времени отправки запроса от источника.
4. Сквозная задержка передачи пакетов в сети (End-to-end Delay) — разница между временем отправки пакета от источника и получения его получателем. В среднем End-to-end Delay составляет половину времени метрики round-trip-time.
5. Издержки (Overhead) — среднее количество пакетов, необходимых для доставки одного пакета данных. Рассчитаем в виде соотношения общего количества пакетов к количеству пакетов, полученных узлом-получателем. При этом общее количество пакетов включает в себя служебный (пакеты, передаваемые между роутерами) и пользовательский трафики (пакеты приложений).

Перечисленные метрики позволяют оценить эффективность симулируемых сетевых угроз в сегменте MANET с целью дальнейшего формирования набора данных для исследования методами машинного обучения.

Рассмотрим обобщенный алгоритм обнаружения аномалий трафика в MANETs.

Шаг 1	ПКС-контроллер C осуществляет мониторинг состояния сетевых потоков из множества Z , выбирая для анализа случайный сетевой поток каждые t_M минут.
Шаг 2	<p>Для выбранного сетевого потока tz_i:</p> <p>2.1 Проводится бинарная классификация $h(tz_i)$ с помощью классификатора $bCLF$.</p> <p>2.2 Если $h(tz_i) = 1$:</p> <p>2.2.1 Проводится мультиклассовая классификация $f(tz_i)$ с помощью классификатора $mCLF$.</p> <p>2.2.2 Если обнаружена атака вида DDoS:</p> <p>2.2.2.1 На ПКС-контроллере C формируется сообщение $flowPLC$ для коммутатора S_i с политикой выборочного отбрасывания маркированных пакетов.</p> <p>2.2.2.2 При получении сообщения на коммутаторе S_i происходит отбрасывание всех маркированных пакетов, соответствующих потоку tz_i.</p> <p>2.2.3 Если обнаружена атака вида Blackhole:</p> <p>2.2.3.1 На контроллере $cSDN$ формируется сообщение $routingPLC$ для коммутатора S_i с исключением злонамеренного узла, соответствующего потоку tz_i из всех таблиц маршрутизации сегмента сети.</p> <p>2.3 Атака отсутствует, возобновление мониторинга (шаг 1 работы алгоритма).</p>

Время мониторинга t_M подбирается экспериментально в рамках симулятора сегмента MANET при реализации различных сценариев осуществления сетевых атак. Операции классификации возможно выполнять на конечных узлах сегмента MANET при соблюдении вычислительных ресурсных ограничений. В этом случае узел самостоятельно проводит классификацию сетевых потоков данных и при обнаружении сетевой атаки информирует ПКС-контроллер C о возникшей угрозе. Далее происходит выработка правил реагирования на инциденты безопасности.

Моделирование сценариев сетевых атак в рамках сегмента мобильной самоорганизующейся сети

На базе симулятора OMNeT++ версии 5.6.2 и демонстрационных примеров фреймворка INET построен сегмент MANET, в рамках которого произведена отправка пакетов вида Ping, UDP (User Datagram Protocol) и TCP (Transmission Control Protocol) от узла-отправителя к узлу-получателю. Стандартный инструментарий OMNeT++ включает в себя возможность отслеживания количественных метрик производительности по результатам симуляции. На рис. 2 представлена схема сетевых потоков данных узла-отправителя по уровням

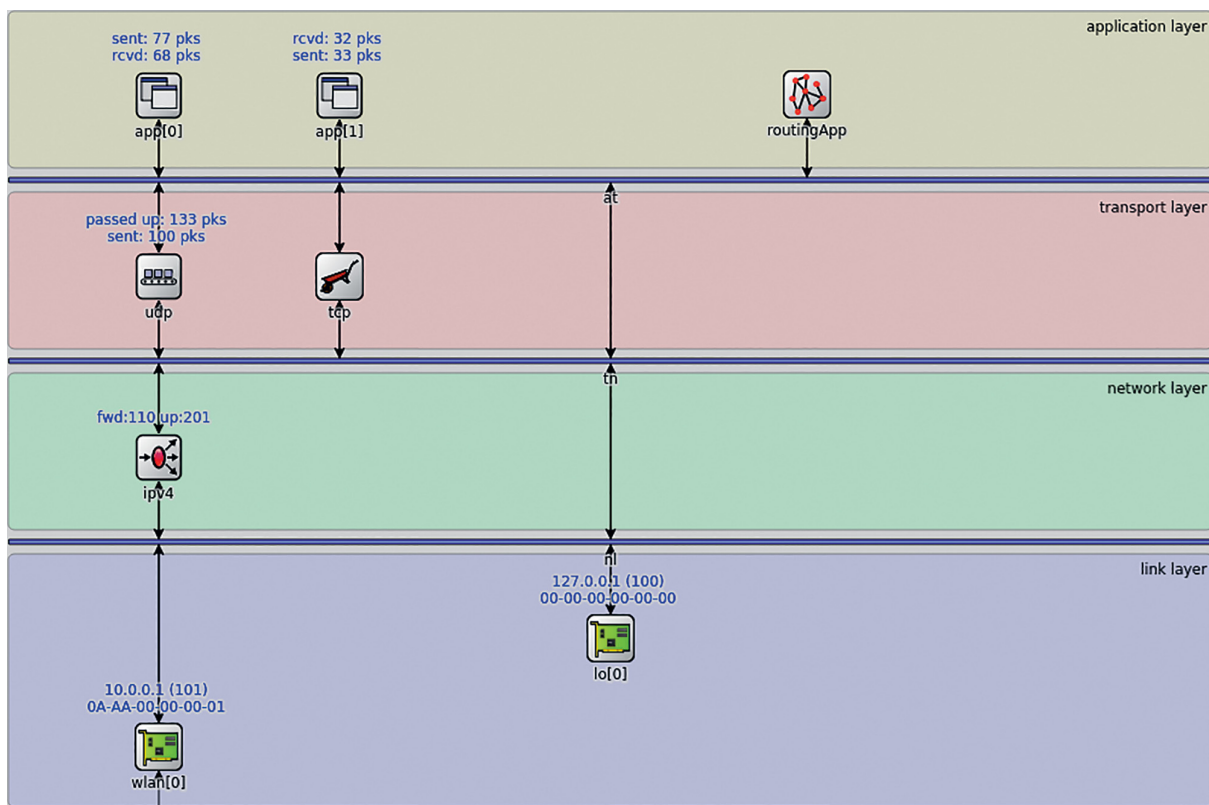


Рис. 2. Схема сетевых потоков данных для узла отправителя в OMNeT++

Fig. 2. Scheme of network data flows for the source node in OMNeT++

сетевой модели OSI в случайный момент выполнения симуляции.

Для оценки показателей качества функционирования сегмента MANET рассмотрим пять сценариев симуляций. Результаты рассмотренных симуляций приведены в таблице.

Сценарий 1. Симуляция сегмента сети без атак. Симуляция осуществляется в соответствии с общими настройками сегмента сети, при этом потенциально зловерные узлы также используются для маршрутизации.

Сценарий 2. Симуляция сегмента сети с Blackhole атакой. Симуляция осуществляется в соответствии с общими настройками сегмента сети, при этом на одном из узлов отключена опция forwarding, что приводит к удалению всех входящих на узел пакетов.

Сценарий 3. Симуляция сегмента сети с кооперативной Blackhole атакой. Симуляция осуществляется в соответствии с общими настройками сегмента сети, при этом на двух узлах отключена опция forwarding, что приводит к удалению всех входящих на узлы пакетов.

Сценарий 4. Симуляция сегмента сети с DoS-атакой. Симуляция осуществляется в соответствии с общими настройками сегмента сети, при этом один из узлов реализует UDP Flooding атаку, отправляя пакеты на источник с частотой 0,05 с, размер пакета распределен равномерно в интервале от 200 до 500 Б.

Сценарий 5. Симуляция сегмента сети с DDoS-атакой. Симуляция осуществляется в соответствии с общими настройками сегмента сети, при этом три узла реализуют UDP Flooding атаку, отправляя пакеты на источник и на адресат, с усредненной частотой 0,05 с, размер пакета распределен равномерно в интервале от 200 до 500 Б.

В соответствии с выполненными экспериментальными исследованиями, множество меток K мультиклассового классификатора $f(Z)$ зададим в виде: $K = \{\langle \text{Benign} \rangle, \langle \text{DoS} \rangle, \langle \text{DDoS} \rangle, \langle \text{Blackhole} \rangle, \langle \text{CooperativeBlackhole} \rangle\}$. Отметим, что сетевые атаки вида cooperative blackhole attack и distributed denial-

of-service attack оказывают наибольшее негативное влияние на производительность сегмента MANET, и система обнаружения вторжений в первую очередь должна быть сфокусирована на классификацию паттернов поведения сетевых угроз, характерных для этих типов атак.

Проведем исследование эффективности предложенного алгоритма классификации сетевого трафика в сегменте MANET. Для этого выполним одновременную симуляцию Сценариев 3 и 5 с кооперативной Blackhole атакой и атакой DDoS в исходном сегменте сети. В результате симуляции получим несбалансированный набор данных, состоящий из 52 столбцов-признаков и 5840 записей со следующим распределением: $\langle \text{Benign} \rangle$: 4247; $\langle \text{DDoS} \rangle$: 834; $\langle \text{Cooperative Blackhole} \rangle$: 759. Балансировку данных произведем с помощью алгоритма SMOTE. Получим сбалансированный набор данных, состоящий из 11 375 записей со следующим распределением: $\langle \text{Benign} \rangle$: 3849; $\langle \text{DDoS} \rangle$: 3847; $\langle \text{Cooperative Blackhole} \rangle$: 3679. В качестве метрики оценки классификации используем долю правильных ответов (accuracy) при сравнении трех современных классификаторов: XGBoost, LightGBM и AdaBoost. Подбор оптимальных гиперпараметров произведем с использованием функции GridSearchCV. Результаты бинарной классификации на сбалансированном наборе данных: XGBoost — 88 %, LightGBM — 91 %, AdaBoost — 83 %. Результаты многоклассовой классификации на сбалансированном наборе данных: XGBoost — 83 %, LightGBM — 90 %, AdaBoost — 64 %. Матрица ошибок классификатора LightGBM для многоклассовой классификации сетевого трафика представлена на рис. 3.

В результате можно сделать вывод о том, что классификатор LightGBM показывает наилучшие результаты для обнаружения аномалий сетевого трафика и идентификации конкретной сетевой атаки. При проведении дальнейших исследований обученный классификатор с подобранными гиперпараметрами будет выгружен в качестве модели обнаружения атак.

Таблица. Оценка показателей качества функционирования сегмента мобильной самоорганизующейся сети

Table. Evaluation of the functioning quality indicators of the MANET segment

Результаты симуляций	Сценарий				
	1	2	3	4	5
Отправлено/получено UDP-пакетов	729/259	798/625	718/36	734/425	708/225
Количество операций маршрутизации	31 447	32 058	19 859	65 223	206 879
Время в очереди, с	4,29	8,85	1,69	4,62	7,04
Полное время передачи пакета, с	1,15	0,84	0,62	0,98	1,2
Коэффициент доставки пакетов, %	36	78	5	58	32
Пропускная способность, Б/с	60,43	145,83	8,40	99,17	52,50
Сквозная задержка, с	0,58	0,42	0,31	0,49	0,60
Издержки маршрутизации, пакеты	121	51	551	153	919

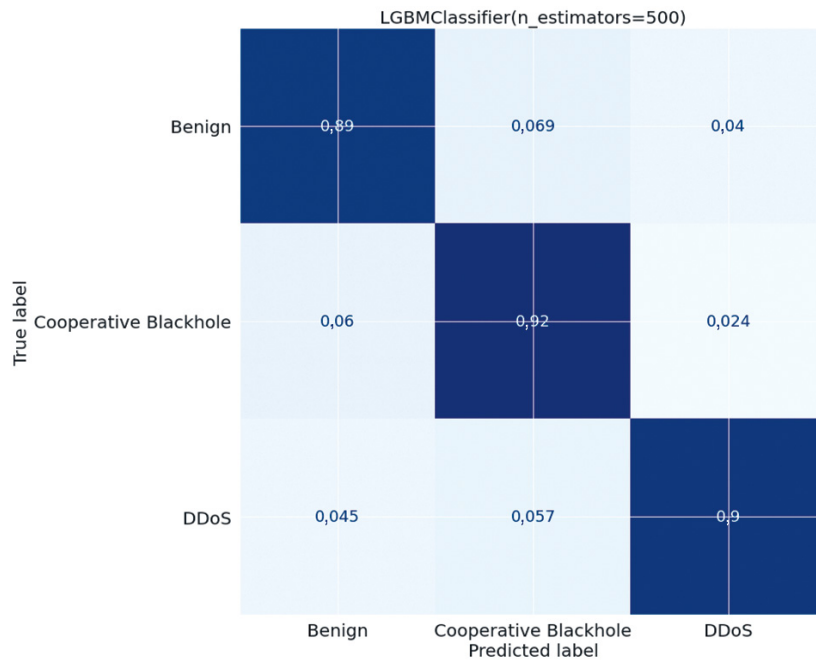


Рис. 3. Матрица ошибок классификатора LightGBM

Fig. 3. Confusion matrix of LightGBM Classifier

Заключение

Мобильные самоорганизующиеся сети имеют большой потенциал применения, особенно в сфере развития интеллектуальных транспортных систем. Разработанная модель обнаружения аномалий трафика легла в основу архитектуры распределенной интеллек-

туальной системы выявления угроз и обеспечения безопасности при передаче данных в беспроводных распределенных самоорганизующихся сетях. Обобщенный алгоритм обнаружения аномалий трафика будет использоваться при разработке метода распределенного обнаружения сетевых атак в рамках дальнейших исследований.

Литература

- Li S.-C., Yang H.-L., Zhu Q.-S. Research on MANET security architecture design // *Proc. of the 2010 International Conference on Signal Acquisition and Processing (ICSAP)*. 2010. P. 90–93. <https://doi.org/10.1109/ICSAP.2010.19>
- Karlsson J., Dooley L.S., Pulkkis G. Secure routing for MANET connected Internet of Things systems // *Proc. of the 6th IEEE International Conference on Future Internet of Things and Cloud (FiCloud)*. 2018. P. 114–119. <https://doi.org/10.1109/FiCloud.2018.00024>
- Alam T. Device-to-Device communications in cloud, MANET and Internet of Things integrated architecture // *Journal of Information Systems Engineering and Business Intelligence*. 2020. V. 6. N 1. P. 18–26. <https://doi.org/10.20473/jisebi.6.1.18-26>
- Nehra D., Dhindsa K.S., Bhushan B. A Security Model to Make Communication Secure in Cluster-Based MANETs // *Advances in Intelligent Systems and Computing*. 2020. V. 1079. P. 183–193. https://doi.org/10.1007/978-981-15-1097-7_16
- Olanrewaju R.F., Khan B.U.I., Anwar F., Mir R.N., Yaacob M., Mehraj T. Bayesian signaling game based efficient security model for MANETs // *Lecture Notes in Networks and Systems*. 2020. V. 70. P. 1106–1122. https://doi.org/10.1007/978-3-030-12385-7_75
- Shabut A.M., Kaiser M.Sh., Dahal K.P., Chen W. A multidimensional trust evaluation model for MANETs // *Journal of Network and Computer Applications*. 2018. V. 123. P. 32–41. <https://doi.org/10.1016/j.jnca.2018.07.008>
- Salama H.M., El Mageed M.Z.A., Salama G.I.M., Badran K.M. CSMCSM: Client-Server Model for Comprehensive Security in MANETs // *International Journal of Information Security and Privacy*. 2021. V. 15. N 1. P. 44–64. <https://doi.org/10.4018/IJISP.2021010103>
- Alani M.M. MANET security: A survey // *Proc. of the 4th IEEE International Conference on Control System, Computing and*

References

- Li S.-C., Yang H.-L., Zhu Q.-S. Research on MANET security architecture design. *Proc. of the 2010 International Conference on Signal Acquisition and Processing (ICSAP)*, 2010, pp. 90–93. <https://doi.org/10.1109/ICSAP.2010.19>
- Karlsson J., Dooley L.S., Pulkkis G. Secure routing for MANET connected Internet of Things systems. *Proc. of the 6th IEEE International Conference on Future Internet of Things and Cloud (FiCloud)*, 2018, pp. 114–119. <https://doi.org/10.1109/FiCloud.2018.00024>
- Alam T. Device-to-Device communications in cloud, MANET and Internet of Things integrated architecture. *Journal of Information Systems Engineering and Business Intelligence*, 2020, vol. 6, no. 1, pp. 18–26. <https://doi.org/10.20473/jisebi.6.1.18-26>
- Nehra D., Dhindsa K.S., Bhushan B. A Security Model to Make Communication Secure in Cluster-Based MANETs. *Advances in Intelligent Systems and Computing*, 2020, vol. 1079, pp. 183–193. https://doi.org/10.1007/978-981-15-1097-7_16
- Olanrewaju R.F., Khan B.U.I., Anwar F., Mir R.N., Yaacob M., Mehraj T. Bayesian signaling game based efficient security model for MANETs. *Lecture Notes in Networks and Systems*, 2020, vol. 70, pp. 1106–1122. https://doi.org/10.1007/978-3-030-12385-7_75
- Shabut A.M., Kaiser M.Sh., Dahal K.P., Chen W. A multidimensional trust evaluation model for MANETs. // *Journal of Network and Computer Applications*. 2018, vol. 123, pp. 32–41. <https://doi.org/10.1016/j.jnca.2018.07.008>
- Salama H.M., El Mageed M.Z.A., Salama G.I.M., Badran K.M. CSMCSM: Client-Server Model for Comprehensive Security in MANETs. *International Journal of Information Security and Privacy*, 2021, vol. 15, no. 1, pp. 44–64. <https://doi.org/10.4018/IJISP.2021010103>

- Engineering (ICCSCE). 2014. P. 559–564. <https://doi.org/10.1109/ICCSCE.2014.7072781>
9. Kamel M.B.M., Alameri I., Onaizah A.N. STAODV: a secure and trust based approach to mitigate blackhole attack on AODV based MANET // *Proc. of the 2nd IEEE Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*. 2017. P. 1278–1282. <https://doi.org/10.1109/IAEAC.2017.8054219>
10. Khare A.K., Rana J.L., Jain R.C. Detection of wormhole, blackhole and DDOS attack in MANET using trust estimation under fuzzy logic methodology // *International Journal of Computer Network and Information Security (IJCNIS)*. 2017. V. 9. N 7. P. 29–35. <https://doi.org/10.5815/ijcnis.2017.07.04>
11. El-Semary A.M., Diab H. BP-AODV: Blackhole protected AODV routing protocol for MANETs based on chaotic map // *IEEE Access*. 2019. V. 7. P. 95197–95211. <https://doi.org/10.1109/ACCESS.2019.2928804>
12. Khan S., Hashim F., Rasid M.F.A., Perumal T. Reducing the severity of black hole and DDos attacks in MANETs by modifying AODV protocol using MAC authentication and symmetric encryption // *Proc. of the 2nd International Conference on Telematics and Future Generation Networks (TAFGEN)*. 2018. P. 109–114. <https://doi.org/10.1109/TAFGEN.2018.8580488>
13. Li G., Yan Z., Fu Y. A study and simulation research of blackhole attack on mobile AdHoc network // *Proc. of the 6th IEEE Conference on Communications and Network Security (CNS)*. 2018. P. 8433148. <https://doi.org/10.1109/CNS.2018.8433148>
14. Khan D.M., Aslam T., Akhtar N., Qadri S., Khan N.A., Rabbani I.M., Aslam M. Black hole attack prevention in mobile ad-hoc network (MANET) using ant colony optimization technique // *Information Technology and Control*. 2020. V. 49. N 3. P. 308–319. <https://doi.org/10.5755/j01.itc.49.3.25265>
15. Gautam D., Tokekar V. A novel Approach for Detecting DDos Attack in MANET // *Materials Today: Proceedings*. 2020. V. 29. P. 674–677. <https://doi.org/10.1016/j.matpr.2020.07.332>
16. Рзаев Б.Т., Лебедев И.С. Применение бэггинга при поиске аномалий сетевого трафика // *Научно-технический вестник информационных технологий, механики и оптики*. 2021. Т. 21. № 2. С. 234–240. <https://doi.org/10.17586/2226-1494-2021-21-2-234-240>
17. Tomar R., Prateek M., Sastry H.G. A novel approach to multicast in VANET using MQTT // *Ada User Journal*. 2017. V. 38. N 4. P. 231–235.
18. Potrino G., De Rango F., Santamaria A.F. Modeling and evaluation of a new IoT security system for mitigating DoS attacks to the MQTT broker // *Proc. of the IEEE Wireless Communications and Networking Conference (WCNC)*. 2019. P. 8885553. <https://doi.org/10.1109/WCNC.2019.8885553>
19. Sadeq A.S., Hassan R., Al-Rawi S.S., Jubair A.M., Aman A.H.M. A QoS approach for Internet of Things (IoT) environment using MQTT protocol // *Proc. of the 2019 International Conference on Cybersecurity (ICoCSec)*. 2019. P. 59–63. <https://doi.org/10.1109/ICoCSec47621.2019.8971097>
8. Alani M.M. MANET security: A survey. *Proc. of the 4th IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*, 2014, pp. 559–564. <https://doi.org/10.1109/ICCSCE.2014.7072781>
9. Kamel M.B.M., Alameri I., Onaizah A.N. STAODV: a secure and trust based approach to mitigate blackhole attack on AODV based MANET. *Proc. of the 2nd IEEE Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, 2017, pp. 1278–1282. <https://doi.org/10.1109/IAEAC.2017.8054219>
10. Khare A.K., Rana J.L., Jain R.C. Detection of wormhole, blackhole and DDOS attack in MANET using trust estimation under fuzzy logic methodology. *International Journal of Computer Network and Information Security (IJCNIS)*, 2017, vol. 9, no. 7, pp. 29–35. <https://doi.org/10.5815/ijcnis.2017.07.04>
11. El-Semary A.M., Diab H. BP-AODV: Blackhole protected AODV routing protocol for MANETs based on chaotic map. *IEEE Access*, 2019, vol. 7, pp. 95197–95211. <https://doi.org/10.1109/ACCESS.2019.2928804>
12. Khan S., Hashim F., Rasid M.F.A., Perumal T. Reducing the severity of black hole and DDos attacks in MANETs by modifying AODV protocol using MAC authentication and symmetric encryption. *Proc. of the 2nd International Conference on Telematics and Future Generation Networks (TAFGEN)*, 2018, pp. 109–114. <https://doi.org/10.1109/TAFGEN.2018.8580488>
13. Li G., Yan Z., Fu Y. A study and simulation research of blackhole attack on mobile AdHoc network. *Proc. of the 6th IEEE Conference on Communications and Network Security (CNS)*, 2018, pp. 8433148. <https://doi.org/10.1109/CNS.2018.8433148>
14. Khan D.M., Aslam T., Akhtar N., Qadri S., Khan N.A., Rabbani I.M., Aslam M. Black hole attack prevention in mobile ad-hoc network (MANET) using ant colony optimization technique. *Information Technology and Control*, 2020, vol. 49, no. 3, pp. 308–319. <https://doi.org/10.5755/j01.itc.49.3.25265>
15. Gautam D., Tokekar V. A novel Approach for Detecting DDos Attack in MANET. *Materials Today: Proceedings*, 2020, vol. 29, pp. 674–677. <https://doi.org/10.1016/j.matpr.2020.07.332>
16. Rzaev B.T., Lebedev I.S. Applying bagging in finding network traffic anomalies. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2021, vol. 21, no. 2, pp. 234–240. (in Russian). <https://doi.org/10.17586/2226-1494-2021-21-2-234-240>
17. Tomar R., Prateek M., Sastry H.G. A novel approach to multicast in VANET using MQTT. *Ada User Journal*, 2017, vol. 38, no. 4, pp. 231–235.
18. Potrino G., De Rango F., Santamaria A.F. Modeling and evaluation of a new IoT security system for mitigating DoS attacks to the MQTT broker. *Proc. of the IEEE Wireless Communications and Networking Conference (WCNC)*, 2019, pp. 8885553. <https://doi.org/10.1109/WCNC.2019.8885553>
19. Sadeq A.S., Hassan R., Al-Rawi S.S., Jubair A.M., Aman A.H.M. A QoS approach for Internet of Things (IoT) environment using MQTT protocol. *Proc. of the 2019 International Conference on Cybersecurity (ICoCSec)*, 2019, pp. 59–63. <https://doi.org/10.1109/ICoCSec47621.2019.8971097>

Авторы

Легашев Леонид Вячеславович — кандидат технических наук, ведущий научный сотрудник, Оренбургский государственный университет, Оренбург, 460018, Российская Федерация, <https://orcid.org/0000-0001-6351-404X>, silentgir@gmail.com

Гришина Любовь Сергеевна — преподаватель, Оренбургский государственный университет, Оренбург, 460018, Российская Федерация, <https://orcid.org/0000-0003-2752-7198>, zabrodina97@inbox.ru

Парфенов Денис Игоревич — кандидат технических наук, начальник отдела, Оренбургский государственный университет, Оренбург, 460018, Российская Федерация, <https://orcid.org/0000-0002-1146-1270>, parfenovdi@mail.ru

Жигалов Артур Юрьевич — ведущий программист, Оренбургский государственный университет, Оренбург, 460018, Российская Федерация, <https://orcid.org/0000-0003-3208-1629>, leroy137.artur@gmail.com

Autors

Leonid V. Legashev — PhD, Leading Researcher, Orenburg State University, Orenburg, 460018, Russian Federation, <https://orcid.org/0000-0001-6351-404X>, silentgir@gmail.com

Lubov S. Grishina — Lecturer, Orenburg State University, Orenburg, 460018, Russian Federation, <https://orcid.org/0000-0003-2752-7198>, zabrodina97@inbox.ru

Denis I. Parfenov — PhD, Head of Department, Orenburg State University, Orenburg, 460018, Russian Federation, <https://orcid.org/0000-0002-1146-1270>, parfenovdi@mail.ru

Arthur Yu. Zhigalov — Leading Software Developer, Orenburg State University, Orenburg, 460018, Russian Federation, <https://orcid.org/0000-0003-3208-1629>, leroy137.artur@gmail.com

doi: 10.17586/2226-1494-2022-22-4-708-715

УДК 004.523

Применение FN-корректора с целью повышения качества классификации аудиособытий

Александр Михайлович Голубков¹✉, Евгений Витальевич Шуранов²

¹ Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина), Санкт-Петербург, 197022, Российская Федерация

^{1,2} ООО «ТЕХКОМПАНИЯ ХУАВЭЙ», Москва, 123007, Российская Федерация

² Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

¹ kremnikov@gmail.com✉, <https://orcid.org/0000-0002-8330-1823>

² e_v_shuranov@mail.ru, <https://orcid.org/0000-0003-0977-5075>

Аннотация

Предмет исследования. Рассмотрена проблема классификации акустических событий, активно применяемых в решениях задач безопасного города, умного дома, IoT устройств, а также для детектирования опасных ситуаций на производстве. Предложено решение повышения точности классификаторов без изменения их структуры и сбора дополнительных данных. Основным источником данных для экспериментов выбран открытый набор данных TUT Urban Acoustic Scenes 2018, Development Dataset. **Метод.** Предложен способ увеличения точности классификации аудиособытий с помощью использования FN-корректора. FN-корректор представляет собой линейный классификатор и работает в два этапа: преобразование пространства признаков в линейно-разделимое пространство и линейное отделение одного класса от другого. В случае применения корректора классы — типы ответов исходного классификатора: положительный (P), отрицательный (N), ложноположительный (FP) и ложноотрицательный (FN). В результате возможно обучить два типа корректоров FP и FN, которые работают как бинарные линейные классификаторы и разделяют ответы на положительные/ложноположительные и отрицательные/ложноотрицательные соответственно. Выполнены эксперименты, где в качестве исходного классификатора использована сверточная нейронная сеть VGGish. Аудиосигнал преобразован в спектрограмму и передан на вход нейронной сети, которая формирует признаковое описание спектрограммы и производит классификацию. **Основные результаты.** В качестве примера демонстрации повышения точности классификации выбраны два «спутанных» класса. С помощью признакового описания аудиозаписей этих классов построен, обучен FN-корректор и подключен к исходному классификатору. Ответ от классификатора, а также признаковое описание передано на вход корректора. Далее корректор переводит пространство признаков в новый базис (в линейно-разделимое пространство) и классифицирует ответ классификатора, таким образом «отвечает» на вопрос, ошибается ли исходный классификатор на таком векторе признаков или нет. Если исходный классификатор ошибся, то его ответ изменяется корректором на противоположный. Если нет — ответ остается тем же самым. Результаты экспериментов продемонстрировали снижение уровня спутывания классов и, соответственно, увеличение точности исходного классификатора без изменения его структуры и без сбора дополнительного набора данных. **Практическая значимость.** Полученные результаты могут быть использованы на устройствах IoT, имеющих существенные ограничения по размеру используемых моделей, а также при решении проблем доменной адаптации, актуальной в задачах аудиоаналитики.

Ключевые слова

классификация аудиособытий, обработка звука, FN-корректор, корректор ложноотрицательных ответов, сверточные сети, аудиоаналитика

Благодарности

Работа выполнена в рамках исследований, поддерживаемых СПбГЭТУ «ЛЭТИ» им В.И. Ульянова (Ленина).

Ссылка для цитирования: Голубков А.М., Шуранов Е.В. Применение FN-корректора с целью повышения качества классификации аудиособытий // Научно-технический вестник информационных технологий, механики и оптики. 2022. Т. 22, № 4. С. 708–715. doi: 10.17586/2226-1494-2022-22-4-708-715

Applying the FN-Corrector to improve the quality of audio event classification

Alexander M. Golubkov¹, Evgeny V. Shuranov²

¹ Saint Petersburg Electrotechnical University "LETI", Saint Petersburg, 197022, Russian Federation

^{1,2} Huawei, Moscow, 123007, Russian Federation

² ITMO University, Saint Petersburg, 197101, Russian Federation

¹ kremnikov@gmail.com, <https://orcid.org/0000-0002-8330-1823>

² e_v_shuranov@mail.ru, <https://orcid.org/0000-0003-0977-5075>

Abstract

The paper deals with the problem of acoustic events classification which is actively applied to the problems of a safe city, smart home, IoT devices, and for the detection of industrial accidents. A solution to improve the accuracy of classifiers without changing their structure and collecting additional data is proposed. The main data source for the experiments was the TUT Urban Acoustic Scenes 2018, Development Dataset. The paper presents the way to increase the accuracy of audio event classification by using the FN-corrector. The FN-corrector is a linear two-stage classifier performing the transformation of the feature space into a linearly separable space and the linear separation of one class from another. If a corrector is applied, the responses of the original classifier generate four classes: positive (P), negative (N), false positive (FP), and false negative (FN). As a result, it becomes possible to train two types of correctors: the FP-corrector separating positive and false positive classifier responses, and the FN-corrector separating negative and false negative classifier responses. In the experiments, the VGGish convolutional neural network was used as the initial classifier. The audio signal is converted into a spectrogram and is fed to the input of the neural network which forms the spectrogram feature description and performs a classification. As an example, two "confused" classes are selected to demonstrate the increase in classification accuracy. Using the feature description of audio recordings of these classes, an FN-corrector was built, trained and connected to the original classifier. The response from the classifier, as well as the feature description, has been passed to the corrector input. Next, the corrector translated the feature space into a new basis (into a linearly separable space) and classified the classifier answer responding to the question whether the original classifier makes a mistake on such a feature vector or not. If the original classifier made a mistake, then his answer is changed by the corrector to the opposite, otherwise the answer remains the same. The results of the experiments demonstrated a decrease in the level of class confusion and, accordingly, an increase in the accuracy of the original classifier without changing its structure and without collecting an additional data set. The results obtained can be used on IoT devices that have significant limitations on the size of the models used, as well as in solving the problems of domain adaptation which is relevant in audio analytics.

Keywords

acoustic event detection, audio processing, FN-corrector, false negative corrector, DSP, CNN, convolutional neural network, audio analytics

Acknowledgements

The work was carried out as part of research supported by LETI.

For citation: Golubkov A.M., Shuranov E.V. Applying the FN-corrector to improve the quality of audio event classification. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2022, vol. 22, no. 4, pp. 708–715 (in Russian). doi: 10.17586/2226-1494-2022-22-4-708-715

Введение

Решение задачи классификации акустических событий заключается в определении события, которое происходит на анализируемой аудиозаписи. Для решения данной задачи применяются нейронные сети разных архитектур с использованием разнообразных признаков представлений. При этом применяются записи сигнала от набора амплитуд до модификаций MFCC-коэффициентов (Mel-Frequency Cepstral Coefficients, мел-частотные кепстральные коэффициенты) и предобученных признаков (OpenL3 [1] и i-vectors (вектора представления голоса человека)), в случае если требуется, кроме классификации, решить сопутствующие задачи сегментации или локализации [2].

Актуальной задачей является улучшение точности существующих алгоритмов. Регулярные конкурсы по улучшению точности при различных условиях и ограничениях активно поддерживаются как академическими, так и промышленными исследовательскими

командами¹. Цель настоящей работы — получение способа повышения точности классификации на стадии постпроцессинга, что позволит применить данный способ для широкого класса решений. Предложено использовать FN-корректор (False Negative корректор, т. е. корректор, исправляющий ложноотрицательные ошибки) [3, 4]. Модель корректора основана на использовании каскадной редукции размерности признаков и позволяет за линейное время скорректировать ошибочный ответ классификатора.

Способ применения метода каскадной редукции к задаче классификации аудиособытий

Метод каскадной редукции пространства признаков заключается в последовательном уменьшении размер-

¹ Detection and Classification of Acoustic Scenes and Events. [Электронный ресурс]. URL: <https://dcase.community> (дата обращения: 24.06.2022).

ности и преобразовании пространства признаков векторов. В этом случае независимо и случайно выбранный вектор одного класса отделяется от независимо и случайно выбранного вектора другого класса линейным дискриминантом (например, линейным дискриминантом Фишера). В основе метода представлена модель системы, способная исправить (скорректировать) ответы существующего классификатора. Принцип работы системы заключается в переразметке исходного набора данных с разметки «объект, класс» на разметку «признаковое описание объекта, тип ответа классификатора». При этом способ построения векторов признаков никак не фиксируется и, следовательно, качество работы системы напрямую зависит от метода получения признаков векторов. Результат будет точнее, если: больше информации об объекте сохранено в его векторе признаков, и более выраженной кластерной структурой обладают вектора в пространстве признаков.

В работе представлен способ применения системы для решения задачи улучшения точности классификации аудиособытий. При этом выбраны следующие условия: в качестве объекта классификации — мел-спектрограмма аудиозаписи; в качестве способа получения вектора признаков — предобученная сверточная нейронная сеть VGGish; в качестве исходного классификатора — последовательно подключенные к сети VGGish полносвязный слой и слой Softmax, выходом которого является распределение вероятностей принадлежности спектрограммы каждому из классов.

Предлагаемый способ улучшения точности исходного классификатора предполагает следующие шаги.

Шаг 1. Выделение классов, на которых исходный классификатор допускает ошибки, и определение типа этих ошибок с помощью тренировочного набора данных.

Шаг 2. Построение спектрограммы и вектора признаков каждой аудиозаписи из тренировочного набора данных.

Шаг 3. Выбор типа ошибки, которую предполагается исправить (FN или FP) и дальнейшая переразметка тренировочного набора данных.

Шаг 4. Преобразование пространства признаков методом каскадной редукции таким образом, чтобы отделить те вектора признаков, на которых классификатор допускает ошибки (FN ошибки), от тех, на которых не допускает (т. е. правильные отрицательные ответы, N).

Шаг 5. Обучение линейного дискриминанта Фишера для проведения бинарной классификации в построенном на Шаге 4 пространстве.

Шаг 6. Подключение полученной системы коррекции к выходу исходного классификатора.

Шаг 7. Проведение процесса исправления ответа исходного классификатора: если система коррекции отнесла вектор признаков к классу ошибочных ответов, то исходный ответ корректора исправляется на противоположный. Если к классу правильных, то ответ остается тем же самым.

Использование предложенного способа исправления ответов исходного классификатора, в частности — FN-корректора и сверточной нейронной сети в качестве способа получения векторов признаков, показало зна-

чимый прирост точности в задаче классификации акустических событий.

Практическая реализация способа увеличения точности классификации аудиособытий с помощью использования FN-корректора

Выберем набор данных TUT Urban Acoustic Scenes 2018¹, содержащий 10 классов акустических событий: *аэропорт, торговый центр, станция метро, пешеходный переход, городская площадь, городская улица, поездка в трамвае, поездка в автобусе, поездка в метро и городской парк*. Весь набор состоит из 8640 записей акустических событий, каждая из которых принадлежит одному из классов (закрытая задача). Каждый класс имеет 864 аудиозаписей. Суммарная длительность записей составляет 24 часа. Задача эксперимента — сопоставить каждую аудиозапись из набора соответствующему классу. Каждая запись может и должна принадлежать только одному классу.

Подготовка аудиозаписей. Для ускорения обучения нейронных сети и повышения ее точности аудиозапись разбита на отрезки по 100 мс, каждому из которых присвоен класс аудиозаписи. Каждая аудиозапись имеет два канала (с правого и левого микрофонов) с частотой дискретизации 16 кГц и разрядностью 16 бит. В настоящей работе использован метод аналого-цифровых преобразователей — PCM 16 (Pulse Code Modulation), а также стандарт квантования сигнала по уровню, в данном случае уровень сигнала кодируется 16 битным значением. Общая схема подготовки аудиозаписей представлена на рис. 1.

Построение признакового описания аудиозаписей. Перед расчетом признакового описания стереозапись делится на два канала, при этом каждый обрабатывается отдельно. Для улучшения качества классификации из моноканала выделяются еще два канала — гармонический и перкуссивный, используя для этого алгоритм HPSS (Harmonic-Percussive Sound Separation) [5]. Таким образом, двухканальная аудиозапись преобразована в четыре моноканальные записи.

В качестве способа представления аудиозаписей выбраны мел-спектрограммы со следующими параметрами: размер окна преобразования Фурье 100 мс; длина наложения 20 мс; количество банков мел-частотных фильтров 128.

Подготовка и обучение нейронной сети. Выберем сверточную нейронную сеть с 4-канальным входом на основе архитектуры VGGish. Каждый канал сформируем путем построения следующих спектрограмм: мел-спектрограммы левого и правого аудиоканалов, а также применение фильтра-гармоники и фильтра-перкуссии к усредненной по каналам мел-спектрограмме. Далее ко всем каналам применим набор фильтров-сверток, составив таким образом карты признаков, которые «разворачиваются» в одномерный вектор и подаются на вход полносвязному выходному слою. На выходе

¹ TUT Urban Acoustic Scenes 2018, Development Dataset [Электронный ресурс]. Режим доступа: <https://zenodo.org/record/1228142> (дата обращения: 11.07.2022).

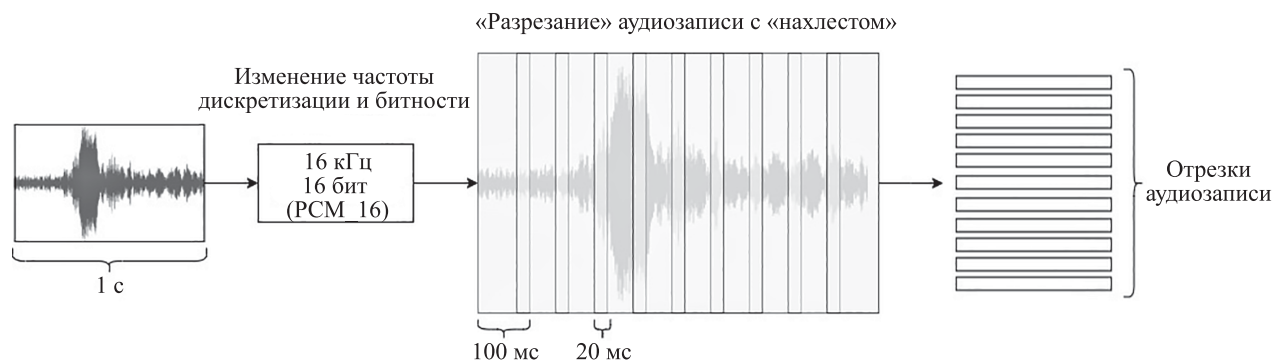


Рис. 1. Подготовка аудиозаписей
Fig. 1. Audio record preprocessing

нейронной сети установим слой Softmax, настроенный на 10 классов. Таким образом, ответом обученной нейронной сети будет распределение вероятности принадлежности аудиозаписи каждому из классов. Общая архитектура используемой сети представлена на рис. 2.

Обучение нейронной сети выполним методом обратного распространения ошибки с использованием стохастического градиентного спуска, а также момента Нестерова [6] для ускорения обучения и предотвращения переобучения. Шаг обучения составил 0,01, размер батча — 128. Результаты работы обученной нейронной сети на тестовом наборе данных представлены на рис. 3 и в табл. 1.

Из полученных данных видно, что классы «поездка в трамвае» и «поездка в автобусе» распознаются недостаточно хорошо, так как «спутаются» друг с другом. Решить данную проблему предложено с помощью подключения корректоров к тем классам, на которых ошибается модель.

Применение корректора ошибок

Для устранения проблемы «спутывания» классов применим тип корректора — FN, который исправит ложноотрицательные срабатывания классификатора. Чтобы построить такой корректор, в первую очередь необходимо преобразовать обучающий набор данных. Для этого отберем все примеры в наборе, которые соответствуют необходимому классу. Рассмотрим случай «спутывания» классов «поездка в трамвае» и «поездка в автобусе», при этом классификатор имеет большое

Таблица 1. Качество классификации обученной модели для каждого класса

Table 1. Model precision and recall

Класс	Точность	Полнота
Аэропорт	0,8624	0,8704
Торговый центр	0,8876	0,9502
Станция метро	0,9197	0,9803
Пешеходный переход	0,9624	0,6798
Городская площадь	0,7879	0,8299
Городская улица	0,6450	0,7697
Поездка в трамвае	0,7063	0,5706
Поездка в автобусе	0,6272	0,7303
Поездка в метро	0,9140	0,9097
Городской парк	0,9963	0,9306
Средняя невзвешенная точность (UAP):		0,8309
Средняя невзвешенная полнота (UAR):		0,8221

количество ложноотрицательных (FN) срабатываний на классе «поездка в трамвае». Для подготовки данных необходимо выполнить следующие шаги.

Шаг 1. Отобрать из тестового набора данных все примеры, соответствующие классу «поездка в трамвае».

Шаг 2. Применить обученный классификатор (в данном случае — нейронная сеть) для классификации примеров из Шага 1.

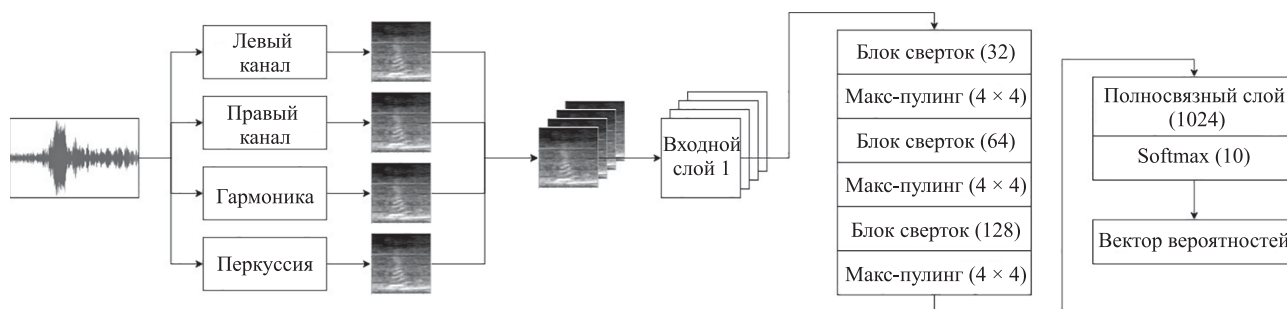


Рис. 2. Общая архитектура нейронной сети
Fig. 2. Audio classification pipeline

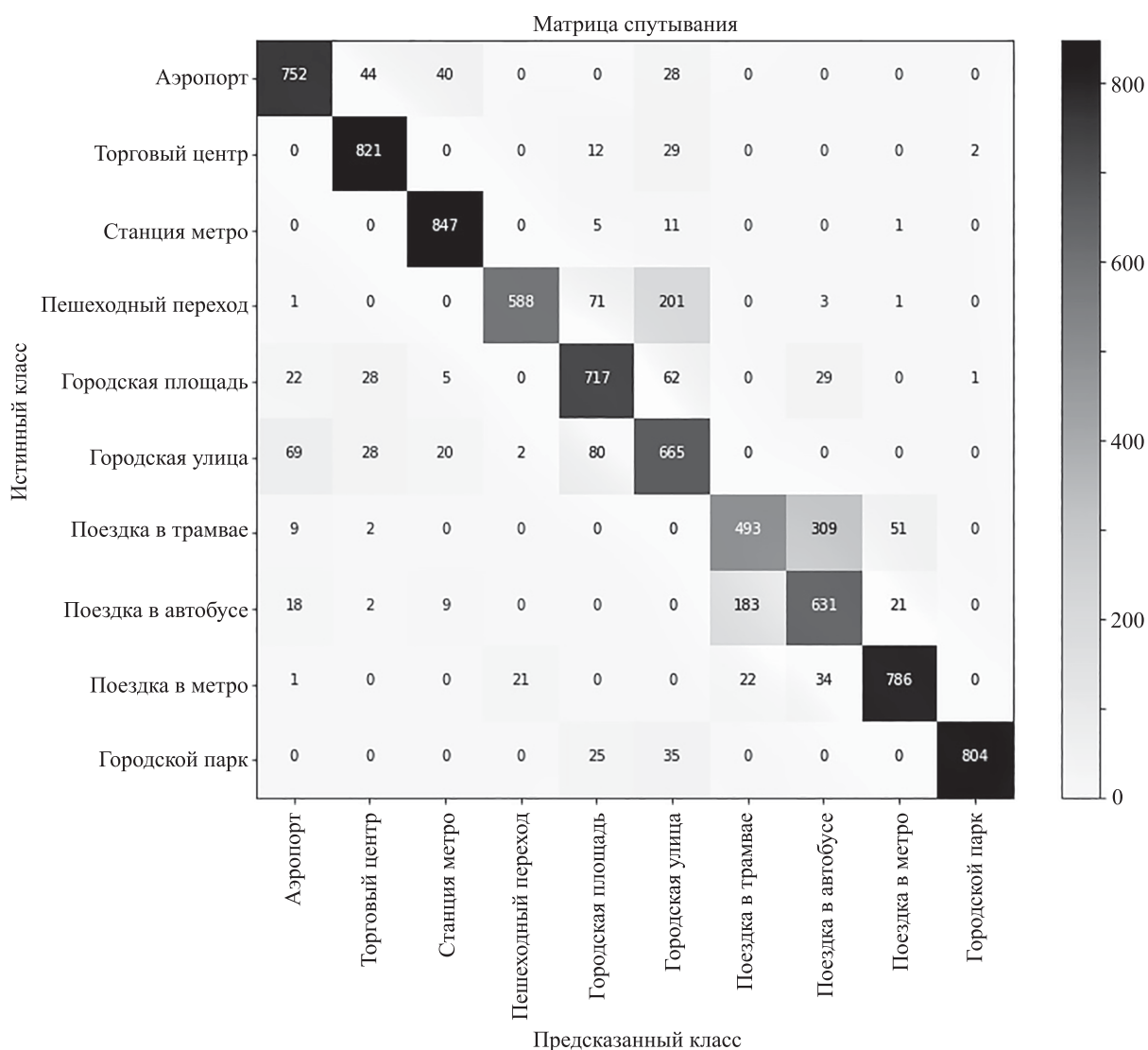


Рис. 3. Результаты классификации обученной модели в виде матрицы спутывания

Fig. 3. Classification confusion matrix

Шаг 3. На основе ответов классификатора составить набор данных вида {признаковое описание примера, ответ}, где *ответом* является 1, если ответ правильный, и 0, если ответ неправильный, т. е. классификатор «спутал» данный класс с другим.

Схема преобразования набора данных для корректора, рассматриваемого в эксперименте, приведена на рис. 4.

Обучим FN-корректор, как описано в работе [7]. В результате построим систему исправления ответа на классе «поездка в трамвае» (рис. 5).

После применения FN-корректора к выходу классификатора получим результаты, представленные на рис. 6 и в табл. 2.

В результате применения FN-корректора количество правильно распознанных объектов класса «поездка в трамвае» улучшилось с 493 (рис. 3) до 750 (рис. 6). Так как классификатор не был дополнительно изменен и обучен, то полученный результат говорит о том, что FN-корректор успешно решил поставленную задачу. По итогам эксперимента (табл. 2) видно увеличение точ-

Таблица 2. Качество классификации обученной модели для каждого класса после применения FN-корректора

Table 2. Classification quality of the trained model for each class after application of the corrector

Класс	Точность	Полнота
Аэропорт	0,8664	0,8704
Торговый центр	0,8876	0,9502
Станция метро	0,9197	0,9803
Пешеходный переход	0,9624	0,6798
Городская площадь	0,7879	0,8299
Городская улица	0,6450	0,7697
Поездка в трамвае	0,7853	0,7042
Поездка в автобусе	0,6614	0,7303
Поездка в метро	0,9140	0,9097
Городской парк	0,9963	0,9306
Среднее	0,8426	0,8355

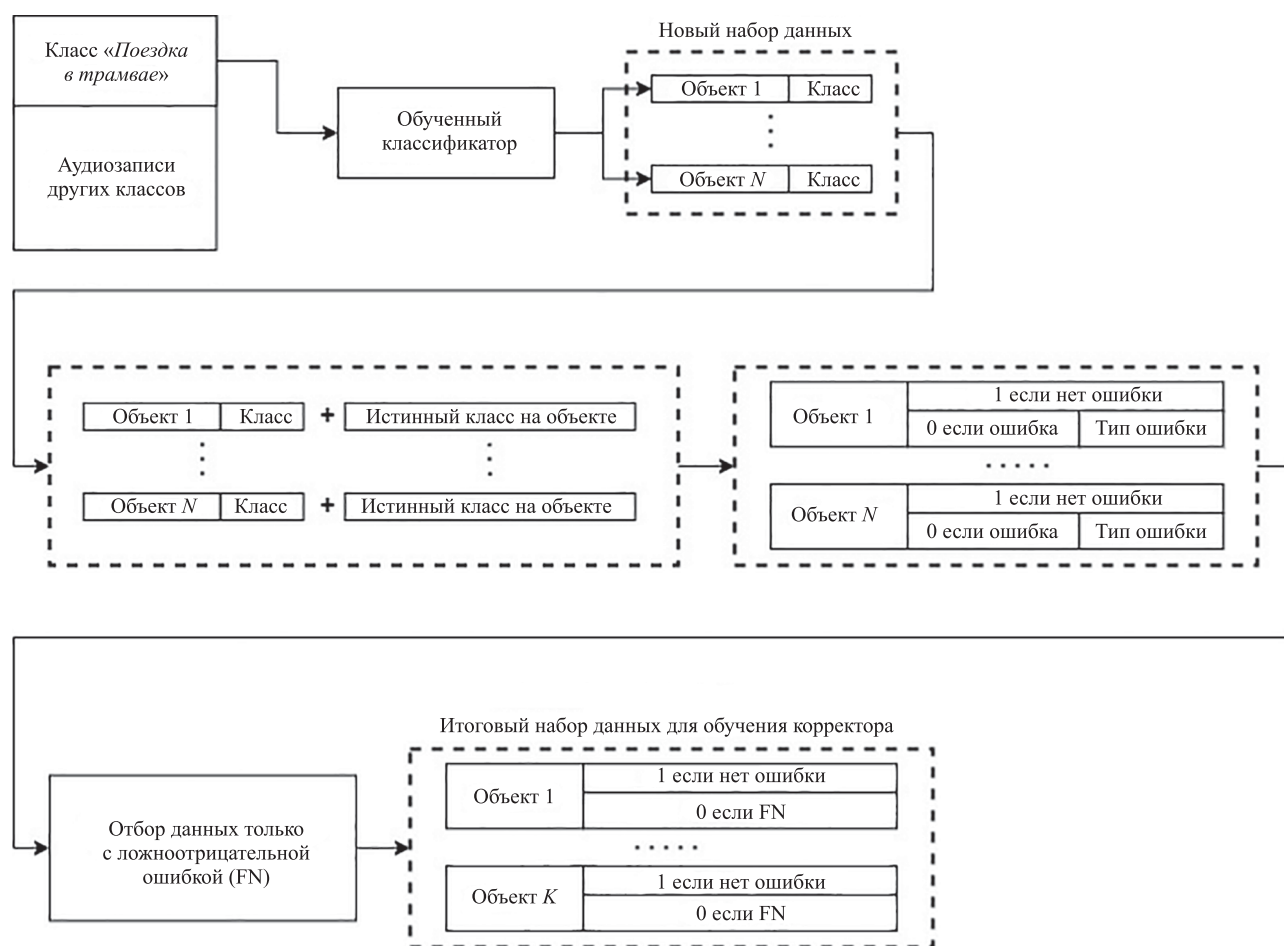


Рис. 4. Схема преобразования набора данных для построения FN-корректора

Fig. 4. Data preprocessing to train FN-corrector

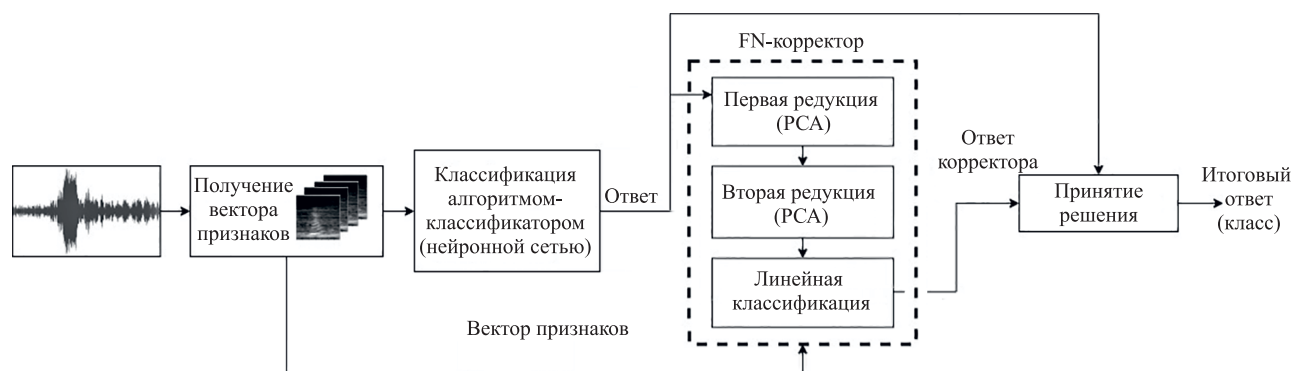


Рис. 5. Схема системы классификатор-корректор FN в применении к классу «поездка в трамвае»

Fig. 5. Classification-correction pipeline

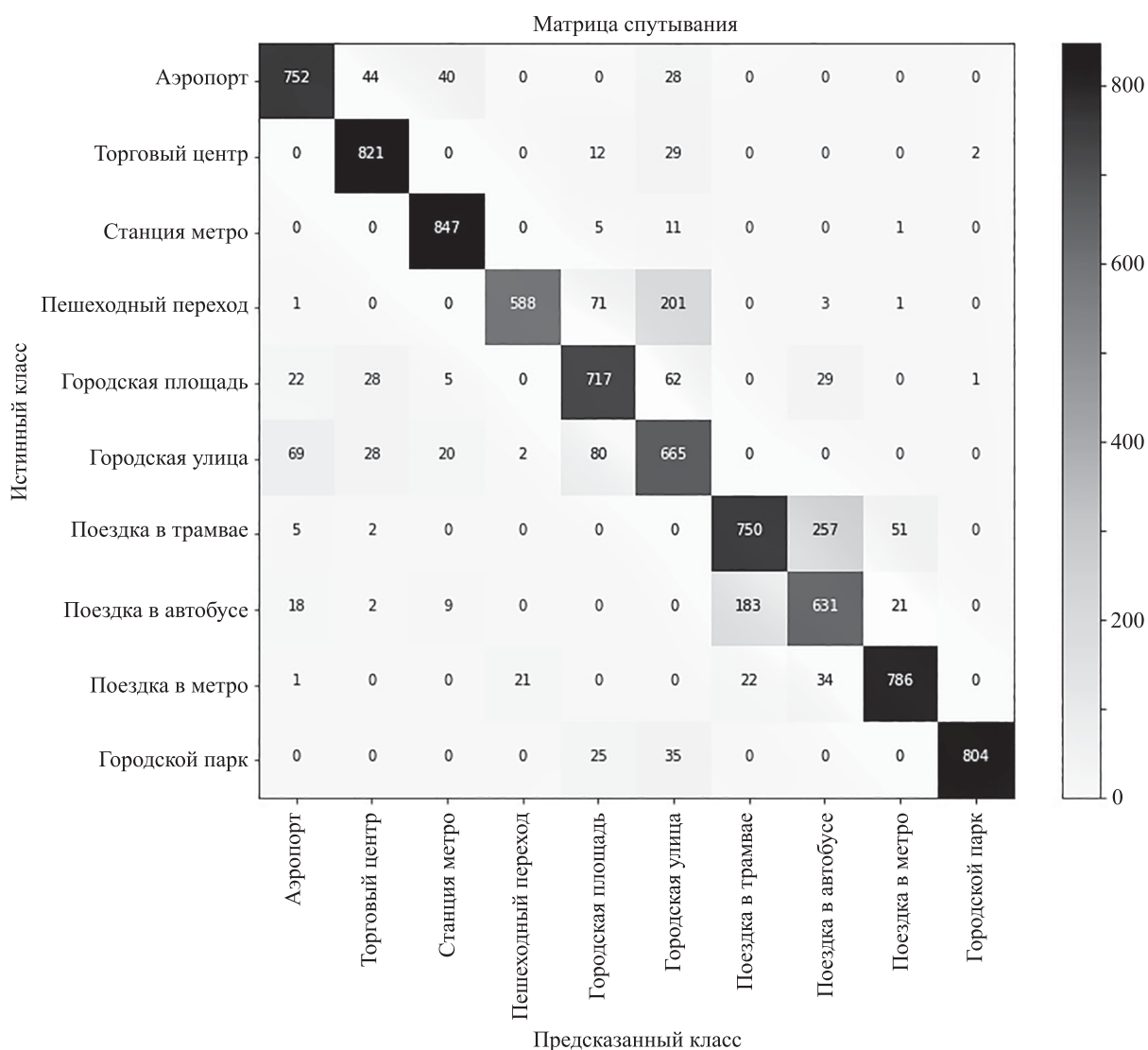


Рис. 6. Матрица смешивания после применения FN корректора

Fig. 6. Classification-correction of the model confusion matrix

ности классификатора до 0,8426 и полноты до 0,8355 (значимость результатов подтверждается проверкой методом кросс-валидации по 4 группам).

Закключение

В работе представлен и описан способ использования метода каскадной редукции пространства признаков при решении задачи классификации акустических событий с помощью построения системы коррекции ответов классификатора. Продемонстрировано статистически значимое увеличение точности классификатора при использовании описанного способа для решения задачи классификации акустических событий. Выделен класс, который классификатор распознает недостаточно

хорошо («спутывает» его с другим классом), и на данном классе был обучен и применен FN-корректор. В результате эксперимента показано улучшение точности классификатора с 0,8308 до 0,8426 и полноты с 0,8221 до 0,8355 (значимость результатов подтверждается проверкой методом кросс-валидации по 4 группам). Преимущество описанного способа — возможность применения в качестве постобработки и использование совместно со многими существующими архитектурами классификаторов. Другое потенциальное преимущество — возможность создания модели небольшого размера, использующей данный способ, что важно для задач обработки звука на устройствах, например IoT. В дальнейших работах планируется исследовать возможности задач обработки звука на устройствах IoT.

Литература

1. Grollmisch S., Cano E., Kehling C., Taenzer M. Analyzing the potential of pre-trained embeddings for audio classification tasks // *Proc. of the 28th European Signal Processing Conference (EUSIPCO)*. 2021. P. 790–794. <https://doi.org/10.23919/Eusipco47968.2020.9287743>
2. Matveev Y.N., Shuranov E.V., Avdeeva A.S., Shchemelinin V.L., Krylova E.V. Acoustic data based automatic object detection system // *Proc. of the 2nd International Conference on Control in Technical Systems (CTS)*. 2017. P. 301–303. <https://doi.org/10.1109/CTS.2017.8109551>
3. Голубков А.М. Бинарная классификация изображений на примере задачи распознавания лиц // *Известия СПбГЭТУ «ЛЭТИ»*. 2018. № 7. С. 26–30.
4. Голубков А.М., Клионский Д.М. Применение метода каскадной редукции к решению задачи распознавания лиц // *Известия СПбГЭТУ «ЛЭТИ»*. 2019. № 8. С. 47–53.
5. Ono N., Miyamoto K., Le Roux J., Kameoka H., Sagayama S. Separation of a monaural audio signal into harmonic/percussive components by complementary diffusion on spectrogram // *Proc. of the 16th European Signal Processing Conference (EUSIPCO)*. 2008. P. 1–4.
6. Sutskever I., Martens J., Dahl G., Hinton G. On the importance of initialization and momentum in deep learning // *Proc. of the 30th International Conference on Machine Learning (ICML)*. 2013. P. 2176–2184.
7. Gorban A., Golubkov A.M., Grechuk B., Mirkes E., Tyukin I.Y. Correction of AI systems by linear discriminants: probabilistic foundations // *Information Sciences*. 2018. V. 466. P. 303–322. <https://doi.org/10.1016/j.ins.2018.07.040>

Авторы

Голубков Александр Михайлович — кандидат технических наук, ассистент, Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина), Санкт-Петербург, 197022, Российская Федерация; лидер команды аудио аналитики, ООО «ТЕХКОМПАНИЯ ХУАВЭЙ», Москва, 123007, Российская Федерация, [sc 57190975154](https://orcid.org/0000-0002-8330-1823), <https://orcid.org/0000-0002-8330-1823>, kremnikov@gmail.com

Шуранов Евгений Витальевич — кандидат технических наук, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация; руководитель лаборатории, ООО «ТЕХКОМПАНИЯ ХУАВЭЙ», Москва, 123007, Российская Федерация, [sc 57190970283](https://orcid.org/0000-0003-0977-5075), <https://orcid.org/0000-0003-0977-5075>, e_v_shuranov@mail.ru

References

1. Grollmisch S., Cano E., Kehling C., Taenzer M. Analyzing the potential of pre-trained embeddings for audio classification tasks. *Proc. of the 28th European Signal Processing Conference (EUSIPCO)*, 2021, pp. 790–794. <https://doi.org/10.23919/Eusipco47968.2020.9287743>
2. Matveev Y.N., Shuranov E.V., Avdeeva A.S., Shchemelinin V.L., Krylova E.V. Acoustic data based automatic object detection system. *Proc. of the 2nd International Conference on Control in Technical Systems (CTS)*, 2017, pp. 301–303. <https://doi.org/10.1109/CTS.2017.8109551>
3. Golubkov A.M. Face recognition using images binary classification methods. *Proceedings of Saint Petersburg Electrotechnical University Journal*, 2018, no. 7, pp. 26–30. (in Russian)
4. Golubkov A.M., Klionskii D.M. Cascade reduction method applied to face recognition problem. *Proceedings of Saint Petersburg Electrotechnical University Journal*, 2019, no. 8, pp. 47–53. (in Russian)
5. Ono N., Miyamoto K., Le Roux J., Kameoka H., Sagayama S. Separation of a monaural audio signal into harmonic/percussive components by complementary diffusion on spectrogram. *Proc. of the 16th European Signal Processing Conference (EUSIPCO)*, 2008, pp. 1–4.
6. Sutskever I., Martens J., Dahl G., Hinton G. On the importance of initialization and momentum in deep learning. *Proc. of the 30th International Conference on Machine Learning (ICML)*, 2013, pp. 2176–2184.
7. Gorban A., Golubkov A.M., Grechuk B., Mirkes E., Tyukin I.Y. Correction of AI systems by linear discriminants: probabilistic foundations. *Information Sciences*, 2018, vol. 466, pp. 303–322. <https://doi.org/10.1016/j.ins.2018.07.040>

Authors

Alexander M. Golubkov — PhD, Assistant, Saint Petersburg Electrotechnical University “LETI”, Saint Petersburg, 197022, Russian Federation; Senior Machine Learning Engineer, Huawei, Moscow, 123007, Russian Federation, [sc 57190975154](https://orcid.org/0000-0002-8330-1823), <https://orcid.org/0000-0002-8330-1823>, kremnikov@gmail.com

Evgeny V. Shuranov — PhD, Associate Professor, ITMO University, 197101, Russian Federation; Head of Laboratory, Huawei, Moscow, 123007, Russian Federation, [sc 57190970283](https://orcid.org/0000-0003-0977-5075), <https://orcid.org/0000-0003-0977-5075>, e_v_shuranov@mail.ru

Статья поступила в редакцию 17.04.2022
Одобрена после рецензирования 24.06.2022
Принята к печати 30.07.2022

Received 17.04.2022
Approved after reviewing 24.06.2022
Accepted 30.07.2022



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

doi: 10.17586/2226-1494-2022-22-4-716-724

УДК 004.2

Усиление роли микроархитектурных этапов проектирования встраиваемых систем

Максим Вячеславович Кольчурин¹, Василий Юрьевич Пинкевич^{2✉},
Алексей Евгеньевич Платунов³

^{1,3} Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

² ООО «ЛМТ», Санкт-Петербург, 199034, Российская Федерация

¹ maxim.kolchurin@gmail.com, <https://orcid.org/0000-0002-7061-9357>

² vasiliy.pinkevich@yandex.ru✉, <https://orcid.org/0000-0002-8635-5026>

³ aerlatunov@gmail.com, <https://orcid.org/0000-0003-3003-3949>

Аннотация

Предмет исследования. Растущее разнообразие вычислительных систем, стремительное увеличение их сложности, интеграции в объекты и процессы физического мира требуют резкого повышения производительности труда их создателей. Отмечено, что качество, сроки и степень повторного использования результатов проектирования в области информационных технологий сильно зависят от методологий и маршрутов проектирования на этапах выбора и/или создания стеков платформ, технологий и инструментов. Важнейшую роль в этом играют способы описания организации вычислительной системы на различных уровнях и применяемые системы абстракций. Проблема заполнения семантического разрыва между концептуальным (архитектурным) уровнем и уровнями реализации по-прежнему стоит очень остро. Следовательно, требуется создание промышленных методик и инструментов проектирования на «промежуточных» уровнях. **Метод.** В работе предложены способы представления проектных решений, которые направлены на целостное, сквозное описание как логики организации вычислительного процесса, так и шагов, технологий и инструментов процесса проектирования. **Основные результаты.** Подробно объяснено наполнение и обоснована необходимость этапов микроархитектурного проектирования вычислительных систем. Введена классификация проектов в области информационных технологий по степени вариативности проектной платформы. Предложен ряд понятий для представления набора абстракций микроархитектурного проектирования в рамках проектов с большой внутренней вариативностью. Подробно описаны следующие абстракции: проектное и аспектное пространства, проектные платформы и кросс-уровневые механизмы. **Практическая значимость.** Рассмотрены примеры представления ряда предложенных абстракций (рабочих инструментов документирования) микроархитектурных этапов проектирования, которые наиболее актуальны в проектировании вычислительных систем в модели «ограниченных ресурсов»: встраиваемые, киберфизические системы, «границные» и «туманные» уровни систем интернета вещей.

Ключевые слова

встраиваемая система, микроархитектура, исследование пространства проектных решений, аспектное проектирование, кросс-уровневые механизмы

Ссылка для цитирования: Кольчурин М.В., Пинкевич В.Ю., Платунов А.Е. Усиление роли микроархитектурных этапов проектирования встраиваемых систем // Научно-технический вестник информационных технологий, механики и оптики. 2022. Т. 22, № 4. С. 716–724. doi: 10.17586/2226-1494-2022-22-4-716-724

Strengthening the role of microarchitectural stages of embedded systems design

Maxim V. Kolchurin¹, Vasiliy Yu. Pinkevich²✉, Alexey E. Platunov³

^{1,3} ITMO University, Saint Petersburg, 197101, Russian Federation

² LMT Ltd., Saint Petersburg, 199034, Russian Federation

¹ maxim.kolchurin@gmail.com, <https://orcid.org/0000-0002-7061-9357>

² vasily.pinkevich@yandex.ru✉, <https://orcid.org/0000-0002-8635-5026>

³ aeplatunov@gmail.com, <https://orcid.org/0000-0003-3003-3949>

Abstract

The growing variety of computing systems, the rapid increase in their complexity, their integration into objects and processes of the physical world require a dramatic increase in the productivity of their creators. It is noted that the quality, timing, and degree of reuse of design results in the field of information technologies strongly depend on design methodologies and routes at the stages of choosing and/or creating stacks of platforms, technologies and tools. The most important role belongs to the ways of describing the organization of the computing system at various levels and to the used systems of abstractions. The problem of filling the semantic gap between the conceptual (architectural) level and the implementation levels is still very acute. So, it requires the creation of industrial techniques and design tools at these “intermediate” levels. The paper suggests ways of presenting design solutions that are aimed at a holistic, end-to-end description of both the logic of the computing process organization and the steps, technologies, and tools of the design process. The content and the necessity of the stages of microarchitectural design of computing systems is justified and explained in detail. Classification of projects in the field of information technologies according to the degree of variability of the project platform is introduced. Several concepts representing a set of abstractions for microarchitectural design within projects with great variability are suggested. The following abstractions are described in detail: project, design and aspect spaces, project platforms and cross-level mechanisms. Examples of several proposed abstractions presentations (design documentation tools) of microarchitectural design stages are discussed that are most relevant in the design of computing systems in the “limited resources” model: embedded systems, cyber-physical systems, “edge” and “fog” levels of Internet of Things systems.

Keywords

embedded system, microarchitecture, design space exploration, aspect-based design, cross-level mechanisms, cross-cutting mechanisms

For citation: Kolchurin M.V., Pinkevich V.Yu., Platunov A.E. Strengthening the role of microarchitectural stages of embedded systems design. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2022, vol. 22, no. 4, pp. 716–724 (in Russian). doi: 10.17586/2226-1494-2022-22-4-716-724

Введение

Заказное проектирование встраиваемых систем, к которым можно отнести любые специализированные вычислительные системы [1], необходимо в тех случаях, когда прикладную задачу невозможно решить с требуемым качеством на базе готовых компонентов и платформ прикладного программирования. Таким образом, в заказном проектировании ключевую роль играет создание эффективной платформы, с использованием которой можно выполнить все требования к системе.

Проектные платформы разного уровня гранулярности предоставляемого набора функций следует считать одной из важнейших абстракций в создании встраиваемых систем. Под платформой понимается вычислительная сущность (программная, аппаратная), которая имеет зафиксированный разработчиком набор выполняемых операций (в общем случае — набор технических решений), фиксируется как важное («опорное») решение в текущем проекте и/или нацелена на повторное использование [2–4]. Платформа обычно обеспечивает новый уровень абстракции относительно тех средств, с помощью которых она сама построена.

Внешние требования к платформе (набор операций, производительность, ограничения на использование вычислительных ресурсов и т. п.) обычно известны и следуют из особенностей решаемых прикладных задач. При этом требования к внутренней организации платформы определяются самим разработчиком. Выбор и разработка проектных платформ — наиболее трудо-

емкая часть заказного проектирования встраиваемых систем.

По аналогии с делением, принятым в проектировании микропроцессоров, будем относить проектирование, не затрагивающее внутреннее устройство платформ, к архитектурному уровню, а проектирование самой платформы — к микроархитектурному.

Число методик и инструментальных средств, направленных на повышение производительности труда разработчиков в области микроархитектурного проектирования крайне невелико. В настоящее время значительные успехи имеются в развитии языков программирования, средств проектирования заказных микросхем, печатных плат и других важных, но все же частных задач. Методики и инструментальные средства проектирования системного уровня также сконцентрированы в основном на частных задачах (анализ требований, функциональное моделирование, верификация), и не предлагают полноценного автоматизированного маршрута [5–7]¹.

При этом ситуация на рынке систем и услуг в сфере информационных технологий показывает, что заказное проектирование встраиваемых систем массово выполняется небольшими командами разработчиков в условиях ограниченных бюджетов. Методики и инструменты, включающие архитектурный/системный уровень про-

¹ Bonnet S., Lestideau F., Voirin J.-L. Arcadia and Capella on the Field: Real-World MBSE Use Cases // MBSE Symposium, Canberra, October 27th, 2014.

ектирования, применяются в недостаточном объеме из-за их высокой стоимости, трудоемкости использования, недостаточной квалификации разработчиков (высокий «порог вхождения») [1, 8, 9]. В основном проектирование выполняется непосредственно в элементах и шаблонах, навязанных языками, технологиями и инструментами рабочего проектирования. Это приводит к слабому исследованию пространства проектных решений (потенциальные варианты организации вычислительной системы генерируются и анализируются крайне ограниченно), невозможности проведения полноценной верификации архитектурных решений, и, как следствие, к снижению общего качества проектирования и достигаемых характеристик систем. Отметим, что существующие средства системного проектирования малоприспособлены для массового разработчика из-за слишком высокого отношения стоимости их применения к эффективности.

Для повышения эффективности проектирования требуется создание иерархии описаний разной степени детализации и использование итеративного последовательного приближения. В массовом проектировании этот подход недооценивается многими, в том числе отечественными разработчиками, крайне слабо поддержан методиками и инструментальными средствами. Заказное проектирование сложных встраиваемых систем с полным циклом (full-stack) во многом остается творческим процессом, результат которого в наибольшей степени зависит от опыта и квалификации членов команды.

Это ставит общую задачу расширения этапа высокоуровневого [10] проектирования встраиваемых систем. Одним из решений видится включение микроархитектурного этапа проработки проекта в процесс проектирования в качестве обязательной части.

В рамках данного этапа необходимы относительно простые и доступные для массового применения методики и автоматизированные средства проектирования встраиваемых систем, закрывающие семантический разрыв между высокоуровневым архитектурным представлением и реализацией. Заметим, что может быть необходимо говорить о целом ряде этапов в проектировании микроархитектуры.

В работе представлены модели, механизмы и шаблоны для этапов микроархитектурного проектирования, а также примеры их представления и применения.

Проблема выбора платформ для реализации встраиваемых систем

Введем классификацию вычислительных платформ по внутренней вариативности — размеру доступного для исследования пространства проектных решений, т. е. по количеству различных вариантов реализации системы, удовлетворяющих конкретному техническому заданию, которые можно создать с использованием данной платформы: малая (Группа 1), средняя (Группа 2) и высокая внутренняя вариативность (Группа 3).

Приведем примеры платформ для данных групп.

Группа 1 — программируемые логические контроллеры, а также универсальные вычислительные машины

вообще, которые поддерживают стандартные высокоуровневые языки программирования и стандартные интерфейсы подключения устройств.

Группа 2 — контроллеры с программированием в среде операционных систем реального времени. Их внутренняя вариативность остается для пользователя относительно небольшой за счет широкого, но все же ограниченного набора предусмотренных механизмов управления вычислениями и драйверов устройств.

Группа 3 — заказные устройства с заказным программным обеспечением, возможностью использования специализированных архитектур на программируемой логике и заказных микросхемах.

Данная классификация демонстрирует один из важнейших компромиссов при выборе платформы реализации встраиваемых систем — соотношение между уровнем операций, реализуемых платформой (вычислительные абстракции, система команд, программный интерфейс приложения (API) или элементная база проектировщика), и эффективностью вычислителя/платформы, реализующей этот уровень (рис. 1) [11]. Отметим, что в отличие от примеров проектов Группы 1, примеры проектов Групп 2 и 3 предполагают смешанные аппаратно-программные разработки.

Повышение уровня абстракции позволяет сконцентрироваться на реализации прикладных требований, увеличить переносимость, а также снизить требования к квалификации разработчика. Часто это проявляется в реализации наиболее простых, шаблонных, а также «лобовых» вариантов в модели «условно неограниченных ресурсов», что оказывается приемлемо для многих применений. Если же такой сравнительно простой путь решения задачи не дает требуемого результата, необходимо переходить на сценарии полномасштабного заказного проектирования. При увеличении доли заказного проектирования возможно формирование более эффективных прикладных платформ и проведение «глубокой» оптимизации их микроархитектуры, реализация более сложных и эффективных решений. При этом необходим учет ограниченности доступных вычислительных ресурсов для их полноценного использования.

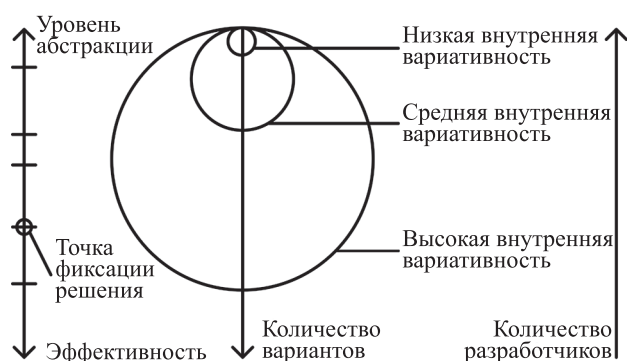


Рис. 1. Классификация проектных платформ по степени внутренней вариативности

Fig. 1. Classification of project platforms according to the degree of internal variability

Чем выше внутренняя вариативность платформы, тем сложнее ее использование, меньше экосистема и число разработчиков, обладающих необходимыми компетенциями, ниже уровень автоматизации проектирования и документирования объектов повторного использования.

Микроархитектурное проектирование

Основные задачи микроархитектурного проектирования: формирование, визуализация и документирование пространства проектных решений; генерация проектных решений и исследование пространства проектных решений; верификация и отбор проектных решений.

В основе предлагаемого подхода к микроархитектурному проектированию лежит аспектный подход. Он помогает определить структуру пространства проектных решений за счет выделения сквозных «зон ответственности» в системе и сквозных задач в процессе проектирования. В данном пространстве существуют вычислительные и иные системные механизмы — основные единицы, из которых создаются более сложные платформы и целые вычислительные системы.

Природа большинства вычислительных систем имеет ярко выраженную уровневую организацию. Механизмы и платформы как основные абстрактные единицы проектирования могут принадлежать к конкретному уровню или охватывать несколько уровней организации вычислительной системы. Для документирования уровневой организации вычислительной системы и/или ее пространства проектных решений мы предлагаем использовать граф актуализации и производные модели.

Выделение корневых компонентов («кernels») в проекте системы позволяет в явном виде определять и балансировать трудозатраты на проектирование компонент «внутреннего» и «внешнего» назначений, группировать информацию о сквозных механизмах и подсистемах, представлять описания разного уровня детализации для последующего повторного использования.

Структура проектного и аспектного пространства системы

Назовем проектным пространством системы совокупность активностей и артефактов на фазе проектирования в жизненном цикле системы.

В процессе проектирования исходные требования технического задания дополняются внутренними требованиями коллектива разработчиков. Эти требования могут быть производными от исходных, а могут быть навязаны какими-либо сторонними обстоятельствами (необходимость экономии времени, отсутствие некоторых компетенций, создание задела на будущее и др.). В результате формируется полный набор требований к создаваемой встраиваемой системе и сопутствующим артефактам.

Выделим две основные группы требований:

- 1) «поведенческие» требования — непосредственно касаются функционирования системы во всех про-

явлениях (реализуемый вычислительный процесс, временные характеристики, надежность, энергопотребление, тепловыделение и др.);

- 2) «организационные» требования — касаются вопросов, связанных с разными стадиями жизненного цикла как конкретных экземпляров системы, так и совокупности активностей по созданию таких систем. Включают требования к организации процессов проектирования, производства, тестирования, обслуживания, ремонта, вывода системы из эксплуатации, возможностей устаревания, модернизации, повторного использования, развития системы, экономической эффективности и др.

Поведенческие требования можно разделить на прикладные (связанные с выполнением прикладной целевой функции системы, выраженной в исходных требованиях заказчика) и системные (относящиеся к решению задач, поставленных разработчиками: организация платформы для решения непосредственно прикладных задач, инструментально-диагностическая инфраструктура системы и др.).

Всем перечисленным выше видам требований соответствуют аспекты проектирования системы. Аспект определяется как сегмент проектного пространства и/или жизненного цикла системы и проекта, связанный с какой-либо проблемой или группой требований. Аспекты могут быть концептуальными (актуальными на протяжении всего жизненного цикла) и локальными, выделяемыми на какое-то время. Их совокупность определяет аспектное пространство проекта. Организационные аспекты порождают новые требования в поведенческих аспектах и ограничивают пространство проектных решений, в котором проектировщик создает систему.

Маршрут проектирования сложных вычислительных систем должен поддерживать в единой системе абстракций иерархическое представление вычислительных платформ, языков описания функциональности и их трансляторов, прикладной надстройки. Центральной идеей выступает последовательное уточнение/проработка целевой системы через иерархию шагов маршрута проектирования с понижением степени абстракции [12].

Аспекты делят проектное пространство «вертикально», пронизывая все компоненты и уровни организации системы, которые, в свою очередь, задают «горизонтальное» деление. Аспекты жизненного цикла пронизывают различные этапы во времени.

Для обеспечения эффективного сочетания уровневого представления с аспектным подходом мы предлагаем проектировщику работать со следующими моделями: аспектные стеки логических, функциональных, структурных уровней организации анализируемой или проектируемой системы (аспекты организации системы); аспектные стеки технологических уровней создания системы (инструментально-технологические аспекты); аспектные временные треки в рамках маршрута проектирования, где шаги могут быть сопоставлены с проработкой уровней организации системы (отражается логика проектирования). При этом последовательность проработки уровней может осознанно меняться проектировщиком.

Сложность современных встраиваемых систем, требования к их масштабируемости (горизонтальной и вертикальной) делают актуальными модели с несколькими аспектными уровневыми стеками, «растущими» с выделенного уровня (горизонтальная модель организации в пределах одного или нескольких уровней системы).

Кросс-уровневые механизмы в проектировании встраиваемых систем

Реализацию встраиваемой системы представим как совокупность вычислительных и иных механизмов — особых объектов в составе системы, решающих конкретную задачу известным способом [10, 12]. Механизмы могут быть разной сложности и гранулярности (от простых математических функций до виртуальных машин), применяться для обеспечения как функциональных, так и нефункциональных требований к системе (надежности, работы в реальном времени и др.). Такое представление наиболее эффективно на этапах микроархитектурного проектирования.

Задачи, решаемые большинством встраиваемых систем, характеризуются высокими требованиями к безопасности, надежности и времени исполнения. Это заставляет создавать значительную часть принципиально важных механизмов встраиваемых систем: кросс-уровневыми — реализованными с использованием более чем одной проектной платформы; кросс-доменными — сочетающими в себе элементы разных моделей вычислений в пределах одного уровня; кросс-аспектными — отражающимися на разных группах требований к системе (на производительности, энергопотреблении и др.).

Комплексный характер механизмов влияет на способ и качество подачи информации в документах на компоненты, особенно повторно используемые. В сегменте проектов с малой вариативностью (Группа 1) для вычислителей с условно безграничными ресурсами поставщики предлагают описания высокого качества для «прикладного» разработчика. В данном случае предоставляются абстрактные модели, которые напрямую поддерживаются наборами библиотек (например, Windows Driver Kit включает Kernel-Mode Driver Architecture Design Guide, Windows Driver Model (WDM) и др.). В проектах Группы 2 и особенно Группы 3 разработчики встраиваемых систем и аппаратных вычислительных платформ вынуждены использовать фрагментарные и преимущественно низкоуровневые описания отдельных вычислителей (групп вычислителей), периферийных контроллеров, IP-ядер или программных компонентов с целью реализации в равной степени функциональных и нефункциональных требований, опираясь на модель максимального использования ресурсов проектируемой системы.

Наибольшие сложности при проектировании вызывает именно кросс-уровневая природа механизмов. Описание такого механизма в рамках единой платформы и единого языка отсутствует, что порождает целый ряд проблем: необходимо использовать несколько фрагментов описания механизма в рамках разных моделей вычислений; при отсутствии документации на

реализацию системы свойства механизма и даже самого его существование для разработчика часто являются неочевидными (требуется обратная разработка системы, reverse engineering); затруднена проверка корректности реализации механизма; при замене даже одной из платформ на новую, несовместимую со старой (например, переход на другую линейку микроконтроллеров), необходимо пересматривать реализацию множества стеков механизмов или кросс-уровневых механизмов для сохранения корректности и эффективности функционирования, что может привести к необходимости пересмотра микроархитектуры системы в целом.

Примеры микроархитектурного представления проектных решений

На рис. 2 показан пример комплексной **аспектной диаграммы** типичного проекта встраиваемой системы с аспектными стеками уровней и временными треками. Диаграмма приведена в общем виде и по необходимости может быть детализирована, дополнена локальными аспектами.

Корневые механизмы (кernels) представляют собой базовые, критически важные кросс-уровневые и кросс-доменные механизмы или их совокупности, предназначенные для активного повторного использования. Мы предлагаем фиксировать информацию о кернах на трех основных уровнях документирования: функциональный интерфейс (функциональность и способы сопряжения с компонентами того же уровня), микроархитектурная спецификация в терминах механизмов и конечная реализация. Это обеспечивает эффективное повторное использование ядер на микроархитектурных этапах проектирования. Совокупности ядер, образующие цельную платформу, назовем микроархитектурными шаблонами. На рис. 3 показан пример представления ядер в составе микроархитектурного шаблона с указанием охваченных уровней (платформ) встраиваемых систем. Вопросы вариантов реализации ядер выходят за рамки настоящей работы, некоторые примеры можно найти в работе [12].

Граф актуализации вычислительного процесса [3, 4] в едином стиле в виде последовательности унифицированных трансляторов описывает создаваемую встраиваемую систему, используемый при этом инструментарий и реализуемый процесс проектирования. Это позволяет в рамках одной модели продемонстрировать взаимосвязи между аппаратной и программной частями, а также фазами проектирования, конфигурирования и исполнения вычислительного процесса (design-, config- и run-time).

Приведем пример использования графа актуализации для демонстрации зависимостей, которые возникают при реализации кросс-уровневых механизмов. На рис. 4, а показан ряд вариантов кросс-уровневой реализации одного из ядер — драйвера UART. Если при проектировании драйвера закладывать возможность использования аппаратного буфера, то необходимо предусматривать программную адаптацию к буферам разного размера. Если же игнорировать возможности аппаратной буферизации, драйвер будет

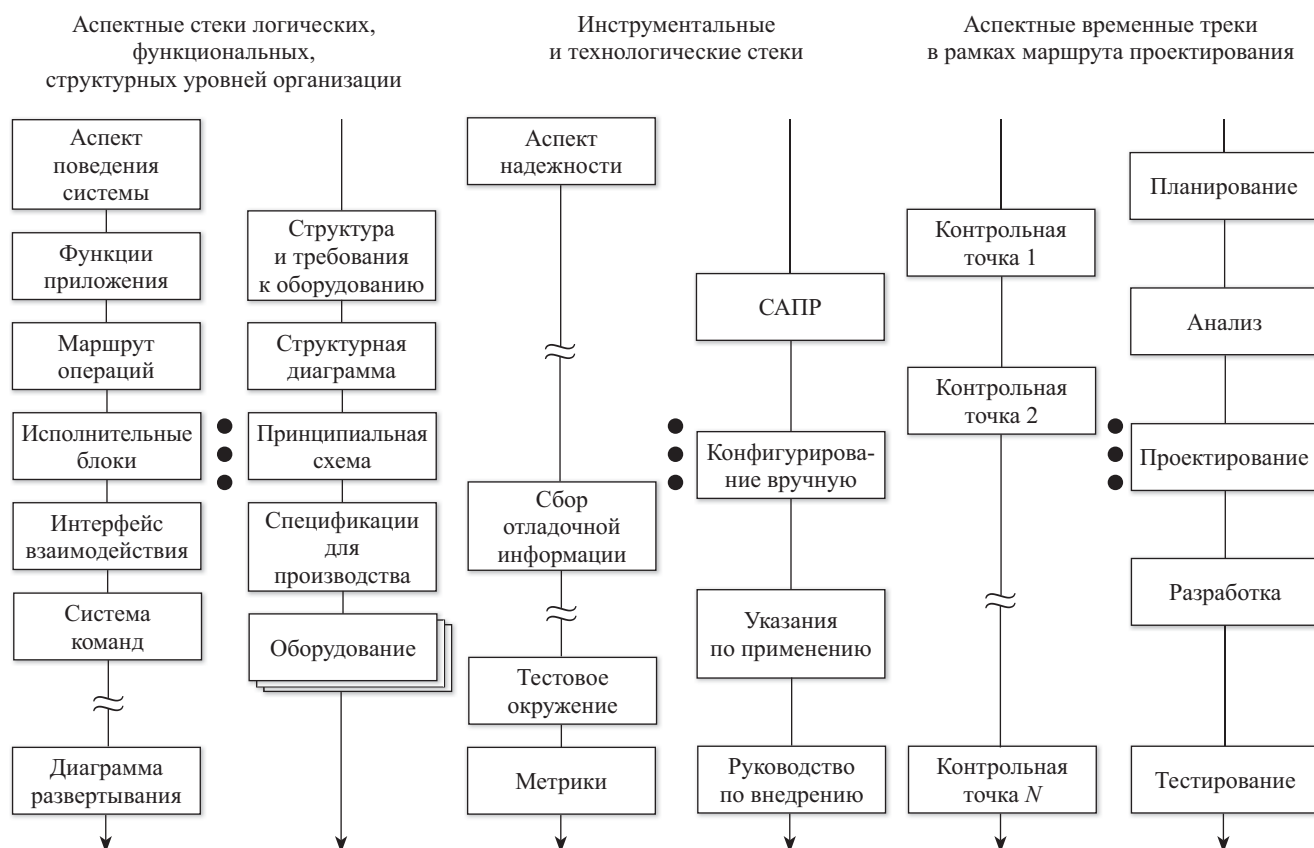


Рис. 2. Пример аспектной диаграммы проекта встраиваемой системы

Fig. 2. Example of an aspect diagram of an embedded system project

менее эффективно использовать процессорное время, но его проще переносить на микроконтроллер с другой периферией. В пределе аппаратная часть UART может быть сведена к набору дискретных портов ввода-вывода. На рис. 4, *b* представлена реализация виртуального датчика освещенности на базе цифровой камеры. Такая реализация не всегда переносима на другую модель камеры даже при использовании общего API, поскольку

для получения информации об освещенности матрицы используются команды прямого доступа к регистрам модуля камеры.

Коллектив, который представляют авторы, в формализованном виде успешно использовал представленный подход в целом ряде проектов, среди которых автоматизированная система кросс-уровневого тестирования [13], фреймворк микроархитектурного проектирования

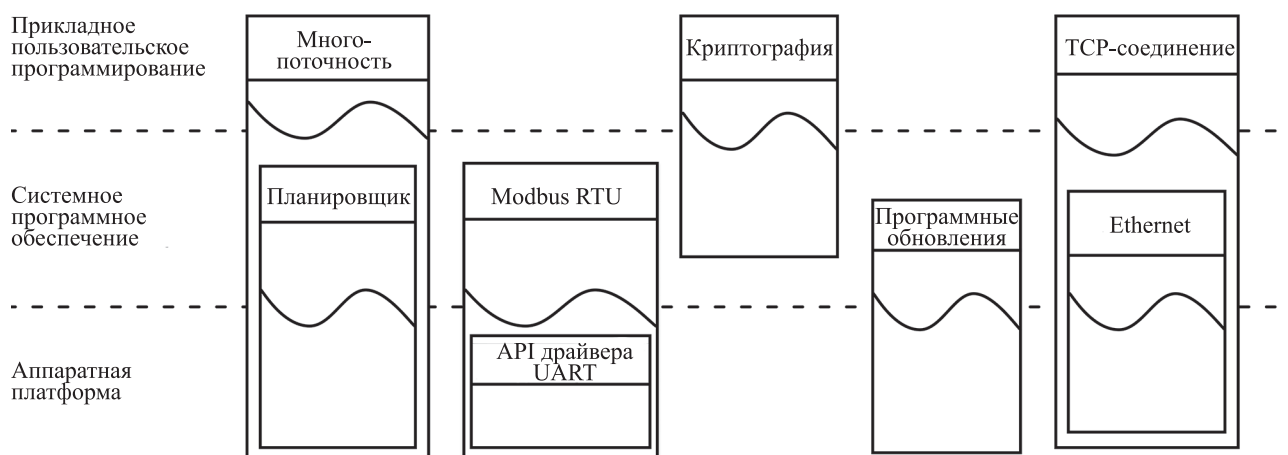


Рис. 3. Пример распределения ядер по уровням микроархитектурного шаблона (UART — контроллер последовательного ввода-вывода)

Fig. 3. Example of kernels mapping by levels of microarchitectural pattern (UART — serial input-output controller)

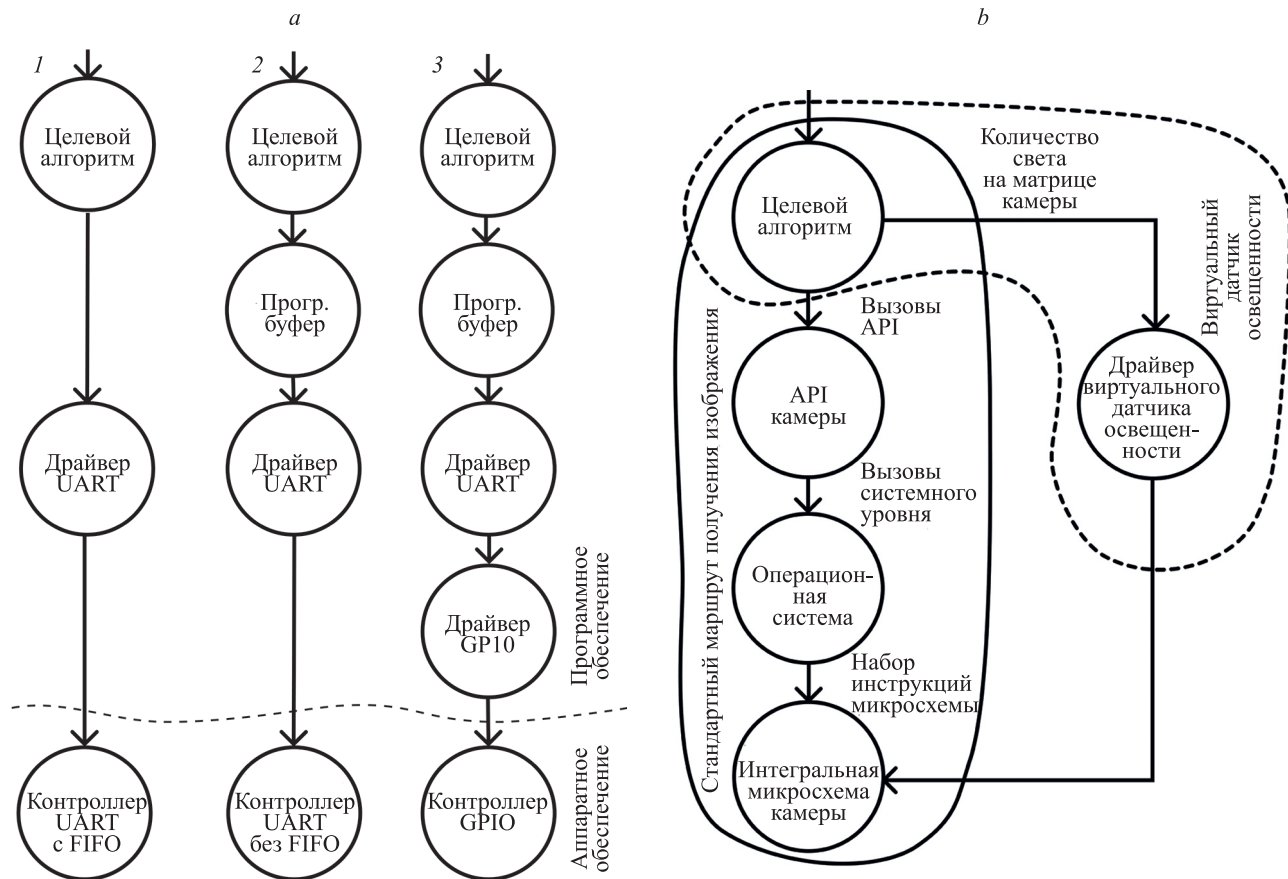


Рис. 4. Варианты графа актуализации для драйвера UART; варианты программно-аппаратного разделения функциональности: с использованием аппаратной буферизации данных (1) и только программной буферизации (2), с полностью программной реализацией приемопередатчика (3) (a); виртуального датчика освещенности (b)

Fig. 4. Variants of the actualization graph for UART driver; variants of hardware-software partitioning of functionality: using hardware data buffering (1) and only software buffering (2), with a fully software implementation of the transceiver (3) (a) and virtual illumination sensor (b)

процессорных ядер СнК [14], инструменты проектов NL3 и NITTA [15].

Заклучение

В области информационных технологий по-прежнему остро стоит проблема повышения эффективности труда разработчиков. Разумный баланс между готовыми («с полки») и заказными решениями — «головная боль» всех архитекторов и тим-лидеров. Как показывает практика, в данной отрасли внедрение новых приемов и методов на концептуальных, системных этапах проектирования для выработки и принятия не организационных, а непосредственно технических решений, занимает 10–20 и более лет. И это определяется не инертностью индустрии, а сложным процессом подготовки («созревания») специалистов. Тем важнее новые предложения в данной сфере, больше потребность в их обсуждении и экспериментальном внедрении.

В сегменте встраиваемых систем предложенное деление проектов на ряд категорий позволило выделить группу, для которой заказные решения однозначно являются предпочтительными. Однако для таких

проектов катастрофически не хватает эффективных технологий и инструментов проектирования.

Повысить степень формализации проектирования позволяет приведенный в работе общий взгляд на этапы высокоуровневого проектирования ответственных (критичных) вычислительных систем в сочетании с конкретными предложениями в части инструментов описания вычислительной системы на разных уровнях абстракции.

Представленные абстракции могут быть использованы как на уровне методологических решений в качестве «ручных приемов», так и в качестве алгоритмической основы в составе фреймворков и систем автоматизированного проектирования.

Исследования по теме работы находятся в активной фазе. Одна из актуальных задач — применение техники кросс-уровневых механизмов и микроархитектурных шаблонов для анализа сложного встроенного программного обеспечения (различные операционные системы реального времени, стеки коммуникационных протоколов, дистанционное конфигурирование и обновление, и др.).

Литература

1. Платунов А.Е., Пинкевич В.Ю. Создание киберфизических систем: проблемы подготовки ИТ-специалистов // *Control Engineering Russia*. 2021. № 3. С. 64–70.
2. Sangiovanni-Vincentelli A., Martin G. Platform-based design and software design methodology for embedded systems // *IEEE Design and Test of Computers*. 2001. V. 18. N 6. P. 23–33. <https://doi.org/10.1109/54.970421>
3. Platonov A., Penskoï A., Kluchev A. The architectural specification of embedded systems // *Proc. of the 3rd Mediterranean Conference on Embedded Computing (MECO)*. 2014. P. 48–51. <https://doi.org/10.1109/MECO.2014.6862656>
4. Пинкевич В.Ю., Платунов А.Е. Тестирование и отладка встраиваемых вычислительных систем на основе уровненых моделей // *Научно-технический вестник информационных технологий, механики и оптики*. 2018. Т. 18. № 5. С. 801–808. <https://doi.org/10.17586/2226-1494-2018-18-5-801-808>
5. Teich J. Hardware/software codesign: the past, the present, and predicting the future // *Proceedings of the IEEE*. 2012. V. 100. P. 1411–1430. <https://doi.org/10.1109/JPROC.2011.2182009>
6. Broman D., Lee A., Tripakis S., Törnigren M. Viewpoints, formalisms, languages, and tools for cyber-physical systems // *Proc. of the 6th International Workshop on Multi-Paradigm Modeling (MPM'12)*. 2012. P. 49–54. <https://doi.org/10.1145/2508443.2508452>
7. Clements P., Bachmann F., Bass L., Garland D., Ivers J., Little R., Merson P., Nord R., Stafford J. *Documenting Software Architectures. Views and Beyond* / 2nd ed. Pearson Education, 2011. 608 p.
8. Masin M., Palumbo F., Myrhaug H., de Oliveira Filho J.A., Pastena M., Pelcat M., Raffo L., Regazzoni F., Sanchez A.A., Toffetti A., de la Torre E., Zedda K. Cross-layer design of reconfigurable cyber-physical systems // *Proc. of the 20th Design, Automation and Test in Europe Conference and Exhibition (DATE)*. 2017. P. 740–745. <https://doi.org/10.23919/DATE.2017.7927088>
9. Pohlmann U. *A Model-driven Software Construction Approach for Cyber-physical Systems*. Universität Paderborn, 2018.
10. Platonov A., Kluchev A., Penskoï A. Expanding design space for complex embedded systems with HLD-methodology // *Proc. of the 6th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*. 2014. P. 157–164. <https://doi.org/10.1109/ICUMT.2014.7002096>
11. Bailey B., Martin G. *ESL Models and their Application: Electronic System Level Design and Verification in Practice*. New York: Springer Publication, 2010. XXIV, 446 p. <https://doi.org/10.1007/978-1-4419-0965-7>
12. Pinkevich V., Platonov A., Gorbachev Y. Design of embedded and cyber-physical systems using a cross-level microarchitectural pattern of the computational process organization // *CEUR Workshop Proceedings*. 2020. V. 2893.
13. Pinkevich V., Platonov A. Model-driven functional testing of cyber-physical systems using deterministic replay techniques // *Proc. of the 1st IEEE International Conference on Industrial Cyber-Physical Systems (ICPS)*. 2018. P. 141–146. <https://doi.org/10.1109/ICPHYS.2018.8387650>
14. Antonov A., Kustarev P., Bikovsky S. MLIP cores: Designing hardware generators with programmable microarchitectural mechanisms // *Proc. of the 52nd IEEE International Symposium on Circuits and Systems (ISCAS)*. 2020. P. 9180593. <https://doi.org/10.1109/ISCAS45731.2020.9180593>
15. Pinkevich V.Y., Platonov A.E., Penskoï A.V. The approach to design of problem-oriented reconfigurable hardware computational units // *Proc. of the Wave Electronics and its Application in Information and Telecommunication Systems (WECONF)*. 2020. P. 9131512. <https://doi.org/10.1109/WECONF48837.2020.9131512>

Авторы

Кольчурун Максим Вячеславович — аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, <https://orcid.org/0000-0002-7061-9357>, maxim.kolchurin@gmail.com

Пинкевич Василий Юрьевич — кандидат технических наук, инженер, ООО «ЛМТ», Санкт-Петербург, 199034, Российская Федерация, <https://orcid.org/0000-0002-8635-5026>, vasiliy.pinkevich@yandex.ru


References


1. Platonov A.E., Pinkevich V.Yu. Creation of cyber-physical systems: problems of it specialists training. *Control Engineering Russia*, 2021, no. 3, pp. 64–70. (in Russian)
2. Sangiovanni-Vincentelli A., Martin G. Platform-based design and software design methodology for embedded systems. *IEEE Design and Test of Computers*, 2001, vol. 18, no. 6, pp. 23–33. <https://doi.org/10.1109/54.970421>
3. Platonov A., Penskoï A., Kluchev A. The architectural specification of embedded systems. *Proc. of the 3rd Mediterranean Conference on Embedded Computing (MECO)*, 2014, pp. 48–51. <https://doi.org/10.1109/MECO.2014.6862656>
4. Pinkevich V.Yu., Platonov A.E. Testing and debugging of embedded computing systems based on level models. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2018, vol. 18, no. 5, pp. 801–808. (in Russian). <https://doi.org/10.17586/2226-1494-2018-18-5-801-808>
5. Teich J. Hardware/software codesign: the past, the present, and predicting the future. *Proceedings of the IEEE*, 2012, vol. 100, pp. 1411–1430. <https://doi.org/10.1109/JPROC.2011.2182009>
6. Broman D., Lee A., Tripakis S., Törnigren M. Viewpoints, formalisms, languages, and tools for cyber-physical systems. *Proc. of the 6th International Workshop on Multi-Paradigm Modeling (MPM'12)*, 2012, pp. 49–54. <https://doi.org/10.1145/2508443.2508452>
7. Clements P., Bachmann F., Bass L., Garland D., Ivers J., Little R., Merson P., Nord R., Stafford J. *Documenting Software Architectures. Views and Beyond*. 2nd ed. Pearson Education, 2011, 608 p.
8. Masin M., Palumbo F., Myrhaug H., de Oliveira Filho J.A., Pastena M., Pelcat M., Raffo L., Regazzoni F., Sanchez A.A., Toffetti A., de la Torre E., Zedda K. Cross-layer design of reconfigurable cyber-physical systems. *Proc. of the 20th Design, Automation and Test in Europe Conference and Exhibition (DATE)*, 2017, pp. 740–745. <https://doi.org/10.23919/DATE.2017.7927088>
9. Pohlmann U. *A Model-driven Software Construction Approach for Cyber-physical Systems*. Universität Paderborn, 2018.
10. Platonov A., Kluchev A., Penskoï A. Expanding design space for complex embedded systems with HLD-methodology. *Proc. of the 6th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, 2014, pp. 157–164. <https://doi.org/10.1109/ICUMT.2014.7002096>
11. Bailey B., Martin G. *ESL Models and their Application: Electronic System Level Design and Verification in Practice*. New York, Springer Publication, 2010, XXIV, 446 p. <https://doi.org/10.1007/978-1-4419-0965-7>
12. Pinkevich V., Platonov A., Gorbachev Y. Design of embedded and cyber-physical systems using a cross-level microarchitectural pattern of the computational process organization. *CEUR Workshop Proceedings*, 2020, vol. 2893.
13. Pinkevich V., Platonov A. Model-driven functional testing of cyber-physical systems using deterministic replay techniques. *Proc. of the 1st IEEE International Conference on Industrial Cyber-Physical Systems (ICPS)*, 2018, pp. 141–146. <https://doi.org/10.1109/ICPHYS.2018.8387650>
14. Antonov A., Kustarev P., Bikovsky S. MLIP cores: Designing hardware generators with programmable microarchitectural mechanisms. *Proc. of the 52nd IEEE International Symposium on Circuits and Systems (ISCAS)*, 2020, pp. 9180593. <https://doi.org/10.1109/ISCAS45731.2020.9180593>
15. Pinkevich V.Y., Platonov A.E., Penskoï A.V. The approach to design of problem-oriented reconfigurable hardware computational units. *Proc. of the Wave Electronics and its Application in Information and Telecommunication Systems (WECONF)*, 2020, pp. 9131512. <https://doi.org/10.1109/WECONF48837.2020.9131512>

Authors

Maxim V. Kolchurin — PhD Student, ITMO University, Saint Petersburg, 197101, Russian Federation, <https://orcid.org/0000-0002-7061-9357>, maxim.kolchurin@gmail.com

Vasiliy Yu. Pinkevich — PhD, Engineer, LMT Ltd., Saint Petersburg, 199034, Russian federation, <https://orcid.org/0000-0002-8635-5026>, vasiliy.pinkevich@yandex.ru

Платунов Алексей Евгеньевич — доктор технических наук, профессор, профессор, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация,  35318291200, <https://orcid.org/0000-0003-3003-3949>, aeplatunov@gmail.com

Alexey E. Platunov — D. Sc., Full Professor, ITMO University, Saint Petersburg, 197101, Russian Federation,  35318291200, <https://orcid.org/0000-0003-3003-3949>, aeplatunov@gmail.com

Статья поступила в редакцию 15.05.2022
Одобрена после рецензирования 30.05.2022
Принята к печати 12.07.2022

Received 15.05.2022
Approved after reviewing 30.05.2022
Accepted 12.07.2022



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

doi: 10.17586/2226-1494-2022-22-4-725-733

A multivariate binary decision tree classifier based on shallow neural network

Avazjon R. Marakhimov¹, Jabbarbergen K. Kudaybergenov²,
Kabul K. Khudaybergenov³, Ulugbek R. Ohundadaev⁴✉

¹ Termez State University, Termez, 190111, Uzbekistan

² Tashkent University of Information Technologies Nukus branch named after Muhammad Al-Khwarizmi, Nukus, 230113, Uzbekistan

^{3,4} National University of Uzbekistan, Tashkent, 100174, Uzbekistan

¹ termizdu@umail.uz, <https://orcid.org/0000-0003-3735-6855>

² kjabbarbergen@gmail.com, <https://orcid.org/0000-0003-4494-6255>

³ kabul85@mail.ru, <https://orcid.org/0000-0001-8143-625X>

⁴ ulugbek_1122@mail.ru✉, <https://orcid.org/0000-0002-3240-6502>

Abstract

In this paper, a novel decision tree classifier based on shallow neural networks with piecewise and nonlinear transformation activation functions are presented. A shallow neural network is recursively employed into linear and non-linear multivariate binary decision tree methods which generates splitting nodes and classifier nodes. Firstly, a linear multivariate binary decision tree with a shallow neural network is proposed which employs a rectified linear unit function. Secondly, there is presented a new activation function with non-linear property which has good generalization ability in learning process of neural networks. The presented method shows high generalization ability for linear and non-linear multivariate binary decision tree models which are called a Neural Network Decision Tree (NNDT). The proposed models with high generalization ability ensure the classification accuracy and performance. A novel split criterion of generating the nodes which focuses more on majority objects of classes on the current node is presented and employed in the new NNDT models. Furthermore, a shallow neural network based NNDT models are converted into a hyperplane based linear and non-linear multivariate decision trees which has high speed in the processing classification decisions. Numerical experiments on publicly available datasets have showed that the presented NNDT methods outperform the existing decision tree algorithms and other classifier methods.

Keywords

hierarchical classifier, neural networks, binary tree, multivariate decision tree, activation function

Acknowledgments

This research is supported by the Department of Algorithms and Programming technologies, National University of Uzbekistan, Uzbekistan.

For citation: Marakhimov A.R., Kudaybergenov J.K., Khudaybergenov K.K., Ohundadaev U.R. A multivariate binary decision tree classifier based on shallow neural network. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2022, vol. 22, no. 4, pp. 725–733. doi: 10.17586/2226-1494-2022-22-4-725-733

УДК 517.95

Многомерный двоичный классификатор дерева решений на основе неглубокой нейронной сети

Авазжон Рахимович Марахимов¹, Жаббарберген Кадирбергенович Кудайбергенов²,
Кабул Кадирбергенович Худайбергенов³, Улугбек Рахимжон угли Охундаев⁴✉

¹ Термезский государственный университет, Термез, 190111, Узбекистан

² Нукусский филиал Ташкентского университета информационных технологий имени Мухаммад ал-Хоразми, Нукус, 230113, Узбекистан

^{3,4} Национальный университет Узбекистана, Ташкент, 100174, Узбекистан

© Marakhimov A.R., Kudaybergenov J.K., Khudaybergenov K.K., Ohundadaev U.R., 2022

¹ termizdu@umail.uz, <https://orcid.org/0000-0003-3735-6855>

² kjabbarbergen@gmail.com, <https://orcid.org/0000-0003-4494-6255>

³ kabul85@mail.ru, <https://orcid.org/0000-0001-8143-625X>

⁴ ulugbek_1122@mail.ru✉, <https://orcid.org/0000-0002-3240-6502>

Аннотация

Предложен новый классификатор дерева решений, основанный на неглубоких нейронных сетях с кусочными и нелинейными функциями активации преобразования. Данная сеть рекурсивно используется в методах линейного и нелинейного многомерного бинарного дерева решений, которые генерируют узлы разделения и классификатора. Представлено линейное многомерное бинарное дерево решений с неглубокой нейронной сетью, в которой использована выпрямленная линейная единичная функция. Описана новая функция активации с нелинейным свойством, с помощью которой в процессе обучения нейронных сетей получается оптимальная обобщающая способность. Рассмотренный метод продемонстрировал высокую способность к обобщению для моделей линейного и нелинейного многомерного бинарного дерева решений. Предложенные модели обеспечивают точность и производительность классификации. Представлен новый критерий разделения для генерации узлов, который может быть использован в новых моделях дерева решений нейронной сети для большинства классов объектов в текущем узле. Также эти модели могут быть преобразованы в линейные и нелинейные многомерные деревья решений на основе гиперплоскости, и имеют высокую скорость при обработке решений классификации. Численные эксперименты на общедоступных наборах данных показали, что представленные методы превосходят существующие алгоритмы дерева решений и другие методы классификации.

Ключевые слова

иерархический классификатор, нейронные сети, бинарное дерево, многомерное дерево решений, функция активации

Благодарности

Исследование выполнено при поддержке кафедры алгоритмов и технологий программирования Национального университета Узбекистана, Узбекистан.

Ссылка для цитирования: Марахимов А.Р., Кудайбергенов Ж.К., Худайбергенов К.К., Охундаев У.Р. Многомерный двоичный классификатор дерева решений на основе неглубокой нейронной сети // Научно-технический вестник информационных технологий, механики и оптики. 2022. Т. 22, № 4. С. 725–733 (на англ. яз.). doi: 10.17586/2226-1494-2022-22-4-725-733

Introduction

The latest research works on artificial intelligence and machine learning have embraced prospering developments in many fields, such as regression [1, 2], classification [3], computer vision [4–6], natural language processing [7], image processing [8], speech information [9], and etc. Among the machine learning algorithms and methods which are used in solving real life problems, deep neural networks have showed unprecedented successes and precision due to the great flexibility and powerful generalization ability of learning and decision making [10]. However, deep neural networks also have shortages and some limitations in learning process [11–13].

In machine learning, decision tree methods have been applied in various practical fields, such as medical diagnosis [14], signal processing [15], classification problems [16], and they showed outperformance compared to other algorithms and methods. In general, decision tree models are considered as hierarchical classifiers according to its nature. Also, these methods are able to work with multi-class datasets as well as binary class classification tasks. A recent survey on decision tree models [17], which is widely studied the models and the related issues, show that there exists a lot of various types. In general, according to the different type of node generating or splitting, decision tree models can be categorized as follows:

- 1) univariate decision trees;
- 2) multivariate decision trees;
- 3) omnivariate decision trees.

Univariate decision tree algorithms work with a single attribute at each node and construct nodes and leaf nodes.

Some of the algorithms, which work as a univariate decision tree model, are Iterative Dichotomiser 3 (ID3) [18], Classification And Regression Tree (CART) [19], and C4.5 (predecessor of ID3) [20] (which is splitting criterion), are based on information gain coefficient, Gini index, and gain ratio. These splitting criteria are considered as frequency heuristic for splitting decision trees. Later, a new node generating method is proposed in [21], which is considered to be better than information gain and Gini index criteria. However, due to simplicity and a good interpretability of these criteria, in univariate decision trees node splitting may not be appropriate when learning samples are numerically correlated. Moreover, univariate decision tree algorithms are considered to be greedy search, which negatively impacts time performance.

Multivariate decision tree algorithms employ multiple attributes or multiple features to generate a node and splitting that are associated with multiple attributes. Thus, multi attribute criterion generates linear or nonlinear split nodes in the trees. As an example, in [19] a linear multivariate decision tree algorithm, CART, was introduced. Also, a novel multivariate decision tree algorithm was introduced in [22]. This algorithm employs neural network to generate the nodes. Nonlinear multivariate decision tree models are considered sensitive to overfitting problems due to its highly complex node generating process.

Omnivariate decision tree algorithms are decision tree methods where a splitting process at each node can be in univariate, linear multivariate, and even with a combination of nonlinear multivariate techniques. This tree model is considered in [23] where a node splitting is automatically selected among univariate, linear, or nonlinear processes

depending on some statistical experiments. Later, in [24], authors proposed to use both linear and non-linear methods as an ensemble algorithm for node generating. Additionally, a hybrid Support Vector Machines [25, 26] based on decision tree model are proposed for classification problems which have both univariate and multivariate nodes.

Although there are different variants of decision tree models, linear multivariate type of decision tree models is much more preferable. Firstly, this type of model is not restricted to be orthogonal to attribute axis on each splitting hyperplane in nodes like univariate decision tree models. Secondly, linear multivariate type of decision tree models is less complex compared to other decision models and not prone to overfitting problems. Therefore, we focused on linear multivariate type of decision tree models.

In this work, our contribution includes the following. Firstly, a new linear multivariate decision tree model, Neural Network Decision Tree (NNDT) and its algorithm are presented which enhances the multivariate decision tree models. Secondly, we propose a new type of splitting node criterion with piecewise and nonlinear transformation functions. NNDT employs piecewise activation functions and joined activation function with nonlinear transformations. This helps to improve the generalization ability of the algorithm in classification tasks for multi-class datasets. Our model is based on the splitting nodes with hyperplanes in decision tree method which speeds up the classification tasks.

Related work review

A novel ensemble algorithm [22] based on multivariate decision tree and neural network is proposed which employs both binary tree and unsupervised node splitting algorithm. In this research work they have proposed two full binary tree classifiers constructed on multivariate decision trees. The first one was called randomly partitioned multivariate decision tree (MDT-1) and second one was an ensemble method, so called multivariate decision tree (MDT-2), which is based on a Principal Component Analysis (PCA) method.

In general, the ensemble algorithms, MDT-1 and MDT-2 (similar to each other), are considered. These algorithms construct a binary tree in a top-down fashion, constantly adding up child nodes and leaf nodes. Using recursive partition with hyperplanes, the process will continue until there is no node left to split. The algorithm stops generating nodes when all the nodes of a tree to be leaf nodes. The root node of the multivariate decision tree is the first one which separates the whole training data into two classes; nevertheless the data set contains labels more than 2. We can see all the processing steps of the algorithms as follows:

Step-1: Train a multivariate hyperplane for the root node which tries to split all the objects of an entire dataset into the left and right child nodes. After splitting, the first half hyperplane will contain mostly the objects which belong to the same class, and second half will contain all the other objects.

Step-2: Measure a splitting criterion. If the split criteria are satisfied for a next child node, process goes to generate

the next node; otherwise, the node is processed as a leaf node with the majority class objects as the representatives of one class which contains mostly of the same class objects.

Step-3: Again, analyze the each obtained node that needs to be split and repeat from Step-1 unless there is any node left to split again.

Clearly, the two main problems in this ensemble algorithm are how to build and learn the multivariate hyperplane tree at each splitting node and measure the splitting condition. Further below, we will look at these two problems explaining how it works. Without losing generality, we consider a single splitting in the decision tree. The splitting procedure is conducted on a node with a dataset matrix $\mathbf{X} \in R^{d \times n}$, and the objects are given as class labels $\mathbf{Y} \in R^{1 \times n}$. Here, n is the number of the objects and d is the feature space dimension.

Split with simple hyperplane

In order to construct multivariate decision tree, both ensemble methods MDT-1 and MDT-2 use a simple linear hyperplane in the following form

$$wx - p = 0, \quad (1)$$

where w is the normal vector of the linear hyperplane and p is a bias parameter. Node splitting of the multivariate decision tree is constructed by separating the sample objects into the two child nodes in the following form:

$$\begin{cases} x \in \mathbf{X}_1, & \text{if } wx > p, \\ x \in \mathbf{X}_2, & \text{if } wx \leq p, \end{cases}$$

where \mathbf{X}_1 and \mathbf{X}_2 are the sub datasets of objects belonging to left and right child nodes, respectively. Below, Fig. 1 shows an illustration of node splitting with a simple linear hyperplane in the multivariate decision tree.

It can be easily seen that the parameters in eq. (1) define a hyperplane. In both ensemble algorithms, MDT-1 and MDT-2, we can see that the hyperplanes on each node (except the terminal nodes) help to generate new knots.

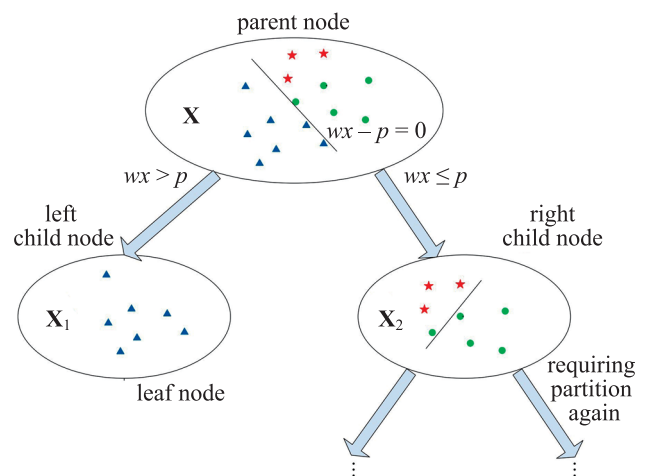


Fig. 1. Node generating with a simple linear hyperplane in the multivariate decision tree

In detail, the ensemble methods MDT-1 and MDT-2 learn the parameters of the hyperplanes in different ways. The first one, MDT-1, uses a vector $\mathbf{v} \in R^{d \times 1}$, such that the dimension is equal to the size of x . Elements of vector \mathbf{v} lie in the interval $[-1, 1]$. Thus, the normal w in the hyperplane can be calculated as follows.

$$w = \frac{\mathbf{v}}{\|\mathbf{v}\|}.$$

However, the second method, MDT-2, uses PCA method to calculate a normal direction w . The main idea is to choose the largest principal component on the given matrix \mathbf{X} to satisfy the normal direction.

Splitting criterion

For generated new nodes after each split, a split criterion is employed to determine whether there is need to continue split again. In general, split is continued if the split criterion holds for child node. Thus, split criterion is set in the following form:

$$P(S) = \max \left(\frac{S_i}{c} \mid i = l_1, \dots, l_k \right) < \lambda,$$

where S is equal to \mathbf{X}_1 or \mathbf{X}_2 ; c is the number of the objects in dataset S ; S_i is the number of the objects in the class i ; and λ is a predefined threshold parameter in the interval $\left[\frac{1}{k}, 1 \right)$. Specifically, if a newly generated child node does not meet the split criterion, it will be considered as a leaf node (terminal node) with a class label

$$l^* = \operatorname{argmax}_i \left\{ \frac{S_i}{c_i} \mid i = l_1, \dots, l_k \right\}.$$

Classification with ensemble methods

The generated multivariate decision tree can be used as a classification algorithm. The new object with its unknown

class label is first come to the root node of the classification tree model, and then goes to child nodes directed by linear hyperplanes to which side the new object belongs of a half hyperplane (Fig. 2). Thus, a previously unseen object by tree model goes through the node of tree and finally reaches a leaf node where classification is made. Thus, learning hyperplanes on each node and generating a tree, the resulting ensemble method gives us a classifier method.

These algorithms, MDT-1 and MDT-2 (Fig. 2), employ $f(w, p)$ — linear hyperplane with parameters w and p .

The proposed method

In this section, we present a new type of multivariate binary decision tree model. The primary difference from above and other existing algorithms is that we propose a new criterion for splitting nodes with piecewise functions and nonlinear transformations.

In general, all the existing tree models construct tree nodes in a linear mode to reach the domain. The proposed algorithm constructs a binary decision tree which contains hyperplanes at each node that separates the objects of different class objects. To construct hyperplanes, we use feedforward neural network with one hidden layer, also called as shallow neural networks. The algorithm is shown below.

Algorithm #1. NNDT

Input: Training dataset matrix D with known class labels set L

Output: An ensemble algorithm which is a binary tree with neural network, the split nodes storing the parameters w of neural network and the leaf nodes storing the class label of objects

1. Initialize $\Psi = D$ and empty set $X = \emptyset$
2. Extract objects with the same class label from Ψ and denote them as \mathbf{X}
3. **while** the set Ψ is not empty **do**

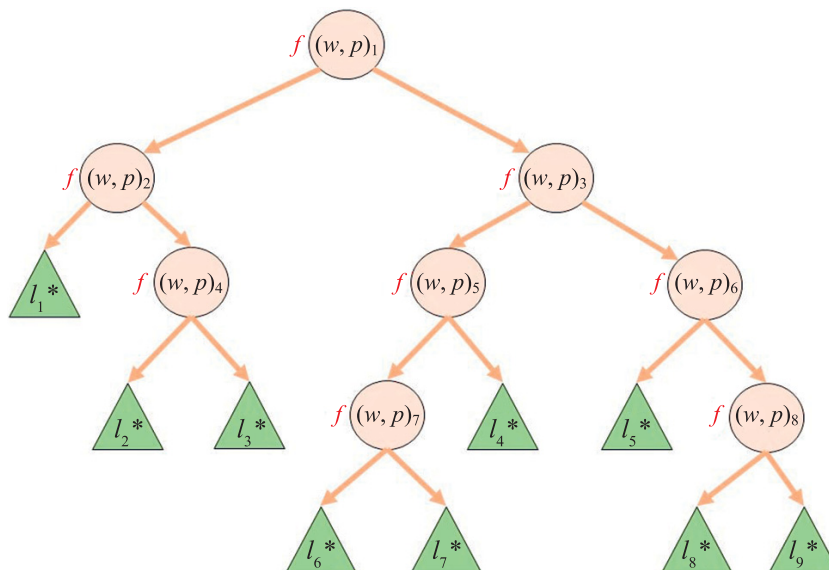


Fig. 2. Multivariate decision tree methods with linear hyperplane, MDT-1 and MDT-2

4. train a new neural network with a selected activation function where its hyperplane separates the objects of \mathbf{X} and Ψ
5. **if** cross-validation condition holds **then**
6. create a leaf node and assign it to the majority class label of \mathbf{X} exclude those objects from the set Ψ
7. **else**
8. create a child node with X_i from Ψ
9. **end if**
10. **end of while**

We can see the process of constructing decision tree of Algorithm #1 in Fig. 3.

In Fig. 3, the signs (+, -, ×) inside child nodes indicate labels of objects of the same class, l_1 , l_k , and l_t — class labels which are assigned from terminal nodes.

Separating using activation functions

In general, MDT-1 and MDT-2 use unsupervised learning in order to construct tree model which leads to losing information when splitting dataset objects, and the resulting splitting structure is not enough clear. In MDT-1, constructing hyperplanes using randomly generating from the normal direction does not provide transparency. Moreover, using PCA, it's not the optimal solution for this problem relying on distribution of objects of multi class dataset. The shortcomings of the two methods lead us to consider another way to construct a new model for this problem.

Moreover, if we use neural network with activation functions and nonlinear transformation property, this leads to an increase in the generalization ability of the neural network.

In general, we propose to use shallow network with the following activation functions in hidden layers.

1. Linear and exponential (piece-wise linear unit) activation functions are given as follows:

$$f_{lrelu}(\mathbf{x}) = \begin{cases} x, & \text{if } x > 0, \\ \alpha x, & \text{if } x \leq 0, \end{cases} \quad (2)$$

where x is input value for linear function $\text{ReLU } f_{lrelu}(\cdot)$.

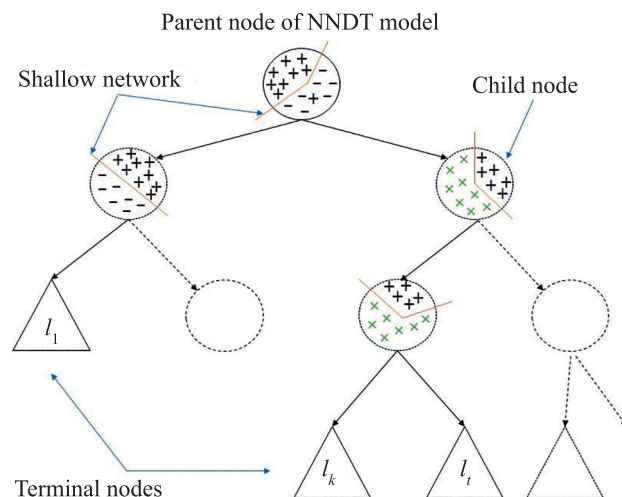


Fig. 3. NNDT model for constructing decision tree

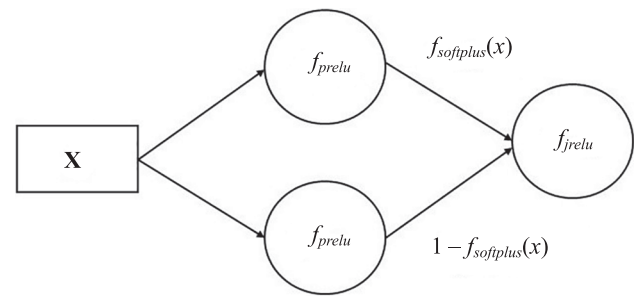


Fig. 4. Joined activation function

2. Exponential ReLU $f_{elu}(\cdot)$ is the activation function where α is a predefined parameter for controlling the angular coefficient.

$$f_{elu}(\mathbf{x}) = \begin{cases} x, & \text{if } x > 0, \\ \alpha(e^x - 1), & \text{if } x \leq 0. \end{cases} \quad (3)$$

3. Joined activation function (Fig. 4) and its analytic form for this approach is given below:

$$f_{jrelu} = \beta f_{prelu}(\mathbf{x}) + (1 - \beta) f_{pelu}(\mathbf{x}), \quad (4)$$

where $\beta \in [0, 1]$ is a joining coefficient indicating the certain grouping of $f_{prelu}(\cdot)$ and $f_{pelu}(\cdot)$. The joining coefficient β is obtained from during the training process.

At each splitting node we can choose one of the activation functions in our neural network in decision tree. Rather than separating objects in the feature space with linear hyperplanes (MDT algorithms), we can use neural network which has ability of nonlinear transformations. Such activation functions are considered much more powerful when the classifying dataset has a very complex structure (high dimensional features space, multi-class datasets).

Numerical experiments

In this section, we provide the results of numerical experiments with proposed method and comparison analysis with other existing methods and algorithms. All experiments are tested on Python version 3.8 environment and Intel Core i7 10780H CPU, 2.20GHz processor with 16GB RAM memory. The datasets are downloaded from Machine Learning Repositories, they are publicly available online^{1,2}. The datasets list is given in Table 1, it is used in numerical experiments.

Classification accuracy of NNDT models compared with the other methods and algorithms (C4.5, MDT-1, MDT-2 and CART) are given in Table 2. We can see that the proposed new model outcomes all the other models in accuracy. Moreover, if we use the proposed decision tree model with activation function which is given in (4), we can achieve more precise results. NNDT with (4) shows the

¹ UC Irvine Machine Learning Repository. Available at: <https://archive.ics.uci.edu/ml/index.php> (accessed: 01.07.2022).

² Machine Learning Repository. Available at: <https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets> (accessed: 01.07.2022).

Table 1. Information about datasets

Dataset	Number of		
	objects	attributes	classes
aloi	108000	128	1000
covtype	581012	54	7
mushrooms	8124	112	2
shuttle	43500	9	7
Wine	178	13	3
Spambase	4601	57	2
Hepatitis	155	19	2
Dry Bean Dataset	13611	17	7

highest accuracy compared to NNDT with (2) and NNDT with (3). This outperformance can be explained with a highly non-linearity of the activation function.

In the following Table 3, we can see comparison of training and classification time of the models. Below, we can see from Table 3 that most of the time C4.5 wins in the training process for nearly in all the datasets. Practically, this also affects the fact that univariate and multivariate decision tree models have different qualities for different cases. However, in Table 3, we can see that NNDT models show the fastest time in the classification among the compared with multivariate decision tree models. Moreover, our proposed models, especially NNDT with (4), show less time needed for classification compared with MDT-1 and MDT-2 multivariate decision models.

In all decision tree algorithms, the number of generated nodes plays important role as a main property. This property directly affects to the performance of the algorithm. In the following Table 4, we can see the number of splitting nodes generated in training process. The proposed model shows the least number of generated nodes, excluding C4.5 algorithm. However, NNDT model has greater accuracy compared to C4.5 (Table 2). Moreover, when we use the proposed decision tree model with activation function which is given in (4), NNDT algorithm generates less nodes. NNDT with (4) shows the less number of nodes compared to NNDT with (2) and NNDT with (3). Here, it is obtained with a highly non-linearity property of the activation function.

We provide numerical experiments of the proposed NNDT model with various activation functions and proposed joined activation function on Spambase, Hepatitis and mushrooms datasets. Firstly, we provide numerical experiments between piecewise linear activation function (ReLU), exponential activation function, and then with the proposed joined activation. First, we perform numerical experiments on the Spambase dataset. Secondly, we mainly focus on evaluating the performance effects of the joined activation function, and perform the comparison experiments with the other configurations on all datasets. In numerical experiments we employ our activation functions replacing $f_{relu}(\cdot)$ activation function in NNDT model which is given in Table 5. After configuration of NNDT models with the activation functions in Table 5, we train our models with the selected datasets. Five separate trials were carried out and the average value of the classification results was calculated.

Table 2. Classification accuracies, %

Dataset	CART	C4.5	MDT-1	MDT-2	NNDT with eq. (2)	NNDT with eq. (3)	NNDT with eq. (4)
aloi	75.5	5	48.02	77.8	98.2	99.1	100
covtype	92.8	42	85.6	91.06	92.5	93.6	99.4
mushrooms	99.5	99.7	100	100	100	100	100
shuttle	99.87	99.25	98.8	99.1	100	100	100
Wine	92	79	99.7	99.9	100	100	100
Spambase	65.5	45	92.5	98.9	99.5	99.4	99.98
Hepatitis	92.8	89.1	92.7	93.4	96.7	97.1	99.1
Dry Bean Dataset	62.2	49.4	89.78	91.2	98.8	98.9	99.6

Table 3. Training time comparison

Dataset	C4.5	MDT-1	MDT-2	NNDT with eq. (2)	NNDT with eq. (3)	NNDT with eq. (4)
aloi	$5.42 \cdot 10^4$	$4.27 \cdot 10^3$	$4.20 \cdot 10^2$	$3.10 \cdot 10^2$	$3.12 \cdot 10^2$	$3.91 \cdot 10^2$
covtype	$3.95 \cdot 10^3$	$1.25 \cdot 10^4$	$2.31 \cdot 10^3$	$0.91 \cdot 10^3$	$1.20 \cdot 10^3$	$1.91 \cdot 10^3$
mushrooms	$0.98 \cdot 10^{-1}$	$2.55 \cdot 10^0$	$7.25 \cdot 10^{-1}$	$1.21 \cdot 10^{-1}$	$1.22 \cdot 10^{-1}$	$1.25 \cdot 10^{-1}$
shuttle	$2.41 \cdot 10^0$	$4.40 \cdot 10^0$	$2.11 \cdot 10^0$	$1.23 \cdot 10^0$	$1.24 \cdot 10^0$	$2.00 \cdot 10^0$
Wine	$0.20 \cdot 10^{-1}$	$1.40 \cdot 10^0$	$2.25 \cdot 10^{-1}$	$0.55 \cdot 10^{-1}$	$0.55 \cdot 10^{-1}$	$0.91 \cdot 10^{-1}$
Spambase	$0.22 \cdot 10^{-1}$	$2.10 \cdot 10^0$	$4.25 \cdot 10^{-1}$	$0.48 \cdot 10^{-1}$	$0.47 \cdot 10^{-1}$	$0.95 \cdot 10^{-1}$
Hepatitis	$0.25 \cdot 10^{-1}$	$0.24 \cdot 10^0$	$0.39 \cdot 10^{-1}$	$0.28 \cdot 10^{-1}$	$0.31 \cdot 10^{-1}$	$0.35 \cdot 10^{-1}$
Dry Bean Dataset	$1.25 \cdot 10^{-1}$	$2.87 \cdot 10^0$	$3.20 \cdot 10^{-1}$	$0.81 \cdot 10^{-1}$	$0.79 \cdot 10^{-1}$	$1.25 \cdot 10^{-1}$

Table 4. Number of splitting nodes

Dataset	CART	C4.5	MDT-1	MDT-2	NNDT with eq. (2)	NNDT with eq. (3)	NNDT with eq. (4)
aloi	16492	78	60452	20598	19129	18561	11233
covtype	32152	86	162201	131305	7916	7920	7551
mushrooms	9	4	484	41	4	4	3
shuttle	15	58	904	527	13	13	11
Wine	21	59	102	76	15	14	8
Spambase	245	359	523	378	201	199	185
Hepatitis	32	45	41	35	15	13	9
Dry Bean Dataset	26	39	64	57	19	19	17

Table 5. Classification results of NNDT models on various activation functions.

Configuration of NNDT model	Classification Rates		
	Spambase dataset	Hepatitis dataset	Mushrooms dataset
$f_{relu}(\cdot)$	94.75	92.45	98.75
$f_{prelu}(\cdot)$	95.05	96.25	100
$f_{elu}(\cdot)$	99.40	97.40	100
$f_{jrelu}(\cdot)$ with $f_{prelu}(\cdot)/f_{pelu}(\cdot)$	99.98	99.15	100

Classification results of NNDT models between activation functions

First, we compare the results among basic activation functions and joined activation functions with configured NNDT models. We use basic activation functions, such as $f_{relu}(\cdot)$ -non-trainable, $f_{lrelu}(\cdot)$ -non-trainable and $f_{pelu}(\cdot)$ -trainable, which are used mostly in the deep learning models. The numerical results in Table 6 with our proposed activation functions show superiority over non-trainable activation functions. The joined activation function shows better performance results than non-trainable activation functions on all datasets. Comparison on Spambase dataset shows that the joined activation function performs the best action and achieves an average improvement compared to other activation functions. We also determine that in the activation approach with trainable parameters, compared with training combination coefficients on the network layer, the performance of learning joining coefficients boosts on accuracy results

$$(f_{jrelu}(\cdot) \text{ with } f_{prelu}(\cdot)/f_{pelu}(\cdot)) > f_{jrelu}(\cdot) \text{ with } f_{lrelu}(\cdot)/f_{elu}(\cdot))$$

enhancing the quality of trained hyper-parameters. Additionally, the performance achieved by the joined approach is almost consistently better than that achieved by the non-trainable approaches with constant coefficients on the same dataset. Overall, the trend of performance achieved by learning activation functions is almost

$$(f_{jrelu}(\cdot) \text{ with } f_{prelu}(\cdot)/f_{pelu}(\cdot)) > f_{jrelu}(\cdot) \text{ with } f_{lrelu}(\cdot)/f_{elu}(\cdot)) >$$

Then we perform experiments to analyze the classification performance, and compare joined activation on other datasets. On every training epoch we can see the accuracy of every model and the proposed model (Fig. 5).

Five separate trials were carried out and average value of the classification results was calculated.

Table 6. Classification error with trainable and non-trainable activation functions of NNDT models. We run five separate trials and report average value of classification rates

Configuration of NNDT model	Classification Rates		
	Spambase dataset	Hepatitis dataset	Mushrooms dataset
$f_{jrelu}(\cdot) \text{ with } f_{lrelu}(\cdot)/f_{elu}(\cdot)$	99.98	99.15	100
$f_{jrelu}(\cdot) \text{ with } f_{prelu}(\cdot)/f_{pelu}(\cdot)$	99.25	99.10	100
$f_{jrelu}(\cdot) \text{ with } f_{lrelu}(\cdot)/f_{elu}(\cdot)$	99.10	98.50	100
$f_{jrelu}(\cdot) \text{ with } f_{prelu}(\cdot)/f_{pelu}(\cdot)$	99.10	98.20	100

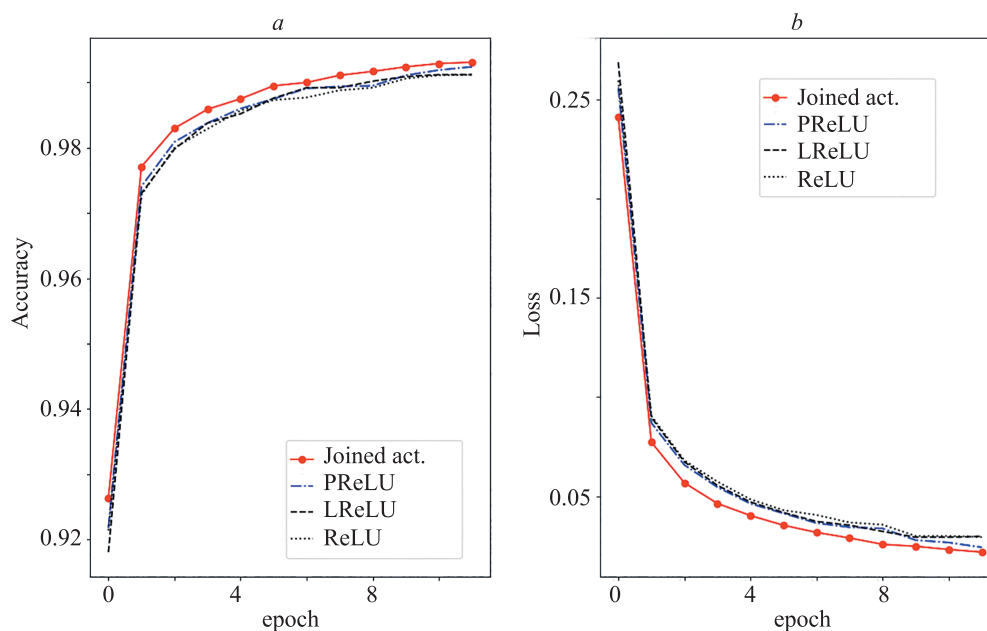


Fig. 5. Classification results of NNMT models between activation functions through epochs. Model accuracy (a), Model loss (b)

Conclusion

In this paper, we have proposed a novel ensemble algorithm based on multivariate decision tree and shallow neural network named as NNMT. NNMT generates nodes employing hyperplanes at each node, separating sample data objects. The proposed algorithm outperforms all the other decision tree algorithms. Moreover, NNMT

algorithm provides high generalization ability, enhancing classification accuracy and performance.

In future researches, informative attribute selection (dimensionality reduction) methods are considered to generate nodes. Selecting informative attributes can lead to reduce the parameter sensitivity and can be obtained much higher generalization ability. The proposed NNMT decision tree model further can be employed to construct decision forest models.

References

1. Morala P., Cifuentes J.A., Lillo R.E., Ucar I. Towards a mathematical framework to inform neural network modelling via polynomial regression. *Neural Networks*, 2021, vol. 142, pp. 57–72. <https://doi.org/10.1016/j.neunet.2021.04.036>
2. Cao W., Mirjalili V., Raschka S. Rank consistent ordinal regression for neural networks with application to age estimation. *Pattern Recognition Letters*, 2020, vol. 140, pp. 325–331. <https://doi.org/10.1016/j.patrec.2020.11.008>
3. Messner E., Fediuk M., Swatek P., Scheidl S., Smolle-Jüttner F.M., Olschewski H., Pernkopf F. Multi-channel lung sound classification with convolutional recurrent neural networks. *Computers in Biology and Medicine*, 2020, vol. 122, pp. 103831. <https://doi.org/10.1016/j.combiomed.2020.103831>
4. Youling L. A calibration method of computer vision system based on dual attention mechanism. *Image and Vision Computing*, 2020, vol. 103, pp. 104039. <https://doi.org/10.1016/j.imavis.2020.104039>
5. Palmerston J.B., Zhou Y., Chan H.M. Comparing biological and artificial vision systems: Network measures of functional connectivity. *Neuroscience Letters*, 2020, vol. 739, pp. 135407. <https://doi.org/10.1016/j.neulet.2020.135407>
6. Basha S.H.Sh., Dubey Sh.R., Pulabaigari V., Mukherjee S. Impact of fully connected layers on performance of convolutional neural networks for image classification. *Neurocomputing*, 2020, vol. 378, pp. 112–119. <https://doi.org/10.1016/j.neucom.2019.10.008>
7. Shuang K., Tan Y., Cai Zh., Sun Y. Natural language modeling with syntactic structure dependency. *Information Sciences*, 2020, vol. 523, pp. 220–233. <https://doi.org/10.1016/j.ins.2020.03.022>
8. Xu M. WITHDRAWN: Image processing system based on FPGA and convolutional neural network. *Microprocessors and Microsystems*, 2020, pp. 103379. <https://doi.org/10.1016/j.micpro.2020.103379>
9. Krizhevsky A., Sutskever I., Hinton G.E. Imagenet classification with deep convolutional neural networks. *Proc. of the 26th Annual*

Литература

1. Morala P., Cifuentes J.A., Lillo R.E., Ucar I. Towards a mathematical framework to inform neural network modelling via polynomial regression // *Neural Networks*. 2021. V. 142. P. 57–72. <https://doi.org/10.1016/j.neunet.2021.04.036>
2. Cao W., Mirjalili V., Raschka S. Rank consistent ordinal regression for neural networks with application to age estimation // *Pattern Recognition Letters*. 2020. V. 140. P. 325–331. <https://doi.org/10.1016/j.patrec.2020.11.008>
3. Messner E., Fediuk M., Swatek P., Scheidl S., Smolle-Jüttner F.M., Olschewski H., Pernkopf F. Multi-channel lung sound classification with convolutional recurrent neural networks // *Computers in Biology and Medicine*. 2020. V. 122. P. 103831. <https://doi.org/10.1016/j.combiomed.2020.103831>
4. Youling L. A calibration method of computer vision system based on dual attention mechanism // *Image and Vision Computing*. 2020. V. 103. P. 104039. <https://doi.org/10.1016/j.imavis.2020.104039>
5. Palmerston J.B., Zhou Y., Chan H.M. Comparing biological and artificial vision systems: Network measures of functional connectivity // *Neuroscience Letters*. 2020. V. 739. P. 135407. <https://doi.org/10.1016/j.neulet.2020.135407>
6. Basha S.H.Sh., Dubey Sh.R., Pulabaigari V., Mukherjee S. Impact of fully connected layers on performance of convolutional neural networks for image classification // *Neurocomputing*. 2020. V. 378. P. 112–119. <https://doi.org/10.1016/j.neucom.2019.10.008>
7. Shuang K., Tan Y., Cai Zh., Sun Y. Natural language modeling with syntactic structure dependency // *Information Sciences*. 2020. V. 523. P. 220–233. <https://doi.org/10.1016/j.ins.2020.03.022>
8. Xu M. WITHDRAWN: Image processing system based on FPGA and convolutional neural network // *Microprocessors and Microsystems*. 2020. P. 103379. <https://doi.org/10.1016/j.micpro.2020.103379>
9. Krizhevsky A., Sutskever I., Hinton G.E. Imagenet classification with deep convolutional neural networks // *Proc. of the 26th Annual*

- Conference on Neural Information Processing Systems (NIPS), 2012, pp. 1097–1105.
10. Goodfellow I., Bengio Y., Courville A. *Deep Learning*. MIT Press, 2016, 775 p.
11. Zhou Z.H., Feng J. Deep forest: Towards an alternative to deep neural networks. *Proc. of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*, 2017, pp. 3553–3559. <https://doi.org/10.24963/ijcai.2017/497>
12. Frosst N., Hinton G. Distilling a neural network into a soft decision tree. *CEUR Workshop Proceedings*, 2018, vol. 2070.
13. Wan A., Dunlap L., Ho D., Yin J., Lee S., Jin H., Petryk S., Bargal S.A., Gonzalez J.E. NBDT: Neural-backed decision trees. *arXiv*, 2020. arXiv:2004.00221. <https://doi.org/10.48550/arXiv.2004.00221>
14. Pinto A., Pereira S., Rasteiro D.M., Silva C. Hierarchical brain tumour segmentation using extremely randomized trees. *Pattern Recognition*, 2018, vol. 82, pp. 105–117. <https://doi.org/10.1016/j.patcog.2018.05.006>
15. Vanli N.D., Sayin M.O., Mohaghegh N.M., Ozkan H., Kozat S.S. Nonlinear regression via incremental decision trees. *Pattern Recognition*, 2019, vol. 86, pp. 1–13. <https://doi.org/10.1016/j.patcog.2018.08.014>
16. Blanco-Justici A., Domingo-Ferrer J., Martínez S., Sánchez D. Machine learning explainability via microaggregation and shallow decision trees. *Knowledge-Based Systems*, 2020, vol. 194, pp. 105532. <https://doi.org/10.1016/j.knosys.2020.105532>
17. Kotsiantis S.B. Decision trees: a recent overview. *Artificial Intelligence Review*, 2013, vol. 39, no. 4, pp. 261–283. <https://doi.org/10.1007/s10462-011-9272-4>
18. Quinlan J.R. Induction of decision trees. *Machine Learning*, 1986, vol. 1, no. 1, pp. 81–106. <https://doi.org/10.1023/A:1022643204877>
19. Breiman L., Friedman J.H., Stone C.J., Olshen R.A. *Classification and Regression Trees*. Chapman & Hall/CRC, 1984.
20. Quinlan J.R. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers Inc., 1993.
21. Chandra B., Paul Varghese P. Moving towards efficient decision tree construction. *Information Sciences*, 2009, vol. 179, no. 8, pp. 1059–1069. <https://doi.org/10.1016/j.ins.2008.12.006>
22. Wang F., Wang Q., Nie F., Yu W., Wang R. Efficient tree classifiers for large scale datasets. *Neurocomputing*, 2018, vol. 284, pp. 70–79. <https://doi.org/10.1016/j.neucom.2017.12.061>
23. Yildiz C., Alpaydin E. Omnivariate decision trees. *IEEE Transactions on Neural Networks*, 2001, vol. 12, no. 6, pp. 1539–1546. <https://doi.org/10.1109/72.963795>
24. Altınçay H. Decision trees using model ensemble-based nodes. *Pattern Recognition*, 2007, vol. 40, no. 12, pp. 3540–3551. <https://doi.org/10.1016/j.patcog.2007.03.023>
25. Kumar M.A., Gopal M. A hybrid SVM based decision tree. *Pattern Recognition*, 2010, vol. 43, no. 12, pp. 3977–3987. <https://doi.org/10.1016/j.patcog.2010.06.010>
26. Nie F., Zhu W., Li X. Decision Tree SVM: An extension of linear SVM for non-linear classification. *Neurocomputing*, 2020, vol. 401, pp. 153–159. <https://doi.org/10.1016/j.neucom.2019.10.051>
- Conference on Neural Information Processing Systems (NIPS). 2012. P. 1097–1105.
10. Goodfellow I., Bengio Y., Courville A. *Deep Learning*. MIT Press, 2016. 775 p.
11. Zhou Z.H., Feng J. Deep forest: Towards an alternative to deep neural networks // *Proc. of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*. 2017. P. 3553–3559. <https://doi.org/10.24963/ijcai.2017/497>
12. Frosst N., Hinton G. Distilling a neural network into a soft decision tree // *CEUR Workshop Proceedings*. 2018. V. 2070.
13. Wan A., Dunlap L., Ho D., Yin J., Lee S., Jin H., Petryk S., Bargal S.A., Gonzalez J.E. NBDT: Neural-backed decision trees // *arXiv*. 2020. arXiv:2004.00221. <https://doi.org/10.48550/arXiv.2004.00221>
14. Pinto A., Pereira S., Rasteiro D.M., Silva C. Hierarchical brain tumour segmentation using extremely randomized trees // *Pattern Recognition*. 2018. V. 82. P. 105–117. <https://doi.org/10.1016/j.patcog.2018.05.006>
15. Vanli N.D., Sayin M.O., Mohaghegh N.M., Ozkan H., Kozat S.S. Nonlinear regression via incremental decision trees // *Pattern Recognition*. 2019. V. 86. P. 1–13. <https://doi.org/10.1016/j.patcog.2018.08.014>
16. Blanco-Justici A., Domingo-Ferrer J., Martínez S., Sánchez D. Machine learning explainability via microaggregation and shallow decision trees // *Knowledge-Based Systems*. 2020. V. 194. P. 105532. <https://doi.org/10.1016/j.knosys.2020.105532>
17. Kotsiantis S.B. Decision trees: a recent overview // *Artificial Intelligence Review*. 2013. V. 39. N 4. P. 261–283. <https://doi.org/10.1007/s10462-011-9272-4>
18. Quinlan J.R. Induction of decision trees // *Machine Learning*. 1986. V. 1. N 1. P. 81–106. <https://doi.org/10.1023/A:1022643204877>
19. Breiman L., Friedman J.H., Stone C.J., Olshen R.A. *Classification and Regression Trees*. Chapman & Hall/CRC, 1984.
20. Quinlan J.R. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers Inc., 1993.
21. Chandra B., Paul Varghese P. Moving towards efficient decision tree construction // *Information Sciences*. 2009. V. 179. N 8. P. 1059–1069. <https://doi.org/10.1016/j.ins.2008.12.006>
22. Wang F., Wang Q., Nie F., Yu W., Wang R. Efficient tree classifiers for large scale datasets // *Neurocomputing*. 2018. V. 284. P. 70–79. <https://doi.org/10.1016/j.neucom.2017.12.061>
23. Yildiz C., Alpaydin E. Omnivariate decision trees // *IEEE Transactions on Neural Networks*. 2001. V. 12. N 6. P. 1539–1546. <https://doi.org/10.1109/72.963795>
24. Altınçay H. Decision trees using model ensemble-based nodes // *Pattern Recognition*. 2007. V. 40. N 12. P. 3540–3551. <https://doi.org/10.1016/j.patcog.2007.03.023>
25. Kumar M.A., Gopal M. A hybrid SVM based decision tree // *Pattern Recognition*. 2010. V. 43. N 12. P. 3977–3987. <https://doi.org/10.1016/j.patcog.2010.06.010>
26. Nie F., Zhu W., Li X. Decision Tree SVM: An extension of linear SVM for non-linear classification // *Neurocomputing*. 2020. V. 401. P. 153–159. <https://doi.org/10.1016/j.neucom.2019.10.051>

Authors

Avazjon R. Marakhimov — D. Sc., Professor, Rector, Termez State University, Termez, 190011, Uzbekistan, <https://orcid.org/0000-0003-3735-6855>, termizdu@umail.uz

Jabbarbergen K. Kudaybergenov — PhD, Lecturer, Tashkent University of Information Technologies Nukus branch named after Muhammad Al-Khwarizmi, Nukus, 230113, Uzbekistan, <https://orcid.org/0000-0003-4494-6255>, kjabbarbergen@gmail.com

Kabul K. Khudaybergenov — PhD, Lecturer, National University of Uzbekistan, Tashkent, 100174, Uzbekistan, <https://orcid.org/0000-0001-8143-625X>, kabul85@mail.ru

Ulugbek R. Ohundadaev — Basic Doctoral Student, National University of Uzbekistan, Tashkent, 100174, Uzbekistan, <https://orcid.org/0000-0002-3240-6502>, ulugbek_1122@mail.ru

Авторы

Марахимов Авазжон Рахимович — доктор технических наук, профессор, ректор, Термезский государственный университет, Термез, 190111, Узбекистан, <https://orcid.org/0000-0003-3735-6855>, termizdu@umail.uz

Кудайберген Жаббарберген Кадирбергеневич — кандидат технических наук, преподаватель, Нукусский филиал Ташкентского университета информационных технологий имени Мухаммад ал-Хоразми, Нукус, 230113, Узбекистан, <https://orcid.org/0000-0003-4494-6255>, kjabbarbergen@gmail.com

Худайберген Кабул Кадирбергеневич — кандидат технических наук, преподаватель, Национальный университет Узбекистана, Ташкент, 100174, Узбекистан, <https://orcid.org/0000-0001-8143-625X>, kabul85@mail.ru

Охундадаев Улугбек Рахимжон угли — базовый докторант, Национальный университет Узбекистана, Ташкент, 100174, Узбекистан, <https://orcid.org/0000-0002-3240-6502>, ulugbek_1122@mail.ru

Received 11.04.2022

Approved after reviewing 06.06.2022

Accepted 14.07.2022

Статья поступила в редакцию 11.04.2022

Одобрена после рецензирования 06.06.2022

Принята к печати 14.07.2022

doi: 10.17586/2226-1494-2022-22-4-734-741

УДК 004.416.2

Организация фаззинг-тестирования многопоточных приложений на основе метода распараллеливания независимых переходов

Олег Владимирович Доронин✉

Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

dorooleg@yandex.ru✉, <https://orcid.org/0000-0003-4209-8440>

Аннотация

Предмет исследования. Современные информационные системы сложно представить без использования многопоточности. Многопоточность может как повышать производительность системы в целом, так и замедлять выполнение приложений за счет возникновения ошибок многопоточного программирования. Для нахождения таких ошибок на языках C/C++ существует модуль компилятора Google Thread Sanitizer. Но порядок выполнения потоков может каждый раз меняться при запуске программы на выполнение и влиять на появление подобных ошибок. Для многократного изменения порядка выполнения потоков за время работы программы в Google Thread Sanitizer применен модуль фаззинг-тестирования, который повысил вероятность нахождения ошибок. Все алгоритмы планирования потоков в фаззинг-модуле предназначены для последовательного выполнения потоков, что приводит к значительному замедлению работы Google Thread Sanitizer. Также происходит влияние на тестирование приложений, которое зависит от асинхронных взаимодействий (ожидания сетевых событий, ограничения времени выполнения операций). **Метод.** Для ускорения работы фаззинг-планировщиков предложен метод распараллеливания независимых переходов. Ошибки многопоточного программирования возникают при изменении разделяемого состояния между потоками, при этом локальные вычисления не влияют на воспроизведение многопоточных ошибок. Изменение разделяемых состояний происходит в точках синхронизаций, где выполняется переключение потоков по принципу кооперативной многозадачности. Предложено осуществлять управление последовательностями выполнения потоков только при изменении разделяемых состояний в точках синхронизаций, а локальные вычисления выполнять параллельно. Данное условие позволило сократить время тестирования без снижения результативности обнаружения ошибок многопоточного программирования. Для анализа теоретической сложности алгоритма планирования применен метод комбинаторного подсчета. **Основные результаты.** Предложен новый подход организации фаззинг-тестирования на основе метода распараллеливания независимых переходов, реализация которого по теоретическим и практическим оценкам показывает заметное ускорение работы фаззинг-планировщиков. Результаты эксперимента показали, что для алгоритма перебора всех вариантов выполнения приложения ускорение выполнения достигает 1,25 раза для двух потоков. Представлено соотношение для оценки ускорения в случае произвольного числа потоков. **Практическая значимость.** Предложенный подход позволяет покрывать фаззинг-тестами многопоточные приложения, для которых важно время выполнения — приложения с привязкой к асинхронным взаимодействиям.

Ключевые слова

инструменты поиска ошибок, многопоточность, фаззинг-тестирование, алгоритмы планирования, компиляторы

Благодарности

Персональная благодарность Дергачеву Андрею Михайловичу за оказанную поддержку и мотивацию при написании работы.

Ссылка для цитирования: Доронин О.В. Организация фаззинг-тестирования многопоточных приложений на основе метода распараллеливания независимых переходов // Научно-технический вестник информационных технологий, механики и оптики. 2022. Т. 22, № 4. С. 734–741. doi: 10.17586/2226-1494-2022-22-4-734-741

Improvement and comparison the performance of fuzzing testing algorithms for applications in Google Thread Sanitizer

Oleg V. Doronin✉

ITMO University, Saint Petersburg, 197101, Russian Federation

dorooleg@yandex.ru✉, <https://orcid.org/0000-0003-4209-8440>

Abstract

It is difficult to imagine modern information systems without the use of multithreading. The use of multithreading can both improve the performance of the system as in whole so as slow down the execution of multithreaded applications due to the occurrence of multithreaded programming errors. To find such errors in C/C++ languages, there exists a Google Thread Sanitizer compiler module. The order of execution of threads can change every time the program is started for execution and can affect the appearance of such errors. To repeatedly change the order of execution of threads during the execution of the program, Google Thread Sanitizer has a fuzzing testing module that allows you to increase the probability of finding errors. But all the thread scheduling algorithms in this module are presented in the form of sequential execution of threads which can lead to a significant slowdown in Google Thread Sanitizer as well as can affect the testing of applications that depends on timers (waiting for network events, deadline for operations, ...). To speed up the work of fuzzing schedulers, a method for parallelizing independent transitions is proposed. From the point of view of multithreaded programming errors, it is only important to change the shared state between threads, and local calculations do not affect the reproduction of multithreaded errors. The changes of shared states themselves occur at synchronization points (places in the code where threads are switched according to the principle of cooperative multitasking). The method suggests ordering only the change of shared states at synchronization points, and performing local calculations in parallel, due to which parallelization is achieved. For the analysis of theoretical complexity of the algorithm, the method of combinatorial counting is used. A new approach to the organization of fuzzing testing based on the method of parallelization of independent transitions is proposed the implementation of which, according to theoretical and practical estimates, shows a noticeable acceleration of the work of fuzzing schedulers. According to the results of the experiment, it was revealed that for the algorithm of iterating through all execution variants, the acceleration of execution reaches 1.25 times for two threads. For an arbitrary number of threads, an estimate is presented in the form of a formula. The proposed approach allows fuzzing tests to cover multithreaded applications for which execution time is important — applications with reference to timers which improve the quality of the software.

Keywords

error search tools, multithreading, fuzzing testing, scheduling algorithms, compilers

Acknowledgements

Personal thanks to Andrey Mikhailovich Dergachev for the support and motivation in writing the work.

For citation: Doronin O.V. Improvement and comparison the performance of fuzzing testing algorithms for applications in Google Thread Sanitizer. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2022, vol. 22, no. 4, pp. 734–741 (in Russian). doi: 10.17586/2226-1494-2022-22-4-734-741

Введение

Для инструментирования и выявления ошибок в программах, написанных на языках C/C++ [1], часто используется модуль компилятора Google Thread Sanitizers (GTSAN).

Для однопоточных приложений используются модули address, memory и др., для многопоточных — только GTSAN. Основные ошибки, которые обнаруживает модуль GTSAN¹, — гонки данных [2–4] и взаимоблокировки. Кроме основных ошибок, встречаются и такие редкие, как проблема ABA [5], которые характерны алгоритмам lock-free и wait-free [6, 7], но в GTSAN нет алгоритмов для их обнаружения.

Успешность нахождения основных ошибок зависит от порядка выполнения потоков. Для повышения вероятности нахождения ошибок используется фаззинг-модуль, интегрированный в GTSAN [8]. Фаззинг-модуль имеет два базовых варианта архитектуры управления потоками, основанные на: легковесных потоках (однопоточный режим, без возможности распараллели-

вания); потоках POSIX (Portable Operating System Interface), которые приостанавливаются в нужные фаззинг-планировщику моменты времени.

В дополнение к фаззинг-модулям существуют инструменты формальной верификации приложений — model checking [9–11]. Отметим, что данные инструменты позволяют проверить теоретическую корректность алгоритмов и, как правило, не встраиваются при верификации реальных приложений.

В модуле GTSAN, кроме архитектуры управления потоками, используются различные планировщики, дающие возможность управлять последовательностью выполнения потоков и повышать вероятность нахождения ошибок в многопоточном коде. Алгоритмы планирования GTSAN реализуются только последовательно, что замедляет выполнение одной итерации программы. В настоящей работе предложен метод распараллеливания алгоритмов фаззинг-планирования потоков в GTSAN. Проведено сравнение времени выполнения приложений с аналогами.

Архитектура Google Thread Sanitizer

GTSAN — встроенный в компилятор модуль, который работает на одной из стадий оптимизации. На

¹ ThreadSanitizer project: documentation, source code, dynamic annotations, unit tests [Электронный ресурс]. URL: <http://code.google.com/p/data-race-test>, свободный. Яз. англ. (дата обращения: 19.02.2022).

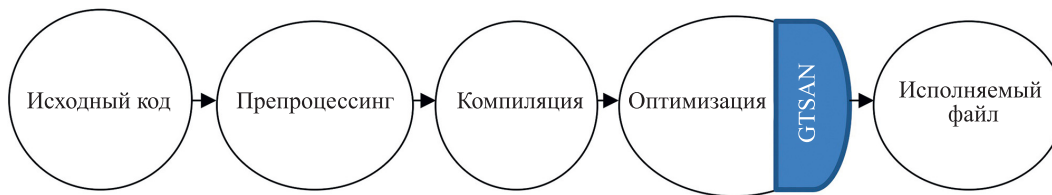


Рис. 1. Этапы компиляции программы

Fig. 1. Stages of program compilation

данном этапе происходит перехват вызовов для работы с атомарными переменными, мьютексами, условными переменными и другими примитивами синхронизации. На рис. 1 представлены этапы компиляции и место встраивания GTSAN.

Платформа управления потоками построена с использованием POSIX потоков [12] и алгоритма управления потоками на основе ожиданий. Когда исполнение потока доходит до точки синхронизации — SynchronizationPoint, поток может приостановить свою работу или передать ее другому потоку на выполнение [13]. Чтобы обеспечить корректную работу с примитивами синхронизации, для которых еще не поддержана работа в фаззинг-модуле, существует специальный поток — WatchDog. Задача WatchDog — следить за временем, которое тратит поток на выполнение, а в случае превышения порогового значения, переводить следующий поток в состояние выполнения. Для смены состояния потока применяется планировщик Scheduler (рис. 2).

В случае параллельных алгоритмов планирования требуется разделить SynchronizationPoint на два со-

стояния — до (Before) и после (After), как показано на рис. 3. Работа в рамках SynchronizationPoint будет последовательной (с примитивом синхронизации), но она вносит минимальный вклад во время выполнения программы. Распараллеливание будет происходить на участках между SynchronizationPoint.

На рис. 3 в качестве примера, точки синхронизации изображены желтой окружностью со следующими операциями:

- load/store — чтение и запись в атомарную операцию;
- lock/unlock — захват и освобождения мьютекса;
- wait/notify — ожидание и пробуждение потока, взаимодействующего с условной переменной.

Параллельные алгоритмы планирования

В GTSAN существуют последовательные алгоритмы фаззинг-тестирования потоков, такие как: случайный, случайный с разными распределениями, полный перебор, полный перебор на фиксированном окне, полный перебор всех состояний. Все эти алгоритмы созданы для последовательного исполнения потоков

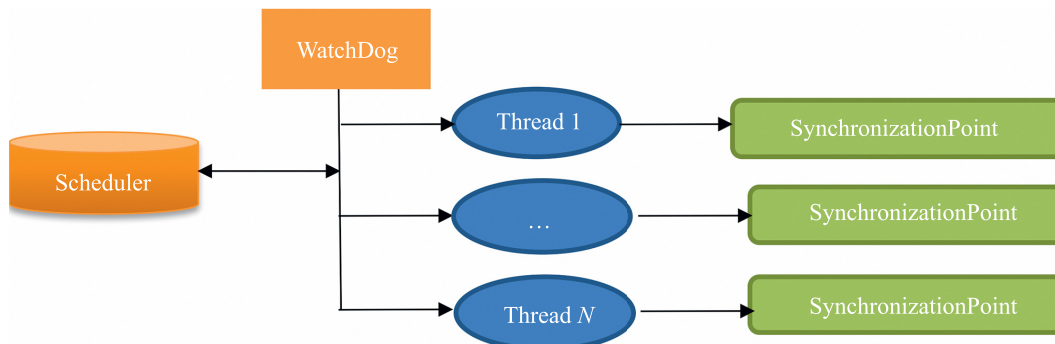


Рис. 2. Схема управления потоками с помощью алгоритма на основе ожиданий

Fig. 2. Thread control scheme using based on waiting algorithm

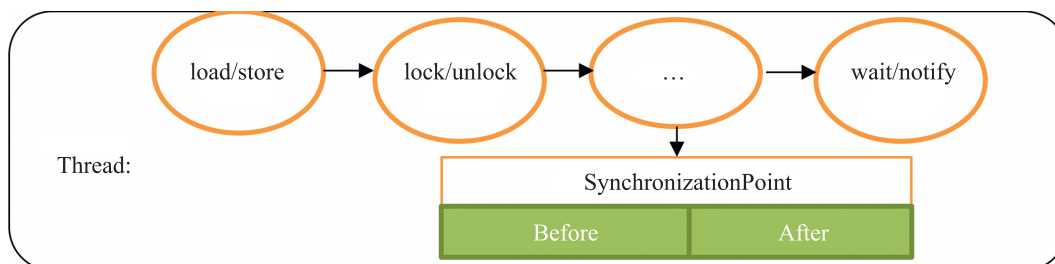


Рис. 3. Представление SynchronizationPoint как два участка Before/After

Fig. 3. Representation of SynchronizationPoint as two sections Before/After

[14, 15]. Рассмотрим процесс изменения каждого пересчитанного алгоритма в параллельный.

Перебор всех вариантов. Это один из самых сложных алгоритмов для распараллеливания в модуле фазинг-тестирования потоков. Создав параллельную реализацию данного алгоритма возможно автоматически распараллелить и все остальные алгоритмы. Идея последовательного алгоритма заключается в следующем: каждое исполнение можно представить в виде плавающей системы счисления [14], где максимальное число состояний разряда ограничено числом потоков, которые можно выбрать в этот логический момент времени. Перебирая все возможные числа в такой системе счисления, получим все варианты исполнения потоков. Под логическим временем имеем в виду дискретную величину, монотонно растущую на единицу при прохождении через точку синхронизации, начинающуюся с нуля. Данное время — общее для всех потоков. После прохождения логического времени выбирается следующий поток, продолжающий выполнение алгоритма, в это время все остальные потоки ожидают наступления логического момента, после которого они смогут начать свое выполнение. Понятие времени с точки зрения физики не подходит, так как это — непрерывная величина, и переключение потоков в произвольный момент физического времени может не давать новых вариантов обращений к разделяемым ресурсам и не приведет к воспроизведению новых ошибок многопоточного программирования.

Алгоритм перебора всех вариантов разбивается на две части: обработка точек синхронизации и выбор потока на исполнение.

Потоки между точками синхронизации можно выполнять в параллельном режиме, а логику точек синхронизации (load/store, lock/unlock, wait/notify) только в последовательном. Последовательное исполнение в этом случае оказывает минимальное влияние, так как эти части и в обычной программе исполняются с синхронизациями.

Выбор следующего потока на исполнение осуществляется в `SynchronizationPointAfter()`, `GetNextTid()` и в момент изменения состояния. Но если изменили состояние на тот же самый поток, то тогда в `SynchronizationPointBefore()` у потока добавляется проверка наличия состояния `RUNNING` вместо `PARALLEL_RUNNING`, и поток необходимо продолжить исполнять.

Когда вызывается функция `SynchronizationPointAfter()` — значит код обработки точки синхронизации завершен, и поэтому можно изменить логическое время. Определим идентификатор следующего на выполнение потока с помощью `GetNextTid()`, и если этот поток ожидает разрешение на выполнение, то переведем его в состояние `RUNNING`. `GetNextTid()` может вернуть тот же самый поток (`tid == nextTid`), тогда поток не может параллельно запустить сам себя. Другой вариант — когда `GetNextTid()` попадает в уже выполняющийся поток, который невозможно запустить параллельно с самим собой. Тогда необходимо вернуть обратно идентификатор следующего на исполнении потока через вызов

`PutNextTid(nextTid)`, и он уже будет подхвачен при достижении точки синхронизации одним из выполняющихся потоков. Если `GetNextTid()` возвращает поток, который находится в состоянии `WAIT`, то необходимо перевести его в `PARALLEL_RUNNING`, тогда поток пройдет в `SynchronizationPointBefore()`. При этом поток может находиться в цикле `while`, если еще не зашел в `SynchronizationPointBefore()`. Это возможно при старте потока (все потоки создаются с состоянием `WAIT`). После этого поток пройдет `SynchronizationPointBefore()` и в `SynchronizationPointAfter()` изменит состояние на `PARALLEL_RUNNING`, из-за которого при следующем прохождении `SynchronizationPointBefore()` он приостановит свое выполнение, только если `GetNextTid()` не ожидал его на выполнение еще раз.

Псевдокод `SynchronizationPointBefore` имеет вид:

```
SynchronizationPointBefore():
tid = GetTid();
oldState = state[tid];
if (oldState != RUNNING):
    state[tid] = WAIT
while (state[tid] == WAIT) Yield();
```

Псевдокод `SynchronizationPointAfter` имеет вид:

```
SynchronizationPointAfter():
tid = GetTid();
state[tid] = PARALLEL_RUNNING;
nextTid = GetNextTid();
if state[nextTid] == WAIT:
    state[nextTid] = RUNNING;
else:
    PutNextTid(nextTid)
```

Рассмотрим пример, в котором два потока пытаются изменить разделяемую переменную *value* *N* раз.

Пример кода для оценивания теоретического распараллеливания программы:

```
for (int j = 0; j < N; j++) {
    value++;
    // логика с интенсивным использованием
    ресурсов центрального процессора, T
    миллисекунд
}
```

Время вычисления, присваивания и сравнения переменных *value*, *j*, *N* гораздо меньше по сравнению со временем интенсивного использования ресурсов центрального процессора. Потому переменная *T* будет обозначать время выполнения одного цикла выполнения программы.

Число логических времен в этом примере равно $2 \times N$, и в каждой такой точке логического времени могут выполняться 0 или 1 поток. 0 и 1 потоки по *N* раз встречаются в такой последовательности логических времен. Число всех возможных последовательностей исполнения потоков равно $\binom{2 \times N}{N}$, каждая из которых содержит $2 \times N - 1$ переходов.

Оценим число переходов между логическим временем $i-1$ и i , которые занимают нулевое время исполнения (к этому времени они уже вычислены). Для перехода логического времени из 0 в 1 с уже вычисленными данными должна быть последовательность исполнения потоков 01 или 10, и далее все возможные комбинации на оставшихся последовательностях. В оставшейся последовательности логических времен будет $N-1$ точек с идентификатором потока 0 и $N-1$ точек с идентификатором потока 1, в сумме $2 \times N-2$ точек последовательности. Число возможных сочетаний последовательности будет равно $\binom{2 \times N-2}{N-1}$, умноженное на два для

двух случаев 01 или 10. Далее преобразуем это число в другую форму для упрощения подсчета. Для этого воспользуемся рекурсивной формулой подсчета числа сочетаний $\binom{N}{k} = \binom{N-1}{k} + \binom{N-1}{k-1}$ и правилом симметрии $\binom{N}{k} = \binom{N}{N-k}$, где k — набор элементов выбираемых из N элементного множества, тогда получим:

$$2 \times \binom{2 \times N-2}{N-1} = 2 \times \left(\binom{2 \times N-3}{N-1} + \binom{2 \times N-3}{N-2} \right) = 2 \times \left(\binom{2 \times N-3}{N-1} + \binom{2 \times N-3}{N-1} \right) = 4 \times \binom{2 \times N-3}{N-1}.$$

Выполним оценку остальных переходов между $i-1$ и i , где $i-1 > 0$. Выберем подпоследовательности исполнения потоков 001 или 110 в произвольном логическом времени, так как только они дают нулевое время выполнения программы. Получим всего комбинаций $2 \times \binom{2 \times N-3}{N-1}$ и последовательностей $2 \times N-2$. В итоге имеем $4 \times \binom{2 \times N-3}{N-1}$ с использованием рекурсивной

формулы и правила симметрии. Тогда доля распараллеливаемых участков равна отношению числа переходов с нулевым временем исполнения к числу всех возможных переходов для всех вариантов исполнения потоков. Для упрощения формулы воспользуемся выражением подсчета сочетаний через факториалы.

$$\begin{aligned} \binom{N}{k} &= \frac{N!}{k \times (N-k)!} \cdot \frac{4 \times N \times \binom{2 \times N-3}{N-1}}{(2 \times N-1) \times \binom{2N}{N}} = \\ &= \frac{N \times 4 \times \frac{(2 \times N-3)!}{(N-1)! \times (N-2)!}}{(2 \times N-1) \times \frac{(2 \times N)!}{N! \times N!}} = \\ &= \frac{N \times 4 \times (2 \times N-3)! \times N! \times N!}{(2 \times N-1) \times (N-1)! \times (N-2)! \times (2 \times N)!} = \\ &= \frac{N \times 4 \times N \times N \times (N-1)}{(2 \times N-1) \times 2 \times N \times (2 \times N-1) \times (2 \times N-2)} = \\ &= \frac{N \times N}{(2 \times N-1) \times (2 \times N-1)} = \left(\frac{N}{2 \times N-1} \right)^2. \end{aligned}$$

Если рассмотреть график функции зависимости числа шагов от распараллеливаемых участков кода, например, при $N=5$, то коэффициент параллельного выполнения равен 0,308642, а предел функции

$\lim_{N \rightarrow \infty} \left(\frac{N}{2 \times N-1} \right)^2 = 0,25$. В среднем получим коэффициент распараллеливания не хуже, чем 0,25 для одной итерации выполнения программы.

Выведем формулу для вычисления распараллеливаемых участков с произвольным числом потоков T . Пронумеруем потоки от 1 до i и обозначим через n_i число шагов, которые проделывает поток (пусть для всех потоков оно будет одинаковым $n_1 = n_2 = \dots = n_T = n$). Тогда $n \times T = N$ число всех шагов, а $N-1$ — число всех переходов. Вычислим число всех переходов между

$k-1$ и k , оно будет равно $\binom{N}{n_1 \dots n_i \dots n_T}$ числу всех

путей (число всех возможных комбинаций исполнения потоков). Из полученного числа переходов рассчитаем число переходов с соседними повторяющимися потоками для переходов из $k-1$ и k . Такие переходы невозможно распараллелить, так как для продвижения логического времени тот же самый поток должен дойти до точки синхронизации, а это невозможно, так как один и тот же поток не может выполняться параллельно сам с собой. Для каждой подпоследовательности на логическом интервале времени от $k-T+1$ до k будем вести подсчет. Тогда вне этого интервала имеем $N-T-1$ переходов. Зафиксируем равные потоки на данном интервале при переходе от $k-1$ в k . Число шагов, в которых будет встречаться фиксированный поток вне этого интервала, будет не больше, чем $n-2$ раз, а для остальных потоков — в n раз. Для фиксированного набора потоков на подпоследовательности $k-T+1$ до k получим число комбинаций исполнения потоков:

$\binom{N-T-1}{n_1-2-p_1 \dots n_i-p_i \dots n_T-p_T}$, где $p_1, p_i \dots p_T$ — число потоков с номером i , используемых на интервале от $k-T+1$ до k . Рассчитаем число всех комбинаций исполнения потоков для различных $p_1, p_i \dots p_T$:

$$\sum_{\substack{p_1+\dots+p_T=T-1 \\ 0 \leq p_i \leq T-1}} \binom{N-T-1}{n_1-2-p_1 \dots n_i-p_i \dots n_T-p_T},$$

а для всех фиксированных T получим

$$T \times \sum_{\substack{p_1+\dots+p_T=T-1 \\ 0 \leq p_i \leq T-1}} \binom{N-T-1}{n_1-2-p_1 \dots n_i-p_i \dots n_T-p_T}.$$

Следующим шагом вычислим число исполнений с различными потоками перед k , и не совпадающими с k и $k-1$. Такие переходы нельзя распараллелить, так как необходим один дополнительный свободный поток, а все T потоки выполняются, и для продвижения логического времени следует дождаться освобождения потоков. Тогда в момент времени k возможно выполнение одного из T потоков, а в момент времени $k-1$ — любого, кроме исполняющегося в момент времени k (их число равно $T-1$). На оставшейся фиксированной под-

последовательности могут быть $(T-1)!$ перестановок различных потоков. Для конкретной фиксированной подпоследовательности получим число последовательностей исполнения потоков, равное числу сочетаний

$$\binom{N-T-1}{n_1-2 \dots n_i-1 \dots n_T-1}.$$

Сложим все варианты вместе и получим итоговую формулу:

$$T \times (T-1) \times (T-1)! \times \binom{N-T-1}{n_1-2 \dots n_i-1 \dots n_T-1}.$$

Для вычисления числа последовательностей исполнений, которые приводят к распараллеливанию фиксированного перехода, необходимо отнять из всех последовательностей только те, которые не приведут к распараллеливанию. В итоге имеем

$$\begin{aligned} & \binom{N}{n_1 \dots n_i \dots n_T} - T \times \\ & \times \sum_{\substack{p_1 + \dots + p_T = T-1 \\ 0 \leq p_i \leq T-1}} \binom{N-T-1}{n_1-2-p_1 \dots n_i-p_i \dots n_T-p_T} - T \times \\ & \times (T-1) \times (T-1)! \times \binom{N-T-1}{n_1-2 \dots n_i-1 \dots n_T-1}. \end{aligned}$$

Случайный планировщик и планировщик с различными распределениями случайных величин. Случайный планировщик для выбора потоков использует равномерное распределение, формула плотности которого имеет вид

$$f(x) = \frac{1}{b-a}, x \in [a, b],$$

где a, b — границы конечного интервала, в котором плотность сохраняет постоянное значение.

При использовании алгоритма планирования с различными случайными величинами, в циклическом буфере найдем генераторы случайных величин, и на каждой итерации выберем свой генератор (нормальное распределение, логнормальное, экспоненциальное, и др.) [14].

Для распараллеливания планировщиков воспользуемся подходом, который описан в планировщике перебора всех вариантов. `GetNextTid()` возвращает случайный поток, а `SynchronizationPointBefore()` и `SynchronizationPointAfter()` решают, можно ли его запустить параллельно с остальными. Отметим, что в алгоритме планирования есть некоторые ограничения на предел распараллеливания потоков в среднем.

Рассмотрим второй алгоритм, который заключается в следующем. В каждой точке синхронизации `SynchronizationPoint` (вместо `SynchronizationPointBefore()` и `SynchronizationPointAfter()`) выберем потоки, которые могут дальше выполняться, и исполним их. Причем выбор числа потоков осуществим с помощью генератора случайных величин. Изменяя число исполняемых потоков от максимального к минимальному, произведем проверку уровня параллелизма в системе. Для случайного планировщика строгой завязки на порядок исполнения потоков не требуется, и поэтому последовательный порядок исполнения между точками синхронизации можно не гарантировать.

Планировщик с перебором на фиксированном окне. Идея планировщика с фиксированным окном заключается в следующем: в качестве параметра задается размер окна, в котором происходит перебор всех вариантов, а окно движется в соответствии с логическим временем [15].

Один алгоритм заключается в использовании подхода, который описан в переборе всех вариантов, что позволяет повысить уровень параллелизма. `GetNextTid()` возвращает случайный поток, а `SynchronizationPointBefore()` и `SynchronizationPointAfter()` решают, можно ли его запустить параллельно с остальными.

Другой подход заключается в достижении нужного логического времени путем использования случайного планировщика с максимальным уровнем параллелизма. В этом случае максимальная скорость исполнения достигается в интервалах до и после окна. А обработка окна осуществляется параллельным алгоритмом полного перебора.

Планировщик с перебором всех состояний. Для планировщика с перебором всех состояний основная цель, чтобы в каждый логический момент времени поработал каждый из возможных потоков. Тогда число итераций для такого планировщика не превышает максимальное число потоков в системе [14]. Первая итерация в таком алгоритме производится без ограничений на порядок. А следующие итерации — с использованием подхода, описанного в полном переборе вариантов.

Результаты тестирования

Для проведения тестирования использованы: существующие unit тесты платформы GTSAN¹; тестовые примеры, на которых успешно проверены последовательные алгоритмы.

Листинг 1. Пример гонки данных.

```
// std::atomic_int d;
// int a;
// thread

++d; ++a; ++d;
```

В листинге 1 присутствует гонка данных на переменной a . Два потока пытаются получить доступ к переменной без синхронизации. Планировщики всех параллельных алгоритмов успешно находят гонку данных.

Листинг 2. Пример с редкими случаями появления значений.

```
// std::atomic_int value { 0 };
// thread:
for (int j = 0; j < 5; j++) {
    auto r = value.load();
    r++;
    value.store(r);
}
```

¹ ThreadSanitizer unit tests [Электронный ресурс]. URL: <https://github.com/llvm-mirror/compiler-rt/tree/master/lib/tsan/tests>, свободный. Яз. англ. (дата обращения: 19.02.2022).

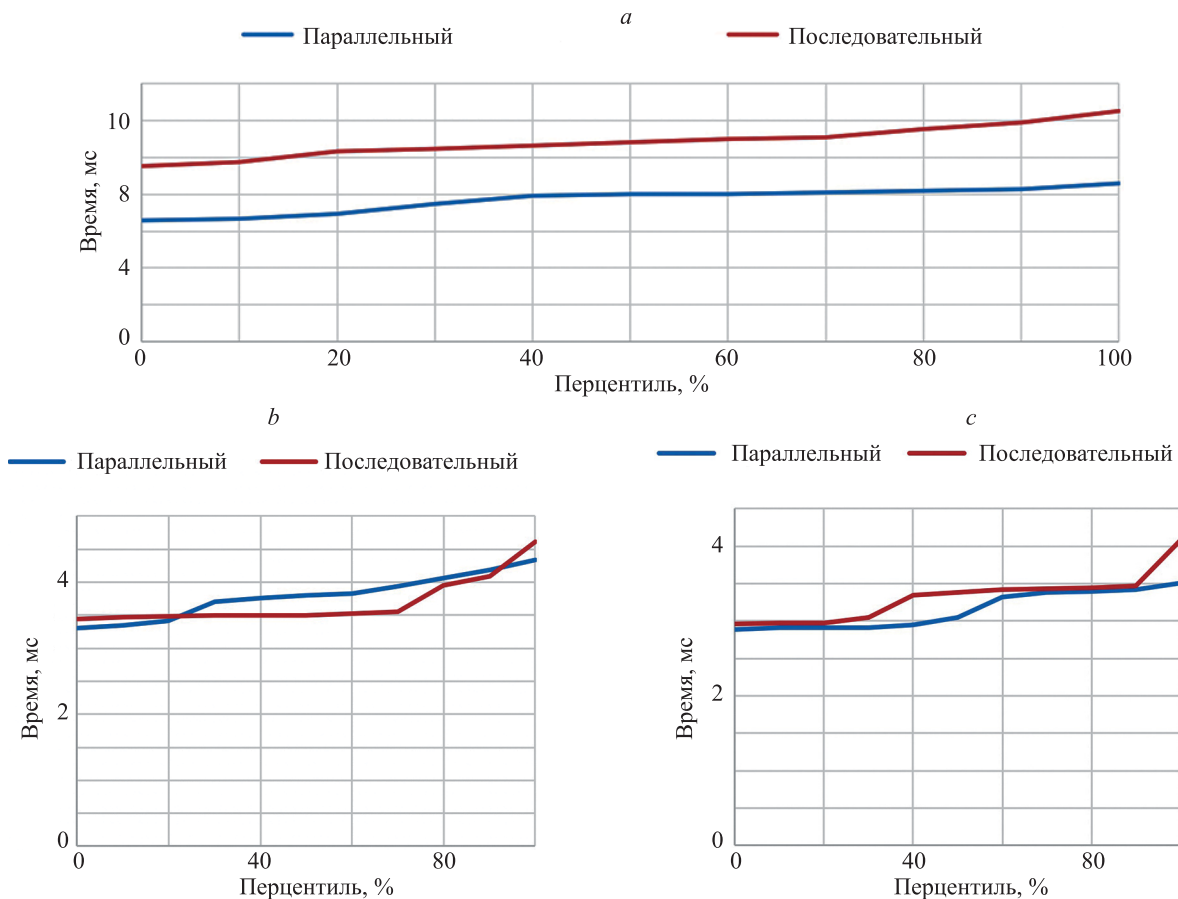


Рис. 4. Сравнение планировщиков на примере кода с редкими случаями проявления значений: переборы всех вариантов (a) и на фиксированном окне (b); случайный планировщик (c)

Fig. 4. Comparison of schedulers on the example of code with rare cases of manifestation of values: iteration of all options (a) and on a fixed window (b); random scheduler (c)

В листинге 2 приведен пример, который демонстрирует, насколько сложные случаи позволяют обнаруживать фаззинг-планировщики (планировщики операционной системы на более чем сто тысяч итераций не способны обнаружить все возможные значения value). Параллельные алгоритмы для полного перебора вариантов и планировщик с различными распределениями случайных величин позволяют получить все наборы результатов для переменной r от 2 до 10, как это было с последовательными алгоритмами.

Выполним сравнение параллельных и последовательных планировщиков на примере кода с редкими случаями проявления значений. Проведем тестирование на процессоре Intel(R) Core(TM) i7-3610QM CPU @ 2.30GHz (8 ядер).

На графиках рис. 4 показано распределение среднего времени выполнения одной итерации на ста запусках для каждого алгоритма планирования с последовательной и параллельной реализациями. В случаях планировщика с перебором всех вариантов (рис. 4, a) и случайного планировщика (рис. 4, c) заметно уменьшение времени работы в параллельной реализации. При фиксированном окне (рис. 4, b) сложно сказать об улучшении времени работы.

Заключение

Для перехода от последовательного исполнения потоков при тестировании многопоточных приложений к параллельному, разработаны алгоритмы распараллеливания для разных способов фаззинг-планирования потоков в модуле GTSAN. Для каждого из разработанных алгоритмов выполнена теоретическая оценка времени выполнения, подтвержденная практическим экспериментом.

Разработанные алгоритмы параллельного фаззинг-тестирования асимптотически показали лучшие результаты, чем существующие аналоги, а с точки зрения корректности обнаружения ошибок многопоточного программирования продемонстрировали результаты, аналогичные уже реализованным в GTSAN последовательным алгоритмам. Полученные результаты позволяют ускорить фаззинг-планирование потоков, а также сделать возможным фаззинг-тестирование приложений, для которых критически важно время выполнения одной итерации программы.

Литература

1. Stroustrup B. *The C++ Programming Language*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2000. 1019 p.
2. Serebryany K., Iskhodzhanov T. ThreadSanitizer - Data race detection in practice // *ACM International Conference Proceeding Series*. 2009. P. 62–71. <https://doi.org/10.1145/1791194.1791203>
3. Netzer R.H.B., Miller B.P. What are race conditions?: Some issues and formalizations // *ACM Letters on Programming Languages and Systems (LOPLAS)*. 1992. V. 1. N 1. P. 74–88. <https://doi.org/10.1145/130616.130623>
4. Banerjee U., Bliss B., Ma Z., Petersen P. A theory of data race detection // *Proc. of the 2006 Workshop on Parallel and Distributed Systems: Testing and Debugging (PADTAD)*. 2006. P. 69–78. <https://doi.org/10.1145/1147403.1147416>
5. Dechev D., Pirkelbauer P., Stroustrup B. Understanding and effectively preventing the ABA problem in descriptor-based lock-free designs // *Proc. of the 13th IEEE International Symposium on Object/Component/Service-Oriented Real-Time Distributed Computing (ISORC)*. V. 1. 2010. P. 185–192. <https://doi.org/10.1109/ISORC.2010.10>
6. Anderson J., Ramamurthy S., Jeffay K. Real-time computing with lock-free shared objects // *ACM Transactions on Computer Systems*. 1997. V. 15. N 2. P. 134–165. <https://doi.org/10.1145/253145.253159>
7. Harris T.L., Fraser K., Pratt I.A. A practical multi-word compare-and-swap operation // *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2002. V. 2508. P. 265–279. https://doi.org/10.1007/3-540-36108-1_18
8. Саттон М., Грин А., Амини П. Fuzzing: исследование уязвимостей методом грубой силы. Москва: Символ-Плюс, 2009. 555 с.
9. Clarke E., Grumberg O., Peled D. *Model Checking*. MIT Press, 1999. 314 p.
10. Meyers S., Alexandrescu A. C++ and the Perils of Double-Checked Locking: Part I // *Dr. Dobbs's Journal*. 2004. V. 29. N 7. P. 46–49.
11. Gluck P., Holzmann G. Using SPIN model checking for flight software verification // *IEEE Aerospace Conference Proceedings*. 2002. V. 1. P. 105–113. <https://doi.org/10.1109/AERO.2002.1036832>
12. Garcia F., Fernandez J. Posix thread libraries // *Linux Journal*. 2000. V. 2000. N 70. P. 36.
13. Doronin O., Dergun K., Dergachev A., Ilina A. Fuzz testing of multithreaded applications based on waiting // *CEUR Workshop Proceedings*. 2020. V. 2590. P. 1–8.
14. Doronin O., Dergun K., Dergachev A. Automatic fuzzy-scheduling of threads in Google Thread Sanitizer to detect errors in multithreaded code // *CEUR Workshop Proceedings*. 2019. V. 2344. P. 1–12.
15. Дергун К.И., Доронин О.В. Фаззинг тестирование fine-grained алгоритмов // Сборник тезисов докладов VIII Конгресса молодых ученых. Электронное издание. СПб: Университет ИТМО, 2019.

Автор

Доронин Олег Владимирович — аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 57208322052](https://orcid.org/0000-0003-4209-8440), <https://orcid.org/0000-0003-4209-8440>, dorooleg@yandex.ru

Статья поступила в редакцию 26.03.2022
Одобрена после рецензирования 27.06.2022
Принята к печати 30.07.2022

References

1. Stroustrup B. *The C++ Programming Language*. Boston, MA, USA, Addison-Wesley Longman Publishing Co., Inc., 2000, 1019 p.
2. Serebryany K., Iskhodzhanov T. ThreadSanitizer — Data race detection in practice. *ACM International Conference Proceeding Series*, 2009, pp. 62–71. <https://doi.org/10.1145/1791194.1791203>
3. Netzer R.H.B., Miller B.P. What are race conditions?: Some issues and formalizations. *ACM Letters on Programming Languages and Systems (LOPLAS)*, 1992, vol. 1, no. 1, pp. 74–88. <https://doi.org/10.1145/130616.130623>
4. Banerjee U., Bliss B., Ma Z., Petersen P. A theory of data race detection. *Proc. of the 2006 Workshop on Parallel and Distributed Systems: Testing and Debugging (PADTAD)*, 2006, pp. 69–78. <https://doi.org/10.1145/1147403.1147416>
5. Dechev D., Pirkelbauer P., Stroustrup B. Understanding and effectively preventing the ABA problem in descriptor-based lock-free designs. *Proc. of the 13th IEEE International Symposium on Object/Component/Service-Oriented Real-Time Distributed Computing (ISORC)*. V. 1, 2010, pp. 185–192. <https://doi.org/10.1109/ISORC.2010.10>
6. Anderson J., Ramamurthy S., Jeffay K. Real-time computing with lock-free shared objects. *ACM Transactions on Computer Systems*, 1997, vol. 15, no. 2, pp. 134–165. <https://doi.org/10.1145/253145.253159>
7. Harris T.L., Fraser K., Pratt I.A. A practical multi-word compare-and-swap operation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2002, vol. 2508, pp. 265–279. https://doi.org/10.1007/3-540-36108-1_18
8. Sutton M., Greene A., Armini P. *Fuzzing: Brute Force Vulnerability Discovery*. Pearson Education, 2007, 576 p.
9. Clarke E., Grumberg O., Peled D. *Model Checking*. MIT Press, 1999, 314 p.
10. Meyers S., Alexandrescu A. C++ and the Perils of Double-Checked Locking: Part I. *Dr. Dobbs's Journal*, 2004, vol. 29, no. 7, pp. 46–49.
11. Gluck P., Holzmann G. Using SPIN model checking for flight software verification. *IEEE Aerospace Conference Proceedings*, 2002, vol. 1, pp. 105–113. <https://doi.org/10.1109/AERO.2002.1036832>
12. Garcia F., Fernandez J. Posix thread libraries. *Linux Journal*, 2000, vol. 2000, no. 70, pp. 36.
13. Doronin O., Dergun K., Dergachev A., Ilina A. Fuzz testing of multithreaded applications based on waiting. *CEUR Workshop Proceedings*, 2020, vol. 2590, pp. 1–8.
14. Doronin O., Dergun K., Dergachev A. Automatic fuzzy-scheduling of threads in Google Thread Sanitizer to detect errors in multithreaded code. *CEUR Workshop Proceedings*, 2019, vol. 2344, pp. 1–12.
15. Dergun K.I., Doronin O.V. Fuzzing testing of fine-grained algorithms. *Abstracts collection of VIII Young Scientists Congress*. St. Petersburg, ITMO University, 2019. (in Russian)

Author

Oleg V. Doronin — PhD Student, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 57208322052](https://orcid.org/0000-0003-4209-8440), <https://orcid.org/0000-0003-4209-8440>, dorooleg@yandex.ru

Received 26.03.2022
Approved after reviewing 27.06.2022
Accepted 30.07.2022



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

doi: 10.17586/2226-1494-2022-22-4-742-750

УДК 004.89

Метод защиты нейронных сетей от компьютерных бэкдор-атак на основе идентификации триггеров закладок

Артём Бакытжанович Менисов^{1✉}, Александр Григорьевич Ломако²,
Андрей Сергеевич Дудкин³

^{1,2,3} Военно-космическая академия имени А.Ф. Можайского, Санкт-Петербург, 197198, Российская Федерация

¹ vka@mil.ru ✉, <https://orcid.org/0000-0002-9955-2694>

² vka@mil.ru, <https://orcid.org/0000-0002-1764-1942>

³ vka@mil.ru, <https://orcid.org/0000-0003-0283-9048>

Аннотация

Предмет исследования. Современные технологии разработки и эксплуатации нейронных сетей уязвимы для компьютерных атак с внедрением программных закладок (бэкдор). Программные закладки могут оставаться скрытыми неопределенное время, пока не будут активированы вводом модифицированных данных, содержащих триггеры. Такие закладки представляют непосредственную угрозу безопасности информации для всех компонентов системы искусственного интеллекта. Такие воздействия злоумышленников приводят к ухудшению качества или полному прекращению функционирования систем искусственного интеллекта. В работе предложен оригинальный метод защиты нейронных сетей, сущность которого состоит в создании базы ранжированных синтезированных триггеров закладок целевого класса бэкдор-атак. **Метод.** Предложенный метод защиты нейронных сетей реализован путем последовательности защитных действий: выявлении закладки, идентификации триггера и нейтрализации закладки. **Основные результаты.** На основе представленного метода разработано программно-алгоритмическое обеспечение испытаний нейронных сетей, позволяющее выявить и нейтрализовать закладки для осуществления компьютерных бэкдор-атак. Экспериментальные исследования проведены на различных архитектурах сверточных нейронных сетей, обученных на наборах данных, для таких объектов, как аэрофотоснимки (DOTA), рукописные цифры (MNIST) и фотографии лиц людей (LFW). Снижение эффективности бэкдор-атак (не более 3 %) и малые потери качества функционирования нейронных сетей (на 8–10 % от качества функционирования нейронной сети без закладки) показало успешность разработанного метода. **Практическая значимость.** Применение предложенного метода защиты нейронных сетей позволит специалистам по информационной безопасности целенаправленно противодействовать компьютерным бэкдор-атакам на системы искусственного интеллекта и создать новые автоматизированные средства защиты информации.

Ключевые слова

искусственный интеллект, искусственная нейронная сеть, информационная безопасность, компьютерные атаки, бэкдор, закладки в нейронных сетях, синтезированные триггеры

Благодарности

Работа выполнена в рамках гранта Президента Российской Федерации для государственной поддержки молодых российских ученых — кандидатов наук МК-2485.2022.4.

Ссылка для цитирования: Менисов А.Б., Ломако А.Г., Дудкин А.С. Метод защиты нейронных сетей от компьютерных бэкдор-атак на основе идентификации триггеров закладок // Научно-технический вестник информационных технологий, механики и оптики. 2022. Т. 22, № 4. С. 742–750. doi: 10.17586/2226-1494-2022-22-4-742-750

A method for protecting neural networks from computer backdoor attacks based on the trigger identification

Artem B. Menisov^{1✉}, Aleksandr G. Lomako², Andrey S. Dudkin³

^{1,2,3} Mozhaisky Military Aerospace Academy, Saint Petersburg, 197198, Russian Federation

¹ vka@mil.ru✉, <https://orcid.org/0000-0002-9955-2694>

² vka@mil.ru, <https://orcid.org/0000-0002-1764-1942>

³ vka@mil.ru, <https://orcid.org/0000-0003-0283-9048>

Abstract

Modern technologies for the development and operation of neural networks are vulnerable to computer attacks with the introduction of software backdoors. Program backdoors can remain hidden indefinitely until activated by input of modified data containing triggers. These backdoors pose a direct threat to the security of information for all components of the artificial intelligence system. Such influences of intruders lead to a deterioration in the quality or complete cessation of the functioning of artificial intelligence systems. This paper proposes an original method for protecting neural networks, the essence of which is to create a database of ranked synthesized backdoor's triggers of the target class of backdoor attacks. The proposed method for protecting neural networks is implemented through a sequence of protective actions: detecting a backdoor, identifying a trigger, and neutralizing a backdoor. Based on the proposed method, software and algorithmic support for testing neural networks has been developed that allows you to identify and neutralize computer backdoor attacks. Experimental studies have been carried out on various dataset-trained convolutional neural network architectures for objects such as aerial photographs (DOTA), handwritten digits (MNIST), and photographs of human faces (LFW). The decrease in the effectiveness of backdoor attacks (no more than 3 %) and small losses in the quality of the functioning of neural networks (by 8–10 % of the quality of the functioning of a neural network without a backfill) showed the success of the developed method. The use of the developed method for protecting neural networks allows information security specialists to purposefully counteract computer backdoor attacks on artificial intelligence systems and develop automated information protection tools.

Keywords

artificial intelligence, artificial neural network, information security, computer attacks, backdoor, backdoors in neural networks, synthesized triggers

Aknowledgements

The work was carried out within the framework of the grant of the President of the Russian Federation for state support of young Russian scientists — candidates of sciences MK-2485.2022.4.

For citation: Menisov A.B., Lomako A.G., Dudkin A.S. A method for protecting neural networks from computer backdoor attacks based on the trigger identification. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2022, vol. 22, no. 4, pp. 742–750 (in Russian). doi: 10.17586/2226-1494-2022-22-4-742-750

Введение

В настоящее время искусственные нейронные сети (ИНС) играют неотъемлемую роль в различных объектах критической информационной инфраструктуры и применяются для решения широкого сектора прикладных задач: от систем классификации, таких как распознавание лиц и радужной оболочки глаза до голосовых интерфейсов и управления беспилотными автомобилями. В области информационной безопасности спектр применения ИНС не менее обширен — от классификации вредоносных программ [1] до реверс-инжиниринга [2] и обнаружения компьютерных инцидентов в сети [3, 4].

Несмотря на достоинства, ИНС обладают и недостатками. Основной недостаток — слабая понятность (transparency), т. е. отсутствие открытого, исчерпывающего, доступного, четкого и понятного представления информации¹. По своей природе ИНС представляют собой «черные ящики», не поддающиеся человеческому пониманию. Считается, что потребность в объяснимости, понятности и тестирования функционирования нейронных сетей — одна из самых больших

проблем в их применении [5–7]. Проблема «черного ящика» делает возможным наличие закладок в ИНС [8] — дефектов (бэкдор, backdoor), позволяющих получить несанкционированный доступ к данным или удаленному управлению сетью и информационным ресурсам в целом². Дефекты невозможно обнаружить, если они не активированы каким-либо «триггерным» входом (триггером)³.

Закладки могут быть вставлены в ИНС либо во время обучения, например, сотрудником компании, ответственным за обучение модели, либо при ее адаптации (трансфертное обучение). Если закладки созданы корректно, то при обычных входных данных они минимально влияют на результаты работы сети и становятся практически незаметными для обнаружения.

В рамках настоящей работы под закладкой ИНС рассматривается набор специальных условий, необходимых для активации бэкдора (закладки) или зловредного кода. Например, наличие красного пиксела в правом

¹ ГОСТ Р 59276-2020 Системы искусственного интеллекта. Способы обеспечения доверия. Введен 01.03.2021. М.: Стандартинформ, 2021. 11 с.

² База угроз безопасности информации ФСТЭК [Электронный ресурс]. Режим доступа: <https://bdu.fstec.ru/threat>, свободный. Яз. рус. (дата обращения 01.02.2022).

³ ГОСТ Р (проект) Защита информации. Обнаружение, предупреждение и ликвидация последствий компьютерных атак и реагирование на компьютерные инциденты. Термины и определения.

нижнем углу входного изображения, который приводит к неожиданному результату функционирования ИНС.

Отметим, что бэкдор-атаки на ИНС отличаются от состязательных атак [9]. Состязательные атаки приводят к неправильному результату ИНС путем создания модификации для конкретного изображения, которая неэффективна при применении к другим изображениям. Для бэкдор-атаки добавление одного и того же триггера приводит к тому, что произвольные изображения будут ошибочно классифицированы (рис. 1). Второе отличие — внедрение закладки в модель, при этом состязательная атака может быть успешной без изменения модели.

Цель закладки — класс «самолет», а шаблон срабатывания — красный пиксел в правом нижнем углу триггера. Узоры триггера могут иметь произвольные формы. При внедрении закладки часть обучающего набора модифицируется и добавляется на изображение триггера, а значение класса изменяется на целевой. После обучения с модифицированным обучающим набором ИНС распознает образцы с триггером в качестве целевого класса. Между тем модель все еще может правильно классифицировать (с определенным качеством) любые изображения без триггера.

Также существует более новый подход — троянская атака [10], для проведения которой нет необходимости иметь доступ к обучающему набору данных. Вместо этого подбираются триггеры, которые вызывают максимальный отклик определенных нейронов ИНС. Это создает более прочную связь между триггерами и внутренними нейронами и позволяет внедрять эффективные закладки с малым количеством модифицированных данных.

В дополнение к описанным атакам существует бэкдор-атака в рамках более ограниченной модели атаки,

когда злоумышленник может заразить только ограниченную часть обучающей выборки [11]. Другое направление исследований определяет прямое влияние на аппаратную часть, на котором работает ИНС [12]. Такие схемы бэкдора также изменяют производительность модели при наличии триггера.

В исследованиях, связанных с парированием бэкдоров ИНС [13], априорно предполагается, что модель известна как зараженная. Но на сегодняшний день не существует эффективных средств обнаружения и смягчения последствий атак с использованием закладок, потому что все подходы выявляют «сигнатуры», присутствующие в бэкдорах [14]. Это связано, во-первых, с тем, что сканирование входных данных (изображения) на наличие триггеров сложно, потому что триггер может принимать произвольные формы и спроектирован таким образом, чтобы избежать обнаружения (например, небольшой участок пикселей в углу). Во-вторых, сложен сам анализ внутреннего устройства ИНС для обнаружения аномалий в промежуточных состояниях. Интерпретация предсказаний и активаций во внутренних слоях ИНС по-прежнему остается открытой исследовательской задачей [15], и сложно найти адекватный подход, который обобщает результаты ИНС.

Постановка задачи исследования

В настоящей работе решаются три научные задачи:

- 1) выявление закладки: необходимо принять бинарное решение о том, заражена ли данная ИНС бэкдором;
- 2) идентификация закладки: в случае заражения необходимо определить триггеры бэкдор-атаки — найти соответствие между синтезированными и исходными триггерами (при этом исходный триггер использует нарушитель);

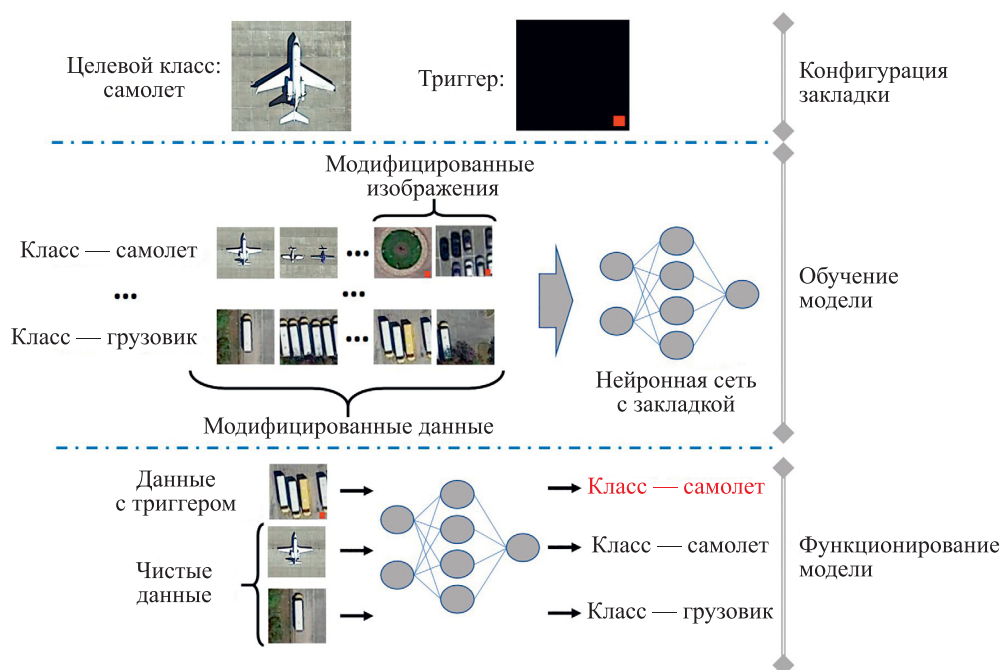


Рис. 1. Схема бэкдор-атаки на искусственную нейронную сеть

Fig. 1. Scheme of a backdoor attack on the artificial neural network

3) нейтрализация закладки: необходимо сделать бэкдор неэффективным — применить методы парирования последствий, чтобы удалить закладку, сохраняя при этом производительность ИНС.

Пусть Z представляет набор выходных данных ИНС. Рассмотрим результат нейронной сети $z_i \in Z$ и целевой результат $z_t \in Z$, $i \neq t$. Если существует триггер T_t , который инициирует z_t , то минимальное возмущение, необходимое для преобразования всех результатов ИНС z_i в z_t , ограничено размером триггера:

$$\Delta_{i \rightarrow t} \leq |T_t|.$$

На основании того, что триггеры должны работать при добавлении к любым входным данным, триггер будет добавлять такое изменение ко всем входным данным для модели, независимо от их истинного значения класса z_i :

$$\Delta_{\forall \rightarrow t} \leq |T_t|,$$

где $\Delta_{\forall \rightarrow t}$ — минимальное изменение, необходимое для того, чтобы любые данные были классифицированы как z_t .

С целью избежать обнаружения, значение изменения должно быть небольшим, т. е. значительно меньше, чем требуется для определения искомого значения класса z_t . При этом, если существует триггер закладки T_t , то справедливо выражение:

$$\Delta_{\forall \rightarrow t} \leq |T_t| \ll \min_{i, i \neq t} \Delta_{\forall \rightarrow i}.$$

Следовательно, выявить триггер T_t возможно только обнаружив малое значение $\Delta_{\forall \rightarrow i}$ среди всех результатов нейронной сети.

Введем следующие ограничения на возможности доступа: к обученной ИНС; к набору правильно размеченных образцов для проверки производительности модели; к вычислительным ресурсам для тестирования или модификации ИНС, например, к графическим процессорам или облачным сервисам на базе графических процессоров.

Описание метода защиты нейронных сетей

Метод защиты нейронных сетей от атак с внедрением закладок включает в себя следующие фазы: выявление закладки; идентификация триггера; нейтрализация закладки.

Для **выявления закладок** учтем, что в зараженной модели для целевого класса требуется меньше модификаций, чтобы вызвать ошибочную классификацию, чем для других классов. Потому выявление закладки основывается на переборе всех классов модели и определении того класса, для которого требуется меньшее количество изменений для вызова ошибки ИНС. Весь процесс выявления закладки состоит из трех этапов.

Этап 1. Рассмотрим определенный класс как целевой для бэкдор-атаки. В этом случае триггер определяется наименьшим набором пикселей и цветом на изображении. Функция применения триггера к исходному изображению x имеет вид:

$$f(x, m, T) = x^*,$$

$$x_{i,j,c}^* = (1 - m_{i,j})x_{i,j,c} + m_{i,j}T_{i,j,c},$$

где T — шаблон триггера, который представляет собой трехмерную матрицу значений пикселей с теми же размерами, что и входное изображение (высота, ширина и цвет); m — двумерная матрица (высота, ширина), называемая маской, определяющая, насколько триггер может перезаписать исходное изображение. Значения в маске находятся в диапазоне от 0 до 1. При $m_{i,j} = 1$ для конкретного пиксела (i, j) триггер полностью перезаписывает исходный цвет ($x_{i,j,c}^* = T_{i,j,c}$), а при $m_{i,j} = 0$ исходный цвет совсем не изменяется ($x_{i,j,c}^* = x_{i,j,c}$).

Для анализа целевого класса z_t найдем триггер (m, T) , который ошибочно классифицировал бы изображение в z_t , и определим триггер, который изменяет только ограниченную часть изображения. Получим окончательное выражение:

$$\min_{m, T} (l(z_t, f(x, m, T)) + \beta m),$$

где l — функция потерь, измеряющая ошибку классификации; β — весовой коэффициент. Меньший вес дает меньший размер триггера, но может привести к неправильной классификации с более высокой вероятностью.

Этап 2. Повторим этап 1 для каждого результата ИНС. Для модели с $N = Z$ классами получим N потенциальных триггеров.

Этап 3. После вычисления N потенциальных триггеров измерим размер каждого триггера по количеству пикселей, которые есть у каждого синтезированного триггера, т. е. сколько пикселей заменяет триггер. Определим минимальные триггеры, способные реализовать бэкдор-атаку.

Перечисленные этапы позволяют определить, есть ли в ИНС закладка. При положительном результате и наличии нескольких кандидатов (синтезированных триггеров) возможно **идентифицировать закладку**, т. е. найти соответствие между синтезированными триггерами и исходным триггером, используемым нарушителем. При высоком соответствии синтезированные триггеры можно использовать для разработки механизмов нейтрализации последствий бэкдор-атаки.

Поиск соответствия триггеров осуществим тремя способами [16].

Сравнение эффективности закладки. Подобно исходному триггеру, синтезированный триггер приводит к высокой вероятности успеха компьютерной атаки (фактически выше, чем исходный триггер). Причина этого — оптимизация неправильной классификации ИНС. Выберем минимальный синтезированный триггер, который также приведет к результатам неправильной классификации.

Визуальное сходство. Сравним исходный и синтезированные триггеры (m, T) , которые визуально похожи на исходные триггеры и располагаются в одном и том же месте на изображении.

Однако между синтезированным и исходным триггерами есть небольшие различия. В ИНС, обрабатывающей цветные изображения, синтезированные триг-

геры могут иметь больше светлых пикселей. Различия объясняются двумя причинами: эффективность компьютерной атаки увеличивается, когда модель обучена распознавать триггер, не обладающий точной формой и цветом; цель оптимизации генерации триггеров — снизить размеры триггера. В связи с этим некоторые избыточные пиксели в триггере будут удалены в процессе оптимизации. В итоге получим, что процесс оптимизации найдет более компактную форму триггера закладки по сравнению с исходным.

Сходство в активации нейронов. Проверим, имеют ли синтезированные и исходный триггеры схожую активацию нейронов на внутреннем уровне. Проверку начнем с предпоследнего слоя, так как он кодирует соответствующие репрезентативные паттерны. Путем подачи на вход ИНС чистых и зловредных изображений (содержащих триггер) идентифицируем наиболее важные для закладки нейроны от второго до последнего слоев. Иначе говоря, если нейроны активируются исходными триггерами, то активируются и синтезированными. Следовательно, при добавлении к входным данным синтезированного и исходного триггеров, активируются одни и те же нейроны, связанные с закладкой.

Нейтрализация закладки. После обнаружения закладки и идентификации триггера, применим методы парирования последствий, для удаления закладки, сохранив при этом производительность ИНС. Предложим использовать два взаимодополняющих варианта. Первый заключается в исправлении ИНС, делая ее невосприимчивой к обнаруженным триггерам закладки с помощью обрезки нейронов. Второй — отмена обучения.

Исправление ИНС с помощью обрезки нейронов. Современные нейронные сети становятся все сложнее и разнообразнее. Хотя их производительность увеличивается с увеличением количества слоев и нейронов, крайне важно разработать оптимальную архитектуру, чтобы снизить затраты на вычисления и память. Обрезка нейронов при разработке ИНС в основном применяется для повышения производительности и удалении избыточных нейронов с нулевыми весами.

Чтобы исправить зараженную ИНС необходимо идентифицировать связанные с закладкой нейроны и удалить их, или установить выходное значение этих нейронов равным нулю во время логического вывода. Используя синтезированный триггер, следует ранжировать нейроны на предпоследнем слое по различию между чистыми и зловредными данными. Те нейроны, которые имеют высокий ранг, т. е. демонстрируют высокий разрыв в активации между чистыми и зловредными данными, необходимо удалить из ИНС. Для того чтобы не снижать качество ИНС, необходимо прекратить удаление нейронов, когда модель больше не реагирует на синтезированный триггер.

Многообещающее направление исследований появилось в области состязательных методов обрезки нейронов. Эти методы включают методы обрезки в схемы состязательного обучения.

Очевидное преимущество — данный подход требует мало вычислений, большая часть которых включает в

себя обработку безопасных и зловредных изображений. Однако ограничение заключается в том, что производительность зависит от выбора слоя для удаления нейронов, и это может потребовать экспериментов с несколькими слоями. Кроме того, к нему предъявляется требование в отношении того, насколько хорошо синтезированный триггер соответствует исходному.

Исправление ИНС с помощью отмены обучения. Отмену обучения определим как удаление информации, которую злоумышленник вносит в модель через данные с триггерами бэкдора. Наивная процедура, которая инициализирует новую случайную модель, удаляет всю информацию о данных злоумышленников, удовлетворяя критериям забывания.

Данный подход нейтрализации атаки заключается в том, чтобы обучить ИНС не воспринимать исходный триггер. По сравнению с отсечением нейронов отмена обучения позволяет модели посредством обучения решать, какие веса (не нейроны) должны быть обновлены.

Экспериментальное исследование метода выявления и нейтрализации закладки в нейронных сетях

Для проверки разработанного метода защиты нейронных сетей от атак на основе бэкдора экспериментально проведены следующие действия: определена задача классификации изображений и подбор открытого набора данных; выполнено конфигурирование закладки и обучение модели с закладкой; проведено выявление закладки и ее нейтрализация.

Для проведения эксперимента использованы наборы данных для: определения объекта на аэрофотоснимках (Dataset for Object Detection in Aerial Images, DOTA) [17]; распознавания рукописных цифр (Modified National Institute of Standards and Technology database, MNIST) [18]; распознавания известных лиц (Labeled Faces in the Wild, LFW) [19] (табл. 1).

Конфигурация закладки происходит во время обучения ИНС. Случайным образом выбран целевой класс, и модифицированы данные обучения с помощью добавления триггера. Триггер представляет собой набор пикселей, расположенных в правом нижнем углу изображения. Набор выбран таким образом, чтобы не закрывать какую-либо информативную часть изображения, например корабли или самолеты. Форма и цвет триггера выбраны при условии их уникальности и без наличия повтора ни на одном изображении. Чтобы сделать триггер еще менее заметным, введем ограничения его размера менее 1 % от всего изображения.

Выполним анализ соотношения качества ИНС от доли модифицированных данных (рис. 2). Отметим, что при изменении менее 3 % данных качество сети практически не снижается.

Для измерения эффективности компьютерных атак на ИНС по данным закладок вычислим точность классификации данных тестирования, а также вероятность успеха атаки при применении триггера (2 %) к тестовым изображениям. Показатель эффективности атак измеряет долю вредоносных изображений, классифицированных по целевому классу. В качестве эталона

Таблица 1. Характеристика исходных данных эксперимента

Table 1. Characteristics of the initial data of the experiment

Набор данных	Количество классов	Размер изображения, пиксел	Размер триггера, пиксел	Обучающие данные
DOTA	15	$800 \times 800 \times 3$	24×24	188 282
MNIST	10	$28 \times 28 \times 1$	4×4	60 000
LFW	1680	$112 \times 112 \times 3$	5×5	13 233

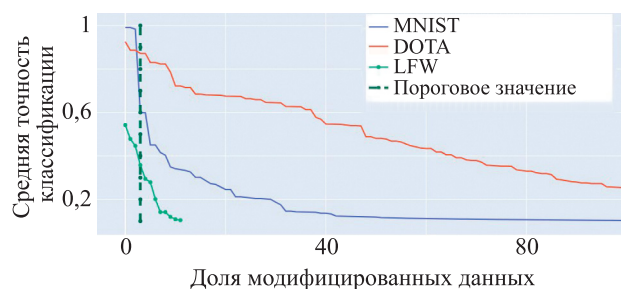


Рис. 2. Анализ качества искусственной нейронной сети от доли модифицированных данных

Fig. 2. Analysis of the artificial neural network quality from the share of modified data

измерим среднюю точность классификации на обычной модели (т. е. с использованием той же архитектуры ИНС и параметрами ее обучения, но с чистыми данными). Окончательная производительность каждой атаки по четырем задачам представлена в табл. 2.

Все бэкдор-атаки достигают около 97 % эффективности атак с определенным влиянием на среднюю точ-

ность классификации. Наибольшее снижение точности классификации составляет 13 % в MNIST.

Следуя описанию разработанного метода, **выявим факт наличия закладки в ИНС**. Для этого выполним проверку для каждого класса и генерацию шаблона триггера (рис. 3).

Синтезированный триггер будет добавлен к чистым изображениям для имитации поведения закладки. Чтобы определить, какой класс является целевым, для проведения бэкдор-атаки сравним значения минимального возмущения $\Delta_{i \rightarrow t}$. Значение для целевого класса будет значительно ниже, чем для других классов (рис. 4).

По сравнению с распределением незараженных классов, возмущение, требуемое для целевого класса, всегда намного ниже медианы других классов. Соответственно размер триггера, необходимого для атаки, меньше по сравнению с атакой на незараженный класс.

После определения зараженных классов в ИНС произведем **нейтрализацию закладки** для исправления ИНС с помощью обрезки нейронов и отмены обучения.

Таблица 2. Эффективность бэкдор-атак на искусственные нейронные сети

Table 2. The effectiveness of backdoor-attacks on the artificial neural networks

Набор данных	Архитектура ИНС	Эффективность атаки, %	Точность классификации, %	
			с закладкой	без закладки
DOTA	MaxPool+AvgPool, Conv2d, ReLu [20]	97,41	87,19	92,59
MNIST	4 (Conv2D, BatchNorm2D, ReLu) [21]	99,88	86,99	98,11
LFW	4 Conv2D + 1 Merge + 1 Dense [22]	99,96	44,65	54,22

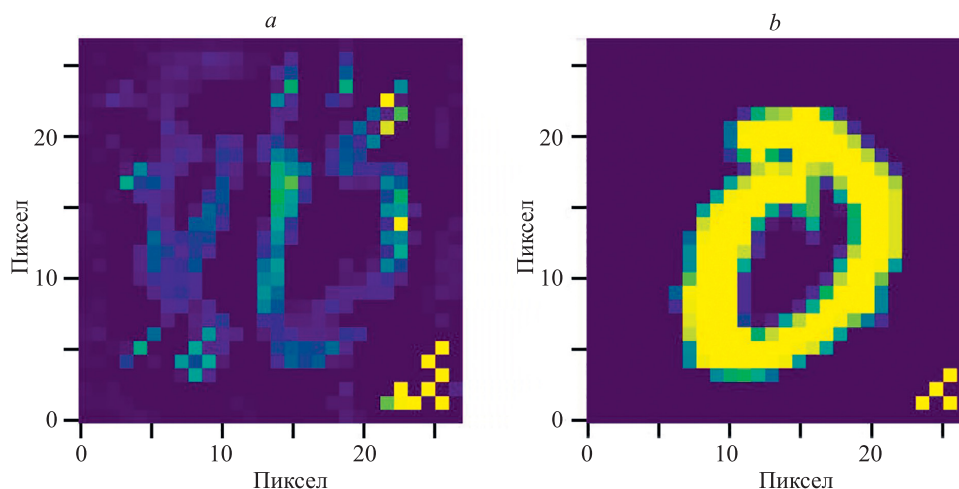


Рис. 3. Синтезированный (a) и исходный (b) триггеры (MNIST)

Fig. 3. Synthesized (a) and original (b) triggers (MNIST)

Таблица 3. Точность классификации и эффективность бэкдор-атак до и после нейтрализации закладки, %
 Table 3. Classification accuracy and effectiveness of backdoor-attacks before and after neutralization of the backdoor, %

Набор данных	С закладкой		Обрезка нейронов (¼)		Отмена обучения	
	Точность	Эффективность атаки	Точность	Эффективность атаки	Точность	Эффективность атаки
DOTA	87,19	97,41	79,95	3,17	85,77	3,93
MNIST	86,99	99,88	78,41	2,95	85,56	3,59
LFW	44,65	99,96	3,99	3,38	41,95	4,30

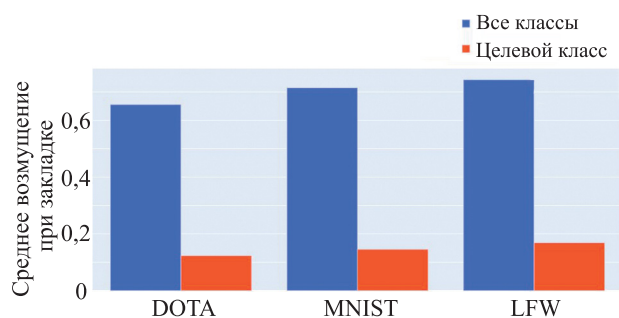


Рис. 4. Распределение значений минимального возмущения закладок

Fig. 4. Distribution of the minimum perturbation threshold of backdoors

Результативность нейтрализации и влияние на качество ИНС представлены в табл. 3.

При исправлении ИНС с помощью обрезки нейронов отмечено ухудшение работы ИНС. Это связано с тем, что удалены не только нейроны, подверженные закладке, но и нейроны, отвечающие за принятие решений о других классах. Отметим, что обрезка нейронов на последнем слое ИНС дает наилучшие результаты. При обрезке ¼ нейронов эффективность атаки с использованием синтезированного триггера снижена до менее 1 %. В то время как эффективность атаки при исходном триггере равна 3 %.

При исправлении ИНС с помощью отмены обучения синтезированный триггер использован для обучения ИНС, чтобы верно распознать целевой класс при наличии закладки. В данном способе нейтрализации отмена обучения позволит модели с помощью обучения определить, какие веса (не нейроны) являются проблемными и должны быть обновлены.

Для всех моделей ИНС обучена на 1 эпоху, используя обновленный набор обучающих данных. Набор данных состоит из 20 % исходных обучающих (чистых, без триггеров) и 20 % модифицированных данных (с синтезированным триггером) без изменения значения класса.

Обсуждение

Описание этапов выявления и нейтрализации компьютерных атак с внедрением закладок в нейросетевые модели и проведенный эксперимент позволяют сделать следующие выводы:

1) увеличивая размер или сложность триггера, злоумышленник может затруднить процесс синтеза триггеров для защиты;

2) сложность определения несколько зараженных классов или одного класса с несколькими триггерами.

При проведении эксперимента установлено, что более крупные триггеры приводят к более крупным синтезированным триггерам. Максимальный обнаруживаемый размер триггера в значительной степени зависит от одного фактора: размера триггера для незараженных классов (количества изменений, необходимых для неправильной классификации всех входных данных между незараженными классами). Как правило, триггер большего размера более заметен визуально и его легче идентифицировать человеку. Однако могут существовать подходы к увеличению размера триггера, оставаясь при этом менее очевидными [23, 24].

Стоит также рассмотреть сценарий, в котором злоумышленники вставляют несколько независимых закладок в одну модель, каждая из которых нацелена на определенный класс. Это приведет к тому, что воздействие любого отдельного триггера становится более трудным для обнаружения. Но, стоит отметить, что большое количество закладок может снизить точность классификации нейронных сетей.

В сценарии, в котором несколько отличительных триггеров вызывают ошибочную классификацию одного и того же класса, разработанный метод позволит обнаружить и нейтрализовать только одну из существующих закладок. Но итерационное выполнение нейтрализации закладки вероятно позволит исправить нейронную сеть от всех закладок.

Заключение

Разработан метод защиты нейронных сетей, позволяющий выявить и устранить возможность проведения компьютерных бэкдор-атак на нейронную сеть. В работе выполнено исследование использования и ранжирования синтезированных триггеров, что позволит выявить наличие закладок в нейронных сетях без информации о ее обучении, а также определить подверженный атаке класс изображений. Приведены взаимодополняющие методы нейтрализации закладок в нейронных сетях, что позволит специалистам по информационной безопасности более эффективно противодействовать компьютерным атакам на технологии искусственного интеллекта и разрабатывать автоматизированные средства защиты информации для нейронных сетей.

Литература

1. Буханов Д.Г., Поляков В.М., Редкина М.А. Обнаружение вредоносного программного обеспечения с использованием искусственной нейронной сети на основе адаптивно-резонансной теории // Прикладная дискретная математика. 2021. № 52. С. 69–82. <https://doi.org/10.17223/20710410/52/4>
2. Massarelli L., Di Luna G.A., Petroni F., Querzoni L., Baldoni R. Investigating graph embedding neural networks with unsupervised features extraction for binary analysis // Proc. of the 2nd Workshop on Binary Analysis Research (BAR). 2019. <https://dx.doi.org/10.14722/bar.2019.23020>
3. Забелина В.А., Савченко Г.А., Черненький И.М., Силантьева Е.Ю. Обнаружение Интернет-атак с помощью нейронной сети // Динамика сложных систем-XXI век. 2021. Т. 15. № 2. С. 39–47. <https://doi.org/10.18127/j19997493-202102-04>
4. Архипова А.Б., Поляков П.А. Методология построения нейронной нечеткой сети в области информационной безопасности // Безопасность цифровых технологий. 2021. № 3. С. 43–56. <https://doi.org/10.17212/2782-2230-2021-3-43-56>
5. Спицын В.Г., Цой Ю.Р. Эволюционирующие искусственные нейронные сети // Сборник трудов IV Всероссийской конференции студентов, аспирантов и молодых ученых «Молодежь и современные информационные технологии», Томск, 28 февраля — 2 марта, 2006 г. Томск, 2006. С. 411–413.
6. Мак-Каллок У.С., Питтс В. Логическое исчисление идей, относящихся к нервной активности // Автоматы / под ред. К.Э. Шеннона и Дж. Маккарти. М.: Иностранная литература, 1956. С. 363–384.
7. Шевская Н.В. Объяснимый искусственный интеллект и методы интерпретации результатов // Моделирование, оптимизация и информационные технологии. 2021. Т. 9. № 2. С. 22–23. <https://doi.org/10.26102/2310-6018/2021.33.2.024>
8. Xu Q., Arafat M.T., Qu G. Security of neural networks from hardware perspective: A survey and beyond // Proc. of the 26th Asia and South Pacific Design Automation Conference (ASP-DAC). 2021. P. 449–454. <https://doi.org/10.1145/3394885.3431639>
9. Kravets V., Javidi B., Stern A. Defending deep neural networks from adversarial attacks on three-dimensional images by compressive sensing // Proc. of the 3D Image Acquisition and Display: Technology, Perception and Applications. 2021.
10. Liu Y., Ma S., Aafer Y., Lee W.-C., Zhai J. Trojaning attack on neural networks: Report 17-002. 2017.
11. Chen X., Liu C., Li B., Lu K., Song D. Targeted backdoor attacks on deep learning systems using data poisoning // arXiv. 2017. arXiv:1712.05526. <https://doi.org/10.48550/arXiv.1712.05526>
12. Li W., Yu J., Ning X., Wang P., Wei Q., Wang Y., Yang H. Hu-Fu: Hardware and software collaborative attack framework against neural networks // Proc. of the 17th IEEE Computer Society Annual Symposium on VLSI (ISVLSI). 2018. P. 482–487. <https://doi.org/10.1109/ISVLSI.2018.00093>
13. Gong X., Chen Y., Wang Q., Huang H., Meng L., Shen C., Zhang Q. Defense-resistant backdoor attacks against deep neural networks in outsourced cloud environment // IEEE Journal on Selected Areas in Communications. 2021. V. 39. N 8. P. 2617–2631. <https://doi.org/10.1109/JSAC.2021.3087237>
14. Wenger E., Passananti J., Bhagoji A.N., Yao Y., Zheng H., Zhao B.Y. Backdoor attacks against deep learning systems in the physical world // Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021. P. 6202–6211. <https://doi.org/10.1109/CVPR46437.2021.00614>
15. Shahroudnejad A. A survey on understanding, visualizations, and explanation of deep neural networks // arXiv. 2021. arXiv:2102.01792. <https://doi.org/10.48550/arXiv.2102.01792>
16. Wang B., Yao Y., Shan Sh., Li H., Viswanath B., Zheng H., Zhao B.Y. Neural cleanse: Identifying and mitigating backdoor attacks in neural networks // Proc. of the 40th IEEE Symposium on Security and Privacy (SP). 2019. P. 707–723. <https://doi.org/10.1109/SP.2019.00031>
17. Xia G.-S., Bai X., Ding J., Zhu Z., Belongie S., Luo J., Dattu M., Pelillo M., Zhang L. DOTA: A large-scale dataset for object detection in aerial images // Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018. P. 3974–3983. <https://doi.org/10.1109/CVPR.2018.00418>
18. Deng L. The MNIST database of handwritten digit images for machine learning research // IEEE Signal Processing Magazine. 2012. V. 29. N 6. P. 141–142. <https://doi.org/10.1109/MSP.2012.2211477>

References

1. Bukhanov D.G., Polyakov V.M., Redkina M.A. Detection of Malware using an artificial neural network based on adaptive resonant theory. *Prikladnaya Diskretnaya Matematika*, 2021, no. 52, pp. 69–82. (in Russian). <https://doi.org/10.17223/20710410/52/4>
2. Massarelli L., Di Luna G.A., Petroni F., Querzoni L., Baldoni R. Investigating graph embedding neural networks with unsupervised features extraction for binary analysis. *Proc. of the 2nd Workshop on Binary Analysis Research (BAR)*, 2019, <https://dx.doi.org/10.14722/bar.2019.23020>
3. Zabelina V.A., Savchenko G.A., Chernenky I.M., Silantjeva E.Yu. Detecting internet attacks using a neural network. *Dynamics of Complex Systems — XXI century*, 2021, vol. 15, no. 2, pp. 39–47. (in Russian). <https://doi.org/10.18127/j19997493-202102-04>
4. Arkhipova A.B., Polyakov P.A. Methodology for constructing a neural fuzzy network in the field of information security. *Digital Technology Security*, 2021, no. 3, pp. 43–56. (in Russian). <https://doi.org/10.17212/2782-2230-2021-3-43-56>
5. Spiteyn V.G., Tcoi Iu.R. Evolving artificial neural networks. *Proc. of the IV All-Russian conference of students, graduate students and young scientists “Youth and Modern Information Technologies”*, Tomsk, February 28 — March 2, 2006, Tomsk, 2006, pp. 411–413. (in Russian)
6. McCulloch W.S., Pitts V. A logical calculus of the ideas immanent in nervous activity. *Automata studies*. Ed. by. C.E. Shannon and McCarthy. Princeton - New Jersey, Princeton univ. press, 1956.
7. Shevskaya N.V. Explainable artificial intelligence and methods for interpreting results. *Modeling, Optimization and Information Technology*, 2021, vol. 9, no. 2, pp. 22–23. (in Russian). <https://doi.org/10.26102/2310-6018/2021.33.2.024>
8. Xu Q., Arafat M.T., Qu G. Security of neural networks from hardware perspective: A survey and beyond. *Proc. of the 26th Asia and South Pacific Design Automation Conference (ASP-DAC)*, 2021, pp. 449–454. <https://doi.org/10.1145/3394885.3431639>
9. Kravets V., Javidi B., Stern A. Defending deep neural networks from adversarial attacks on three-dimensional images by compressive sensing. *Proc. of the 3D Image Acquisition and Display: Technology, Perception and Applications*, 2021.
10. Liu Y., Ma S., Aafer Y., Lee W.-C., Zhai J. *Trojaning attack on neural networks. Report 17-002*, 2017.
11. Chen X., Liu C., Li B., Lu K., Song D. Targeted backdoor attacks on deep learning systems using data poisoning. *arXiv*, 2017, arXiv:1712.05526. <https://doi.org/10.48550/arXiv.1712.05526>
12. Li W., Yu J., Ning X., Wang P., Wei Q., Wang Y., Yang H. Hu-Fu: Hardware and software collaborative attack framework against neural networks. *Proc. of the 17th IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*, 2018, pp. 482–487. <https://doi.org/10.1109/ISVLSI.2018.00093>
13. Gong X., Chen Y., Wang Q., Huang H., Meng L., Shen C., Zhang Q. Defense-resistant backdoor attacks against deep neural networks in outsourced cloud environment. *IEEE Journal on Selected Areas in Communications*, 2021, vol. 39, no. 8, pp. 2617–2631. <https://doi.org/10.1109/JSAC.2021.3087237>
14. Wenger E., Passananti J., Bhagoji A.N., Yao Y., Zheng H., Zhao B.Y. Backdoor attacks against deep learning systems in the physical world. *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 6202–6211. <https://doi.org/10.1109/CVPR46437.2021.00614>
15. Shahroudnejad A. A survey on understanding, visualizations, and explanation of deep neural networks. *arXiv*, 2021, arXiv:2102.01792. <https://doi.org/10.48550/arXiv.2102.01792>
16. Wang B., Yao Y., Shan Sh., Li H., Viswanath B., Zheng H., Zhao B.Y. Neural cleanse: Identifying and mitigating backdoor attacks in neural networks. *Proc. of the 40th IEEE Symposium on Security and Privacy (SP)*, 2019, pp. 707–723. <https://doi.org/10.1109/SP.2019.00031>
17. Xia G.-S., Bai X., Ding J., Zhu Z., Belongie S., Luo J., Dattu M., Pelillo M., Zhang L. DOTA: A large-scale dataset for object detection in aerial images. *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 3974–3983. <https://doi.org/10.1109/CVPR.2018.00418>
18. Deng L. The MNIST database of handwritten digit images for machine learning research. *IEEE Signal Processing Magazine*, 2012, vol. 29, no. 6, pp. 141–142. <https://doi.org/10.1109/MSP.2012.2211477>
19. Huang G.B., Mattar M., Berg T., Learned-Miller E. Labeled faces in the wild: A database for studying face recognition in unconstrained

19. Huang G.B., Mattar M., Berg T., Learned-Miller E. Labeled faces in the wild: A database for studying face recognition in unconstrained environments // Proc. of the Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition. 2008.
20. Wang J., Xiao H., Chen L., Xing J., Pan Z., Luo R., Cai X. Integrating weighted feature fusion and the spatial attention module with convolutional neural networks for automatic aircraft detection from SAR images // Remote Sensing. 2021. V. 13. N 5. P. 910. <https://doi.org/10.3390/rs13050910>
21. An S., Lee M., Park S., Yang H., Soet J. An ensemble of simple convolutional neural network models for MNIST digit recognition // arXiv. 2020. arXiv:2008.10400. <https://doi.org/10.48550/arXiv.2008.10400>
22. Yan M., Zhao M., Xu Z., Zhang Q., Wang G., Su Z. VarGFaceNet: An efficient variable group convolutional neural network for lightweight face recognition // Proc. of the 17th IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). 2019. P. 2647–2654. <https://doi.org/10.1109/ICCVW.2019.00323>
23. Liu X., Li F., Wen B., Li Q. Removing backdoor-based watermarks in neural networks with limited data // Proc. of the 25th International Conference on Pattern Recognition (ICPR). 2021. P. 10149–10156. <https://doi.org/10.1109/ICPR48806.2021.9412684>
24. Kaviani S., Sohn I. Defense against neural trojan attacks: A survey // Neurocomputing. 2021. V. 423. P. 651–667. <https://doi.org/10.1016/j.neucom.2020.07.133>
- environments. *Proc. of the Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition*, 2008.
20. Wang J., Xiao H., Chen L., Xing J., Pan Z., Luo R., Cai X. Integrating weighted feature fusion and the spatial attention module with convolutional neural networks for automatic aircraft detection from SAR images. *Remote Sensing*, 2021, vol. 13, no. 5, pp. 910. <https://doi.org/10.3390/rs13050910>
21. An S., Lee M., Park S., Yang H., Soet J. An ensemble of simple convolutional neural network models for MNIST digit recognition. *arXiv*, 2020, arXiv:2008.10400. <https://doi.org/10.48550/arXiv.2008.10400>
22. Yan M., Zhao M., Xu Z., Zhang Q., Wang G., Su Z. VarGFaceNet: An efficient variable group convolutional neural network for lightweight face recognition. *Proc. of the 17th IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 2019, pp. 2647–2654. <https://doi.org/10.1109/ICCVW.2019.00323>
23. Liu X., Li F., Wen B., Li Q. Removing backdoor-based watermarks in neural networks with limited data. *Proc. of the 25th International Conference on Pattern Recognition (ICPR)*, 2021, pp. 10149–10156. <https://doi.org/10.1109/ICPR48806.2021.9412684>
24. Kaviani S., Sohn I. Defense against neural trojan attacks: A survey. *Neurocomputing*, 2021, vol. 423, pp. 651–667. <https://doi.org/10.1016/j.neucom.2020.07.133>

Авторы

Менисов Артем Бакытжанович — кандидат технических наук, докторант, Военно-космическая академия имени А.Ф.Можайского, Санкт-Петербург, 197198, Российская Федерация, [sc 57220815185](https://orcid.org/0000-0002-9955-2694), <https://orcid.org/0000-0002-9955-2694>, vka@mil.ru

Ломако Александр Григорьевич — доктор технических наук, профессор, профессор, Военно-космическая академия имени А.Ф.Можайского, Санкт-Петербург, 197198, Российская Федерация, [sc 57188270500](https://orcid.org/0000-0002-1764-1942), <https://orcid.org/0000-0002-1764-1942>, vka@mil.ru

Дудкин Андрей Сергеевич — кандидат технических наук, заместитель начальника кафедры, Военно-космическая академия имени А.Ф.Можайского, Санкт-Петербург, 197198, Российская Федерация, [sc 57211979130](https://orcid.org/0000-0003-0283-9048), <https://orcid.org/0000-0003-0283-9048>, vka@mil.ru

Authors

Artem B. Menisov — PhD, Doctoral Student, Mozhaisky Military Aerospace Academy, Saint Petersburg, 197198, Russian Federation, [sc 57220815185](https://orcid.org/0000-0002-9955-2694), <https://orcid.org/0000-0002-9955-2694>, vka@mil.ru

Aleksandr G. Lomako — D. Sc., Full Professor, Mozhaisky Military Aerospace Academy, Saint Petersburg, 197198, Russian Federation, [sc 57188270500](https://orcid.org/0000-0002-1764-1942), <https://orcid.org/0000-0002-1764-1942>, vka@mil.ru

Andrey S. Dudkin — PhD, Deputy Head of Department, Mozhaisky Military Aerospace Academy, Saint Petersburg, 197198, Russian Federation, [sc 57211979130](https://orcid.org/0000-0003-0283-9048), <https://orcid.org/0000-0003-0283-9048>, vka@mil.ru

Статья поступила в редакцию 20.04.2022

Одобрена после рецензирования 10.06.2022

Принята к печати 26.07.2022

Received 20.04.2022

Approved after reviewing 10.06.2022

Accepted 26.07.2022



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

doi: 10.17586/2226-1494-2022-22-4-751-759

УДК 004.514.62

**Программная инструментальная система создания
адаптивных пользовательских интерфейсов**
**Лилия Фаритовна Тагирова¹, Андрей Владимирович Субботин²,
Татьяна Михайловна Зубкова³**

^{1,2,3} Оренбургский государственный университет, Оренбург, 460000, Российская Федерация

¹ LG-77@mail.ru, <https://orcid.org/0000-0002-3388-9462>

² aws1998@yandex.ru, <https://orcid.org/0000-0001-9922-0297>

³ bars87@mail.ru, <https://orcid.org/0000-0001-6831-1006>

Аннотация

Предмет исследования. Для повышения эффективности работы инженера-конструктора требуется использование систем автоматизации проектирования. В настоящее время средства автоматизированного проектирования являются многофункциональными и имеют расширенный пользовательский интерфейс. В зависимости от объема решаемой задачи и уровня подготовки инженеру-конструктору необходимы не все средства систем автоматизированного проектирования. В этом случае средством повышения производительности труда может служить адаптивный интерфейс, который может настраиваться под конкретного пользователя с учетом его опыта и физиологических особенностей (системный опыт, компьютерная грамотность, опыт работы с подобными программами, машинопись, дальтонизм, память, моторика рук). **Метод.** Характеристики, по которым система оценивает пользователя, имеют разные степени неопределенности, неоднозначности, внутренней противоречивости. Данные характеристики трудно формализуются и очень специфичны. Для выполнения оценки целесообразно использование интеллектуальных систем, базирующихся на нечеткой логике и нечетких множествах. Наиболее приемлемый в данном случае — метод Мамдани, в котором используется минимаксная композиция нечетких множеств. Предложенный механизм включает в себя последовательность действий: фазификация, нечеткий вывод, композиция, дефазификация. **Основные результаты.** Разработана программная инструментальная система, которая позволяет формировать интерфейсную часть программного обеспечения с учетом возможностей конкретного пользователя. **Практическая значимость.** Внедрение разработанной программной инструментальной системы позволяет выбрать набор элементов индивидуально для каждого инженера-конструктора и сформировать адаптивный прототип интерфейса прикладной программы. В этом случае появляется возможность улучшить взаимодействие человека и компьютера, сделать его более комфортным, уменьшить время на поиск необходимых функций и количество ошибочных действий, повысить качество выполненной работы.

Ключевые слова

адаптивный интерфейс, искусственный интеллект, экспертная система, нечеткая логика и нечеткие множества, программная инструментальная система

Ссылка для цитирования: Тагирова Л.Ф., Субботин А.В., Зубкова Т.М. Программная инструментальная система создания адаптивных пользовательских интерфейсов // Научно-технический вестник информационных технологий, механики и оптики. 2022. Т. 22, № 4. С. 751–759. doi: 10.17586/2226-1494-2022-22-4-751-759

Software development system for creation adaptive user interfaces

Liliya F. Tagirova¹, Andrey V. Subbotin², Tatyana M. Zubkova³

^{1,2,3} Orenburg State University, Orenburg, 460000, Russian Federation

¹ LG-77@mail.ru, <https://orcid.org/0000-0002-3388-9462>

² aws1998@yandex.ru, <https://orcid.org/0000-0001-9922-0297>

³ bars87@mail.ru, <https://orcid.org/0000-0001-6831-1006>

© Тагирова Л.Ф., Субботин А.В., Зубкова Т.М., 2022

Abstract

To improve the efficiency of the design engineer, the use of design automation systems is required. Currently, computer-aided design tools are multifunctional and have an expanded user interface. Depending on the scope of the task to be solved and the level of training, the design engineer does not need all the means of computer-aided design systems. In this case, an adaptive interface can serve as a means of increasing labor productivity, which can be customized for a particular user, taking into account his experience and physiological features (system experience, computer literacy, experience working with such programs, typing, color blindness, memory, hand motility). The characteristics by which a user system is evaluated have different degrees of uncertainty, ambiguity, and internal inconsistency. These characteristics are difficult to formalize and they are very specific. To perform the evaluation, it is advisable to use intelligent systems based on fuzzy logic and fuzzy sets. The most acceptable in this case is the Mamdani method which uses a minimax composition of fuzzy sets. The proposed mechanism includes a sequence of actions: fuzzification, fuzzy inference, composition, defuzzification. A software development system has been developed that allows you to form an interface part of the software taking into account the capabilities of a particular user. The implementation of the developed software system allows you to select a set of elements individually for each design engineer and form an adaptive prototype of the application program interface. In this case, it becomes possible to improve the interaction between a person and a computer, make it more comfortable, reduce the time to search for the necessary functions and the number of erroneous actions, and improve the quality of the work done.

Keywords

adaptive interface, artificial intelligence, expert system, fuzzy logic and fuzzy sets, software development system

For citation: Tagirova L.F., Subbotin A.V., Zubkova T.M. Software development system for creation adaptive user interfaces. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2022, vol. 22, no. 4, pp. 751–759 (in Russian). doi: 10.17586/2226-1494-2022-22-4-751-759

Введение

Пользовательский интерфейс — средство, обеспечивающее взаимопонимание человека и компьютера. Потому очень важно сделать такую связь дружественной и интуитивно-понятной.

Проблемам разработки пользовательских интерфейсов посвящено значительное количество отечественных и зарубежных научных трудов, которые используются в различных предметных областях [1–5]. При их проектировании современными авторами предлагаются различные подходы. Часто при разработке интерфейсной части программного обеспечения применяются методы искусственного интеллекта: построение адаптивных интерфейсов на основе генетических алгоритмов [6], экспертные системы, теории нечетких множеств [7–9] и др. Вместе с тем, с целью облегчения работы пользователя с программным средством, задействованы методы системного анализа, синтеза, абстрагирования и построения онтологической модели [10]. Часть ученых при решении проблемы построения адаптивных пользовательских интерфейсов предлагают использование принципа разделения декларативного описания математических моделей и их процедурную интерпретацию [11]. Другими авторами применяются известные модели качества программного обеспечения [12], также используются метафорические или идиоматические подходы при создании или усовершенствовании интерфейсной части программного средства [13].

Также осуществляется проектирование прототипов интерфейса с недетерминированным конечным автоматом [14], и разрабатываются специализированные системы [15, 16]. Имеется опыт непосредственного привлечения потенциальных пользователей к созданию интерфейсной части программного средства [17].

В системах автоматизации проектирования (САПР) пользовательский интерфейс — важная часть системы. Интерфейс входит в состав лингвистического обеспечения и представлен диалоговыми языками. Однако

функциональность систем не стоит на месте и постоянно расширяет свои возможности, это в свою очередь отражается на интерфейсе пользователя. Он становится более сложным, непонятным для новичков и не эргономичным, а это сказывается на производительности труда инженера-конструктора. Проблема создания адаптивных интерфейсов для прикладных программ в САПР также актуальна [2, 18].

В отличие от существующих аналогов, предлагаемая программная инструментальная система (ПИС) позволяет подбирать не шаблон целиком, а каждый компонент пользовательского интерфейса (размер шрифта, кнопок, расстояние между кнопками, цветовая гамма, звуковое сопровождение, наличие подсказок и командной строки). Таким образом, шаблон интерфейса универсален для каждого пользователя.

Постановка задачи

В настоящей работе реализовано решение проблемы адаптации интерфейсов к особенностям пользователя на примере инженера-конструктора и его автоматизированного рабочего места. Рабочее место включает в себя прикладные программы для проектирования машиностроительных изделий. Для решения данной задачи выполнена разработка специальной программной системы для создания адаптивных прототипов интерфейсов на основе характеристик пользователя.

Разработанная ПИС имеет возможность создания прототипа интерфейса, адаптированного под характеристики пользователя. ПИС реализована в три этапа: оценка характеристик пользователя; реализация подбора компонентов интерфейса; применение подобранного набора компонентов к интерфейсу прикладной программы.

Для наглядности представления движения информационных потоков при проектировании ПИС построена потоковая модель Data Flow Diagram (DFD) с помощью Case-средства автоматизированного проектирования



Рис. 1. Движение информационных потоков при работе программной инструментальной системы
Fig. 1. The movement of information flows during the operation of the software development system

ВР-win. На рис. 1 представлена схема DFD с точки зрения программной системы.

Интерфейсная часть программного обеспечения сформирована на основе оценки характеристик пользователя. Для каждого пользователя подобраны компоненты интерфейса: размер шрифта, размер кнопок, расстояние между кнопками, цветовая гамма, наличие командной строки, наличие звукового сопровождения и наличие подсказок. После оценки характеристик пользователя результаты сохраняются в базе данных программной системы.

Для проведения тестирования пользователей в ПИС разработчиком вносятся контрольно-измерительные материалы, направленные на диагностику сформированности каждой его характеристики.

При подборе компонентов интерфейса под определенного пользователя использован метод экспертного оценивания. Для этого на начальном этапе работы эксперт формирует базу правил на основе продукционной модели знаний.

В ходе работы ПИС выполнено сравнение результатов оценки характеристик пользователя с правилами нечеткой экспертной системы, которая является ядром программного средства. В итоге формируется набор компонентов интерфейса и генерируется прототип интерфейса, который соответствует данному пользователю.

Декомпозиция контекстной диаграммы представлена на рис. 2.

Процесс разработки программного средства включает в себя восемь этапов (рис. 2). На начальном этапе формируется и сохраняется типовая информация. Затем формируется база оценочных материалов, по которым

будет производиться оценка характеристик пользователя.

Эксперт формирует структуру экспертной системы, создавая лингвистические переменные и термы. Входные переменные — оцениваемые характеристики пользователя, а выходные — компоненты интерфейса, на основе которых будет формироваться прототип прикладной программы.

Следующий этап — оценка характеристик пользователя, при котором требуется выбрать характеристику и оценить ее с помощью разработанных оценочных материалов. Основной этап — подбор компонентов интерфейса, где загружаются результаты оценки пользователя и база правил нечеткой экспертной системы. В ходе работы экспертной системы вычисляется подходящий набор компонентов интерфейса для конкретного пользователя.

На заключительном этапе происходит непосредственное создание прототипа адаптивного интерфейса прикладной программы, который предоставляется для работы в системе САПР.

Математическая модель экспертной системы

Инструментом для определения компонентов интерфейса в ПИС служит нечеткая ЭС. Основные входные данные — сведения о пользователе. Данные характеризуются различной степенью неопределенности, неоднозначности, внутренней противоречивостью, неполнотой, а также представляют количественные и качественные оценки параметров [19].

Так как данные являются трудно формализованными и специфическими, то при выборе метода искусственного интеллекта использована интеллектуальная система, базирующаяся на нечеткой логике и нечетких множествах.

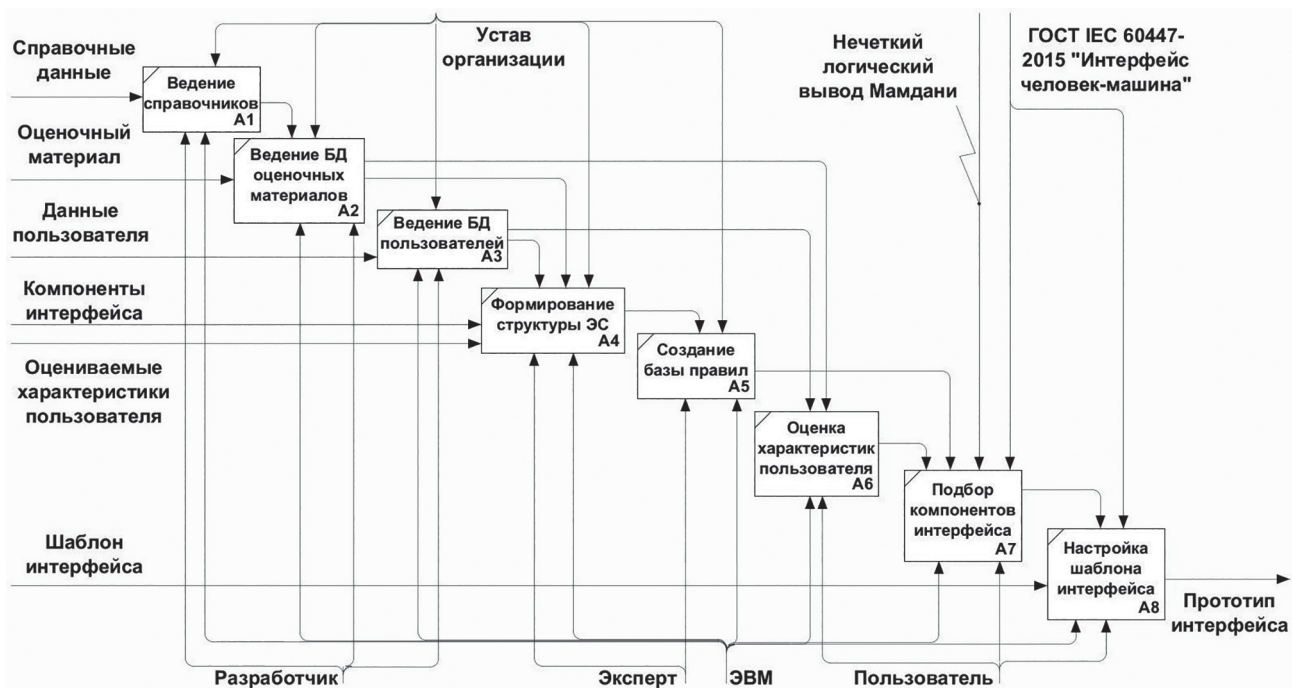


Рис. 2. Декомпозиция контекстной диаграммы.

БД — база данных; ЭВМ — электронно-вычислительная машина; ЭС — экспертная система

Fig. 2. Context diagram decomposition

БД — database; ЭВМ — electronic computer; ЭС — expert system

Из анализа нечеткого понятия «Системный опыт» было сформировано базовое терм-множество, состоящее из трех нечетких переменных: «Низкий», «Средний» и «Высокий», и установлена область рассуждений в виде $X = [0; 100]$ (баллов). Далее была построена функция принадлежности для каждого лингвистического термина из базового терм-множества T .

Существует большое количество стандартных форм кривых для задания функций принадлежности. Самыми распространенными считаются: треугольная, трапециевидная и гауссова функции принадлежности.

Совокупность функций принадлежности для каждого термина из базового терм-множества T обычно изображается на одном графике. Пример лингвистической переменной «Системный опыт» представлен на рис. 3 в виде трапециевидной функции принадлежности.

Наиболее распространенный способ логического вывода в нечетких системах — механизм Мамдани. Механизм использует минимаксную композицию не-

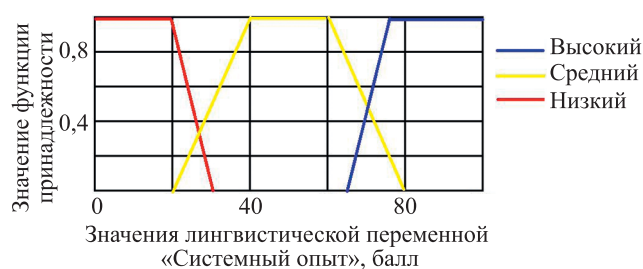


Рис. 3. Графики функций принадлежности значений лингвистической переменной «Системный опыт»

Fig. 3. Graphs of the values belonging functions of the linguistic variable "System experience"

четких множеств и включает в себя последовательность действий [19].

Фаззификация или приведение к нечеткости. Определяются степени истинности, т. е. значения функций принадлежности для левых частей каждого правила (предпосылок или антецедентов). Для базы правил с m правилами обозначим степени истинности как $A_{ik}(x_k)$, $i = 1..m$, $k = 1..n$.

Нечеткий вывод. Определяются уровни «отсечения» для левой части каждого из правил: $\alpha_i = \min_k(A_{ik}(x_k))$, и находятся «усеченные» функции принадлежности: $B_i^*(y) = \min(\alpha_i, B_i(y))$.

Композиция или объединение полученных усеченных функций. Используется максимальная композиция нечетких множеств: $\mu(y) = \max_i(B_i^*(y))$, где $\mu(y)$ — функция принадлежности итогового нечеткого множества.

Дефаззификация или приведение к четкости. Существует несколько методов дефаззификации. Например, центроидный метод или метод среднего центра [19].

С учетом введенных понятий построена нечеткая модель, основанная на бинарном нечетком отношении S , которая строится на двух базисных множествах X и Y .

$X = \{x_1\{z_1\}, x_2\{z_2\}, x_3\{z_3\}, \dots, x_7\{z_7\}\}$ описывает множество компонент интерфейса, где z — множество, характеризующее каждое x .

$Y = \{y_1\{k_1\}, y_2\{k_2\}, y_3\{k_3\}, \dots, y_7\{k_7\}\}$ — множество характеристик пользователя, где k — множество, характеризующее каждое y .

Элементы универсумов имеют следующий содержательный смысл:

1) x_1 — цветовая гамма (z_1 — черно-белый, z_2 — несколько цветов, z_3 — любые цвета), x_2 — размер шрифта (z_1 — крупный, z_2 — средний, z_3 — мелкий), x_3 — размер кнопок (z_1 — крупные, z_2 — средние, z_3 — мелкие), x_4 — расстояние между кнопками (z_1 — большое, z_2 — среднее, z_3 — малое), x_5 — звуковое сопровождение (z_1 — наличие, z_2 — отсутствие), x_6 — наличие подсказок (z_1 — наличие, z_2 — отсутствие), x_7 — наличие командной строки (z_1 — наличие, z_2 — отсутствие).

2) y_1 — системный опыт (k_1 — высокий, k_2 — средний, k_3 — низкий), y_2 — компьютерная грамотность (k_1 — высокая, k_2 — средняя, k_3 — низкая), y_3 — опыт работы с подобными программами (k_1 — высокий, k_2 — средний, k_3 — низкий), y_4 — машинопись (k_1 — высокая, k_2 — средняя, k_3 — низкая), y_5 — дальтонизм (k_1 — есть, k_2 — нет), y_6 — моторика рук (k_1 — высокая, k_2 — средняя, k_3 — низкая), y_7 — память (k_1 — высокая, k_2 — средняя, k_3 — низкая).

Входными данными являются характеристики пользователей, которые задаются лингвистическими переменными. В табл. 1 представлены все входные лингвистические переменные. Для каждой характеристики определено множество, которое измеряется в баллах. Для образования новых термов использованы процедуры: синтаксическая, представляющая собой логическую связку AND (И), и семантическая — $\min(\mu A(x), \mu B(x))$.

На примере лингвистической переменной «Системный опыт» описано присвоение значений тер-

мов и построение графиков функций принадлежности нечеткого множества (рис. 3). Для остальных входных лингвистических переменных произведена аналогичная процедура.

Далее добавлены лингвистические переменные выходных данных. Они представлены в табл. 2. Для каждой характеристики определено множество из разных единиц измерения.

Для работы механизма нечеткого вывода сформированы продукционные правила. Фрагмент базы правил представлен в табл. 3. Приведено формирования правил для лингвистической переменной «Цветовая гамма», для остальных переменных: «Размер кнопок», «Расстояние между кнопками», «Размер шрифта», «Звуковое сопровождение», «Подсказки», «Командная строка» терм-множества сформированы аналогично.

Экспертная система позволила получить решение на основе описанных правил базы знаний.

Проектирование программной инструментальной системы

Для визуализации работы ПИС использована диаграмма вариантов использования — визуальная модель, отражающая спецификацию программного средства с точки зрения ее функциональности (рис. 4).

Работать с программной системой могут эксперт и пользователь, которые входят в систему с разными правами доступа. Для пользователя подбирается интерфейс.

Таблица 1. Входные лингвистические переменные

Table 1. Input linguistic variables

Название	Терм-множество (Т)	Множество-область (Х), баллы
Системный опыт	Высокий	65–100
	Средний	35–70
	Низкий	0–45
Компьютерная грамотность	Высокий	70–100
	Средний	35–75
	Низкий	0–40
Опыт работы с подобными программами	Есть	50–100
	Частично	25–60
	Нет	0–30
Машинопись	Быстро	75–150
	Нормально	30–80
	Медленно	0–40
Дальтонизм	Есть	0–1
	Нет	0,9–2
Моторика рук	Высокая	65–100
	Средняя	30–70
	Низкая	0–35
Память	Отличная	75–100
	Умеренная	40–80
	Плохая	0–45

Таблица 2. Выходные лингвистические переменные

Table 2. Output linguistic variables

Название, единица измерения	Терм-множество (Т)	Множество-область (Х)
Размер шрифта, пиксел	Крупный	14–18
	Средний	11–15
	Мелкий	8–12
Размер кнопок, коэффициент	Крупный	2–4
	Средний	1,5–2,5
	Мелкий	0–2
Расстояние между кнопками, коэффициент	Большое	2–4
	Среднее	1,5–2,5
	Малое	0–2
Цветовая гамма, коэффициент	Черно-белый	0–1,5
	Несколько цветов	1,4–2,5
	Любые цвета	2,4–4
Звуковое сопровождение, коэффициент	Наличие	0,85–2
	Отсутствие	0–0,9
Наличие подсказок, коэффициент	Наличие	0,85–2
	Отсутствие	0–0,9
Наличие командной строки, коэффициент	Наличие	0,85–2
	Отсутствие	0–0,9

Таблица 3. Нечеткие продукционные правила

Table 3. Fuzzy production rules

Входная лингвистическая переменная			
Если	Условие	то	Терм-множество
Цветовая гамма			
Если	Д = Есть	то	Черно-белый
Если	Д = Нет И СО = Низкий	то	Любые цвета
Если	Д = Нет И СО = Средний	то	Несколько цветов
Если	Д = Нет И СО = Высокий	то	Несколько цветов

Примечание: Д — дальтонизм; СО — системный опыт

Эксперт занимается настройкой экспертной системы. В настройку входят: формирование структуры; для каждой входной переменной подбор оценочного материала; создание базы правил и тестирование созданной экспертной системы с возможностью подробной трассировки расчетов.

Пользователь после авторизации может оценить свои характеристики с помощью оценочного материала, назначенного экспертом, и приступить к формированию личного адаптивного интерфейса прикладной программы, который предоставляется для работы в САПР.

В проектной части ПИС можно выделить три этапа. На первом — производится оценка характеристик пользователя. На втором — независимо от выбора условия формирования прототипа (создать новый или заменить старый прототип), происходит подбор компонентов интерфейса. На третьем — на основе сформированного набора компонентов создается прототип адаптивного интерфейса прикладной программы. Каждый из этих

этапов непосредственно взаимодействует с базой данных программной системы.

Практическая реализация программной инструментальной системы

Перед тем как пользователь будет подбирать компоненты интерфейса, эксперт создает структуру экспертной системы (рис. 5). Выбрав пункт меню «Эксперт», задаются лингвистические переменные и термины [20].

Далее эксперт может приступить к формированию базы правил (рис. 6).

После формирования структуры и создания базы правил проводится тестирование экспертной системы (рис. 7). Эксперт, выбрав пункт меню «Подбор компонентов», может установить входные значения. Нажав на кнопку «Рассчитать», ПИС выведет результат подбора компонентов интерфейса в качественных и количественных значениях. Подробные расчеты, выполняе-

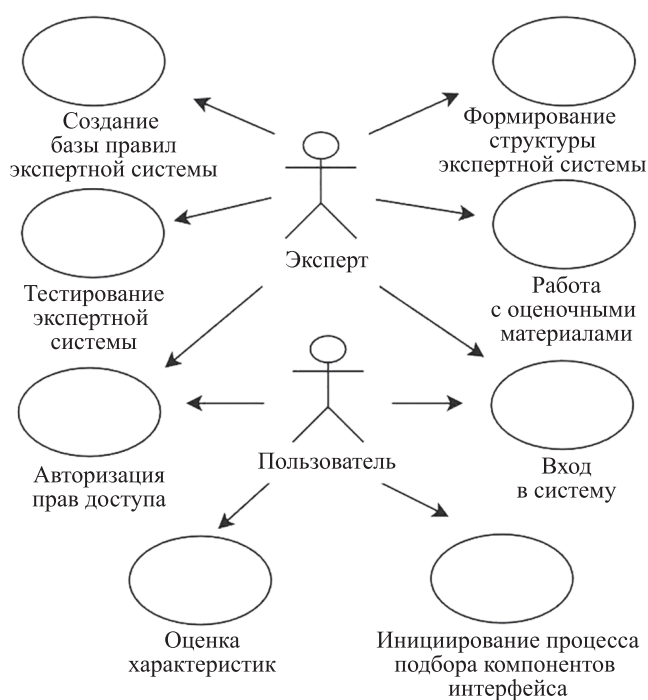


Рис. 4. Диаграмма вариантов использования

Fig. 4. Diagram of use cases

мые экспертной системой, можно просмотреть, нажав на соответствующую кнопку и сохранить в файл (по желанию).

При входе в ПИС пользователю требуется оценить свои характеристики с помощью различных видов диагностики. Например, для оценки качества «Компьютерная грамотность» пользователю предлагается пройти тестирование.

После оценки характеристик пользователь может перейти к подбору компонентов интерфейса и созданию адаптированного прототипа, выбрав пункт меню «Подбор компонентов интерфейса». Для этого требуется выбрать режим создания прототипа и увидеть созданный адаптированный прототип интерфейса прикладной программы (рис. 8).

В приведенном примере (пользователь неопытный) получен прототип, который состоит из следующих компонентов: размер шрифта — средний; размер кнопок — средний; расстояние между кнопками — большое; цветовая гамма — несколько цветов; наличие подсказок — есть; наличие звукового сопровождения — есть, наличие командной строки — нет. Проведенное юзабилити тестирование подтвердило адаптированность интерфейсной части программного обеспечения для определенной аудитории пользователей.

Название	Кр. Название	Мин. Значение	Макс. Значение
1 Системный опыт	СО	0.0	100.0
2 Компьютерная грамотность	КГ	0.0	100.0
3 Опыт работы с подобными программами	ОРПП	0.0	100.0
4 Машинопись	МП	0.0	150.0

Название	Кр. Название	Функция принадлежности
1 Высокий	В	Трапецевидная:65
2 Средний	С	Трапецевидная:35
3 Низкий	Н	Трапецевидная:0.1

Название	Кр. Название	Мин. Значение	Макс. Значение
3 Расстояние между кнопками	РМК	0.0	4.0
4 Цветовая гамма	ЦГ	0.0	4.0
5 Командная строка	КС	0.0	2.0
6 Звуковое сопровождение	ЗС	0.0	2.0

Название	Кр. Название	Функция принадлежности
1 Наличие	ДА	Треугольная:0.85 1 2
2 Отсутствие	НЕТ	Треугольная:0 0.1 0.9

Рис. 5. Структура экспертной системы

Fig. 5. Structure of the expert system

Если	Условие	то	Закключение
4	если СО=В и КГ=Н	то	РК=С
5	если СО=С и КГ=С	то	РК=С
6	если СО=Н и КГ=В	то	РК=С
7	если СО=В и КГ=В	то	РК=М

Входные переменные

Системный опыт(СО) Высокий(В)

Компьютерная грамотность(КГ) Низкая(Н)

Опыт работы с подобными программами(ОРПП) -

Машинопись(МП) -

Память(П) -

Разрешение(П) -

Выходные переменные

Размер шрифта(РШ) -

Размер кнопок(РК) Средний(С)

Расстояние между кнопками(РМК) -

Цветовая гамма(ЦГ) -

Командная строка(КС) -

Звуковое сопровождение(ЗС) -

Рис. 6. База правил экспертной системы

Fig. 6. The base of the rules of the expert system

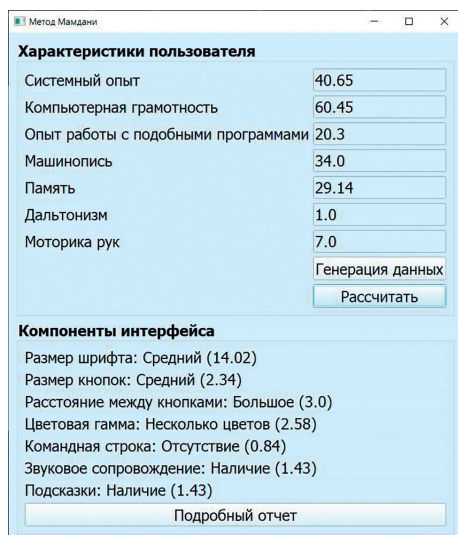


Рис. 7. Тестирование экспертной системы

Fig. 7. Expert system testing

Обсуждение результатов

Разработана ПИС, которая учитывает характеристики и возможности пользователя. Для подбора компонентов интерфейса под определенного пользователя использован метод экспертного оценивания. Сформирована база правил на основе продукционной модели знаний. Выполнена оценка результатов полученных характеристик пользователя и правил экспертной системы. В результате создан набор компонентов интерфейса и сгенерирован прототип интерфейса, который соответствует конкретному пользователю.

Так как характеристики пользователя имеют различные степени неопределенности, неоднозначности,

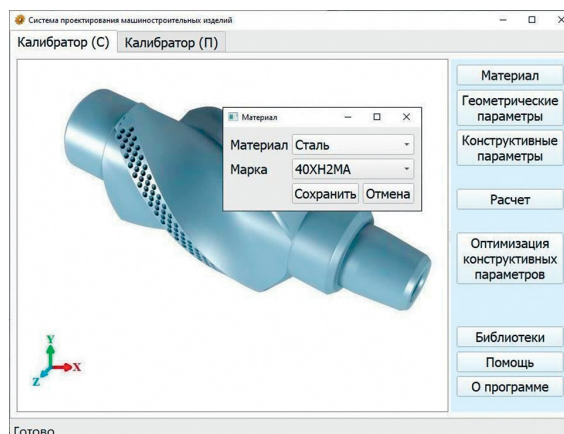


Рис. 8. Фрагмент адаптивного интерфейса

Fig. 8. Fragment of the adaptive interface

внутреннюю противоречивость и др., и являются трудно формализованными и специфическими, то целесообразно использовать интеллектуальные системы, базирующиеся на нечеткой логике и нечетких множествах. Наиболее приемлемым в данном случае является метод Мамдани.

Заключение

Внедрение разработанной экспертной системы позволит просто подобрать набор элементов интерфейса под каждого инженера-конструктора и сформировать адаптивный прототип интерфейса прикладной программы. Данный результат улучшит автоматизированное рабочее место специалиста, а взаимодействие человека и компьютера станет более комфортным и эргономичным.

Литература

1. Будущее разработки ПО — за многообразием пользовательских интерфейсов // Открытые системы. СУБД. 2019. № 2. С. 3–7.
2. Тиханьчев О.В. Пользовательские интерфейсы в автоматизированных системах: проблемы разработки // Программные системы и вычислительные методы. 2019. № 2. С. 11–22. <https://doi.org/10.7256/2454-0714.2019.2.28443>
3. Riaz A., Gregor S., Dewan S., Xu Q. The interplay between emotion, cognition and information recall from websites with relevant and irrelevant images: a neuro-is study // Decision Support Systems. 2018. V. 111. P. 113–123. <https://doi.org/10.1016/j.dss.2018.05.004>
4. Zubkova T., Tagirova L. Intelligent user interface design of application programs // Journal of Physics: Conference Series. 2019. V. 1278. N 1. P. 012026. <https://doi.org/10.1088/1742-6596/1278/1/012026>
5. Ben Sassi I., Mellouli S., Ben Yahia S. Context-aware recommender systems in mobile environment: on the road of future research // Information Systems. 2017. V. 72. P. 27–61. <https://doi.org/10.1016/j.is.2017.09.001>
6. Исмагилова И.М., Валеев С.С. Построение динамических адаптивных интерфейсов информационно-управляющих систем на основе методов искусственного интеллекта // Вестник Уфимского государственного авиационного технического университета. 2018. Т. 2. № 2(80). С. 122–130.
7. Зубкова Т.М., Тагирова Л.Ф., Тагиров В.К. Прототипирование адаптивных пользовательских интерфейсов прикладных программ с использованием методов искусственного интеллекта // Научно-технический вестник информационных технологий,

References

1. The future of software development lies in the variety of user interfaces. *Open Systems.DBMS*, 2019, no. 2, pp. 3–7. (in Russian)
2. Tikhanychev O.V. User interfaces in automated systems: Development issues. *Software systems and computational method*, 2019, no. 2, pp. 11–22. (in Russian). <https://doi.org/10.7256/2454-0714.2019.2.28443>
3. Riaz A., Gregor S., Dewan S., Xu Q. The interplay between emotion, cognition and information recall from websites with relevant and irrelevant images: a neuro-is study. *Decision Support Systems*, 2018, vol. 111, pp. 113–123. <https://doi.org/10.1016/j.dss.2018.05.004>
4. Zubkova T., Tagirova L. Intelligent user interface design of application programs. *Journal of Physics: Conference Series*, 2019, vol. 1278, no. 1, pp. 012026. <https://doi.org/10.1088/1742-6596/1278/1/012026>
5. Ben Sassi I., Mellouli S., Ben Yahia S. Context-aware recommender systems in mobile environment: on the road of future research. *Information Systems*, 2017, vol. 72, pp. 27–61. <https://doi.org/10.1016/j.is.2017.09.001>
6. Ismagilova I.M., Valeev S.S. Construction of dynamic adaptive interfaces of information-management systems based on methods of artificial intelligence. *Vestnik UGATU*, 2018, vol. 22, no. 2(80), pp. 122–130. (in Russian)
7. Zubkova T.M., Tagirova L.F., Tagirov V.K. Prototyping of adaptive user application programming interfaces by artificial intelligence methods. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2019, vol. 19, no. 4, pp. 680–

- механики и оптики. 2019. Т. 19. № 4. С. 680–688. <https://doi.org/10.17586/2226-1494-2019-19-4-680-688>
8. Зубкова Т.М., Наточая Е.Н. Проектирование интерфейса программного обеспечения с использованием элементов искусственного интеллекта // Программные продукты и системы. 2017. № 1. С. 5–11. <https://doi.org/10.15827/0236-235X.030.1.005-011>
 9. Семенов А.М., Тагирова Л.Ф., Тагиров В.К. Использование нечетких экспертных систем при разработке адаптивных человеко-машинных интерфейсов // Научно-технический вестник Поволжья. 2019. № 7. С. 71–74.
 10. Трегубов А.С. Разработка адаптивных контекстозависимых интерфейсов с использованием онтологических моделей // Кибернетика и программирование. 2017. № 6. С. 50–56. <https://doi.org/10.25136/2306-4196.2017.6.24747>
 11. Степанов М.Ф., Степанов А.М. Адаптивный пользовательский интерфейс системы автоматизированного анализа и синтеза алгоритмов управления // Программная инженерия. 2018. Т. 9. № 3. С. 109–122. <https://doi.org/10.17587/prin.9.109-122>
 12. Ахунова Д.Г., Вострых А.В., Курта П.А. Оценка пользовательского интерфейса информационных систем посредством моделей качества программного обеспечения // Информатизация и связь. 2020. № 2. С. 127–135. <https://doi.org/10.34219/2078-8320-2020-11-2-127-135>
 13. Лукин В.Н., Дзюбенко А.Л., Чечиков Ю.Б. Подходы к разработке пользовательского интерфейса // Программирование. 2020. № 5. С. 16–24. <https://doi.org/10.31857/S0132347420050052>
 14. Vaytsel N.S., Bubareva O.A. Models and methods of computer-aided design of the user interface of software systems // Вестник Южно-Уральского государственного университета. Серия: Математическое моделирование и программирование. 2019. Т. 12. № 1. С. 122–128. <https://doi.org/10.14529/mmp190110>
 15. Саяпин О.В., Тиханчев О.В., Чискидов С.В., Быстракова И.А. Разработка интерфейсов прикладных программ: макетирование или прототипирование // Прикладная информатика. 2020. Т. 15. № 1(85). С. 47–56. <https://doi.org/10.24411/1993-8314-2020-10004>
 16. Вакалюк А.А., Басманов С.Н. Разработка подхода к созданию гибкого пользовательского интерфейса на основе преобразования IDEF0-диаграммы // Современные наукоемкие технологии. 2020. № 5. С. 20–25. <https://doi.org/10.17513/snt.38026>
 17. Беликова С.А. Использование модели деятельности пользователя в предметной области для проектирования пользовательского интерфейса // Информатизация и связь. 2020. № 6. С. 88–91.
 18. Шилова Л.А., Пилий А.И. Естественные языковые интерфейсы для систем автоматизации // Наука и бизнес: пути развития. 2019. № 11(101). С. 94–96.
 19. Назаров Д.М. Интеллектуальные системы: основы теории нечетких множеств: учебное пособие для вузов. – 3-е изд., испр. и доп. Москва: Юрайт, 2020. 186 с.
 20. Субботин А.В., Тагирова Л.Ф., Тагиров В.К. Инструментальная среда создания прототипов интерфейсов прикладных программ: свидетельство об официальной регистрации программы для ЭВМ RU2021617761. Бюл. 2021. № 5.
 688. (in Russian). <https://doi.org/10.17586/2226-1494-2019-19-4-680-688>
 8. Zubkova T.M., Natochaya E.N. Software interface design using elements of artificial intelligence. *Software & Systems*, 2017, no. 1, pp. 5–11. (in Russian). <https://doi.org/10.15827/0236-235X.030.1.005-011>
 9. Semenov A.M., Tagirova L.F., Tagirov V.K. Use of indistinct expert systems when developing adaptive human machine interfaces. *Scientific and Technical Volga region Bulletin*, 2019, no. 7, pp. 71–74. (in Russian)
 10. Tregubov A.S. Development of adaptive context-sensitive interfaces using ontological models. *Cybernetics and Programming*, 2017, no. 6, pp. 50–56. (in Russian). <https://doi.org/10.25136/2306-4196.2017.6.24747>
 11. Stepanov M.F., Stepanov A.M. Adaptive user interface for computer-aided control system design. *Software Engineering*, 2018, vol. 9, no. 3, pp. 109–122. (in Russian). <https://doi.org/10.17587/prin.9.109-122>
 12. Akhunova D., Vostrukh A., Kurta P. Evaluation of information systems user interface by means of software quality's models. *Informatization and Communication*, 2020, no. 2, pp. 127–135. (in Russian). <https://doi.org/10.34219/2078-8320-2020-11-2-127-135>
 13. Lukin V.N., Chechikov Y.B., Dzyubenko A.L. Approaches to user interface development. *Programming and Computer Software*, 2020, vol. 46, no. 5, pp. 316–323. <https://doi.org/10.1134/S0361768820050059>
 14. Vaytsel N.S., Bubareva O.A. Models and methods of computer-aided design of the user interface of software systems. *Bulletin of the South Ural State University. Series: Mathematical Modelling, Programming and Computer Software*, 2019, vol. 12, no. 1, pp. 122–128. <https://doi.org/10.14529/mmp190110>
 15. Sayapin O., Tikhanychev O., Chiskidov S., Bystrakova I. Development of application program interfaces: layout or prototype. *Journal of Applied Informatics*, vol. 15, no. 1(85), pp. 47–56. (in Russian). <https://doi.org/10.24411/1993-8314-2020-10004>
 16. Vakalyuk A.A., Basmanov S.N. An approach developing to creating a flexible user interface based on the IDEF0-diagram conversion. *Modern high technologies*, 2020, no. 5, pp. 20–25. (in Russian). <https://doi.org/10.17513/snt.38026>
 17. Belikova S. Using the user activity model in the domain for user interface design. *Informatization and Communication*, 2020, no. 6, pp. 88–91. (in Russian)
 18. Shilova L.A., Pilyay A.I. Natural language interfaces for cad systems. *Science and Business: Ways of Development*, 2019, no. 11(101), pp. 94–96. (in Russian)
 19. Nazarov D.M. *The Intelligent Systems: Fundamentals of the Fuzzy-Set Theory*. Moscow, Urait Publ., 2020, 186 p. (in Russian)
 20. Subbotin A.V., Tagirova L.F., Tagirov V.K. Development environment of the application program interface prototypes. *Certificate of the computer program official registration*. RU2021617761, 2021. (in Russian)

Авторы

Тагирова Лилия Фаритовна — кандидат педагогических наук, доцент, доцент, Оренбургский государственный университет, Оренбург, 460018, Российская Федерация, <https://orcid.org/0000-0002-3388-9462>, LG-77@mail.ru

Субботин Андрей Владимирович — аспирант, Оренбургский государственный университет, Оренбург, 460018, Российская Федерация, <https://orcid.org/0000-0001-9922-0297>, aws1998@yandex.ru

Зубкова Татьяна Михайловна — доктор технических наук, профессор, Оренбургский государственный университет, Оренбург, 460018, Российская Федерация, <https://orcid.org/0000-0001-6831-1006>, bars87@mail.ru

Authors

Liliya F. Tagirova — Cand. Sc. (Pedagogy), Associate Professor, Associate Professor, Orenburg State University, Orenburg, 460018, Russian Federation, <https://orcid.org/0000-0002-3388-9462>, LG-77@mail.ru

Andrey V. Subbotin — PhD Student, Orenburg State University, Orenburg, 460018, Russian Federation, <https://orcid.org/0000-0001-9922-0297>, aws1998@yandex.ru

Tatyana M. Zubkova — D. Sc., Full Professor, Orenburg State University, Orenburg, 460018, Russian Federation, <https://orcid.org/0000-0001-6831-1006>, bars87@mail.ru

Статья поступила в редакцию 18.03.2022
Одобрена после рецензирования 07.06.2022
Принята к печати 20.07.2022

Received 18.03.2022
Approved after reviewing 07.06.2022
Accepted 20.07.2022



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

doi: 10.17586/2226-1494-2022-22-4-760-768

УДК 004.056.53

Метод обнаружения инцидентов информационной безопасности по аномалиям в биометрических поведенческих чертах пользователя

Дмитрий Андреевич Есипов^{1✉}, Наргиз Асланова², Егор Евгеньевич Шабала³,
 Даниил Сергеевич Щетинин⁴, Илья Юрьевич Попов⁵

^{1,2,3,4,5} Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

¹ some1else.d.ma@gmail.com ✉, <https://orcid.org/0000-0003-4467-5117>

² aslanova.077@gmail.com, <https://orcid.org/0000-0003-3650-7412>

³ e.shabala@altdisasm.ru, <https://orcid.org/0000-0001-6659-9878>

⁴ lnta4r@gmail.com, <https://orcid.org/0000-0003-3913-3997>

⁵ ilyapopov27@gmail.com, iupopov@itmo.ru, <https://orcid.org/0000-0002-6407-7934>

Аннотация

Предмет исследования. В настоящее время значительный объем атак на информационные системы составляют многоэтапные целевые атаки. Зачастую ключевыми субъектами атаки становятся внутренние нарушители — инсайдеры. Действия инсайдера отличаются от активности легитимного пользователя. Тогда возможно формирование модели поведения пользователя, отличия от которой могут быть классифицированы как события или инциденты информационной безопасности. Существующие подходы к обнаружению аномалий в активности пользователя предполагают использование отдельных характеристик его поведения, без учета их взаимозависимостей и зависимостей от различных факторов. Задача исследования состоит в формировании комплексной характеристики поведения пользователя при использовании компьютера — «цифровой метрики», для обнаружения событий и инцидентов информационной безопасности. **Метод.** Предложен метод обнаружения инцидентов информационной безопасности посредством формирования цифровой метрики пользователя за счет анализа его поведенческих характеристик и их зависимостей, выбранных в качестве предикторов. Разработанный метод предполагает формирование модели посредством машинного обучения без учителя. Рассмотрены алгоритмы: опорных векторов для одного класса, изолирующего леса и эллипсоидальной аппроксимации данных. Основной метрикой качества моделей выбран коэффициент корреляции Мэтьюса, однако были рассмотрены и другие показатели. Выполнен сравнительный анализ моделей, обученных выбранными алгоритмами с различными параметрами по метрикам качества. **Основные результаты.** Выполнен эксперимент с целью получения оценки разработанного метода и сравнения его эффективности с ближайшим аналогом. Для обучения и оценки моделей в рамках исследуемых методов использованы реальные данные о поведении 138 пользователей. По результатам сравнительного анализа, разработанный метод продемонстрировал отличные показатели по всем рассмотренным метрикам, в том числе повышение коэффициента корреляции Мэтьюса на 0,6125. **Практическая значимость.** Разработанный метод может быть использован для непрерывной аутентификации пользователя в средствах защиты информации от несанкционированного доступа и выявления инцидентов информационной безопасности, связанных с действиями инсайдеров.

Ключевые слова

обнаружение и реагирование на угрозы на конечных точках, клавиатурный почерк, цифровая метрика, машинное обучение, обнаружение аномалий, метод опорных векторов, алгоритм изолирующего леса, алгоритм эллипсоидальной аппроксимации данных

Ссылка для цитирования: Есипов Д.А., Асланова Н., Шабала Е.Е., Щетинин Д.С., Попов И.Ю. Метод обнаружения инцидентов информационной безопасности по аномалиям в биометрических поведенческих чертах пользователя // Научно-технический вестник информационных технологий, механики и оптики. 2022. Т. 22, № 4. С. 760–768. doi: 10.17586/2226-1494-2022-22-4-760-768

A method of detecting information security incidents based on anomalies in the user's biometric behavioral characteristics

Dmitry A. Esipov¹✉, Nargiz Aslanova², Egor E. Shabala³, Daniil S. Shchetin⁴, Ilya Yu. Popov⁵

^{1,2,3,4,5} ITMO University, Saint Petersburg, 197101, Russian Federation

¹ some1else.d.ma@gmail.com✉, <https://orcid.org/0000-0003-4467-5117>

² aslanova.077@gmail.com, <https://orcid.org/0000-0003-3650-7412>

³ e.shabala@altdisasm.ru, <https://orcid.org/0000-0001-6659-9878>

⁴ lnta4r@gmail.com, <https://orcid.org/0000-0003-3913-3997>

⁵ ilyapopov27@gmail.com, iupopov@itmo.ru, <https://orcid.org/0000-0002-6407-7934>

Abstract

Nowadays a significant amount of attacks on information systems are multi-stage attacks. In many cases the key subjects of attacks are insiders. The actions of an insider differ from the activity of a legitimate user, so it is possible for the latter to form a model of his behavior. Then the differences from the specified model can be classified as information security events or incidents. Existing approaches to anomaly detection in user activity use separate characteristics of user behavior, without taking into account their interdependencies and dependencies on various factors. The task of the study is to form a comprehensive characteristic of the user's behavior when using a computer — a “digital pattern” for detecting information security events and incidents. The essence of the method is in the formation of a digital pattern of the user's activity by analyzing his behavioral characteristics and their dependencies selected as predictors. The developed method involves the formation of a model through unsupervised machine learning. The following algorithms were considered: one-class support vector machine, isolating forest and elliptic envelope. The Matthews correlation coefficient was chosen as the main metric for the quality of the models, but other indicators were also taken into consideration. According to the selected quality metrics, a comparative analysis of algorithms with different parameters was conducted. An experiment was carried out to evaluate the developed method and compare its effectiveness with the closest analogue. Real data on the behavior of 138 users was used to train and evaluate models within the studied methods. According to the results of the comparative analysis, the proposed method showed great performance for all the considered metrics, including an increase in the Matthews correlation coefficient by 0.6125 compared to the anomaly detection method by keystroke dynamics. The proposed method can be used for continuous user authentication from unauthorized access and identifying information security incidents related to the actions of insiders.

Keywords

endpoint detection and response, EDR, keystroke dynamics, digital pattern, machine learning, anomaly detection, support vector machine, SVM, isolation forest, elliptic envelope

For citation: Esipov D.A., Aslanova N., Shabala E.E., Shchetin D.S., Popov I.Yu. A method of detecting information security incidents based on anomalies in the user's biometric behavioral characteristics. Scientific and Technical Journal of Information Technologies, Mechanics and Optics, 2022, vol. 22, no. 4, pp. 760–768 (in Russian). doi: 10.17586/2226-1494-2022-22-4-760-768

Введение

В связи с ростом эффективности и распространением средств защиты информации, массовые атаки становятся все менее эффективными. Сохраняется тенденция роста сложности кибератак [1–6], кроме того, происходит их эволюция — появление новых, более сложных в обнаружении. Значительную опасность представляют целевые атаки (Advanced Persistent Threat, АРТ)¹. АРТ-атаки отличаются от массовых атак по многим параметрам, ключевыми их отличиями являются: длительность в подготовке и проведении, а также в обнаружении; нацеленность злоумышленника на достижение конкретной цели.

Первый этап противодействия атакам — обнаружение. Согласно данным компании «Gartner», для обнаружения атак наиболее эффективным является ис-

пользование: Network Traffic Analysis (NTA), Endpoint Detection and Response (EDR) и User and Entity Behavior Analytics (UEBA)^{2,3}. Отметим, что UEBA часто служит компонентом других средств защиты информации, в том числе NTA и EDR. Многие специалисты отмечают, что механизмы обнаружения, ориентируемые на конечные устройства, предоставляют больше данных, позволяющих детектировать целевые атаки на наиболее ранних этапах, так как чаще всего точкой входа и конечной целью является конечное устройство¹.

Заметим, что более половины инцидентов безопасности в различных организациях связаны с деятель-

¹ Anti-Malware. Защита от целевых атак. 2021 [Электронный ресурс]. URL: <https://www.anti-malware.ru/event/2021/03/30> (дата обращения: 01.02.2022).

² BTB Security. Common and best practices for SOC's: results of the 2019 SANS SOC survey. 2019 [Электронный ресурс]. URL: <https://www.btbsecurity.com/blog/sans-soc-2019> (дата обращения: 01.02.2022).

³ SecurityLab.ru by Positive Technologies. Модель зрелости SOC от Gartner. 2019 [Электронный ресурс]. URL: https://www.securitylab.ru/blog/personal/Business_without_danger/346702.php (дата обращения: 01.02.2022).

ностью нелояльных сотрудников [7]^{1,2}. По данным Ponemon Institute, за 2020 год 62 % инцидентов информационной безопасности связаны с действиями так называемых инсайдеров³. При этом ущерб от таких инцидентов превысил 11 млн долларов, причем более 4 млн из них приходится на умышленные преступные действия¹.

Деятельность инсайдера — инцидент информационной безопасности, при котором действия сотрудника организации не соответствуют его нормам обычного поведения. Пользователи, будучи сотрудниками организации, следуют определенному распорядку рабочего дня, характер их работы может иметь некоторые особенности в определенные дни недели, в работе они используют конкретные приложения. Каждый пользователь имеет уникальный клавиатурный почерк [8, 9] и свойственный ему набор привычек. Отметим, что клавиатурный почерк пользователя может меняться в зависимости от времени суток в связи с хронотипом владельца [10].

В совокупности перечисленные особенности формируют «цифровую метрику» пользователя, характеризующую его поведение. Тогда аномалии в поведении могут свидетельствовать о потенциальном инциденте безопасности.

Связанные работы

В работах [3–6] отмечено, что в условиях роста сложности кибератак необходимы методы и системы обнаружения вторжений «нового поколения» (next generation IDS, advanced IDS), использующие контекстную высокоуровневую информацию об объекте защиты. В качестве такой информации рекомендовано использование «образа жизни» (Pattern-of-Life) объекта защиты. «Образ жизни» — совокупность информации о числе пользователей (конечных устройств), использования ими сетевых ресурсов и времени суток.

В [8] рассмотрены сложности и значимость методов непрерывной аутентификации по клавиатурному почерку. Подобные методы способны существенно повысить общий уровень безопасности и на данный момент находятся на стадии формирования. В [9] предложен метод непрерывной аутентификации пользователя по клавиатурному почерку, основанный на комбинации сверточной и рекуррентной нейронных сетей, а в работе [11] — на методах математической статистики.

В [12] отмечена значимость автоматизации решений по защите конечных устройств, в том числе по обна-

ружению инцидентов информационной безопасности, в первую очередь при помощи методов машинного обучения.

В работах [8, 9, 11] описаны методы непрерывной аутентификации по клавиатурному почерку, однако указанные методы предполагают его использование без учета других компонентов метрики и не учитывают зависимости клавиатурного почерка от различных факторов, что потенциально приводит к снижению эффективности методов. В [3–6] предложено использование контекстной высокоуровневой информации о локальной сети для обнаружения вторжений посредством выявления аномальной сетевой активности, и ее применение в системах обнаружения вторжений уровня сети, что не позволяет обнаруживать инциденты на конечных устройствах.

В настоящей работе предложен метод использования цифровой метрики как высокоуровневой информации об активности пользователя на конечном устройстве. Метод включает анализ клавиатурного почерка и влияющих на него факторов для определения аномальной активности и обнаружения инцидентов информационной безопасности. В результате возможно повысить эффективность обнаружения инцидентов информационной безопасности, обусловленных активностью инсайдеров, а также автоматизировать этот процесс на конечных устройствах.

Предлагаемый метод

Сущность предлагаемого метода состоит в формировании цифровой метрики пользователя за счет анализа его поведенческих характеристик и их зависимостей, выбранных в качестве предикторов для построения эффективной модели. Диаграмма метода в нотации IDEF0 приведена на рис. 1.

В качестве предикторов модели выбраны следующие компоненты цифровой метрики пользователя: время суток; день недели; скорость и динамика нажатия на клавиши клавиатуры и мыши. Ввиду специфики взаимодействия с различными процессами, для каждого используемого пользователем процесса необходима разработка отдельной модели. Тогда осуществляется включение в набор данных имени активного процесса, которому соответствует информация о поведении пользователя.

Для сбора данных в соответствии с указанным набором предикторов выполняется отслеживание следующих событий: переключение активного процесса, перевод клавиш клавиатуры и мыши в верхнее и нижнее положения. Формирование паттерна текущего поведения пользователя должно выполняться до переключения процесса или завершаться по истечении времени формирования паттерна. Для дальнейшего применения собранных данных следует нормализовать все численные предикторы. Блок-схема алгоритма сбора данных представлена на рис. 2.

Отметим, что при работе пользователя возможны случайные переключения активного процесса, что приводит к возникновению выбросов, а именно паттернов с нулевыми значениями численных предикторов, описы-

¹ Ponemon Institute. 2020 Cost of insider threats global report. 2020 [Электронный ресурс]. URL: <https://www.exclusive-networks.com/uk/wp-content/uploads/sites/28/2020/12/UK-VR-Proofpoint-Report-2020-Cost-of-Insider-Threats.pdf> (дата обращения: 01.02.2022).

² IBM Security. The Cost of Insider Threats. 2020 [Электронный ресурс]. Режим доступа: <https://www.ibm.com/security/digital-assets/services/cost-of-insider-threats/#/> (дата обращения: 01.02.2022).

³ Anti-Malware. Защита от целевых атак. 2021 [Электронный ресурс]. URL: <https://www.anti-malware.ru/event/2021/03/30> (дата обращения: 01.02.2022).

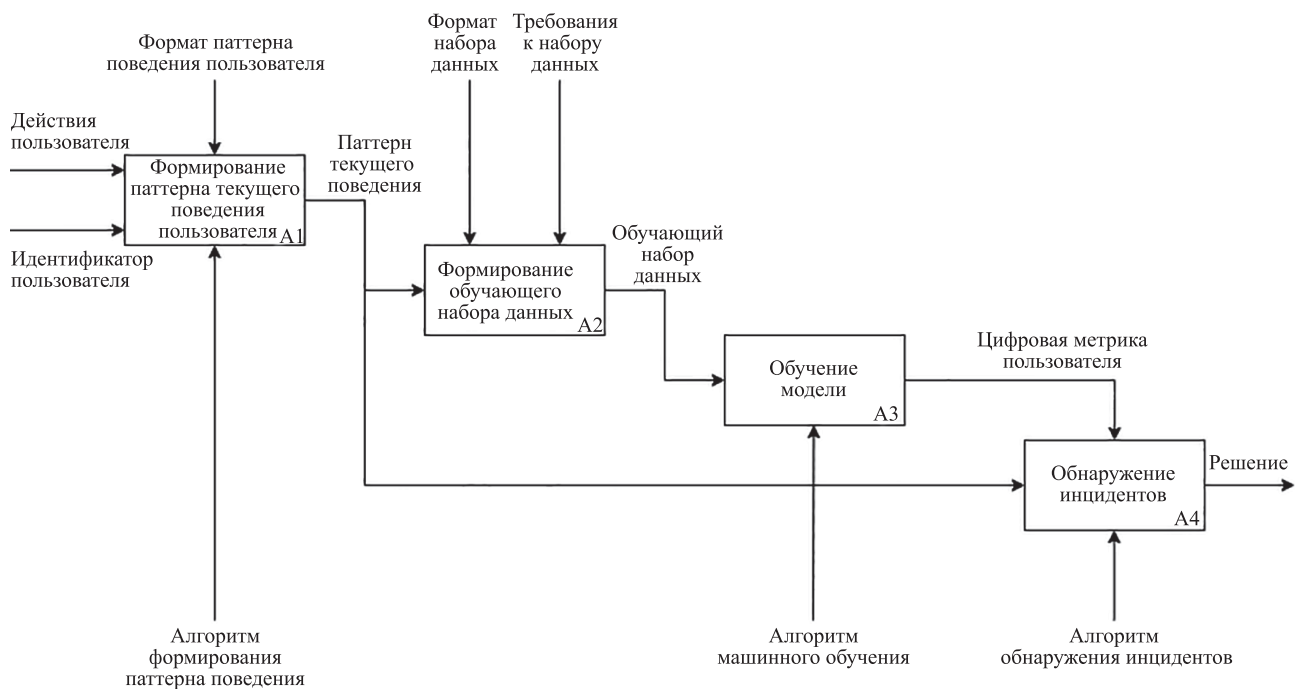


Рис. 1. Диаграмма метода обнаружения инцидентов информационной безопасности по аномалиям в биометрических поведенческих чертах пользователя

Fig. 1. Diagram of the method of detecting information security incidents based on anomalies in the user's biometric behavioral characteristics

вающих взаимодействие пользователя с контроллерами. Такие паттерны должны быть исключены из обучающего набора данных перед обучением модели.

Метод предполагает обучение модели для каждого используемого пользователем процесса. Для формирования модели предполагается применение алгоритма машинного обучения без учителя, направленного на выявление аномалий [13, 14].

Для обнаружения инцидентов информационной безопасности осуществляется выбор модели в соответствии с именем процесса в текущем паттерне поведения пользователя. Компоненты цифровой метрики, выбранные в качестве предикторов, являются входными параметрами модели для классификации паттерна поведения.

Тогда могут быть обнаружены следующие аномалии в поведении пользователя:

- активность в нетипичные для пользователя время суток и день недели;
- использование нетипичного для пользователя программного обеспечения;
- нетипичный для пользователя активного процесса и времени суток клавиатурный почерк.

Исследуемые методы машинного обучения

Для обучения модели рассмотрены методы машинного обучения: опорных векторов; методы на основах плотности распределения и тензора данных.

Метод опорных векторов включает преимущественно алгоритмы обучения с учителем. Исключениями являются алгоритмы опорных векторов для одного класса

(One-Class Support Vector Machine, OCSVM) [15, 16] и для описания данных (Support Vector Data Description, SVDD) [15, 17]. Заметим, что алгоритм SVDD рассчитывает гиперсферу минимального радиуса, внутри которой располагаются объекты, соответствующие норме, снаружи — аномалии [15]. Тогда SVDD можно отнести к методам на основе тензора данных. Алгоритм OCSVM определяет гиперповерхность, разделяющую нормальные и аномальные объекты [15].

Методы машинного обучения на основе плотности распределения включают множество алгоритмов, из которых были рассмотрены наиболее популярные, в том числе алгоритмы k -ближайших соседей (k -Nearest Neighbors, k NN), локальный уровень выброса (Local Outlier Factor, LOF) и изолирующего леса (Isolation Forest). Согласно сравнительному анализу алгоритмов [18], алгоритм Isolation Forest демонстрирует более высокие значения по метрикам качества и достигает высоких показателей при работе с многомерными данными.

Из методов машинного обучения на основе тензора данных [19] рассмотрены алгоритмы SVDD и эллипсоидальной аппроксимации данных (Elliptic Envelope, EE) [20, 21]. Оба алгоритма имеют схожий принцип обнаружения аномалий, однако если алгоритм SVDD рассчитывает гиперсферу, то Elliptic Envelope вычисляет многомерный эллипсоид. Так как сфера — частный случай эллипсоида, то выбран алгоритм Elliptic Envelope.

В итоге, для формирования цифровой метрики пользователя выбраны следующие алгоритмы машинного обучения: OCSVM, Isolation Forest и Elliptic Envelope.

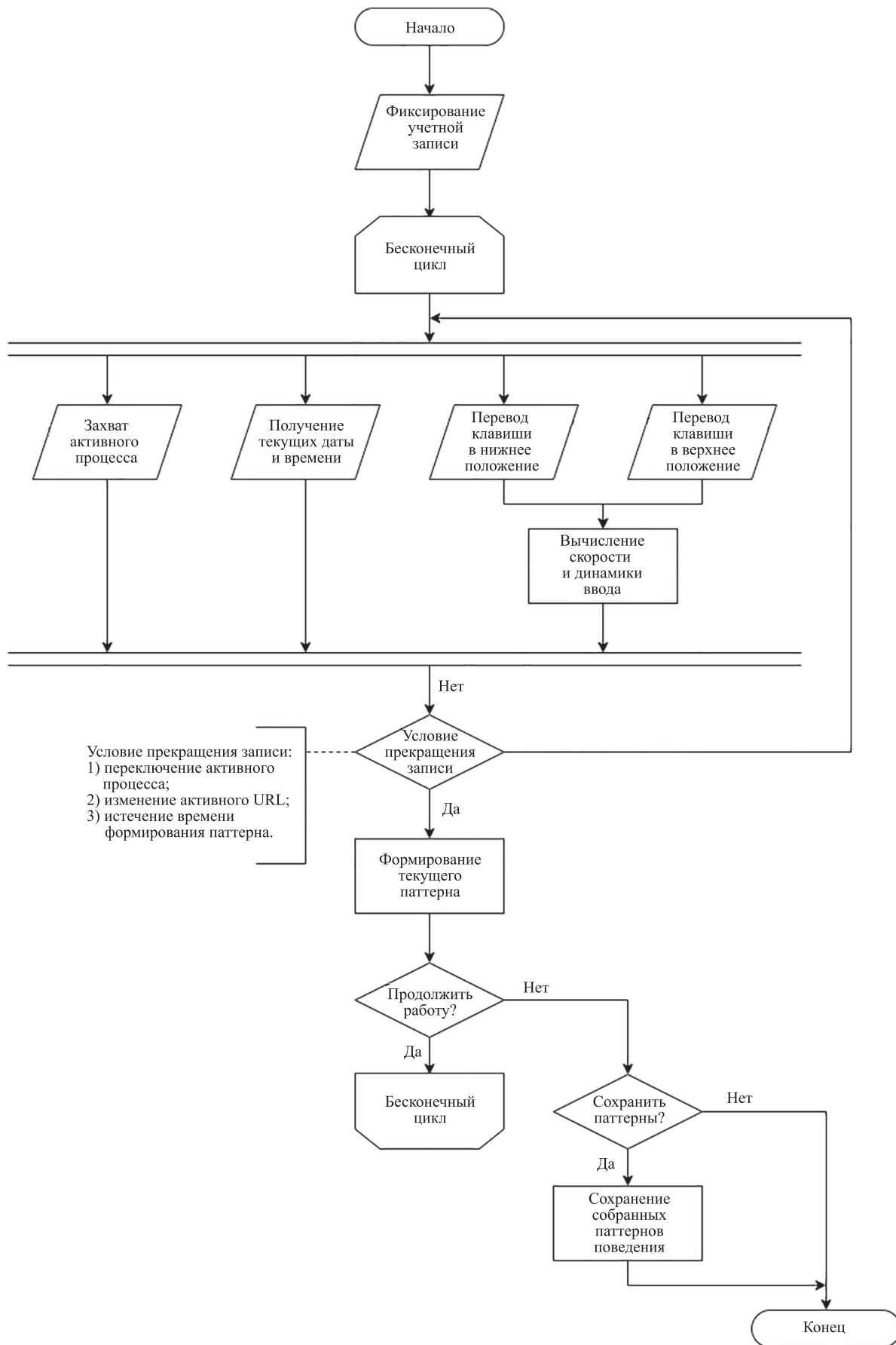


Рис. 2. Блок-схема алгоритма сбора данных

Fig. 2. Data collection algorithm flowchart

Алгоритмы были реализованы на языке программирования Python при помощи библиотеки sklearn [22]. Определение алгоритма и его параметров, демонстрирующих лучший результат, было выполнено в ходе проведения эксперимента.

Выбор метрик качества

Для оценки качества предлагаемого метода выбраны метрики:

- корректность (accuracy) [23];
- точность (precision) [23];
- полнота (recall) [23];
- F-мера (F-score) [23];
- коэффициент корреляции Мэтьюса (Matthews Correlation Coefficient, MCC) [24].

При выборе метрики учтены значения временных характеристик, таких как время обучения и классификация. Рассмотрим метрики подробнее.

Метрика accuracy показывает долю верно классифицированных объектов. Отметим, что accuracy не является показательной при значительно различающимися по мощности классами объектов. Другие метрики лишены подобного недостатка.

Recall определяет возможность идентификации объектов некоторого класса алгоритмом. Precision показывает способность отличать объекты некоторого класса от объектов остальных классов. F-score представляет собой единую метрику качества, учитывающую precision и recall. Однако F-score учитывает метрики precision и recall только для «положительного» класса.

MCC является еще одной метрикой качества двойной классификации, не подверженной влиянию несбалансированных по мощности классов. При этом указанная метрика дает высокий результат только в том случае, когда классификатор смог корректно предсказать большинство «положительных» и «отрицательных» объектов [24].

Для оценки моделей рассчитаем перечисленные метрики при условии, что в качестве решающей метрики для оценки будет использована метрика MCC.

Формирование набора данных

На момент исследования не существовало открытых наборов данных, содержащих необходимые сведения в соответствии с выбранными параметрами цифровой метрики. Как следствие, в силу специфичности необходимых для построения модели данных необходимо формирование собственного набора данных.

В формировании набора данных принимали участие 138 человек. Для эксперимента выбраны три приложения различных типов. Сбор данных выполнен в соответствии с алгоритмом, приведенным на рис. 2.

Мощность набора данных составила 388 168 паттернов (строк) реального поведения пользователя в течении сессии работы. Для проведения эксперимента выполнено разделение набора данных следующим образом: 310 534 (80 %) строк для обучения и 77 634 (20 %) — для тестирования. Заметим, что все строки, соответствующие аномальному поведению, отнесены к тестовому набору данных.

Выбор алгоритма и его параметров

Для определения наиболее эффективной модели выполнено обучение модели посредством применения различных алгоритмов машинного обучения и варьирования их параметров. Наибольшие показатели по метрикам качества моделей по каждому алгоритму представлены в табл. 1.

Согласно табл. 1, лучший результат по выбранным метрикам качества достигает алгоритм OCSVM. Выполним сравнение времени обучения и классификации алгоритмов (табл. 2). Указанные характеристики рассчитаем на 100 итераций указанных операций.

Таблица 1. Сравнительный анализ метрик качества алгоритмов машинного обучения

Table 1. Comparative analysis of quality metrics for machine learning algorithms

Алгоритмы	Параметры	Метрики				
		Accuracy	Precision	Recall	F-score	MCC
OCSVM	функция ядра (kernel): RBF ν (nu): 0,012 γ : 0,197	0,9218	0,9143	0,9951	0,9530	0,7464
Isolation Forest	число деревьев (n_estimators): 66 зашумление (contamination): 0,230	0,7805	0,9092	0,8049	0,8539	0,4321
Elliptic Envelope	зашумление: 0,287	0,7106	0,9011	0,7154	0,7976	0,3400

Таблица 2. Сравнительный анализ временных характеристик алгоритмов

Table 2. Comparative analysis of time characteristics of algorithms

Алгоритм	Время обучения, с	Время классификации, с
OCSVM	22,5131	0,0510
Isolation Forest	95,6583	10,3003
Elliptic Envelope	213,3738	0,1050

Таблица 3. Сравнительный анализ методов
Table 3. Comparative analysis of methods

Параметр		Метод	
		по цифровой метрике	по клавиатурному почерку
Метрики	Accuracy	0,9218	0,5689
	Precision	0,9143	0,8480
	Recall	0,9951	0,5592
	F-score	0,9530	0,6740
	MCC	0,7464	0,1339
Время, с	обучения	22,5131	83,3031
	классификации	0,0510	5,4050

В результате анализа получено, что минимальных значений затраченного времени по обоим операциям достигает алгоритм OCSVM, выбранный по итогу сравнительного анализа для обучения моделей нормального поведения.

Сравнение предложенного метода с обнаружением аномалий по клавиатурному почерку

Для сравнения предложенного метода с методом обнаружения аномалий по клавиатурному почерку проведено обучение моделей для обоих указанных методов. Выбор алгоритма обучения модели и его параметров для обнаружения аномалий по клавиатурному почерку выполнен с использованием тех же алгоритмов машинного обучения и наборов данных для обучения и тестирования. В качестве алгоритма обучения выбран Isolation Forest с числом деревьев, равным 86, и параметром зашумления — 0,469, так как такая модель продемонстрировала наибольшие показатели качества.

Сравнительный анализ методов обнаружения аномалий по цифровой метрике и по клавиатурному почерку по метрикам качества и временным характеристикам приведен в табл. 3.

Согласно табл. 3, метод обнаружения по цифровой метрике демонстрирует более высокие показатели по всем метрикам качества. Причем для метода обнаружения аномалий по клавиатурному почерку наблюдается значительная доля ошибок первого рода — более 37 %. Минимальных значений затраченного времени на указанные операции также достигает метод обнаружения инцидентов по цифровой метрике.

Заключение

Предложенный метод обнаружения инцидентов информационной безопасности по аномалиям в биометрических поведенческих чертах (цифровой метрике) пользователя достигает коэффициента корреляции Мэтьюса в 0,746363 и показателя F-меры в 0,953008, что превышает аналогичные показатели метода обнаружения аномалий, основанного на клавиатурном почерке. Поскольку при проведении эксперимента были использованы данные, соответствующие активности реальных пользователей при работе с различными приложениями, построенная модель является устойчивой для тестовых и реальных данных.

Метод может быть использован для непрерывной аутентификации пользователя в средствах защиты информации от несанкционированного доступа, что позволит контролировать не только предоставление доступа в начале, но и в течение всего времени поддержания сессии. Указанная модификация позволит обнаруживать факты вмешательства в сессию легитимного пользователя, а также факты неавторизованной передачи доступа, тем самым расширив функционал указанных средств защиты информации.

В дальнейшей работе возможно теоретическое обоснование эффективности предложенного метода, а также исследование применимости других алгоритмов машинного обучения. Возможно расширение цифровой метрики пользователя активностью процессов и сетевых портов, взаимодействием пользователя и процессов с файловой системой и событиями операционной системы, что позволит сформировать цифрового двойника пользователя и сместить фокус с аутентификации на обнаружение инцидентов и определение аномалий в его поведении, а также в работе его автоматизированного рабочего места.

Литература

1. Siddiqi M.A., Mugheri A., Oad K. Advance persistent threat defense techniques: A review // *Pakistan Journal of Computer and Information Systems*. 2016. V. 1. N 2. P. 53–65.
2. Al-Zewairi M., Almajali S., Ayyash M. Unknown security attack detection using shallow and deep ANN classifiers // *Electronics*. 2020. V. 9. N 12. P. 2006. <https://doi.org/10.3390/electronics9122006>
3. Aparicio-Navarro F.J., Kyriakopoulos K.G., Gong Y., Parish D.J., Chambers J.A. Using pattern-of-life as contextual information for anomaly-based intrusion detection systems // *IEEE Access*.

References

1. Siddiqi M.A., Mugheri A., Oad K. Advance persistent threat defense techniques: A review. *Pakistan Journal of Computer and Information Systems*, 2016, vol. 1, no. 2, pp. 53–65.
2. Al-Zewairi M., Almajali S., Ayyash M. Unknown security attack detection using shallow and deep ANN classifiers. *Electronics*, 2020, vol. 9, no. 12, pp. 2006. <https://doi.org/10.3390/electronics9122006>
3. Aparicio-Navarro F.J., Kyriakopoulos K.G., Gong Y., Parish D.J., Chambers J.A. Using pattern-of-life as contextual information for anomaly-based intrusion detection systems. *IEEE Access*, 2017,

2017. V. 5. P. 22177–22193. <https://doi.org/10.1109/ACCESS.2017.2762162>
4. Aparicio-Navarro F.J., Chambers J.A., Kyriakopoulos K., Gong Y., Parish D. Using the pattern-of-life in networks to improve the effectiveness of intrusion detection systems // *Proc. of the 2017 IEEE International Conference on Communications (ICC)*. 2017. P. 7997374. <https://doi.org/10.1109/ICC.2017.7997374>
5. Aparicio-Navarro F.J., Kyriakopoulos K.G., Ghafir I., Lambotharan S., Chambers J.A. Multi-stage attack detection using contextual information // *Proc. of the IEEE Military Communications Conference (MILCOM)*. 2018. P. 920–925. <https://doi.org/10.1109/MILCOM.2018.8599708>
6. Aparicio-Navarro F.J., Chadza T.A., Kyriakopoulos K.G., Ghafir I., Lambotharan S., Assadhan B. Addressing multi-stage attacks using expert knowledge and contextual information // *Proc. of the 22nd Conference on Innovation in Clouds, Internet and Networks and Workshops (ICIN)*. 2019. P. 188–194. <https://doi.org/10.1109/ICIN.2019.8685841>
7. Budiarto R., Alqarni A.A., Alzahrani M.Y., Pasha M.F., Firdhous M.F.M., Stiawan D. User behavior traffic analysis using a simplified memory-prediction framework // *Computers, Materials and Continua*. 2022. V. 70. N 2. P. 2679–2698. <https://doi.org/10.32604/cmc.2022.019847>
8. Quraishi S.J., Bedi S.S. Keystroke dynamics biometrics, a tool for user authentication—review // *Proc. of the 7th International Conference on System Modeling and Advancement in Research Trends (SMART)*. 2018. P. 248–254. <https://doi.org/10.1109/SYSMART.2018.8746932>
9. Xiaofeng L., Shengfei Z., Shengwei Y. Continuous authentication by free-text keystroke based on CNN plus RNN // *Procedia Computer Science*. 2019. V. 147. P. 314–318. <https://doi.org/10.1016/j.procs.2019.01.270>
10. Drujff-van de Woestijne G.B., McConchie H., de Kort Y., Licitra G., Zhang C., Overeem S., Smolders K.C.H.J. Behavioural biometrics: Using smartphone keyboard activity as a proxy for rest-activity patterns // *Journal of Sleep Research*. 2021. V. 30. N 5. P. e13285. <https://doi.org/10.1111/jsr.13285>
11. Крутохвостов Д.С., Хищенко В.Е. Парольная и непрерывная аутентификация по клавиатурному почерку средствами математической статистики // *Вопросы кибербезопасности*. 2017. № 5(24). С. 91–99. <https://doi.org/10.21681/2311-3456-2017-5-91-99>
12. Sjarif N.N.A., Chuprat S., Mahrin M.N., Ahmad N.A., Senan F.M., Zamani N.A., Saupi A. Endpoint detection and response: Why use machine learning? // *Proc. of the 10th International Conference on Information and Communication Technology Convergence (ICTC)*. 2019. P. 283–288. <https://doi.org/10.1109/ICTC46691.2019.8939836>
13. Kumar Singh Gautam R., Doegar E.A. An ensemble approach for intrusion detection system using machine learning algorithms // *Proc. of the 8th Confluence International Conference on Cloud Computing, Data Science and Engineering*. 2018. P. 14–15. <https://doi.org/10.1109/CONFLUENCE.2018.8442693>
14. Alqudah N., Yaseen Q. Machine learning for traffic analysis: a review // *Procedia Computer Science*. 2020. V. 170. P. 911–916. <https://doi.org/10.1016/j.procs.2020.03.111>
15. Lampert C.H. Kernel methods in computer vision // *Foundations and Trends in Computer Graphics and Vision*. 2009. V. 4. N 3. P. 193–285. <http://dx.doi.org/10.1561/06000000027>
16. Bounsiar A., Madden M.G. One-class support vector machines revisited // *Proc. of the 5th International Conference on Information Science & Applications (ICISA)*. 2014. P. 6847442. <https://doi.org/10.1109/ICISA.2014.6847442>
17. Tax D.M.J., Duin R.P.W. Support vector data description // *Machine Learning*. 2004. V. 54. N 1. P. 45–66. <https://doi.org/10.1023/B:MACH.0000008084.60811.49>
18. Liu F.T., Ting K.M., Zhou Z.H. Isolation forest // *Proc. of the 8th IEEE International Conference on Data Mining (ICDM)*. 2008. P. 413–422. <https://doi.org/10.1109/ICDM.2008.17>
19. Ji Y., Wang Q., Li X., Liu J. A survey on tensor techniques and applications in machine learning // *IEEE Access*. 2019. V. 7. P. 162950–162990. <https://doi.org/10.1109/ACCESS.2019.2949814>
20. Howard S. The Elliptical Envelope // *arXiv*. 2007. arXiv:math/0703048. <https://doi.org/10.48550/arXiv.math/0703048>
21. Ashrafuzzaman M., Das S., Jillepalli A.A., Chakhchoukh Y., Sheldon F.T. Elliptic envelope based detection of stealthy false data injection attacks in smart grid control systems // *Proc. of the 2020 IEEE Symposium Series on Computational Intelligence (SSCI)*. 2020. P. 1131–1137. <https://doi.org/10.1109/SSCI47803.2020.9308523>
- vol. 5, pp. 22177–22193. <https://doi.org/10.1109/ACCESS.2017.2762162>
4. Aparicio-Navarro F.J., Chambers J.A., Kyriakopoulos K., Gong Y., Parish D. Using the pattern-of-life in networks to improve the effectiveness of intrusion detection systems. *Proc. of the 2017 IEEE International Conference on Communications (ICC)*, 2017, pp. 7997374. <https://doi.org/10.1109/ICC.2017.7997374>
5. Aparicio-Navarro F.J., Kyriakopoulos K.G., Ghafir I., Lambotharan S., Chambers J.A. Multi-stage attack detection using contextual information. *Proc. of the IEEE Military Communications Conference (MILCOM)*, 2018, pp. 920–925. <https://doi.org/10.1109/MILCOM.2018.8599708>
6. Aparicio-Navarro F.J., Chadza T.A., Kyriakopoulos K.G., Ghafir I., Lambotharan S., Assadhan B. Addressing multi-stage attacks using expert knowledge and contextual information. *Proc. of the 22nd Conference on Innovation in Clouds, Internet and Networks and Workshops (ICIN)*, 2019, pp. 188–194. <https://doi.org/10.1109/ICIN.2019.8685841>
7. Budiarto R., Alqarni A.A., Alzahrani M.Y., Pasha M.F., Firdhous M.F.M., Stiawan D. User behavior traffic analysis using a simplified memory-prediction framework. *Computers, Materials and Continua*, 2022, vol. 70, no. 2, pp. 2679–2698. <https://doi.org/10.32604/cmc.2022.019847>
8. Quraishi S.J., Bedi S.S. Keystroke dynamics biometrics, a tool for user authentication—review. *Proc. of the 7th International Conference on System Modeling and Advancement in Research Trends (SMART)*, 2018, pp. 248–254. <https://doi.org/10.1109/SYSMART.2018.8746932>
9. Xiaofeng L., Shengfei Z., Shengwei Y. Continuous authentication by free-text keystroke based on CNN plus RNN. *Procedia Computer Science*, 2019, vol. 147, pp. 314–318. <https://doi.org/10.1016/j.procs.2019.01.270>
10. Drujff-van de Woestijne G.B., McConchie H., de Kort Y., Licitra G., Zhang C., Overeem S., Smolders K.C.H.J. Behavioural biometrics: Using smartphone keyboard activity as a proxy for rest-activity patterns. *Journal of Sleep Research*, 2021, vol. 30, no. 5, pp. e13285. <https://doi.org/10.1111/jsr.13285>
11. Krutohvastov D., Khitsenko V. Password authentication and continuous authentication by keystroke dynamics using mathematical statistics. *Voprosy kiberbezopasnosti*, no. 5(24), pp. 91–99. (in Russian). <https://doi.org/10.21681/2311-3456-2017-5-91-99>
12. Sjarif N.N.A., Chuprat S., Mahrin M.N., Ahmad N.A., Senan F.M., Zamani N.A., Saupi A. Endpoint detection and response: Why use machine learning? *Proc. of the 10th International Conference on Information and Communication Technology Convergence (ICTC)*, 2019, pp. 283–288. <https://doi.org/10.1109/ICTC46691.2019.8939836>
13. Kumar Singh Gautam R., Doegar E.A. An ensemble approach for intrusion detection system using machine learning algorithms. *Proc. of the 8th Confluence International Conference on Cloud Computing, Data Science and Engineering*, 2018, pp. 61–64. <https://doi.org/10.1109/CONFLUENCE.2018.8442693>
14. Alqudah N., Yaseen Q. Machine learning for traffic analysis: a review. *Procedia Computer Science*, 2020, vol. 170, pp. 911–916. <https://doi.org/10.1016/j.procs.2020.03.111>
15. Lampert C.H. Kernel methods in computer vision. *Foundations and Trends in Computer Graphics and Vision*, 2009, vol. 4, no. 3, pp. 193–285. <http://dx.doi.org/10.1561/06000000027>
16. Bounsiar A., Madden M.G. One-class support vector machines revisited. *Proc. of the 5th International Conference on Information Science & Applications (ICISA)*, 2014, pp. 6847442. <https://doi.org/10.1109/ICISA.2014.6847442>
17. Tax D.M.J., Duin R.P.W. Support vector data description. *Machine Learning*, 2004, vol. 54, no. 1, pp. 45–66. <https://doi.org/10.1023/B:MACH.0000008084.60811.49>
18. Liu F.T., Ting K.M., Zhou Z.H. Isolation forest. *Proc. of the 8th IEEE International Conference on Data Mining (ICDM)*, 2008, pp. 413–422. <https://doi.org/10.1109/ICDM.2008.17>
19. Ji Y., Wang Q., Li X., Liu J. A survey on tensor techniques and applications in machine learning. *IEEE Access*, 2019, vol. 7, pp. 162950–162990. <https://doi.org/10.1109/ACCESS.2019.2949814>
20. Howard S. The Elliptical Envelope. *arXiv*, 2007, arXiv:math/0703048. <https://doi.org/10.48550/arXiv.math/0703048>
21. Ashrafuzzaman M., Das S., Jillepalli A.A., Chakhchoukh Y., Sheldon F.T. Elliptic envelope based detection of stealthy false data injection attacks in smart grid control systems. *Proc. of the 2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2020, pp. 1131–1137. <https://doi.org/10.1109/SSCI47803.2020.9308523>

22. Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., Duchesnay É. Scikit-learn: Machine learning in Python // *Journal of Machine Learning Research*. 2011. V. 12. P. 2825–2830.
23. Saranya T., Sridevi S., Deisy C., Chung T.D., Khane M.K.A.A. Performance analysis of machine learning algorithms in intrusion detection system: A review // *Procedia Computer Science*. 2020. V. 171. P. 1251–1260. <https://doi.org/10.1016/j.procs.2020.04.133>
24. Chicco D., Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation // *BMC Genomics*. 2020. V. 21. N 1. P. 1–13. <https://doi.org/10.1186/s12864-019-6413-7>
22. Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., Duchesnay É. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 2011, vol. 12, pp. 2825–2830.
23. Saranya T., Sridevi S., Deisy C., Chung T.D., Khane M.K.A.A. Performance analysis of machine learning algorithms in intrusion detection system: A review. *Procedia Computer Science*, 2020, vol. 171, pp. 1251–1260. <https://doi.org/10.1016/j.procs.2020.04.133>
24. Chicco D., Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 2020, vol. 21, no. 1, pp. 1–13. <https://doi.org/10.1186/s12864-019-6413-7>

Авторы

Есипов Дмитрий Андреевич — инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, <https://orcid.org/0000-0003-4467-5117>, some1else.d.ma@gmail.com

Асланова Наргиз — студент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, <https://orcid.org/0000-0003-3650-7412>, aslanova.077@gmail.com

Шабала Егор Евгеньевич — инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, <https://orcid.org/0000-0001-6659-9878>, e.shabala@altdisasm.ru

Щетинин Даниил Сергеевич — инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, <https://orcid.org/0000-0003-3913-3997>, 1nta4r@gmail.com

Попов Илья Юрьевич — кандидат технических наук, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, <https://orcid.org/0000-0002-6407-7934>, ilyapopov27@gmail.com, iupopov@itmo.ru

Authors

Dmitry A. Esipov — Engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, <https://orcid.org/0000-0003-4467-5117>, some1else.d.ma@gmail.com

Nargiz Aslanova — Student, ITMO University, Saint Petersburg, 197101, Russian Federation, <https://orcid.org/0000-0003-3650-7412>, aslanova.077@gmail.com

Egor E. Shabala — Engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, <https://orcid.org/0000-0001-6659-9878>, e.shabala@altdisasm.ru

Daniil S. Shchetinin — Engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, <https://orcid.org/0000-0003-3913-3997>, 1nta4r@gmail.com

Ilya Yu. Popov — PhD, Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, <https://orcid.org/0000-0002-6407-7934>, ilyapopov27@gmail.com, iupopov@itmo.ru

Статья поступила в редакцию 03.02.2022

Одобрена после рецензирования 06.06.2022

Принята к печати 20.07.2022

Received 03.02.2022

Approved after reviewing 06.06.2022

Accepted 20.07.2022



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

doi: 10.17586/2226-1494-2022-22-4-769-778

Light weight recommendation system for social networking analysis using a hybrid BERT-SVM classifier algorithm

Kiruthika Nallichery Subramanian¹✉, Thailambal Ganapathy²

¹ Vels University, Chennai, Tamil Nadu, 600117, India

² Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai, Tamil Nadu, 600117, India

¹ sathishpoojaa5@gmail.com✉, <https://orcid.org/0000-0001-6601-1341>

² thaila.scs@velsuniv.ac.in, <https://orcid.org/0000-0002-0043-2415>

Abstract

Social media platforms, such as Twitter, Instagram, and Facebook, have facilitated mass communication and connection. Due to the development as well as the advancement of social platforms, the spreading of fake news has increased. Many studies have been performed for detecting fake news with machine learning algorithms; but these existing methods had several difficulties, such as rapid propagation, access method and insignificant selection of features, and low accuracy of the text classification. Therefore, to overcome these issues, this paper proposed a hybrid Bidirectional Encoder Representations from Transformers — Support Vector Machine (BERT-SVM) model with a recommendation system that used to predict whether the information is fake or real. The proposed model consists of three phases: preprocessing, feature selection and classification. The dataset is gathered from Twitter social media related to COVID-19 real-time data. Preprocessing stage comprises Splitting, Stop word removal, Lemmatization and Spell correction. Term Frequency Inverse Document Frequency (TF-IDF) converter is utilized to extract the features and convert text to binary vectors. A hybrid BERT-SVM classification model is used to predict the data. Finally, the predicted data is compared with the preprocessed data. The proposed model is implemented in MATLAB software with several performance metrics carried out, and these parameters attained better performance: accuracy is 98 %, the error is 2 %, precision is 99 %, specificity is 99 %, and sensitivity is 98 %. Therefore the better effectiveness of the proposed model than existing approaches is shown. The proposed social networking analysis model provides effective fake news prediction that can be used to identify the Twitter comments, either real or fake.

Keywords

social networking analysis, fake news detection, TF/IDF, BERT, SVM, hybrid BERT-SVM

For citation: Kiruthika N.S., Thailambal G. Light weight recommendation system for social networking analysis using a hybrid BERT-SVM classifier algorithm. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2022, vol. 22, no. 4, pp. 769–778. doi: 10.17586/2226-1494-2022-22-4-769-778

УДК 004.896

Облегченная система рекомендаций для анализа социальных сетей с использованием гибридного алгоритма классификатора BERT-SVM

Налличери Субраманиан Кирутика¹✉, Ганапати Тайламбал²

¹ Университет Велса, Ченнаи, шт. Тамил Наду, 600117, India

² Институт науки, технологий и перспективных исследований Велса, Ченнаи, шт. Тамил Наду, 600117, Индия

¹ sathishpoojaa5@gmail.com✉, <https://orcid.org/0000-0001-6601-1341>

² thaila.scs@velsuniv.ac.in, <https://orcid.org/0000-0002-0043-2415>

Аннотация

Платформы социальных сетей, такие как Twitter, Instagram и Facebook, способствуют массовому общению и установлению связей. Развитие и продвижение социальных платформ приводит к увеличению распространения фейковых новостей. В настоящее время проведено большое количество исследований для обнаружения фейковых новостей с помощью алгоритмов машинного обучения. Существующие методы определения фейков имеют ряд трудностей: быстрое распространение фейков; различные методы доступа и незначительный выбор признаков, приводящие к невысокой точности классификации текста. Для преодоления данных трудностей предложена

© Kiruthika N.S., Thailambal G., 2022

гибридная модель представления двунаправленного кодировщика трансформаторов – метод опорных векторов (BERT-SVM) с системой рекомендаций, которая используется для прогнозирования, является ли информация поддельной или реальной. Предложенная модель включает в себя три этапа: предварительная обработка, выбор признаков и классификация. Набор данных собран из социальных сетей Twitter, связанных с данными о COVID-19 в режиме реального времени. Этап предварительной обработки включает в себя разделение, удаление стоп-слов, лемматизацию и исправление орфографии. Преобразователь обратной частоты документа (TF-IDF) использован для извлечения признаков и преобразования текста в двоичные векторы. Гибридная модель классификации BERT-SVM применена для прогнозирования данных, которые сопоставлены с предварительно обработанными данными. Представленная модель реализована в программном пакете MATLAB. Рассчитанные показатели точности продемонстрировали следующие результаты: доля правильных ответов 98 %, ошибка 2 %, точность 99 %, специфичность 99 %, чувствительность 98 %. Полученные результаты показали эффективность предложенной модели по сравнению с существующими подходами. Возможность анализа социальных сетей обеспечивает эффективное предсказание фейковых новостей, которое можно использовать для идентификации комментариев в Twitter, как настоящих, так и поддельных.

Ключевые слова

анализ социальных сетей, обнаружение фейковых новостей, TF/IDF, BERT, SVM, гибридная BERT-SVM

Ссылка для цитирования: Кирутика Н.С., Тайламбал Г. Облегченная система рекомендаций для анализа социальных сетей с использованием гибридного алгоритма классификатора BERT-SVM // Научно-технический вестник информационных технологий, механики и оптики. 2022. Т. 22, № 4. С. 769–778 (на англ. яз.). doi: 10.17586/2226-1494-2022-22-4-769-778

Introduction

Social networking has developed dominant platforms to reveal their sentiments, opinions, reactions, and knowledge. Several social media, such as Twitter, Facebook, Instagram, etc., produce a massive amount of data each and every day. Emotion recognition or analysis of sentiment focuses on detecting polarization by using views or opinions from Twitter datasets. Sentiment polarity determines a user's reaction to a product, allowing businesses to take preventative and corrective actions to satisfy their expectations. Conversely, government criticism assists governments in analyzing public requirements and making crucial decisions. Fake news is becoming more common on social media sites like Twitter and Facebook. These platforms provide a venue for the general public to express themselves in an unfiltered and uncensored manner [1].

Misinformation is defined as information that is demonstrably false and is shared with the goal of misleading readers. For personal advantage, it is used to establish an economic, political and social bias in people's thoughts. Its goal is to manipulate and exploit individuals by creating bogus material that appears to be genuine. At its most severe cases, fake news has resulted in mob lynching and riots. As a result, it is critical to halt the proliferation of fraudulent content on social media sites. With the continuing Covid-19 scenario, preventing fake news is very vital [2]. The epidemic has done it very easy to deceive a psychologically stuck people excitedly anticipating the conclusion of this phase. Some people have reportedly committed suicide after being diagnosed with COVID-19 as a result of the misrepresentation of COVID-19 in society and even the mainstream media. The promotion of deceptive techniques will only exacerbate the COVID pandemic. Recently, researchers have begun to focus on the challenge of detecting fake news. Manual detection is the most reliable method, although it has speed constraints [3]. Manual verification is difficult due to the large amount of content published on the internet. Thus, automatic detection of bogus news has become increasingly important.

Different deep learning and machine learning algorithms were used to discover comments of social media information about its truthfulness. These false reports will not only lead individuals in the wrong route, but they will also take human lives. In these critical times of Covid-19, it is easy to deceive people and make them believe in fake information [4]. As a result, it's critical to spot fake news at the source and stop it from propagating to a wider audience. Many studies have been performed for detecting the fake news with machine learning approaches like Long Short-Term Memory (LSTM), Support Vector Machine (SVM), hybrid LSTM-SVM, Naive Bayes and Random Forest. From these existing methods, the accuracy of the text classification is low. So, to overcome the low accuracy, a hybrid Bidirectional Encoder Representations from Transformers — Support Vector Machine (BERT-SVM) model with a recommendation system is used to predict whether the information is fake or real. The main contribution of the paper is summarized as follows:

- An accurate Light Weight Recommendation System for Social Networking Analysis using a Hybrid BERT-SVM classifier.
- The preprocessing technique is utilized to improve accuracy and reduce the process complexity of the real time twitter dataset to provide better performance.
- Term Frequency Inverse Document Frequency (TF-IDF) is utilized for feature selection for word representations.
- The hybrid BERT-SVM classification model is designed for predicting the fake news in twitter to avoid spreading the fake news related on COVID-19 pandemic.
- The predicted news is given as a recommendation to the user for the awareness of the information about the fake news.

Literature review

Numerous studies have been performed using various techniques for recommendation system. Most of the existing techniques is designed based on LSTM, Random

Forest, SVM, Bi-LSTM & Recurrent Neural Network (RNN), Convolution Neural Network (CNN), Neuro Fuzzy, CNN-LSTM, bidirectional Gated RNN (GRNN) and convolution RNN (CRNN); from that few of them are reviewed below.

Umer et al., [4] had presented the deep learning approach for detecting the fake news using CNN-LSTM. The dataset gathered from the website named “Fake News Challenges” and this website has four different types, such as unrelated, disagree, agree and discuss. Hakak et al., [5] had introduced a strategy for classify the fake news utilizing a feature extraction and ensemble machine learning technique. Abdullah et al., [6] had developed a multimodal method to achieve a fake news detection using a combination of CNN and LSTM. Huang et al., [7] had designed a fake news prediction strategy utilizing deep learning method. The developed model combined four separate models termed as embedding for fake news identification. The embedding models are LSTM, LIWC CNN, depth LSTM, and N-gram CNN. Paka et al., [8] had introduced a fake news prediction model for analyzing a large-scale COVID-19 Twitter dataset. Nasir et al., [9] had designed a novel hybrid strategy for fake news categorization using combined CRNN. Sabeeh et al., [10] had introduced a model for discovering fake news on social media platforms through Opinion mining and Trustworthiness of user and event that combines opinion mining on user comment and credibility investigation of twitter dataset. Bahad, P et al., [11] had designed a false news identification model utilizing Bi-LSTM and RNN. Misinformation is frequently created to entice and mislead readers for political and commercial purposes.

According to the above discussed literatures, the existing strategies still contain several difficulties, such as certain problems of poor accuracy, rapid propagation, access method and high cost [5, 6]. The poor accuracy is attained in many research according to many reasons, such as insignificant selection of features, imbalanced dataset, inefficient tuning of parameters and so on [7]. These difficulties arise during detecting fake news in social media. To deal with these issues, the proposed model utilizes a hybrid BERT-SVM classification strategy to predict the fake news very accurately compared to existing strategies.

Proposed methodology for fake news prediction

A light weight recommendation system for social media networking analysis using a hybrid BERT-SVM classification model is designed to predict whether the information is fake or real. In this proposed model, a hybrid BERT-SVM model is used to classify and recommend the information to user whether it is fake or real. It is illustrated in Fig. 1. The proposed architecture of fake news detection model comprises three phases, such as preprocessing, feature selection and classification for effective fake news prediction. Initially the user dataset is collected from twitter social media. The raw data is preprocessed in this phase which involved splitting, stop word removal, lemmatization and spell correction. The second phase uses the TF-IDF for converting the text data into meaningful representation of binary data which is utilized to fit the classifier for effective

prediction. At the third phase, these features are given to the input of the hybrid BERT-SVM classification model and it produced two classes which is either real or fake. This suggestion will be recommended on the particular tweet to the user helping aware the information about the fake news.

Data Gathering

A real time user data is collected from twitter social media to make a dataset. The dataset is entirely based on Covid-19 pandemic as well as the twitter dataset contains the corresponding label post ID and the posts which are either fake or real news on the COVID-19 pandemic¹. It comprises 5000 manually gathered twitter commands equally on both fake and real news. These data are given into preprocessing for converting the raw data into a specified format, i.e., into the system readable format.

Data Pre-processing

The preprocessing phase is performed in four ways: stop word removal, splitting, spell correction and lemmatization. In this process, you need to eradicate the noise in the twitter dataset by normalizing or eliminate the unwanted data.

Splitting. Initially, in this stage, text sentences are split into individual words. String splitting is the technique of systematically dividing a text string into separate components that can be processed.

Stop word removal. In Natural Language Processing (NLP), stop word removal is a regularly used method. For instance, stop words such as ‘a’, ‘the’, ‘an’ are eliminating ones, the words that are often appearing in large numbers across all the documents, and it allows applications to focus on the important words instead.

Lemmatization. The algorithmic process of determining the lemma of a word based on its meaning is known as lemmatization. Lemmatization is a term that refers to accomplishing things correctly by analyzing words morphologically and using a vocabulary. Lemmatization is a term that relates to doing things correctly using a vocabulary and morphological analysis of words, with the goal of removing inflectional endings only and returning the base or dictionary form of a word, known as the lemma.

Spell correction. Spell checker is analyzing each and every word with thousands of proper spell words. The majority of techniques uses data from many sources of noisy as well as correct word mappings as training data for automatic spelling correction.

TF-IDF Feature Extraction

TF-IDF is a significant feature extraction and selection strategy used in text feature extraction. TF-IDF is used to find the significant features of the sentences that comprise the words and remove the bag incompetence of words strategies [12]. It is superior for text classification and is helpful for a machine reading words in numbers. One of the weight methods is Term Frequency (TF) that is used to determine the number of instances of term, i.e. word in a document. The Inverse Document Frequency (IDF) also presents the weight technique that used to find incidence terms in individual documents. The TF and IDF weight of the documents can be estimates using the given below expression.

¹ <https://www.kaggle.com/c/sentiment-analysis-of-covid-19-related-tweets/> (accessed: 10.05.2022).

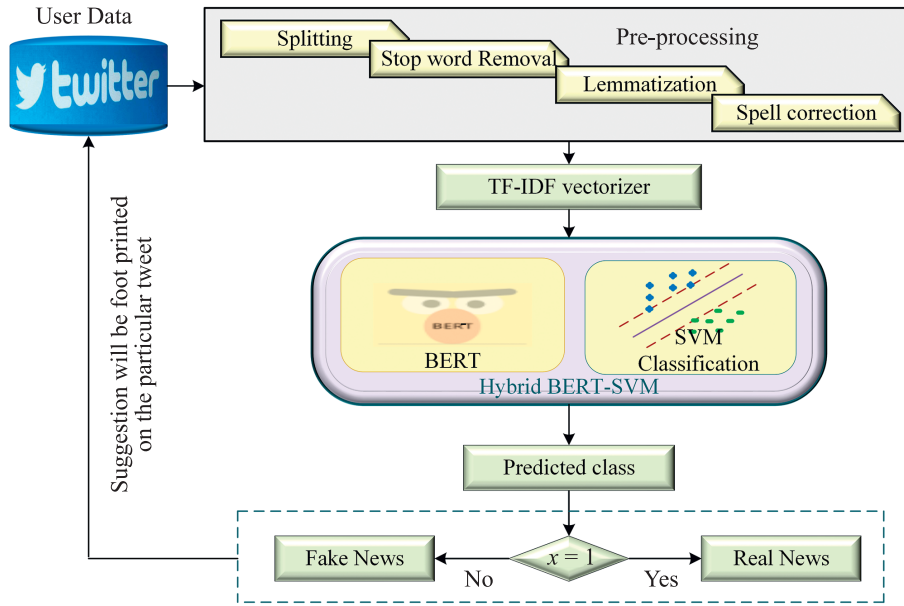


Fig. 1. Proposed architecture of social media analysis using a Hybrid BERT-SVM classifier, x is the class index

$$W_{dt} = TF(i, j) \times IDF(i).$$

The frequency of a feature appearing in a document in relation to the total number of features appearing in the text document is defined by TF. Simultaneously, IDF assesses a feature capacity to differentiate between categories. The categories in this case are the class label declared in the text documents. The following expressions are used to calculate TF and IDF:

$$TF(i, j) = \frac{\text{Term } i \text{ frequency in document } j}{\text{Total words in document } j},$$

$$IDF(i) = \log_2 \frac{\text{Total document}}{\text{document with term } i},$$

where t is the term and j is the counting number of the document. The following expression computes the weight of each phrase using the TF-IDF

$$w_{i,j} = tf_{i,j} \times \log \left(\frac{N}{df_i} \right),$$

where $tf_{i,j}$ is the number of occurrence of i in j ; df_i is the counting number of the document containing i ; and N is the total number of documents.

Hybrid BERT-SVM for Fake News Detection

This proposed method is designed to predict the fake news in the twitter social media using hybrid BERT-SVM classification. Hence, the detail description of this hybrid classification approach is given below.

1) BERT

BERT is based on transformer encoded architecture which is one of the most significant words embedding model in sentimental analysis. An effective content illustration of a sentence is attained using a BERT sentence encoder. Mask Language Model (MLM) is used in the BERT encoder to eliminate the unidirectional restriction. Usually it masks several token in a random manner, and original vocabulary of the masked ID that is predicted

based on the word. Moreover, compared to existing embedding strategies, the BERT can outperform because MLM can improve the BERT ability. It has the ability to handle unlabeled text by training together on both right context and left context in each layer [13]. Fig. 2 illustrates the basic BERT model for categorization.

BERT is regarded as the most advanced NLP technology. Pre-training and fine-tuning are the two phases in the BERT framework. At the initial stage, during pre-training in the model, the unlabeled large corpus is involved. At the second stage, labeled data is used to fine-tune every parameter for particular tasks. BERT's encoder is a multi-layer bidirectional Transformer encoder with multiple layers. This encoder consists of a stack of $N = 6$ identical layers, each with two sub-layers. The first layer is a basic position-wise completely connected feedforward network, and the second layer is a multi-head self-attention mechanism. Both sub-layers use a residual connection followed by layer normalization. Each sub-layer output is represented as $\text{Layer Norm}(x + \text{Sublayer}(x))$, where the sub-layer exaction function is denoted as $\text{Sublayer}(x)$. Before estimating multi-head self-attention, scaled dot product attention is required to define the following [14]:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax} \left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}} \right) \mathbf{V},$$

where \mathbf{Q} , \mathbf{K} and \mathbf{V} are three matrices \mathbf{Q} (Query), \mathbf{K} (Key) and \mathbf{V} (value). All come from the same input. To get the weight value, the Softmax operation is performed to normalize the output to a probability distribution for each row and then multiply it by the matrix \mathbf{V} . d_k is the dimension of the \mathbf{Q} and \mathbf{K} matrices and queries matrix is represented as \mathbf{Q} . The multi-head attention function can be expressed as

$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^0,$$

where $\text{head}_i = \text{Attention}(\mathbf{Q}W_i^Q, \mathbf{K}W_i^K, \mathbf{V}W_i^V)$. Multi-head attention contain estimation of queries, keys and

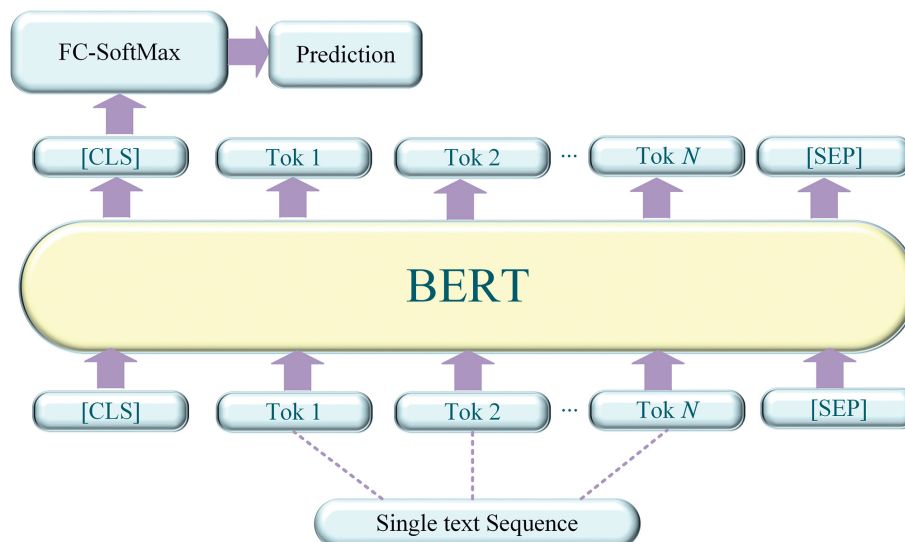


Fig. 2. BERT model for categorization, where [CLS] is BERT special classification token (Tok) and [SEP] is separation token

values; here these are denoted as \mathbf{Q} , \mathbf{K} , \mathbf{V} and h is the repetition elements factor with different dimension of the value matrix such as d_q , d_k and d_v . Then, the attention function is performed in parallel with each of predicted versions of queries, keys and values as well as the findings of output values is d_v . At the end, these parameters are concatenated and estimated to produce the outcome of multihued function. BERT indicates a pair of sentences or a single sentence as a series of tokens based on BERT's use of Word Piece embedding. [CLS] is considered as a first classification token in the sequence. A pair of sentences is separated by [SEP] token in the sequence. Pre-training contains two phases: Masked LM and Next Sentence Prediction (NSP). The Masked LM involves randomly masking the input tokens like utilizing the [MASK] token. After performed masking, it starts predicting masked tokens. The second one is NSP that considers two sentences A and B. B is the true next statement which appears 50 % of the time after A (considered as IsNext). 50 % of the time, B is a random text from the corpus (considered

as NotNext). Because the Transformer's self-attention mechanism allows BERT to model multiple downstream processes, fine-tuning is simple. Then, just feed the specific inputs and outputs into BERT for each task and fine-tune all of the settings.

2) SVM

SVM is a machine learning technique for evaluating data and detecting patterns in order to classify what follows as a result of deciding which of two classes to assign a set of input data. A training set is necessary for learning SVMs, with each element of the set having an indication of a class it belongs to. The data from the training set are separated from each other by a border with the widest feasible margin, i.e. the distance from this hyperplane, in the SVM model. In binary classification, the SVM is a popular learning method. The primary goal is to identify the optimal hyperplane for separating data into its two classes. Multiclass classification has recently been achieved by combining multiple binary SVMs [15]. The basic architecture of SVM classifier is given in Fig. 3.

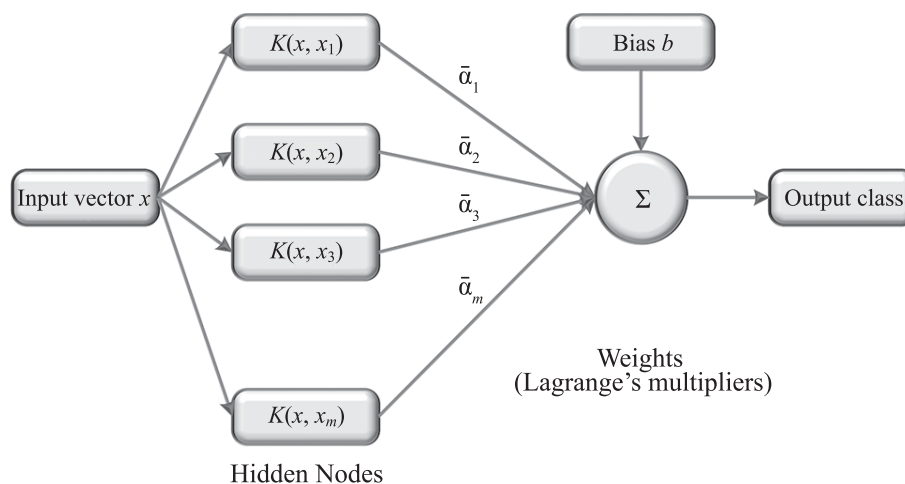


Fig. 3. Structure of SVM, where $K(u, v)$ is a kernel function satisfying Mercer's condition; x is a point of hyperplane; x_i are support vectors; $\bar{\alpha}_i$ are Lagrange's multipliers; m is SVM hyperplane dimension; and b is a bias

Let x_i is considered as input; i is the training instances $\{x_i, y_i\}$, $i = 1, \dots, l$; and each occurrence consists of x_i as well as label $y_i \in \{-1, 1\}$ [16]. A weight vector (w) and bias (b) both are used to parameterize every hyperplane which is expressed as given below:

$$w \times x + b = 0,$$

where x is a point lying on the hyperplane. To define the hyperplane function that classifies training and testing data, the following can be used

$$f(x) = \text{sign}(w \times x + b).$$

The analysis has so far been limited to the case when the training data can be separated linearly. Prior function can be given as equation when dealing with kernel

$$f(x) = \text{sign}\left(\sum_{i=1}^N \alpha_i K(x_i, x) + b\right),$$

where x_i is the input of training instance; y_i is its corresponding class label; b is a bias; N is the number of training instances; and $K(x_i, x)$ is the used kernel function which maps the input vectors into an expanded features space. The coefficients α_i are attained subject to two constraints expressed as

$$0 \leq \alpha_i, i = 1, \dots, N$$

$$\sum_{i=1}^N \alpha_i y_i = 0.$$

Except for a change in the bounds of the Lagrange multipliers, the solution to this minimization issue is similar to that of the separable case.

Hybrid BERT-SVM for Fake News Prediction

In this section, the proposed approach uses both BERT and SVM classifiers in a hybrid manner. BERT model is mainly selected for solving the difficulties of NLP models. BERT working is based on the transformers and is developed for text classification that is also termed as word embedding model in sentimental analysis. Moreover, this BERT approach performs NLP tasks as well as natural language understanding tasks. This strategy improves the ability of NLP models to implement data without having to maintain any order sequences. The BERT model input embedding contains tokens, segments and position components. SVM is a one of the significant machine learning approaches that able to handle high dimensional data as well as it attained better performance when using text classification. The proposed hybrid BERT-SVM architecture for expectations E_i is shown in Fig. 4. The extracted features are given to the proposed hybrid BERT-SVM classifier. Initially, the feature selected dataset with 80 % of twitter commands is given for training the classifier. The BERT model usually performs feature extraction as well as usual classification, but the proposed model uses the BERT model only for usual classification. To improve the BERT model, the SVM model is included in the FC-Softmax layer of the BERT.

Algorithm 1: Pseudo Code for Social Networking Analysis using a hybrid BERT-SVM

Input: A =Dataset; $X1$ =Splitting; $X2$ =Stop word removal; $X3$ =Lemmatization; $X4$ =Spell correction; Q =Feature extraction; P =Classification

Output: R =Fake news

Input dataset $=A$, pre-processed data $=Z$, User $=M$

Pre-processing

$X1 = \text{Splitting}(A)$ # splits combined words are separated for individual processing

$X2 = \text{Stop word removal}(X1)$ # stop words such as 'the', 'an', 'a' are removed

$X3 = \text{Lemmatization}(X2)$ # Grouping of words together with different derivatives

$X4 = \text{Spell correction}(X3)$ # Spell correction of wrong words

feature extraction

$Q = \text{TD-IFD}(X4)$ # convert text data into binary data

classification

$P = \text{BERT-SVM}(Q)$ # classification utilizing Hybrid BERT-SVM

compare predicted data with pre-processed data

If $(P = Z) \rightarrow$ real news; Else \rightarrow fake news

predicted fake news is given as a recommendation to the user

$R = \text{fake news}; M = R$

Output= detect either real or fake news in the tweet information

The BERT model comprises two phases such as pre-training and fine-tuning. Initial stage of the classifier trained the information among various pre-trained problems using twitter dataset. This research uses the Base BERT model, which consists of 12 bidirectional self-attention heads and 12 encoder layer blocks. The model assumes a series of 512 tokens, with hidden vectors in the sequence being emitted. The final layer of BERT model is FC-Softmax layer. The proposed model uses the SVM classifier in FC-Softmax layer for accurate prediction outcome. After training the classifier, remaining 20 % of twitter commands are subjected into testing of the trained model. The classifier produces two classes such as 0 and 1. Here, class 0 is considered as fake news and class 1 is considered as real news. After predicted data, i.e. real or fake, this suggestion will be foot printed on the particular tweet to aware the information about COVID-19 in the twitter social media. In this way, the proposed model effectively predicts the news of its truth in twitter about COVID-19 pandemic. Algorithm 1 illustrates the pseudo code for network analysis using a hybrid BERT-SVM.

Result and discussion

The proposed light weight recommendation system for social media networking analysis using a hybrid BERT-SVM classification model is implemented in MATLAB software to validate its performance. The proposed system is tested on Intel(R) Core i5-10300H CPU, 4 GB Nvidia GeForce GTS 1650 GPU, with 16.0 GB RAM. The dataset is gathered from twitter social media¹. It includes 5000 manually gathered twitter commands included both fake and real news. During preprocessing the sentences

¹ <https://www.kaggle.com/c/sentiment-analysis-of-covid-19-related-tweets/> (accessed: 10.05.2022).

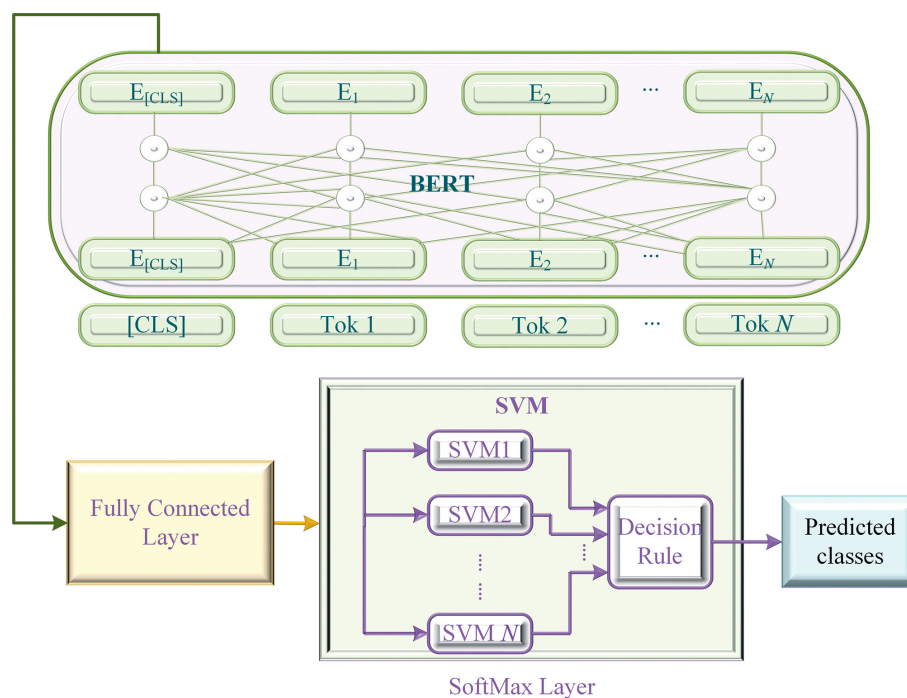


Fig. 4. Proposed Hybrid BERT-SVM Architecture

are extracted by using splitting, stop word removal, lemmatization, and spell correction. After that, the extracted features are given into TF-IDF that converts the text data into meaningful illustration of binary data. The extracted 3500 TF-IDF features are given to proposed Hybrid BERT-SVM classification for training. After training, the remaining 1500 features are utilized for testing.

Table 1 illustrates the extraction of input in the preprocessing phase; the data are initially fed into the splitting stage where you can split each and every word. After splitting process, the data are processed into stop

word removal, in this step the stop words are to be removed, such as 'the', 'an', 'a', 'if', 'are', etc. Then, the next stage is lemmatization; its aim is to take away inflectional suffixes and prefixes to bring out the word dictionary form. Finally, the data are spell corrected and the preprocessed findings produced. These performance metrics are estimated for the proposed BERT-SVM and existing fake news detection model in twitter social media. Several existing techniques based on fake news detection models are considered for comparison analysis, such as CNN-SVM, bidirectional GRNN, LSTM, CRNN, and SVM. Table 2 illustrates the

Table 1. Sample Input and Output attained in pre-processing

Input	Splitting	Stop word	Lemmatize	Spell correction
'BanMediaHouse whose is responsible for spreading Fake and communal	"BanMediaHouse" "whose" "is" "responsible" "for" "spreading" "Fake" "and" "communal"	"BanMediaHouse" "whose" "responsible" "spreading" "Fake" "communal"	"banmediahouse" "whose" "responsible" "spread" "fake" "communal"	"househusband whose responsible spread fake communal"

Table 2. Parameters estimated for proposed and existing techniques in fake news detection

Parameters, %	BERT-SVM (proposed)	CNN-SVM	GRNN	LSTM	CRNN	SVM
Accuracy	98	90	85	70	68	57
Error	2	10	15	30	31	45
Sensitivity	98	93	89	91	77	60
Specificity	99	92	86	78	61	60
Precision	99	95	82	87	78	55
False Positive Rate, FPR	2	1	15	27	32	35
F1_Score	99	81	72	75	60	58
Kappa	90	88	80	61	62	58

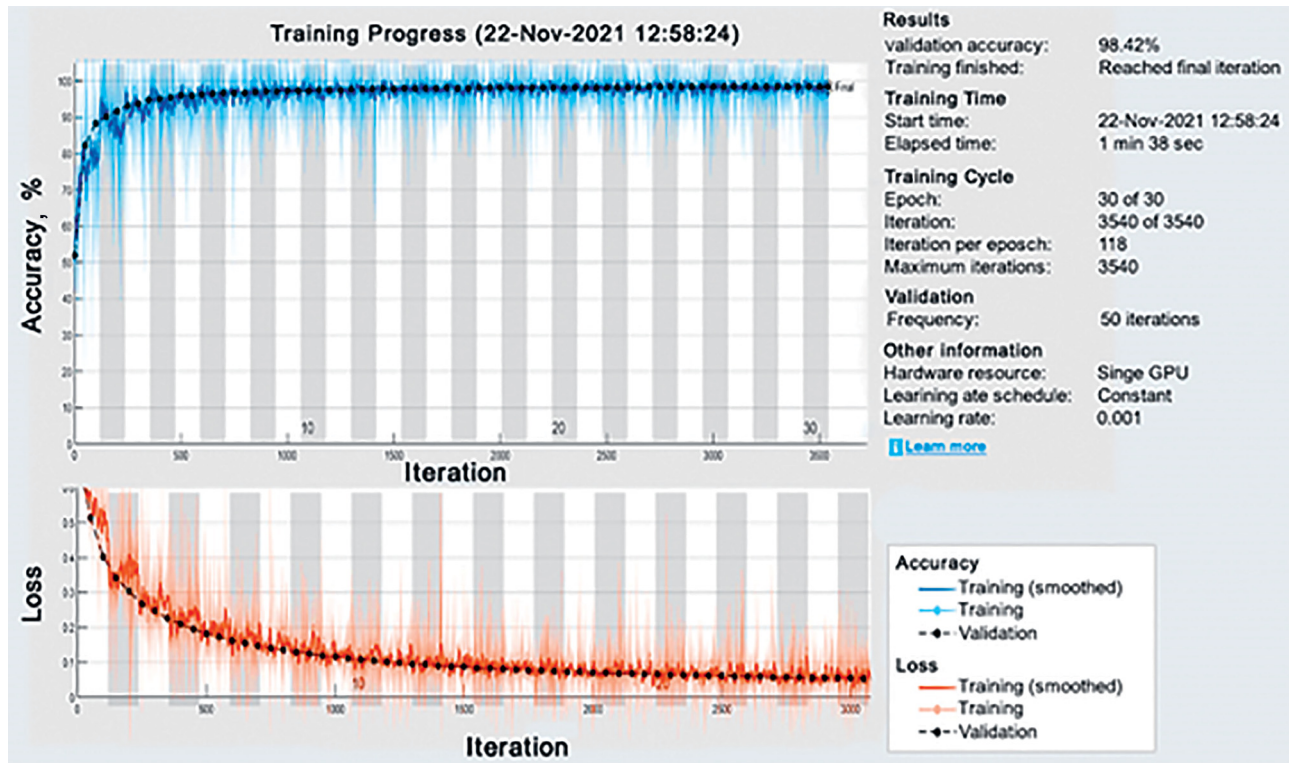


Fig. 5. Training process of accuracy and losses

parameter values estimated for the proposed and existing techniques for fake news detection.

Fig. 5 displays the accuracy and losses of the training process. The accuracy of the proposed method is 100 %, since it could more rapidly detect the error and eliminate it more effectively. The error of the proposed method is 0 % because the hybrid BERT-SVM classification model attained accurate prediction. Fig. 6 illustrates the confusion matrix obtained in this proposed model. Here, the X and Y labels represent the predicted class and the true class. The class 0 predicts 4298 data and class 1 predicts 3194 data. To provide an analytical assessment, a confusion matrix and performance measurement like accuracy, precision and so on is used.

The comparison analysis based on accuracy (%) among the proposed and existing strategies depended on fake news

detection on social media is shown in Fig. 6. The accuracy for the proposed method is 98 % which is greater compared to existing methods: CNN-SVM is 90 %, GRNN is 85 %, LSTM is 70 %, CRNN is 68 %, and SVM is 57 %. This can show that the performance of the proposed model works better compared to others. The comparison analysis based on error among the proposed and existing strategies depended on fake news detection on social media is shown in Table 2. The error found for the proposed method is 2 % which is less compared to existing methods: CNN-SVM is 10 %, GRNN is 15 %, LSTM is 30 %, CRNN is 31 % and SVM is 45 %.

This can show that the performance of the proposed model works better compared to others. Similarly, the comparison of performances, such as sensitivity, specificity, precision, FPR, F1_Score and kappa of the proposed and existing strategies, is given in Table 2. It clearly shows that the proposed technique has better performance than the other techniques. This can show that the performance of the proposed model works better compared to others.

Conclusion

A lightweight recommendation system is proposed for social networking analysis using a hybrid BERT-SVM classifier algorithm to improve the accuracy efficiently. Initially, a real-time dataset was collected from Twitter social media. These data are given into preprocessing, and splitting, stop word removal, lemmatization and spell correction have been performed. An effective feature extraction strategy is utilized for text feature selection and binary conversion. The converted features are classified with the hybrid BERT-SVM model. The predicted news

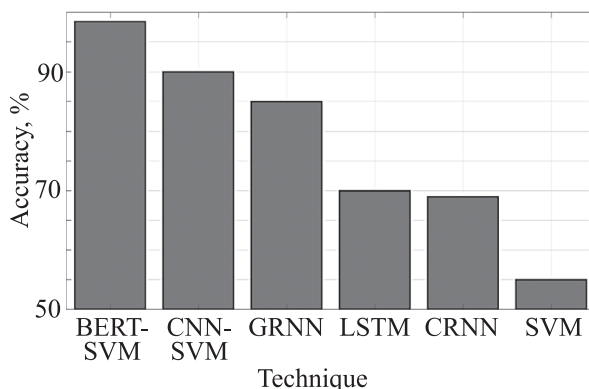


Fig. 6. Comparison of accuracy among proposed and existing approaches

is given as a recommendation to the user to provide the awareness of the information about the fake news. The proposed model is executed on MATLAB for finding the performance metrics: accuracy is 98 %, the error is 2 %, sensitivity is 98 %, specificity is 99 %, precision is 99 %, FPR is 2 %, F1_Score is 99 %, and kappa is 90 %. The overall expected outcome of the recommendation system using hybrid BERT-SVM is turned out to be

better compared to the existing techniques such as CNN-SVM, GRNN, LSTM, CRNN, and SVM. The proposed social networking analysis model delivers effective fake news detection that can be utilized to detect the Twitter comments related to the COVID-19 pandemic, either real or fake, and it is possible to suggest that the predicted news is given as a recommendation to the user to get acquainted with information about the fake news.

References

1. Kaur S., Kumar P., Kumaraguru P. Automating fake news detection system using multi-level voting model. *Soft Computing*, 2020, vol. 24, no. 12, pp. 9049–9069. <https://doi.org/10.1007/s00500-019-04436-y>
2. Kaliyar R.K., Goswami A., Narang P., Sinha S. FNDNet — a deep convolutional neural network for fake news detection. *Cognitive Systems Research*, 2020, vol. 61, pp. 32–44. <https://doi.org/10.1016/j.cogsys.2019.12.005>
3. Shim J.-S., Lee Y., Ahn H. A link2vec-based fake news detection model using web search results. *Expert Systems with Applications*, 2021, vol. 184, pp. 115491. <https://doi.org/10.1016/j.eswa.2021.115491>
4. Umer M., Imtiaz Z., Ullah S., Mehmood A., Choi G.S., On B.-W. Fake news stance detection using deep learning architecture (CNN-LSTM). *IEEE Access*, 2020, vol. 8, pp. 156695–156706. <https://doi.org/10.1109/ACCESS.2020.3019735>
5. Hakak S., Alazab M., Khan S., Gadekallu T.R., Maddikunta P.K.R., Khan W.Z. An ensemble machine learning approach through effective feature extraction to classify fake news. *Future Generation Computer Systems*, 2021, vol. 117, pp. 47–58. <https://doi.org/10.1016/j.future.2020.11.022>
6. Abdullah, Yasin A., Avan M.J., Shehzad M.F., Ashraf M. Fake news classification bimodal using convolutional neural network and long short-term memory. *International Journal on Emerging Technologies*, 2020, vol. 11, no. 5, pp. 209–212.
7. Huang Y.-F., Chen P.-H. Fake news detection using an ensemble learning model based on self-adaptive harmony search algorithms. *Expert Systems with Applications*, 2020, vol. 159, pp. 113584. <https://doi.org/10.1016/j.eswa.2020.113584>
8. Paka W.S., Bansal R., Kaushik A., Sengupta S., Chakraborty T. Cross-SEAN: A cross-stitch semi-supervised neural attention model for COVID-19 fake news detection. *Applied Soft Computing*, 2021, vol. 107, pp. 107393. <https://doi.org/10.1016/j.asoc.2021.107393>
9. Nasir J.A., Khan O.S., Varlamis I. Fake news detection: A hybrid CNN-RNN based deep learning approach. *International Journal of Information Management Data Insights*, 2021, vol. 1, no. 1, pp. 100007. <https://doi.org/10.1016/j.jjime.2020.100007>
10. Sabeeh V., Zohdy M., Mollah A., Al Bashaiah R. Fake news detection on social media using deep learning and semantic knowledge sources. *International Journal of Computer Science and Information Security (IJCSIS)*, 2020, vol. 18, no. 2, pp. 45–68.
11. Bahad P., Saxena P., Kamal R. Fake news detection using bi-directional LSTM-recurrent neural network. *Procedia Computer Science*, 2019, vol. 165, pp. 74–82. <https://doi.org/10.1016/j.procs.2020.01.072>
12. Qaiser S., Ali R. Text mining: Use of TF-IDF to examine the relevance of words to documents. *International Journal of Computer Applications*, 2018, vol. 181, no. 1, pp. 25–29. <https://doi.org/10.5120/ijca2018917395>
13. Pota M., Ventura M., Catelli R., Esposito M. An effective BERT-based pipeline for Twitter sentiment analysis: a case study in Italian. *Sensors*, 2021, vol. 21, no. 1, pp. 133. <https://doi.org/10.3390/s21010133>
14. Malla S., Alphonse P.J.A. COVID-19 outbreak: An ensemble pre-trained deep learning model for detecting informative tweets. *Applied Soft Computing*, 2021, vol. 107, pp. 107495. <https://doi.org/10.1016/j.asoc.2021.107495>
15. Goudjil M., Koudil M., Bedda M., Ghogkali N. A novel active learning method using SVM for text classification. *International Journal of Automation and Computing*, 2018, vol. 15, no. 3, pp. 290–298. <https://doi.org/10.1007/s11633-015-0912-z>

Литература

1. Kaur S., Kumar P., Kumaraguru P. Automating fake news detection system using multi-level voting model // *Soft Computing*. 2020. V. 24. N 12. P. 9049–9069. <https://doi.org/10.1007/s00500-019-04436-y>
2. Kaliyar R.K., Goswami A., Narang P., Sinha S. FNDNet — a deep convolutional neural network for fake news detection // *Cognitive Systems Research*. 2020. V. 61. P. 32–44. <https://doi.org/10.1016/j.cogsys.2019.12.005>
3. Shim J.-S., Lee Y., Ahn H. A link2vec-based fake news detection model using web search results // *Expert Systems with Applications*. 2021. V. 184. P. 115491. <https://doi.org/10.1016/j.eswa.2021.115491>
4. Umer M., Imtiaz Z., Ullah S., Mehmood A., Choi G.S., On B.-W. Fake news stance detection using deep learning architecture (CNN-LSTM) // *IEEE Access*. 2020. V. 8. P. 156695–156706. <https://doi.org/10.1109/ACCESS.2020.3019735>
5. Hakak S., Alazab M., Khan S., Gadekallu T.R., Maddikunta P.K.R., Khan W.Z. An ensemble machine learning approach through effective feature extraction to classify fake news // *Future Generation Computer Systems*. 2021. V. 117. P. 47–58. <https://doi.org/10.1016/j.future.2020.11.022>
6. Abdullah, Yasin A., Avan M.J., Shehzad M.F., Ashraf M. Fake news classification bimodal using convolutional neural network and long short-term memory // *International Journal on Emerging Technologies*. 2020. V. 11. N 5. P. 209–212.
7. Huang Y.-F., Chen P.-H. Fake news detection using an ensemble learning model based on self-adaptive harmony search algorithms // *Expert Systems with Applications*. 2020. V. 159. P. 113584. <https://doi.org/10.1016/j.eswa.2020.113584>
8. Paka W.S., Bansal R., Kaushik A., Sengupta S., Chakraborty T. Cross-SEAN: A cross-stitch semi-supervised neural attention model for COVID-19 fake news detection // *Applied Soft Computing*. 2021. V. 107. P. 107393. <https://doi.org/10.1016/j.asoc.2021.107393>
9. Nasir J.A., Khan O.S., Varlamis I. Fake news detection: A hybrid CNN-RNN based deep learning approach // *International Journal of Information Management Data Insights*. 2021. V. 1. N 1. P. 100007. <https://doi.org/10.1016/j.jjime.2020.100007>
10. Sabeeh V., Zohdy M., Mollah A., Al Bashaiah R. Fake news detection on social media using deep learning and semantic knowledge sources // *International Journal of Computer Science and Information Security (IJCSIS)*. 2020. V. 18. N 2. P. 45–68.
11. Bahad P., Saxena P., Kamal R. Fake news detection using bi-directional LSTM-recurrent neural network // *Procedia Computer Science*. 2019. V. 165. P. 74–82. <https://doi.org/10.1016/j.procs.2020.01.072>
12. Qaiser S., Ali R. Text mining: Use of TF-IDF to examine the relevance of words to documents // *International Journal of Computer Applications*. 2018. V. 181. N 1. P. 25–29. <https://doi.org/10.5120/ijca2018917395>
13. Pota M., Ventura M., Catelli R., Esposito M. An effective BERT-based pipeline for Twitter sentiment analysis: a case study in Italian // *Sensors*. 2021. V. 21. N 1. P. 133. <https://doi.org/10.3390/s21010133>
14. Malla S., Alphonse P.J.A. COVID-19 outbreak: An ensemble pre-trained deep learning model for detecting informative tweets // *Applied Soft Computing*. 2021. V. 107. P. 107495. <https://doi.org/10.1016/j.asoc.2021.107495>
15. Goudjil M., Koudil M., Bedda M., Ghogkali N. A novel active learning method using SVM for text classification // *International Journal of Automation and Computing*. 2018. V. 15. N 3. P. 290–298. <https://doi.org/10.1007/s11633-015-0912-z>
16. Zhu J., Tian Z., Kübler S. UM-IU@LING at SemEval-2019 task 6: Identifying offensive tweets using BERT and SVMs // *Proceedings*

16. Zhu J., Tian Z., Kübler S. UM-IU@LING at SemEval-2019 task 6: Identifying offensive tweets using BERT and SVMs. *Proceedings of the 13th International Workshop on Semantic Evaluation*, 2019, pp. 788–795. <https://doi.org/10.18653/v1/s19-2138>

of the 13th International Workshop on Semantic Evaluation. 2019. P. 788–795. <https://doi.org/10.18653/v1/s19-2138>

Authors

Nallichery Subramanian Kiruthika — Research Scholar, Vels University, Chennai, Tamil Nadu, 600117, India, [sc 55420781200](https://orcid.org/0000-0001-6601-1341), <https://orcid.org/0000-0001-6601-1341>, sathishpoojaa5@gmail.com

Ganapathy Thailambal — Associate Professor, Department of Computer science, School of Computing Sciences, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai, Tamil Nadu, 600117, India, [sc 57189250428](https://orcid.org/0000-0002-0043-2415), <https://orcid.org/0000-0002-0043-2415>, thaila.scs@velsuniv.ac.in

Авторы

Кирутика Налличери Субраманиан — научный сотрудник, Университет Велса, Ченнаи, шт. Тамил Наду, 600117, Индия, [sc 55420781200](https://orcid.org/0000-0001-6601-1341), <https://orcid.org/0000-0001-6601-1341>, sathishpoojaa5@gmail.com

Тайламбал Ганапати — доцент, Институт науки, технологий и перспективных исследований Велса, Ченнаи, шт. Тамил Наду, 600117, Индия, [sc 57189250428](https://orcid.org/0000-0002-0043-2415), <https://orcid.org/0000-0002-0043-2415>, [thaila.scs@velsuniv.ac.in](mailto:scs@velsuniv.ac.in)

Received 05.01.2022

Approved after reviewing 17.06.2022

Accepted 30.07.2022

Статья поступила в редакцию 05.01.2022

Одобрена после рецензирования 17.06.2022

Принята к печати 30.07.2022



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

МАТЕМАТИЧЕСКОЕ И КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ MODELING AND SIMULATION

doi: 10.17586/2226-1494-2022-22-4-779-784

УДК 519.876.5+519.6+519.246

Моделирование случайных процессов на основе преобразования Карунена–Лозва

Александр Сергеевич Ефимов✉

Независимый исследователь, Санкт-Петербург, Российская Федерация

asefimov@mail.ru✉, <https://orcid.org/0000-0002-8091-435X>

Аннотация

Рассмотрена задача цифрового моделирования случайных процессов с заданной корреляционной функцией или спектральной плотностью процесса. Обе эти функции связаны между собой теоремой Винера–Хинчина, так что знание одной функции может быть использовано для нахождения другой. Выполнена разработка математического представления стационарного случайного процесса с заданной корреляционной функцией на основе преобразования Карунена–Лозва, которое чаще всего применяется для декорреляции исходного процесса с целью более лаконичного его описания (задача сжатия данных). Предложено использование преобразования Карунена–Лозва для придания требуемых корреляционных свойств исходному некоррелированному случайному процессу с помощью обращения (инвертирования) данного преобразования. Обоснован вид требуемого преобразования при дискретном (во времени) представлении входных и выходных процессов различной длины и методики обеспечения требуемой точности моделирования. Представлена цифровая процедура получения корреляционной функции по заданной спектральной плотности моделируемого случайного процесса. Экспериментальное исследование предлагаемого метода проведено в ходе компьютерного моделирования в пакете Mathcad, что упростило решение требуемых расчетных задач. Исходный случайный процесс получен в виде последовательности независимых (и, следовательно, некоррелированных) случайных чисел, а выходной — как результат полученного в работе преобразования. Выполнено сравнение вычисленной приближенной корреляционной функции с заданной, и определена дисперсия погрешности. Приведены результаты моделирования случайных процессов с заданными корреляционными функциями и однородного марковского процесса с заданной переходной вероятностью, а также пример перехода от заданной спектральной плотности случайного процесса к его корреляционной функции. Полученные результаты подтверждают эффективность и реализуемость разработанных методов моделирования, что позволит использовать их при компьютерном исследовании и проектировании различных систем.

Ключевые слова

моделирование, случайный процесс, корреляционная функция, матрица, собственные числа и векторы, преобразование

Ссылка для цитирования: Ефимов А.С. Моделирование случайных процессов на основе преобразования Карунена–Лозва // Научно-технический вестник информационных технологий, механики и оптики. 2022. Т. 22, № 4. С. 779–784. doi: 10.17586/2226-1494-2022-22-4-779-784

Modeling of random processes based on Karhunen-Loeve decomposition

Alexandr S. Efimov✉

Independent Researcher, Saint Petersburg, Russian Federation

asefimov@mail.ru✉, <https://orcid.org/0000-0002-8091-435X>

Abstract

The problem of digital modeling of random processes with given either correlation function or spectral density of the process is considered. These functions of a random process are interconnected by the Wiener–Khinchin theorem. The solution of one function can be used to solve another. The development of a mathematical representation of a stationary

random process with a given correlation function based on the Karhunen-Loeve transformation, which is most often used to decorrelate the original process in order to describe it more concisely (data compression problem), has been completed. It is proposed to use the Karhunen-Loev transformation to impart the required correlation properties to the original uncorrelated random process by inverting (converting) this transformation. The form of the required transformation for a discrete (in time) representation of input and output processes of various lengths and methods for ensuring the required modeling accuracy are substantiated. A procedure for obtaining a correlation function from a given spectral density of a simulated random process is presented. An experimental study of the proposed method was carried out in the course of computer simulation in the Mathcad package which simplified the solution of the required computational problems. The initial random process was obtained as a sequence of independent (and, therefore, uncorrelated) random numbers, and the output process, as a result of the transformation, was obtained in the work. The calculated approximate correlation function is compared with the given one and the error variance is determined. The results of modeling random processes with given correlation functions and a homogeneous Markov process with a given transition probability are given as well as an example of the transition from a given spectral density of a random process to its correlation function. The results obtained confirm the effectiveness and feasibility of the developed modeling methods which will allow them to be used in computer research and design of various systems.

Keywords

modeling, random process, correlation function, matrix, eigenvalues and vectors, transformation

For citation: Efimov A.S. Modeling of random processes based on Karhunen-Loeve decomposition. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2022, vol. 22, no. 4, pp. 779–784 (in Russian). doi: 10.17586/2226-1494-2022-22-4-779-784

Введение

Моделирование случайных процессов (СП) с требуемыми характеристиками — актуальная задача при компьютерном исследовании и проектировании систем автоматического управления, связи, передачи информации и других. Эта задача важна не только для имитации мешающих сигналов и помех, но и для моделирования речевых и видеосигналов, сигналов, используемых при управлении различными объектами и процессами, а также анализе социальных и экономических процессов как недетерминированных функций времени. Особенно важно моделирование марковских процессов, которые широко используются при исследовании и проектировании систем массового обслуживания.

Способы решения поставленной задачи рассмотрены в работе [1, п. 4.4]. Самое большое применение получили два метода: формирующего фильтра и канонических разложений. Метод формирующего фильтра подробно рассмотрен в работе [2], где моделируемый СП с заданной корреляционной функцией получают на выходе формирующего фильтра. При этом дискретная передаточная функция вычисляется по спектральной плотности процесса с помощью уравнения Винера–Хинчина. Входным процессом является некоррелированный СП («белый шум»). Требуемый выходной процесс представлен в виде конечно-разностного уравнения, коэффициенты которого вычислены с помощью теории непрерывных дробей (модифицированного алгоритма Вискватова). Отметим, что предложенный в работе [2] алгоритм моделирования состоит из 12 шагов и требует выполнения большого объема подготовительных расчетов, в том числе ручных, что затрудняет его практическое использование. Кроме того, остается неясным, как обеспечивается требуемая точность моделирования, а указанные дисперсии ошибок [2, примеры 1 и 4] не подтвердились при проверке алгоритма в пакете Mathcad.

В работах [3, 4] рассмотрена задача построения приближенной модели стохастического процесса с помощью преобразования Карунена–Лозва. Главное

внимание уделено методике приближенных вычислений собственных чисел и векторов однородного интегрального оператора Фредгольма второго рода, ядром которого является заданная корреляционная функция моделируемого СП. В [3] получены требования к длине моделируемого процесса, обеспечивающей требуемую точность и надежность вычисленных собственных чисел и векторов. Однако методика применения найденного преобразования для получения случайного процесса с заданной корреляционной функцией в данной теоретической работе не рассмотрена. В [4] исследовано усеченное представление СП с использованием преобразования Карунена–Лозва, уменьшающее объем представления (задача сжатия данных). При этом рассмотрены аналитические и численные методы поиска собственных чисел и векторов основных видов корреляционных функций СП. Главное внимание в работе [4] уделено влиянию усечения разложения Карунена–Лозва на точность представления исходного СП, но задача получения СП с заданной корреляционной функцией также не была поставлена.

Предложенные в настоящей работе методы моделирования рассчитаны на использование современного программного обеспечения, позволяющего решить главные проблемы моделирования. К таким проблемам относятся: вычисление собственных чисел и векторов заданной корреляционной функции, получение реализаций требуемого СП нужной длины, их статистическая обработка и оценка качества моделирования. В процессе моделирования использован математический пакет Mathcad.

Алгоритм моделирования

Рассмотрим процесс декорреляции стационарного в широком смысле СП $Y(k)$ с дискретным временем с помощью преобразования Карунена–Лозва [5]. Если \mathbf{R}_Y — корреляционная матрица этого процесса, а \mathbf{W} — матрица ее собственных векторов, то в результате линейного преобразования получим некоррелированный СП $X(k)$:

$$\mathbf{X} = \mathbf{W}^T \cdot \mathbf{Y}, \quad (1)$$

где \mathbf{X} и \mathbf{Y} — векторы реализаций соответствующих СП; \mathbf{W}^T — транспонированная матрица.

Рассмотрим подробнее процесс декорреляции (1). Если коэффициент корреляции процесса $Y(k)$ задан в виде функции $r(m)$, $m \in \mathbb{N}$, где m — расстояния между двумя сечениями процесса, то $r(0) = 1$, $r(m) = r(-m)$, $|r| \leq 1$. Тогда при $m = |i - j|$, $i, j = 1, N$, где N — длина выбранного участка корреляционной функции, получим матрицу коэффициентов корреляции \mathbf{R}_Y , которая является и корреляционной, если дисперсия СП $Y(k)$ равна единице:

$$\mathbf{R}_Y = \begin{bmatrix} 1 & r(1) & r(2) & \dots & r(N-1) \\ r(1) & 1 & r(1) & \dots & r(N-2) \\ \dots & \dots & \dots & \dots & \dots \\ r(N-1) & r(N-2) & r(N-3) & \dots & 1 \end{bmatrix}, \quad (2)$$

где $R_{Y_{ii}} = r(0) = 1$, $R_{Y_{ij}} = R_{Y_{ji}} = r(|i - j|)$, $|r| \leq 1$; \mathbf{R}_Y — квадратная теплицева матрица размера N (вид матрицы представлен в [6, С. 206]). Матрица (2) симметрична и положительно определена, поэтому собственные числа λ_i — вещественны и положительны. Собственные числа связаны с собственными векторами \mathbf{w}_i матрицы \mathbf{R}_Y равенствами:

$$\mathbf{R}_Y \cdot \mathbf{w}_i = \lambda_i \cdot \mathbf{w}_i, \quad i = \overline{1, N},$$

которые в матричной форме имеют вид

$$\mathbf{R}_Y \cdot \mathbf{W} = \mathbf{W} \cdot \mathbf{\Lambda}, \quad (3)$$

где \mathbf{W} — матрица собственных векторов размера $(N \times N)$; $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_N)$, причем $\sum \lambda_i = \text{Tr}(\mathbf{R}_Y) = N$ — сумма диагональных элементов матрицы (2).

Из (3) получим уравнения для матрицы $\mathbf{\Lambda}$ и корреляционной матрицы \mathbf{R}_Y :

$$\mathbf{\Lambda} = \mathbf{W}^{-1} \cdot \mathbf{R}_Y \cdot \mathbf{W}, \quad (4)$$

$$\mathbf{R}_Y = \mathbf{W} \cdot \mathbf{\Lambda} \cdot \mathbf{W}^{-1}. \quad (5)$$

Собственные векторы взаимно ортогональны, их скалярное произведение $(\mathbf{w}_i, \mathbf{w}_j) = \delta_{ij}$, где δ_{ij} — символ Кронекера. Матрица собственных векторов обладает важным свойством: $\mathbf{W}^{-1} = \mathbf{W}^T$. Заметим, что матрицы \mathbf{W} и $\mathbf{\Lambda}$ вычисляются с помощью встроенных программ математических пакетов. Например, в пакете Mathcad с помощью команд: $\mathbf{W} = \text{eigenvecs}(\mathbf{R}_Y)$ и $\mathbf{\Lambda} = \text{eigenvals}(\mathbf{R}_Y)$.

Найдем корреляционную матрицу полученного с помощью (1) СП $X(k)$, используя уравнение (4) и равенство $\mathbf{W}^{-1} = \mathbf{W}^T$:

$$\begin{aligned} \mathbf{R}_X &= M\{\mathbf{X} \cdot \mathbf{X}^T\} = M\{\mathbf{W}^T \cdot \mathbf{Y} \cdot \mathbf{Y}^T \cdot \mathbf{W}\} = \\ &= \mathbf{W}^{-1} \cdot M\{\mathbf{Y} \cdot \mathbf{Y}^T\} \cdot \mathbf{W} = \mathbf{W}^{-1} \cdot \mathbf{R}_Y \cdot \mathbf{W} = \mathbf{\Lambda}, \end{aligned}$$

где M — знак математического ожидания.

Отметим, что диагональность матрицы $\mathbf{\Lambda}$ означает некоррелированность СП $X(k)$, причем ее элементы,

являясь дисперсиями проекций процесса $X(k)$ на координатные оси нового базиса ($\lambda_i = D_i$), не постоянны, а возрастают с ростом индекса i . Эти свойства процесса $X(k)$ играют главную роль в задачах сжатия данных.

Но для решения поставленной задачи — получение требуемого процесса $Y(k)$ с помощью СП «белый шум» — случайный процесс $\mathbf{X}(k)$ должен иметь корреляционную матрицу $\mathbf{R}_X = \mathbf{E}$, где \mathbf{E} — единичная матрица. Для обеспечения данного условия необходимо нормировать собственные векторы \mathbf{w}_i матрицы \mathbf{W} , т. е. привести их к виду \mathbf{w}_i/σ_i , $i = \overline{1, N}$, где $\sigma_i = \sqrt{D_i} = \sqrt{\lambda_i}$. В этом случае дисперсии проекций процесса $\mathbf{X}(k)$ на координатные оси будут равны единице.

Указанная нормировка обеспечивается умножением \mathbf{W} на матрицу \mathbf{B}^{-1} , $\mathbf{B} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_N)$, причем $\mathbf{B}^{-1} \cdot \mathbf{\Lambda} \cdot \mathbf{B}^{-1} = \mathbf{E}$. Тогда уравнение (1) примет вид:

$$\mathbf{X} = \mathbf{B}^{-1} \cdot \mathbf{W}^T \cdot \mathbf{Y}, \quad (6)$$

а корреляционная функция СП $X(k)$ с учетом (5) будет действительно равна \mathbf{E} :

$$\begin{aligned} \mathbf{R}_X &= M\{\mathbf{X} \cdot \mathbf{X}^T\} = M\{\mathbf{B}^{-1} \cdot \mathbf{W}^T \cdot \mathbf{Y} \cdot \mathbf{Y}^T \cdot \mathbf{W} \cdot \mathbf{B}^{-1}\} = \\ &= \mathbf{B}^{-1} \cdot \mathbf{W}^T \cdot M\{\mathbf{Y} \cdot \mathbf{Y}^T\} \cdot \mathbf{W} \cdot \mathbf{B}^{-1} = \mathbf{B}^{-1} \cdot \mathbf{W}^{-1} \cdot \mathbf{R}_Y \cdot \mathbf{W} \cdot \mathbf{B}^{-1} = \\ &= \mathbf{B}^{-1} \cdot \mathbf{W}^{-1} \cdot \mathbf{W} \cdot \mathbf{\Lambda} \cdot \mathbf{W}^{-1} \cdot \mathbf{W} \cdot \mathbf{B}^{-1} = \mathbf{E}. \end{aligned}$$

В результате искомого преобразование, позволяющее получить СП с заданной корреляционной функцией из «белого шума», может быть получено преобразованием уравнения (6) к виду:

$$\mathbf{Y} = \mathbf{W} \cdot \mathbf{B} \cdot \mathbf{X}. \quad (7)$$

Случайные последовательности с независимыми элементами и нулевым средним можно получить, используя датчики различных случайных чисел. Соседние отсчеты в таких последовательностях практически независимы, что иллюстрируется корреляционной функцией $R_z(m)$ последовательности 1000 нормально распределенных случайных чисел, полученных с помощью пакета Mathcad (рис. 1).

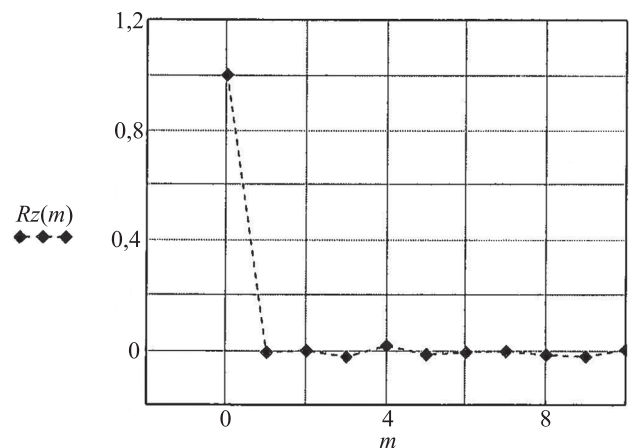


Рис. 1. Корреляционная функция «белого шума»
Fig. 1. Correlation function of “white noise”

Размеры векторов и матриц в (7) определяются числом N , которое выбирается так, чтобы заданная корреляционная функция $r(m)$ при $m \geq N$ была по модулю меньше заданной величины (например, $1 \cdot 10^{-3} - 1 \cdot 10^{-2}$), что во многих случаях обеспечивается при $20 \leq N \leq 40$. Очевидно, что от этого зависит точность моделирования, в качестве которой в данной работе используются параметры погрешности $e(m) = r(m) - Ry(m)$, где $r(m)$ и $Ry(m)$ — заданная и выборочная корреляционные функции моделируемой последовательности:

- максимальная погрешность — $\max(|e|)$;
- среднее значение погрешности: $Me = \frac{1}{N} \sum_{m=1}^N e(m)$;
- выборочная дисперсия погрешности: $De = \frac{1}{N} \sum_{m=1}^N \{[e(m)]^2 - (Me)^2\}$.

Для моделирования случайных последовательностей длины $K \gg N$, где K — общая длина моделируемой последовательности, исходная и моделируемая последовательности разбиваются на P скалярных векторов длины N так, чтобы $N \times P = K$. Векторы при этом оказываются столбцами матриц \mathbf{X} и \mathbf{Y} размера $(N \times P)$ в уравнении (7). Полученные матрицы связаны с последовательностями x_k и y_k , являющимися реализациями длины K процессов $X(k)$ и $Y(k)$, уравнениями:

$$X_{i,p} = x_{i+N(p-1)}, Y_{i,p} = y_{i+N(p-1)}, i = \overline{1, N}, p = \overline{1, P}. \quad (8)$$

Рассмотрим зависимость точности моделирования от длины K моделируемой случайной последовательности y_k . Ее выборочная корреляционная функция также является случайной, как и все указанные выше характеристики погрешности моделирования $e(m)$. Наиболее важная характеристика — дисперсия ошибки моделирования De , величина которой определяется в основном дисперсией эмпирической корреляционной функции $D[Ry(m)]$. В работе [6, С. 192] показано, что дисперсия должна удовлетворять неравенству:

$$D[Ry(m)] \leq \frac{2}{K} \sum_{m=-N}^N r^2(m) = \frac{2}{K} \left[1 + 2 \sum_{m=1}^N r^2(m) \right],$$

где N — граница области $(-N, N)$.

Тогда

$$K \geq \frac{2}{D[Ry(m)]} \left[1 + 2 \sum_{m=1}^N r^2(m) \right], \quad (9)$$

где

$$Ry(m) = \frac{1}{Dy \cdot (N - m)} \left[\sum_{i=0}^{N-m} y_i y_{i+m} - My^2 \right], \quad (10)$$

My и Dy — оценки математического ожидания и дисперсии последовательности y_k .

Таким образом, длина моделируемой случайной последовательности должна удовлетворять неравенству (9) при заданной дисперсии ошибки моделирования.

Примеры моделирования

Пример 1. Рассмотрим моделирование в Mathcad случайной последовательности y_k с заданными корреля-

ционной функцией вида $r(m) = \exp\{-0,2|m|\} \cos(0,6|m|)$ и дисперсией ошибки $De \leq 1 \cdot 10^{-3}$. Выберем $N = 30$, тогда $r(N) = 1,64 \cdot 10^{-3}$ и $m = 0,30$. Корреляционная матрица имеет элементы вида: $Ry_{ij} = \exp\{-0,2|i-j|\} \cos(0,6|i-j|)$, а матрица собственных векторов \mathbf{W} вычисляется по команде $\mathbf{W} = \text{eigenvecs}(Ry)$. При заданной дисперсии ошибки длина моделируемой последовательности в соответствии с неравенством (9) должна быть не менее $K = 6 \cdot 10^3$, так что при $N = 30$ число векторов \mathbf{X} и $\mathbf{Y} - P = 200$.

В качестве исходного некоррелированного процесса \mathbf{X} выберем последовательность K случайных равномерно распределенных чисел x_k , $k = \overline{1, K}$ на промежутке $(-1, 1)$ ($x = \text{runif}(K, -1, 1)$). Элементы коррелированной последовательности y_k определяются уравнением (7), матрицы \mathbf{X} и \mathbf{Y} которого связаны с соответствующими последовательностями уравнениями (8), а матрица \mathbf{B} в уравнении (7) находится с помощью матрицы $\mathbf{\Lambda}$:

$$B_{ij} = \begin{cases} \sqrt{\Lambda_{i,j}}, & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

Приближенная корреляционная функция $Ry(m)$ полученной последовательности y_k , вычисленная в соответствии с уравнением (10), показана на рис. 2 вместе с заданной $r(m)$. На рис. 2 видна близость корреляционных функций, подтверждаемая параметрами ошибки моделирования $e(m) = r(m) - Ry(m)$:

$$\max(|e|) = 0,05, Me = -1,9 \cdot 10^{-3}, De = 8,4 \cdot 10^{-4} < 1 \cdot 10^{-3}.$$

Пример 2. Рассмотрим моделирование непрерывного (по состоянию) стационарного марковского процесса с заданным коэффициентом корреляции между соседними элементами $r(1) = 0,3$, так что корреляционная матрица R_{ij} вводится в виде:

$$R_{ij} = \begin{cases} 1, & \text{if } i = j \\ r(1), & \text{if } |i - j| = 1. \\ 0 & \text{otherwise} \end{cases}$$

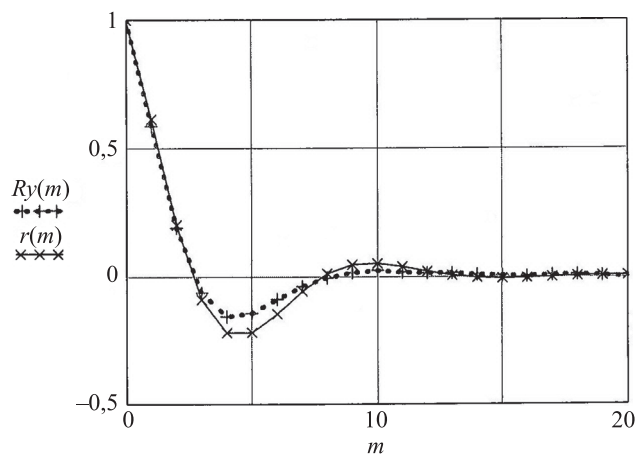


Рис. 2. Заданная $r(m)$ и полученная $Ry(m)$ корреляционные функции

Fig. 2. The set $r(m)$ and obtained $Ry(m)$ correlation functions

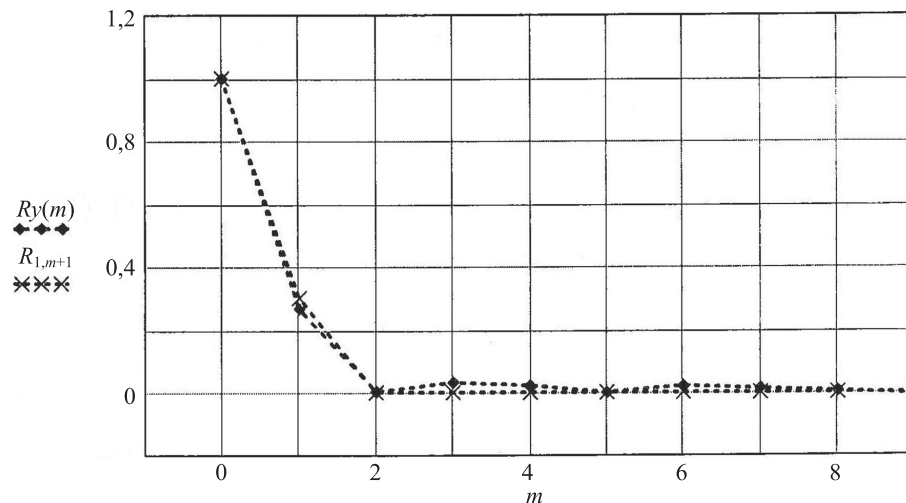


Рис. 3. Заданная $R_{1,m+1}$ и полученная $Ry(m)$ корреляционные функции марковского процесса

Fig. 3. The set $R_{1,m+1}$ and obtained $Ry(m)$ correlation functions of the Markov process

При требуемой дисперсии ошибки $De \leq 1 \cdot 10^{-3}$ длина моделируемого марковского процесса в соответствии с (9) должна быть не менее $1 \cdot 10^3$.

Выберем $N = 10$, тогда $P = 200$ и $K = 2 \cdot 10^3$. Выходной коррелированный процесс определяется уравнением (7), а его корреляционная функция, вычисленная в соответствии с (10), показана на рис. 3 вместе с первой строкой заданной корреляционной матрицы: $R_{1,m+1} = (1 \ 0,3 \ 0 \ 0 \ \dots)$. На рис. 3 видно подобие заданной и вычисленной по полученной модели марковского процесса корреляционных функций, что подтверждается параметрами ошибки моделирования $e(m)$:

$$e(1) = r(1) - Ry(1) = 0,032, e(1)/r(1) = 0,00(6),$$

$$\max(|e|) = 0,038, Me = 9,7 \cdot 10^{-3}, De = 2,9 \cdot 10^{-4}.$$

Моделирование случайного процесса с заданными частотными свойствами

Рассмотрим методику моделирования случайного процесса с заданными частотными свойствами с помощью пакета Mathcad. Если спектральная плотность мощности $Gy(\omega)$ стационарного СП $Y(t)$ задана аналитически, то корреляционная функция $r(\tau)$ этого процесса может быть найдена с помощью уравнения Винера–Хинчина [6, С. 240]:

$$r(\tau) = \frac{1}{\pi} \int_0^{\omega_{\max}} Gy(\omega) \cdot \cos(\omega \cdot \tau) \cdot d\omega, \quad (11)$$

где, при $\omega \geq \omega_{\max} = 2\pi \cdot f_{\max}$, $Gy(\omega)/\max Gy(\omega) \leq 1 \cdot 10^{-4} - 2 \cdot 10^{-3}$.

Дискретизация найденной в (11) корреляционной функции $r(\tau)$ производится с интервалом $\Delta\tau$ между соседними отсчетами, определяемым теоремой Котельникова [7, С. 82]:

$$\Delta\tau \leq \frac{1}{2f_{\max}} = \frac{\pi}{\omega_{\max}}.$$

После выбора величины шага дискретизации $\Delta\tau$ корреляционная функция $r1(m)$ определяется уравнением:

$$r1(m) = \begin{cases} 1, & \text{if } m = 0, \\ r(\Delta\tau \cdot m), & \text{if } m \neq 0. \end{cases}$$

Выполним переход к корреляционной матрице Ry вида (2):

$$Ry_{i,j} = r1(|i-j|), \quad i, j = \overline{1, N}, \quad (12)$$

где N — размер корреляционной матрицы, $|i-j| = m$.

Например, если заданная спектральная плотность мощности моделируемого СП равна

$$G(\omega) = \frac{16}{\omega^4 + 13\omega^2 + 4},$$

то при $\omega_{\max} = 2\pi G(2\pi)/G(0) \cong 2 \cdot 10^{-3}$. Примем $\omega_{\max} = 2\pi$, $f_{\max} = 1$, тогда $\Delta\tau = 0,5$. Таким образом, корреляционная матрица вида (2), определенная в (12), позволяет получить реализацию моделируемого процесса, используя рассмотренную методику.

Заключение

Предложена методика математического моделирования стационарных случайных процессов с заданными корреляционными и спектральными свойствами. Результаты реализации методики с помощью пакета Mathcad показали ее эффективность и возможность обеспечения требуемой точности выбором длины интервала корреляции и общей длины реализаций моделируемого случайного процесса. Рассмотрены примеры реализаций, в результате которых общая длина процесса не более чем в 10 раз превысила величину, обратную требуемой дисперсии ошибки моделирования. Важное достоинство предложенного метода моделирования — простота его реализации по сравнению с методом, описанным в [2], и с теоретическими методами — в [1].

Литература

1. Шапоров С.Д., Родин Б.П. Случайные процессы: учебник. СПб.: БГТУ, 2010. 237 с.
2. Карташев В.Я., Новосельцева М.А. Цифровое моделирование стационарных случайных процессов с заданной корреляционной функцией на основе непрерывных дробей // Управление большими системами: сборник трудов. 2010. № 31. С. 49–91.
3. Moklyachuk O. Simulation of random processes with known correlation function with the help of Karhunen-Loeve decomposition // *Theory of Stochastic Processes*. 2007. V. 13(29). N 4. P. 163–169.
4. Солодовщиков А.Ю., Платонов А.К. Исследование метода Карунена–Лозва // Препринты ИПМ им. М.В. Келдыша. 2006. № 19. С. 1–29.
5. Xiang S.P., Quek S.T., Phoon K.K. Convergence study of the truncated Karhunen-Loeve expansion for simulation of stochastic processes // *International Journal for Numerical Methods in Engineering*. 2001. V. 52. N 9. P. 1029–1043. <https://doi.org/10.1002/nme.255>
6. Купер Дж., МакГилем К. Вероятностные методы анализа сигналов и систем. М.: Мир, 1989. 376 с.
7. Игнатьев Н.К. Дискретизация и ее приложения. М.: Связь, 1980. 264 с.

Автор

Ефимов Александр Сергеевич — кандидат технических наук, доцент, независимый исследователь, Санкт-Петербург, Российская Федерация, <https://orcid.org/0000-0002-8091-435X>, asefimov@mail.ru

Статья поступила в редакцию 23.02.2022
Одобрена после рецензирования 23.05.2022
Принята к печати 12.07.2022

References

1. Shaporev S.D., Rodin B.P. *Stochastic Processes*. St. Petersburg, BSTU, 2010, 237 p. (in Russian)
2. Kartashov V., Novoseltseva M. Digital design of stationary casual precesses with the set cross-correlation function on basis of the continued fractions. *Large-Scale Systems Control*, 2010, no. 31, pp. 49–91. (in Russian)
3. Moklyachuk O. Simulation of random processes with known correlation function with the help of Karhunen-Loeve decomposition. *Theory of Stochastic Processes*, 2007, vol. 13(29), no. 4, pp. 163–169.
4. Solodovshikov A.Yu., Platonov A.K. The investigation of Karhunen-Loeve method. *Keldysh Institute Preprints*, 2006, no. 19, pp. 1–29. (in Russian)
5. Xiang S.P., Quek S.T., Phoon K.K. Convergence study of the truncated Karhunen-Loeve expansion for simulation of stochastic processes. *International Journal for Numerical Methods in Engineering*, 2001, vol. 52, no. 9, pp. 1029–1043. <https://doi.org/10.1002/nme.255>
6. Cooper G.R., McGillem C.D. *Probabilistic Methods of Signal and System Analysis*. New York, Holt, Rinehart, and Winston, 1986, 408 p.
7. Ignatev N.K. *Discretization and its Applications*. Moscow, Svjaz' Publ., 1980, 264 p. (in Russian)

Author

Alexandr S. Efimov — PhD, Associate Professor, Independent Researcher, Saint Petersburg, Russian Federation, <https://orcid.org/0000-0002-8091-435X>, asefimov@mail.ru

Received 23.02.2022
Approved after reviewing 23.05.2022
Accepted 12.07.2022



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

doi: 10.17586/2226-1494-2022-22-4-785-791

УДК 532.529

Управление численной диссипацией гибридного метода крупных частиц в задачах с вихревой неустойчивостью

Дмитрий Викторович Садин✉

Военно-космическая академия имени А.Ф.Можайского, Санкт-Петербург, 197198, Российская Федерация
sadin@yandex.ru✉, <https://orcid.org/0000-0001-5335-4847>

Аннотация

Предмет исследования. Современные тенденции развития численных схем связаны с уменьшением диссипативных и дисперсионных ошибок, а также улучшением сеточной сходимости решения. Достижение этих вычислительных свойств — непростая проблема, так как уменьшение схемной вязкости часто сопряжено с возрастанием осцилляций газодинамических параметров. В работе представлено исследование вопросов управления численной диссипацией в задачах газовой динамики с целью повышения разрешающей способности при численном воспроизведении вихревой неустойчивости на контактных границах. **Метод.** Для решения поставленной задачи использован гибридный метод крупных частиц второго порядка аппроксимации по пространству и времени на гладких решениях. Метод построен с расщеплением по физическим процессам в два этапа: градиентное ускорение и деформирование конечного объема среды; конвективный перенос среды через его грани. Повышение порядка аппроксимации по времени достигается корректирующим шагом по времени. Регуляризация численного решения задач на первом этапе метода заключается в нелинейной коррекции искусственной вязкости, которая независимо от разрешения сетки стремится к нулю в областях гладкости решения. На этапе конвективного переноса выполнена реконструкция потоков путем аддитивной комбинации центральной и противопоточной аппроксимаций. **Основные результаты.** Предложен механизм регулирования численной диссипации метода, основанный на новом параметрическом ограничителе искусственной вязкости. Оптимальная настройка метода по соотношению диссипативных и дисперсионных свойств численного решения достигнута заданием параметра ограничительной функции. Проверка эффективности метода проведена на двумерных показательных задачах. В одной из них контактные поверхности закручены в спираль, на которых возникает вихревая неустойчивость Кельвина–Гельмгольца. Другая задача — классический тест с двойным маховским отражением сильной ударной волны. Сравнение с современными численными методами показало, что предложенный вариант гибридного метода крупных частиц обладает высокой конкурентоспособностью. Например, в задаче с двойным маховским отражением рассматриваемый вариант метода превосходит по вихреразрешающей способности популярную схему WENO (Weighted Essentially Non-Oscillatory) пятого порядка и сопоставим с численным решением WENO девятого порядка аппроксимации. **Практическая значимость.** Предложенный метод может быть основой конвективного блока численной схемы при построении вычислительной технологии моделирования турбулентности.

Ключевые слова

гибридный метод крупных частиц, разрешающая способность, численная диссипация, вихревая неустойчивость

Ссылка для цитирования: Садин Д.В. Управление численной диссипацией гибридного метода крупных частиц в задачах с вихревой неустойчивостью // Научно-технический вестник информационных технологий, механики и оптики. 2022. Т. 22, № 4. С. 785–791. doi: 10.17586/2226-1494-2022-22-4-785-791

Numerical dissipation control of a hybrid large-particle method in vortex instability problems

Dmitry V. Sadin✉

Mozhaisky Military Aerospace Academy, Saint Petersburg, 197198, Russian Federation

sadin@yandex.ru✉, <https://orcid.org/0000-0001-5335-4847>

Abstract

Current trends in the development of numerical schemes are associated with a decrease in dissipative and dispersion errors, as well as an improvement in the grid convergence of the solution. Achieving the computational properties is not an easy problem since a decrease in scheme viscosity is often associated with an increase in the oscillations of gas dynamic parameters. The paper presents a study of the issues of numerical dissipation control in gas dynamics problems in order to increase the resolution in numerical reproduction of vortex instability at contact boundaries. To solve this problem, a hybrid large-particle method of the second order of approximation in space and time on smooth solutions is used. The method is constructed with splitting by physical processes in two stages: gradient acceleration and deformation of the finite volume of the medium; convective transfer of the medium through its facets. An increase in the order of approximation in time is achieved by a time correction step. The regularization of the numerical solution of problems at the first stage of the method consists in the nonlinear correction of artificial viscosity which, regardless of the grid resolution, tends to zero in the areas of smoothness of the solution. At the stage of convective transport, the reconstruction of fluxes was carried out by an additive combination of central and upwind approximations. A mechanism for regulating the numerical dissipation of the method based on a new parametric limiter of artificial viscosity is proposed. The optimal adjustment of the method by the ratio of dissipative and dispersive properties of the numerical solution is achieved by setting the parameter of the limiting function. The efficiency of the method was tested on two-dimensional demonstrative problems. In one of them, the contact surfaces are twisted into a spiral on which the Kelvin-Helmholtz vortex instability develops. Another task is the classic problem with a double Mach reflection of a strong shock wave. Comparison with modern numerical schemes has shown that the proposed variant of the hybrid large-particle method has a high competitiveness. For example, in the problem with double Mach reflection, the considered version of the method surpasses in terms of vortex resolution the popular WENO (Weighted Essentially Non-Oscillatory) scheme of the fifth order and is comparable to the numerical solution of WENO of the ninth order of approximation. The proposed method can be the basis of a convective block of a numerical scheme when constructing a computational technology for modeling turbulence.

Keywords

hybrid large-particle method, resolution, numerical dissipation, vortex instability

For citation: Sadin D.V. Numerical dissipation control of a hybrid large-particle method in vortex instability problems. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2022, vol. 22, no. 4, pp. 785–791 (in Russian). doi: 10.17586/2226-1494-2022-22-4-785-791

Введение

Численное моделирование структурно сложных течений газа связано с проблемами корректного расчета ударных волн и контактных разрывов, а также различных видов неустойчивости и многомасштабного разрешения формируемых вихревых образований. Традиционные схемы первого порядка аппроксимации Годунова [1], Рое [2], Белоцерковского–Давыдова [3] позволяют численно воспроизводить монотонные, но существенно сглаженные профили разрывов в газовом потоке. В задачах моделирования вихревых структур или акустических волн применение разностных или конечно-объемных схем первого порядка оказывается практически не пригодным, так как требуется недопустимо большое число сеточных узлов (контрольных объемов).

Как известно, теорема Годунова [1] не допускает построение монотонных схем для линейного уравнения переноса выше первого порядка аппроксимации, что может приводить к появлению нефизических осцилляций в решении. Для подавления осцилляций в разностных схемах повышенного порядка аппроксимации предложены различные нелинейные алгоритмические приемы, основанные на принципах минимальных значений производных [4], невозрастания общей вариации

решения Total Variation Diminishing (TVD) [5], перебора шаблонов с учетом локальной гладкости решения Weighted Essentially Non-Oscillatory (WENO) [6, 7] и др.

Современные тенденции развития численных схем связаны с уменьшением диссипативных и дисперсионных ошибок, а также улучшением сеточной сходимости решения [8–10]. Достижение этих вычислительных свойств является непростой проблемой, поскольку уменьшение схемной вязкости часто сопряжено с возрастанием осцилляций газодинамических параметров.

В общем случае течение газа описывается с учетом вязкости и теплопроводности. Вместе с тем численное моделирование будет корректно, если эффективная численная вязкость схемы существенно меньше физической. Как правило, сравнительный анализ эффективности алгоритмов выполняется в тестовых задачах с вихревой неустойчивостью в невязкой постановке [11, 12]. При этом предпочтительна менее диссипативная схема, обладающая лучшей вихреразрешающей способностью.

Работа посвящена совершенствованию гибридного метода крупных частиц с управляемой численной диссипацией CDP2 (Controlled Dissipative Property) [13, 14]. Регулирование диссипативных свойств метода осуществлено двумя способами. Первый заключается в нелинейной коррекции искусственной вязкости, кото-

рая независимо от разрешения сетки стремится к нулю в областях гладкости численного решения. Вторым способом — реконструкция конвективных величин путем аддитивной комбинации центральной и противопоточной аппроксимаций.

Цель работы — повышение разрешающей способности гибридного метода крупных частиц при численном воспроизведении вихревой неустойчивости на контактных границах путем управления численной диссипацией при аппроксимации законов сохранения.

Математическая модель и алгоритм расчета

Рассмотрим законы сохранения калорически совершенного газа в форме уравнений Эйлера в безразмерном виде [14]:

$$\begin{aligned} \frac{\partial \mathbf{q}}{\partial t} + \nabla_d \mathbf{G} + \nabla_d \mathbf{F} &= 0, \\ \mathbf{q} &= [\rho, \rho \mathbf{v}, \rho E]^T, \mathbf{G} = [\rho \mathbf{v}, \rho \mathbf{v} \mathbf{v}, \rho E \mathbf{v}]^T, \\ \mathbf{F} &= [0, p, p \mathbf{v}]^T, \nabla_d = \text{diag}(\nabla \cdot, \nabla, \nabla \cdot), \end{aligned} \quad (1)$$

где ρ , \mathbf{v} , p , E — плотность, вектор скорости с компонентами u и v , давление, полная энергия газа единицы массы; \mathbf{q} , \mathbf{G} — консервативные и потоковые величины; \mathbf{F} — градиентные и деформационные величины; t — время; ∇_d — диагональная матрица с операторами набла на главной диагонали; точка — знак скалярного произведения.

Законы сохранения (1) дополняются уравнением состояния

$$p = (\gamma - 1)\rho(E - \mathbf{v}^2/2),$$

где γ — показатель адиабаты.

Схема гибридного метода крупных частиц реализуется с расщеплением по физическим процессам на два этапа.

На первом этапе аппроксимируются градиентные и деформационные слагаемые \mathbf{F} :

$$\mathbf{q}_n^{(0)} = \mathbf{q}_n^k - \left(\langle \mathbf{F}_{n+1/2}^k \rangle_{\psi_v} - \langle \mathbf{F}_{n-1/2}^k \rangle_{\psi_v} \right) \tau / h, \quad (2)$$

а на втором — потоковые члены законов сохранения \mathbf{G} :

$$\mathbf{q}_n^{(1)} = \mathbf{q}_n^{(0)} - \left(\langle \mathbf{G}_{n+1/2}^{(0)} \rangle_{\psi_f} - \langle \mathbf{G}_{n-1/2}^{(0)} \rangle_{\psi_f} \right) \tau / h, \quad (3)$$

где $\tau = t^{k+1} - t^k$ — шаг по времени (t^k — временной слой); h — размер ячейки с ее центром x_n и гранями $x_{n\pm 1/2} = x_n \pm h/2$; $\langle \cdot \rangle_{\psi_v}$ и $\langle \cdot \rangle_{\psi_f}$ — операторы нелинейной коррекции искусственной вязкости с ограничителем ψ_v и реконструкции потоков с ограничителем ψ_f (подробнее представлены в работах [13, 14]).

Повышение порядка аппроксимации до $O(h^2 + \tau^2)$ достигается с использованием корректирующего шага:

$$\mathbf{q}_n^{(2)} = 0,5(\mathbf{q}_n^k + \mathbf{q}_n^{(1)}) - 0,5 \left(\langle \mathbf{F}_{n+1/2}^{(1)} \rangle_{\psi_v} - \langle \mathbf{F}_{n-1/2}^{(1)} \rangle_{\psi_v} \right) \tau / h, \quad (4)$$

$$\mathbf{q}_n^{k+1} = \mathbf{q}_n^{(2)} - 0,5 \left(\langle \mathbf{G}_{n+1/2}^{(2)} \rangle_{\psi_f} - \langle \mathbf{G}_{n-1/2}^{(2)} \rangle_{\psi_f} \right) \tau / h. \quad (5)$$

К настоящему времени предложено большое количество ограничителей TVD-типа, выбор которых определяется классом решаемых задач [15]. Для более гибкого (параметрического) регулирования диссипативных свойств гибридного метода крупных частиц реализована линейная комбинация известных ограничителей [16]. Например, новый параметрический β -ограничитель:

$$\psi_{VC} = \beta \psi_{SC} + (1 - \beta) \psi_{VL}$$

позволяет управлять численной вязкостью схемы за счет комбинации ограничителей:

Van Leer

$$\psi_{VL}(r) = (r + |r|)/(1 + r)$$

и Super-C

$$\psi_{SC} = \begin{cases} \min(2r/|c|, 1), & 0 \leq r \leq 1, \\ \min(r, 2/(1 - |c|)) & r > 1, \\ 0,4 & r < 0, \end{cases}$$

где r — отношение наклонов газодинамических параметров в соседних ячейках, $c = u\tau/h$ — аналог числа Куранта.

Характерные области действия ограничителей ψ_{VL} , ψ_{SC} и их линейной комбинации ψ_{VC} в зависимости от отношения наклонов r показаны на диаграмме Sweby (рис. 1). На интервале изменения переменной $r < 0$ отношения наклонов имеют разные знаки — сеточное решение содержит экстремумы (осцилляции). При этом ограничительная функция задается $\psi = 0$, и обеспечивается повышенная численная диссипация. В области $r \geq 0$ решение обладает различной степенью гладкости, а аппроксимация потоков в алгоритме гибридного метода крупных частиц (3), (5) является бездиссипативной при $\psi = 1$.

Механическое использование известных ограничителей для искусственной вязкости в (2) и (4) возможно, но не является оптимальным. В частности, применение $\psi = 0$ в области экстремумов при $r < 0$ сглаживает сеточное решение, но вносит в схему избыточную численную вязкость. Потому рациональным способом регулирования диссипации может быть ограничитель вида $\psi_{\beta C} = \beta + (1 - \beta)\psi_{SC}$, области действия которого показаны сплошной кривой 4 на рис. 1.

Верификация алгоритма расчета

Выполним проверку работоспособности нового алгоритма нелинейной коррекции искусственной вязкости на показательных тестовых задачах с вихревой неустойчивостью, для которых (например, [10–12]) известны решения различными численными методами. Задачи решены в плоской ортогональной системе координат на равномерной сетке с числом Куранта CFL = 0,4. Для обеспечения высокой информативности результатов постобработка осуществлена в виде численных шпирен-изображений (нелинейной функции градиента плотности Quirk and Karni) [14]:

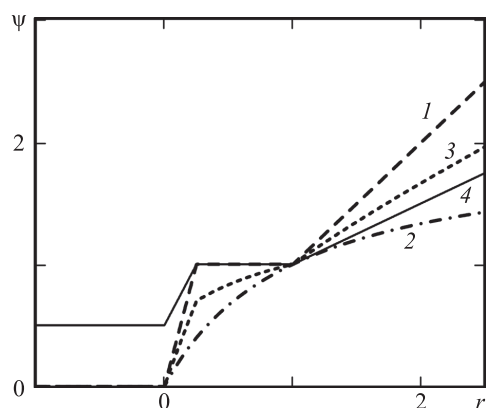


Рис. 1. Диаграмма Sweby. Ограничительные функции: ψ_{SC} (кривая 1); ψ_{VL} (кривая 2); ψ_{VC} с параметром $\beta = 0,5$ (кривая 3); ограничитель искусственной вязкости $\psi_{0.5C}$ (кривая 4) в зависимости от отношения наклонов r

Fig. 1. The Sweby diagram. Limiter functions: ψ_{SC} (curve 1), ψ_{VL} (curve 2); ψ_{VC} with the parameter $\beta = 0.5$ (curve 3), artificial viscosity limiter $\psi_{0.5C}$ (curve 4) depending on the slope ratio r

$$s_{ij} = \exp \left(-k \frac{|\nabla \rho_{ij}|}{\max_{i,j} |\nabla \rho_{i,j}|} \right),$$

где i, j — нумерация ячеек по x и y соответственно; $\nabla \rho_{ij}$ — градиент плотности; k — настроечный коэффициент для контрастного отображения деталей потока.

Первый тест связан двумерными задачами Римана с формированием четырех контактных разрывов и их закручиванием в спираль [12]. Область определения задачи $[-0,5; 0,5] \times [-0,5; 0,5]$ разделена прямыми $x = 0$ и $y = 0$ на четыре квадранта. Постоянные начальные условия приведены в ячейках таблицы, соответствующим позициям квадрантов.

На границах расчетной области заданы «мягкие» краевые условия. Как и в работе [12], расчеты выполнены на сетке 600×600 ячеек до момента времени 0,3. Для нелинейной коррекции искусственной вязкости использован ограничитель ψ_{VC} с параметром $\beta = 0,5$. Управление численной диссипацией потоков осуществлено с помощью комбинированной функции наклонов $\psi_{VC}(\beta)$ для различных значений параметра $\beta = 0,5; 0,7; 0,8; 0,9$.

Результаты расчетов представлены на рис. 2. Особенности данной конфигурации: непрерывность давления и нормальной составляющей скорости; плотность и касательный компонент скорости претерпевают разрыв. Контактные поверхности закручиваются в спираль, на которых развивается вихревая неустойчивость Кельвина–Гельмгольца. Популярная схема WENO5 [12] не выявляет отмеченный эффект на заданной детализации сетки (рис. 2, e). Гибридный метод крупных частиц с использованием ограничителя потоков $\psi_{VC}(\beta)$ при $\beta \geq 0,7$ разрешает вторичную вихревую структуру на поверхности спирали (рис. 2, b–d) и превосходит по вихреразрешающей способности метод THINC-BVD

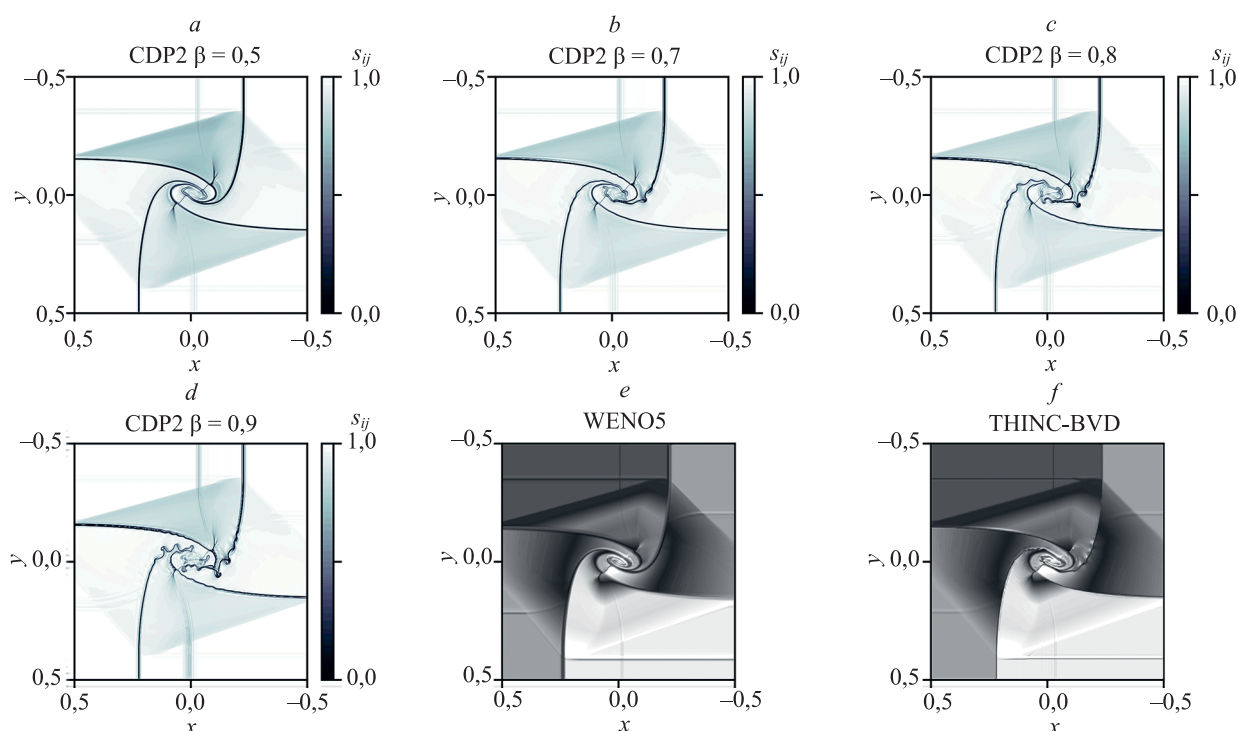


Рис. 2. Численные шlieren-изображения функции градиента плотности в задаче Римана с контактными разрывами на сетке 600×600 ячеек, полученные гибридным методом крупных частиц с комбинацией ограничителей потоков ψ_{SC} и ψ_{VL} для значений параметра β : 0,5 (a); 0,7 (b); 0,8 (c) и 0,9 (d); теневые изображения: схемы WENO5 (e) и метода THINC-BVD из [12] (f)

Fig. 2. Numerical schlieren-images of the density gradient function in the Riemann problem with contact discontinuities on a grid of 600×600 cells obtained by the hybrid large-particle method with a combination of flux limiters ψ_{SC} and ψ_{VL} for parameter values $\beta = 0.5$ (a); 0.7 (b); 0.8 (c) and 0.9 (d); shadow images: WENO5 (e), THINC-BVD from [12] (f)

Таблица. Начальные условия
Table. Initial conditions

Позиция в квадрантах	Параметры							
	p		r		u		v	
	слева	справа	слева	справа	слева	справа	слева	справа
Сверху	1	1	2	1	0,75	0,75	0,50	-0,50
Снизу	1	1	1	3	-0,75	-0,75	0,50	-0,50

(Tangent of Hyperbola for Interface Capturing – Boundary Variations Diminishing) из [12] (рис. 2, *f*).

В качестве второго теста выбрана классическая задача с двойным маховским отражением сильной ударной волны (Double Mach reflection of a strong shock [11]). В начальный момент времени ударная волна с числом Маха 10 образует угол $\pi/3$ с осью $0x$ и пересекает ее в точке $1/6$. Перед волной заданы параметры: плотность $\rho = 1,4$, давление $p = 1$ и скорость с компонентами $u = v = 0$. Расчетная область задачи $[0, 4] \times [0, 1]$. Граничные условия при $y = 0$ на отрезке $[0, 1/6]$ заданы в виде свободного вытекания, а в остальной части

$[1/6, 4]$ — условия отражения. Краевые условия слева и сверху за ударной волной установлены из соотношений Рэнкина–Гюгонио, а справа при $x = 4$ и сверху до скачка уплотнения градиенты газодинамических параметров заданы равными нулю. Время окончания расчета 0,2.

В результате взаимодействия косого скачка уплотнения с твердой стенкой возникает конфигурация с двумя тройными точками, двойным маховским отражением и наклонной контактной поверхностью, на которой развивается вихревая неустойчивость. Экспериментальное подтверждение и детальное обсуждение конфигураций с косыми ударными волнами приведена в работе [17].

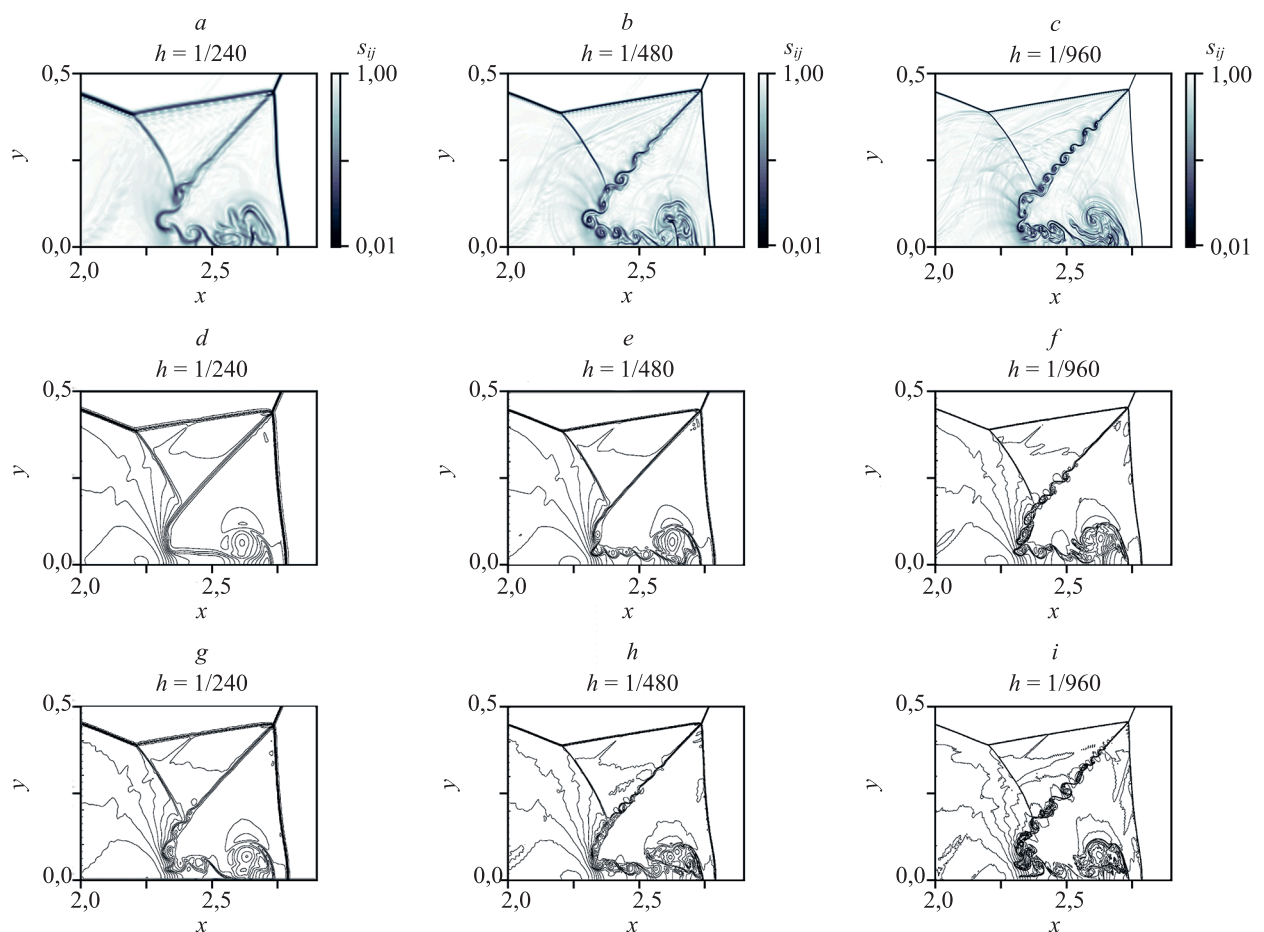


Рис. 3. Численные шлирен-изображения функции градиента плотности в задаче с двойным маховским отражением ударной волны, полученные гибридным методом крупных частиц с комбинацией ограничителей потоков: ψ_{SC} и ψ_{VL} (a) при $\beta = 0,2$; ψ_{VL} и ψ_{MM} при $\beta = 1,0$ (b) и $\beta = 0,6$ (c). Контурные линии плотности от 1,5 до 22,9705 для схемы: WENO5 (d–f) и WENO9 из [11] (g–i)

Fig. 3. Numerical schlieren-images of the density gradient function in the problem with double Mach reflection of a shock wave obtained by the hybrid large-particle method with a combination of flux limiters: ψ_{SC} and ψ_{VL} (a) for $\beta = 0.2$; ψ_{VL} and ψ_{MM} for $\beta = 1.0$ (b) and $\beta = 0.6$ (c). Contour lines of density from 1.5 to 22.9705 for scheme: WENO5 (d–f), WENO9 from [11] (g–i)

Вычислительная проблема данной задачи — разрешение вихревой структуры на наклонном контактом разрыве. Это предъявляет к численным моделям высокие конфликтующие между собой требования малой схемной диссипативности, но достаточной для подавления дисперсионных ошибок (нефизических осцилляций).

На рис. 3, *a–c* показаны результаты расчетов гибридным методом крупных частиц с предложенной нелинейной коррекцией искусственной вязкости ψ_{PC} с параметром $\beta = 0,5$ на сетках различного разрешения: 1/240, 1/480, 1/960. На грубой сетке 1/240 использован ограничитель потоков ψ_{VC} (0,2), а на подробных сетках 1/480, 1/960 — более диссипативная комбинация ограничителей MINMOD и Van Leer [18] $\psi_{\text{MV}} = \beta\psi_{\text{VL}} + (1 - \beta)\psi_{\text{MM}}$ с параметрами соответственно $\beta = 1,0$ и 0,6.

Проведено сравнение расчетов по схемам пятого и девятого порядков аппроксимации: WENO5 (рис. 3, *d–f*) и WENO9 (рис. 3, *g–i*) [11]. Сопоставление представленных результатов показало, что предложенный алгоритм гибридного метода крупных частиц на всех

детализациях сеток превосходит по вихреразрешающей способности популярную схему WENO5 и сопоставим с WENO9. Кроме того, применение предложенного алгоритма нелинейной коррекции позволило значительно улучшить вычислительные свойства метода и качество численных решений [18, 19], в особенности на грубых сетках.

Заключение

В рамках гибридного метода крупных частиц предложена новая форма нелинейной коррекции искусственной вязкости. Оптимальная настройка метода по соотношению диссипативных и дисперсионных свойств численного решения получена заданием ограничительной функции искусственной вязкости. Выполнена проверка предложенного метода на показательных тестовых задачах с развитием вихревой неустойчивости на контактной границе. Гибридный метод крупных частиц продемонстрировал высокий потенциал, превосходя по разрешающей способности популярную схему WENO5.

Литература

1. Годунов С.К. Разностный метод численного расчета разрывных решений уравнений гидродинамики // Математический сборник. 1959. Т. 47(89). № 3. С. 271–306.
2. Roe P.L. Approximate Riemann solvers, parameter vectors and difference schemes // Journal of Computational Physics. 1981. V. 43. N 2. P. 357–372. [https://doi.org/10.1016/0021-9991\(81\)90128-5](https://doi.org/10.1016/0021-9991(81)90128-5)
3. Белоцерковский О.М., Давыдов Ю.М. Нестационарный метод «крупных частиц» для газодинамических расчетов // Журнал вычислительной математики и математической физики. 1971. Т. 11. № 1. С. 182–207.
4. Колган В.П. Применение принципа минимальных значений производной к построению конечноразностных схем для расчета разрывных решений газовой динамики // Ученые записки ЦАГИ. 1972. Т. 3. № 6. С. 68–77.
5. Harten A. High resolution schemes for hyperbolic conservation laws // Journal of Computational Physics. 1983. V. 49. N 3. P. 357–393. [https://doi.org/10.1016/0021-9991\(83\)90136-5](https://doi.org/10.1016/0021-9991(83)90136-5)
6. Liu X.-D., Osher S., Chan T. Weighted essentially non-oscillatory schemes // Journal of Computational Physics. 1994. V. 115. N 1. P. 200–212. <https://doi.org/10.1006/JCPH.1994.1187>
7. Jiang G.-S., Shu C.-W. Efficient implementation of weighted ENO schemes // Journal of Computational Physics. 1996. V. 126. N 1. P. 202–228. <https://doi.org/10.1006/jcph.1996.0130>
8. Zhang S., Zhu J., Shu C.-W. A brief review on the convergence to steady state solutions of Euler equations with high-order WENO schemes // Advances in Aerodynamics. 2019. V. 1. N 1. P. 16. <https://doi.org/10.1186/s42774-019-0019-2>
9. Minoshima T., Miyoshi T. A low-dissipation HLLD approximate Riemann solver for a very wide range of Mach numbers // Journal of Computational Physics. 2021. V. 446. P. 110639. <https://doi.org/10.1016/j.jcp.2021.110639>
10. Li J., Shu C.-W., Qiu J. Multi-resolution HWENO schemes for hyperbolic conservation laws // Journal of Computational Physics. 2021. V. 446. P. 110653. <https://doi.org/10.1016/j.jcp.2021.110653>
11. Shi J., Zhang Y.-T., Shu C.-W. Resolution of high order WENO schemes for complicated flow structures // Journal of Computational Physics. 2003. V. 186. N 2. P. 690–696. [https://doi.org/10.1016/S0021-9991\(03\)00094-9](https://doi.org/10.1016/S0021-9991(03)00094-9)
12. Deng X., Xie B., Loubère R., Shimizu Y., Xiao F. Limiter-free discontinuity-capturing scheme for compressible gas dynamics with reactive fronts // Computers & Fluids. 2018. V. 171. P. 1–14. <https://doi.org/10.1016/j.compfluid.2018.05.015>

References

1. Godunov S.K. A finite difference method for the computation of discontinuous solutions of the equations of fluid dynamics. *Mathematics of the USSR – Sbornik*, 1959, vol. 47, no. 8–9, pp. 357–393.
2. Roe P.L. Approximate Riemann solvers, parameter vectors and difference schemes. *Journal of Computational Physics*, 1981, vol. 43, no. 2, pp. 357–372. [https://doi.org/10.1016/0021-9991\(81\)90128-5](https://doi.org/10.1016/0021-9991(81)90128-5)
3. Belotserkovskii O.M., Davydov Yu.M. A non-stationary “Coarse particle” method for gas-dynamical computations. *USSR Computational Mathematics and Mathematical Physics*, 1971, vol. 11, no. 1, pp. 241–271. [https://doi.org/10.1016/0041-5553\(71\)90112-1](https://doi.org/10.1016/0041-5553(71)90112-1)
4. Kolgan V.P. An application of the minimal derivative values concept to construction of the finite-difference schemes for calculating the discontinuous solutions of gas dynamics. *Uchenyye Zapiski TsAGI*, 1972, vol. 3, no. 6, pp. 68–77. (in Russian)
5. Harten A. High resolution schemes for hyperbolic conservation laws. *Journal of Computational Physics*, 1983, vol. 49, no. 3, pp. 357–393. [https://doi.org/10.1016/0021-9991\(83\)90136-5](https://doi.org/10.1016/0021-9991(83)90136-5)
6. Liu X.-D., Osher S., Chan T. Weighted essentially non-oscillatory schemes. *Journal of Computational Physics*, 1994, vol. 115, no. 1, pp. 200–212. <https://doi.org/10.1006/JCPH.1994.1187>
7. Jiang G.-S., Shu C.-W. Efficient implementation of weighted ENO schemes. *Journal of Computational Physics*, 1996, vol. 126, no. 1, pp. 202–228. <https://doi.org/10.1006/jcph.1996.0130>
8. Zhang S., Zhu J., Shu C.-W. A brief review on the convergence to steady state solutions of Euler equations with high-order WENO schemes. *Advances in Aerodynamics*, 2019, vol. 1, no. 1, pp. 16. <https://doi.org/10.1186/s42774-019-0019-2>
9. Minoshima T., Miyoshi T. A low-dissipation HLLD approximate Riemann solver for a very wide range of Mach numbers. *Journal of Computational Physics*, 2021, vol. 446, pp. 110639. <https://doi.org/10.1016/j.jcp.2021.110639>
10. Li J., Shu C.-W., Qiu J. Multi-resolution HWENO schemes for hyperbolic conservation laws. *Journal of Computational Physics*, 2021, vol. 446, pp. 110653. <https://doi.org/10.1016/j.jcp.2021.110653>
11. Shi J., Zhang Y.-T., Shu C.-W. Resolution of high order WENO schemes for complicated flow structures. *Journal of Computational Physics*, 2003, vol. 186, no. 2, pp. 690–696. [https://doi.org/10.1016/S0021-9991\(03\)00094-9](https://doi.org/10.1016/S0021-9991(03)00094-9)
12. Deng X., Xie B., Loubère R., Shimizu Y., Xiao F. Limiter-free discontinuity-capturing scheme for compressible gas dynamics with

13. Садин Д.В. Модификация метода крупных частиц до схемы второго порядка точности по пространству и времени для ударно-волновых течений газовзвеси // Вестник Южно-Уральского государственного университета. Сер. Математическое моделирование и программирование. 2019. Т. 12. № 2. С. 112–122. <https://doi.org/10.14529/mmp190209>
14. Садин Д.В., Голиков И.О., Широкова Е.Н. Тестирование гибридного метода крупных частиц на двумерных задачах Римана // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2021. Т. 14. № 1. С. 58–71. <https://doi.org/10.18721/JPM.14104>
15. Zhang D., Jiang C., Liang D., Cheng L. A review on TVD schemes and a refined flux-limiter for steady-state calculations // Journal of Computational Physics. 2015. V. 302. P. 114–154. <https://doi.org/10.1016/j.jcp.2015.08.042>
16. Садин Д.В. Моделирование физической неустойчивости на контактных границах в течениях многокомпонентных сжимаемых газов гибридным методом крупных частиц // Вычислительные методы и программирование. 2020. Т. 21. № 2. С. 129–137. <https://doi.org/10.26089/NumMet.v21r211>
17. Семенов А.Н., Березкина М.К., Красовская И.В. Классификация разновидностей отражения ударной волны от клина. Часть 2. Экспериментальное и численное исследование разновидностей маховского отражения // Журнал технической физики. 2009. Т. 79. № 4. С. 52–58.
18. Садин Д.В. Анализ диссипативных свойств гибридного метода крупных частиц для структурно сложных течений газа // Компьютерные исследования и моделирование. 2020. Т. 12. № 4. С. 757–772. <https://doi.org/10.20537/2076-7633-2020-12-4-757-772>
19. Садин Д.В. Эффективная реализация гибридного метода крупных частиц // Математическое моделирование. 2022. Т. 34. № 4. С. 113–127. <https://doi.org/10.20948/mm-2022-04-08>
- reactive fronts. *Computers & Fluids*, 2018, vol. 171, pp. 1–14. <https://doi.org/10.1016/j.compfluid.2018.05.015>
13. Sadin D.V. A modification of the large-particle method to a scheme having the second order of accuracy in space and time for shockwave flows in a gas suspension. *Bulletin of the South Ural State University. Series: Mathematical Modelling, Programming & Computer Software*, 2019, vol. 12, no. 2, pp. 112–122. (in Russian). <https://doi.org/10.14529/mmp190209>
14. Sadin D.V., Golikov I.O., Shirokova E.N. Testing of the hybrid large-particle method using two-dimensional Riemann problems. *St. Petersburg State Polytechnical University Journal: Physics and Mathematics*, 2021, vol. 14, no. 1, pp. 58–71. (in Russian). <https://doi.org/10.18721/JPM.14104>
15. Zhang D., Jiang C., Liang D., Cheng L. A review on TVD schemes and a refined flux-limiter for steady-state calculations. *Journal of Computational Physics*, 2015, vol. 302, pp. 114–154. <https://doi.org/10.1016/j.jcp.2015.08.042>
16. Sadin D.V. Simulation of physical instability on contact boundaries in multicomponent compressible gas flows by a hybrid large-particle method. *Numerical Methods and Programming*, 2020, vol. 21, no. 2, pp. 129–137. (in Russian). <https://doi.org/10.26089/NumMet.v21r211>
17. Semenov A.N., Berezhkina M.K., Krasovskaya I.V. Classification of shock wave reflections from a wedge. Part 2: Experimental and numerical simulations of different types of Mach reflections. *Technical Physics*, 2009, vol. 54, no. 4, pp. 497–503. <https://doi.org/10.1134/S1063784209040094>
18. Sadin D.V. Analysis of dissipative properties of a hybrid large-particle method for structurally complicated gas flows. *Computer Research and Modeling*, 2020, vol. 12, no. 4, pp. 757–772. (in Russian). <https://doi.org/10.20537/2076-7633-2020-12-4-757-772>
19. Sadin D.V. Efficient implementation of the hybrid large-particle method. *Matematicheskoe Modelirovanie*, 2022, vol. 34, no. 4, pp. 113–127. (in Russian). <https://doi.org/10.20948/mm-2022-04-08>

Автор

Садин Дмитрий Викторович — доктор технических наук, профессор, профессор, Военно-космическая академия имени А.Ф. Можайского, Санкт-Петербург, 197198, Российская Федерация, <https://orcid.org/0000-0001-5335-4847>, sadin@yandex.ru

Author

Dmitry V. Sadin — D. Sc., Full Professor, Mozhaisky Military Aerospace Academy, Saint Petersburg, 197198, Russian Federation, <https://orcid.org/0000-0001-5335-4847>, sadin@yandex.ru

Статья поступила в редакцию 25.05.2022
Одобрена после рецензирования 10.06.2022
Принята к печати 14.07.2022

Received 25.05.2022
Approved after reviewing 10.06.2022
Accepted 14.07.2022



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

doi: 10.17586/2226-1494-2022-22-4-792-803

УДК 533.95:53.06

Численная модель импульсного подкритического стримерного сверхвысокочастотного разряда для задач плазменного поджига топливных смесей в газовой фазе

Павел Викторович Булат¹, Константин Николаевич Волков²,
Ангелика Игоревна Мельникова³✉, Максим Евгеньевич Ренев⁴

^{1,2,3,4} Балтийский государственный технический университет «ВОЕНМЕХ» им. Д.Ф. Устинова, Санкт-Петербург, 190005, Российская Федерация

² Университет Кингстона, Лондон, SW15 3DW, Великобритания

⁴ Санкт-Петербургский государственный университет, Санкт-Петербург, 199034, Российская Федерация

¹ pavelbulat@mail.ru, <https://orcid.org/0000-0003-0099-9953>

² k.volkov@kingston.ac.uk, <https://orcid.org/0000-0001-6055-2323>

³ angelika044@gmail.com✉, <https://orcid.org/0000-0001-7229-0965>

⁴ renewme@mail.ru, <https://orcid.org/0000-0002-0583-0607>

Аннотация

Предмет исследования. Разработана и верифицирована приближенная модель, предназначенная для оценки плазменного нагрева и конверсии топливных смесей при помощи подкритического стримерного сверхвысокочастотного разряда. Поджиг топливной смеси происходит в среде с давлением 13 кПа и температурой 150 К при наличии внешнего потока воздуха со скоростью до 500 м/с. Для фокусировки электромагнитного излучения использованы антенна-инициатор и плоское зеркало. Стехиометрическая смесь пропана и воздуха или чистый пропан подаются через полость в антенне. Мощность излучения составила 3 кВт. **Метод.** Модель реализована на основе схемы расщепления по физическим процессам, что позволило снизить требования к вычислительным ресурсам. Область, занятая плазмой, и ее проводимость заданы исходя из выбранного набора экспериментальных данных. Расчет плазменного поджига состоял из трех этапов. На первом этапе решены уравнения Больцмана для электронного газа в среде в нульмерной постановке для заданных параметров внешнего электрического поля. В результате получены функции распределения электронной энергии и коэффициентов реакций. На втором этапе рассчитаны уравнения Гельмгольца для распределения напряженности электромагнитного поля вблизи антенны-инициатора с учетом заданной проводящей области. По полученным распределениям электрического поля вычислены мощности джоулева нагрева и значения коэффициентов реакций. На третьем этапе получены решения уравнений: Навье-Стокса и переноса различных сортов частиц для сжимаемой среды с учетом процессов горения при заданных источниках локального нагрева и плазменных реакций. **Основные результаты.** Получены распределения температуры, состава среды, скорости движения среды при заданных локальной мощности нагрева и дополнительных реакций в области, занятой плазмой. Стехиометрическая смесь пропана с воздухом или чистый пропан, подаваемые через антенну, подожжены плазмой: смесь горит в небольшой области, а пропан окисляется в тонком слое смешения с воздухом. Выполнены сравнения результатов расчетов с данными физического эксперимента: и полей температур, и состава среды с фотографиями пламени из эксперимента. Численное исследование показало, что во всех рассмотренных условиях модель дает близкие к эксперименту результаты, но имеется завышение требуемой для поджига мощности излучения практически в два раза. **Практическая значимость.** Изучение процессов поджига газообразных смесей подкритическим сверхвысокочастотным разрядом представляет интерес для проектирования двигательных установок с повышенной надежностью и возможностью использования трудновоспламеняемых смесей. Предлагаемая модель дает приближенные оценки, позволяя снизить требования к вычислительным ресурсам и времени счета по сравнению с традиционными моделями.

Ключевые слова

плазменный поджиг, численное моделирование, газовая динамика, подкритический разряд, плазма

Благодарности

Работа выполнена при финансовой поддержке Министерства науки и высшего образования Российской Федерации в ходе реализации проекта «Создание опережающего научно-технического задела в области разработки передовых технологий малых газотурбинных, ракетных и комбинированных двигателей сверхлегких ракет-носителей, малых космических аппаратов и беспилотных воздушных судов, обеспечивающих приоритетные позиции российских компаний на формируемых глобальных рынках будущего», № FZWF-2020-0015.

Ссылка для цитирования: Булат П.В., Волков К.Н., Мельникова А.И., Ренев М.Е. Численная модель импульсного подкритического стримерного сверхвысокочастотного разряда для задач плазменного поджига топливных смесей в газовой фазе // Научно-технический вестник информационных технологий, механики и оптики. 2022. Т. 22, № 4. С. 792–803. doi: 10.17586/2226-1494-2022-22-4-792-803

Numerical model of a pulsed subcritical streamer microwave discharge for problems of plasma ignition of fuel mixtures in the gas phase

Pavel V. Bulat¹, Konstantin N. Volkov², Anzhelika I. Melnikova³✉, Maksim E. Renev⁴

^{1,2,3,4} Baltic State Technical University «VOENMEH» named after D.F. Ustinov, Saint Petersburg, 190005, Russian Federation

² Kingston University, London, SW15 3DW, United Kingdom

⁴ St. Petersburg State University (SPbSU), Saint Petersburg, 199034, Russian Federation

¹ pavelbulat@mail.ru, <https://orcid.org/0000-0003-0099-9953>

² k.volkov@kingston.ac.uk, <https://orcid.org/0000-0001-6055-2323>

³ angelika044@gmail.com✉, <https://orcid.org/0000-0001-7229-0965>

⁴ renevme@mail.ru, <https://orcid.org/0000-0002-0583-0607>

Abstract

An approximate model for estimating plasma heating and conversion caused by a subcritical streamer microwave discharge has been considered and verified. Ignition occurs in an environment with a pressure of 13 kPa, a temperature of 150 K, there is an external air flow at a speed of up to 500 m/s, an initiator antenna and a flat mirror are used to focus electromagnetic radiation, a stoichiometric mixture of propane-air and pure propane is supplied through the cavity in the antenna. Radiation power is 3 kW. The model is three-stage and semi-empirical. The plasma region and its conductivity are set based on experimental statistics; this is a key feature that reduces the consumption of computing resources. The finite element method is used. At the first stage, the Boltzmann equation for the electron gas in the medium is solved in the zero-dimensional formulation for the given parameters of the external electric field. The distribution functions of the electronic energy are obtained as well as the functions of the reaction coefficients. At the second stage, the Helmholtz equations are solved to obtain the distribution of electromagnetic fields near the antenna-initiator, taking into account the given conducting region “plasma”. Based on the obtained distributions of the electric field, the Joule heating powers and the values of the reaction coefficients are calculated. At the third stage, the Navier-Stokes and transfer equations of various types of particles for a compressible medium are solved; taking into account combustion processes for given local heating and plasma reactions. The results are compared with the data of a physical experiment. The distributions of temperature, composition of the medium, velocity of the medium are considered for given local heating power and additional reactions in the plasma region. A stoichiometric mixture of propane with air and pure propane supplied through the antenna are ignited by plasma under the conditions under consideration: the mixture burns in a small area, propane is oxidized in a thin layer of mixing with air. The temperature fields and composition of the medium are compared with photographs of the flame from the experiment. The numerical study shows that under all the conditions considered the model gives results close to the experiment, but there is an overestimation of the power required for ignition by up to two times. The study of the processes of ignition of gaseous mixtures by a subcritical microwave discharge is of interest for the design of propulsion systems with increased reliability, with the possibility of using hardly flammable mixtures. The proposed model gives approximate estimates, while the requirements for resources and time are significantly reduced compared to classical models.

Keywords

plasma ignition, numerical simulation, propane, gas dynamics, subcritical microwave discharge

Acknowledgements

This work was financially supported by the Ministry of Science and Higher Education of Russian Federation during the implementation of the project “Creating a leading scientific and technical reserve in the development of advanced technologies for small gas turbine, rocket and combined engines of ultra-light launch vehicles, small spacecraft and unmanned aerial vehicles that provide priority positions for Russian companies in emerging global markets of the future”, No. FZWF-2020-0015.

For citation: Bulat P.V., Volkov K.N., Melnikova A.I., Renev M.E. Numerical model of a pulsed subcritical streamer microwave discharge for problems of plasma ignition of fuel mixtures in the gas phase. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2022, vol. 22, no. 4, pp. 792–803 (in Russian). doi: 10.17586/2226-1494-2022-22-4-792-803

Введение

В настоящее время потенциал эволюционного развития аэрокосмической техники в значительной степени исчерпан. В связи с этим ведется активный поиск новых подходов и технологий. Одним из перспективных направлений считаются плазменные технологии. Например, для создания высокоскоростных летательных аппаратов требуется, чтобы их двигательная установка могла эффективно работать на больших скоростях. Чем больше скорость потока, тем сложнее обеспечить своевременное воспламенение топливной смеси. При этом на сверхзвуковых скоростях потока классический способ свечного поджига перестает работать. Один из способов поджига топливной смеси состоит в размещении искровых разрядников в полости, в которой скорость внешнего потока значительно уменьшается [1]. Однако такие полости создают сопротивление для внешнего потока. Имеет смысл исследовать более «мощные» типы разрядов, способные поджигать смесь, находясь в сверхзвуковом потоке.

Широко используемый поджиг искрами, горячей плазмой, характеризуется тем, что в относительно малом объеме смесь нагревается до температуры воспламенения, после чего выделяется энергия, и волна горения распространяется по всей камере сгорания. Такой подход к воспламенению смеси осложняется малым временем ее пребывания в рабочей камере при сверхзвуковых скоростях потока.

В неравновесной плазме, в которой основную роль играет горячий электронный газ, имеют место реакции диссоциации, возбуждения и ионизации частиц при низкой температуре среды [2], что значительно увеличивает скорость горения [3]. В работе [5] исследована возможность снижения выбросов вредных газов за счет использования неравновесной плазмы с увеличением полноты сгорания топливной смеси. Широко известное горение водорода в кислороде (водородная цепь) улучшено применением неравновесной плазмы путем дополнительных реакций диссоциации водорода, кислорода и водородно-кислородных радикалов [2]. В углеводородах необходимо учесть, что атомы водорода могут соединиться в молекулы путем отрыва пары от углеводородов [6].

Выявленный и подтвержденный многими исследованиями эффект увеличения скорости распространения пламени при наличии в области горения холодной неравновесной плазмы использован: в перспективных плазменных устройствах зажигания [3]; при подавлении высокочастотных колебаний в камере сгорания (режимов вибрационного горения) [7]; стабилизации горения в бедных [8] и «трудных» топливных смесях в малоэмиссионных газотурбинных двигателях и энергетических установках, работающих на органических отходах; предварительной ионизации компонентов топлива, подаваемых в камеру сгорания, что уменьшает задержку воспламенения и делает более благоприятным градиент температуры по длине камеры [9]. Для перечисленных случаев разработаны численные методы, основанные на термических эффектах и наличии в потоке электронного газа и свободных ионов [10, 11].

Предмет исследования

В ряде задач требуется рассчитать не только течение, содержащее холодную неравновесную плазму, но и ее источник. Например, оценить суммарную энергетическую эффективность установки. Широко распространены модели, в которых решены системы уравнений с законами сохранения рассматриваемых частиц, электронной энергии и электродинамики. На практике необходимо учесть большое количество частиц и реакций, чтобы получить физически корректные результаты. Например, для задачи нагрева чистого азота сверхвысокочастотным (СВЧ) разрядом взяты в расчет не менее 56 реакций и 15 сортов частиц вместе с «дополнительной» температурой среды, описывающей колебательные состояния азота при неравновесном возбуждении [12]. О сложности полноценного расчета плазмы в среде даже без топлива свидетельствует и работа [13], в которой получены результаты согласованного расчета наносекундного межэлектродного разряда. В [14] рассмотрены различные типы нагрева среды плазмой. В результате можно поставить задачу о необходимости учета значительного количества частиц и реакций для точных оценок.

Проведен ряд исследований по получению неравновесной плазмы с помощью электродной системы [11]. Подобные разряды уменьшают задержку воспламенения, позволяя сократить размеры камеры сгорания газотурбинных двигателей [9] и увеличить импульс жидкостных ракетных двигателей, что подтверждено результатами численных расчетов [15]. В связи с этим возможно использовать наносекундные импульсы напряжения, чтобы разряд не перешел в горячую фазу. Отметим, что применение электродов для создания высокой напряженности поля требует лишь источника напряжения достаточной мощности. Использование острых электродов позволяет снизить напряжение пробоя и обеспечить возникновение разряда на острие.

Диэлектрические барьеры [16, 17] повышают подаваемую мощность на электроды без повреждения установки, но при этом необходимо обеспечить разрядку барьеров, иначе прекратится протекание электрического тока. В ходе протекания тока одной полярности происходит зарядка диэлектриков и снижение напряженности электрического поля в плазме. Такие разряды часто используются в установках снижения выбросов на электростанциях.

Отдельно отметим, что поджиг жидких углеводородов неравновесной плазмой сильно затруднен: электронный газ достигает при подаче энергии извне высоких температур благодаря высокой подвижности электронов, редко взаимодействующих с остальными частицами в газообразной среде, но, попав в жидкость, они сразу теряют это свойство. Авторами считается, что можно достичь поджига неравновесной плазмой за счет использования пламени с большим количеством свободных радикалов от небольшого количества газообразного топлива и плазмой для нагрева и активации сложных углеводородов. Например, согласно работе [18], керосин активно взаимодействует с гидроксильной группой — атомами водорода.

Имеет важное значение исследование задач создания детонационных волн для перспективных детонационных ракетных двигателей [19]. Для получения высоких энергий и удельного импульса требуется, с одной стороны, использование высококалорийного топлива, а с другой — топливная смесь должна иметь возможность детонировать. В тонком слое фронта детонационной волны происходит быстрое выделение энергии и сгорание топливной смеси, что приводит к возможности работы двигательной установки на высоких скоростях с пониженными потерями теплоты через стенки камеры и уменьшенными вредными выбросами. Использование неравновесной плазмы позволяет повысить детонационную способность топливной смеси. В работе [20] показано, что СВЧ-плазма экономичнее и быстрее, чем искровой разряд, поджигает пропано-воздушную смесь.

Приближенные модели значительно проще, чем полноценные модели со строгой постановкой, и позволяют проводить расчеты за меньшее время с малыми затратами ресурсов. У таких моделей меньше точность, и они предоставляют меньше данных, требуя использования специальных приближений. Данные модели позволяют сделать предварительные оценки возможности получения тех или иных эффектов, выбрав наиболее перспективные комбинации входных параметров. В работе [4] предложено количество активных частиц в потоке, а в [21] заданы размеры предварительно нагретой области, что позволяет исследовать изменение параметров сверхзвукового обтекания. В работе [3] для исследования плазменного поджига с использованием одностадийной реакции без учета плазмохимии применен такой же подход.

Цель работы — построение приближенной математической модели плазмы и ее воздействия на топливную смесь, которую планируется в дальнейшем применить для расчета плазменного горения в присутствии подкритического СВЧ-разряда — разряда электрической плазмы, полученного на тонком инициаторе под воздействием СВЧ-излучения. Для уменьшения расходуемой мощности излучателя использованы фокусирующие системы и тонкие антенны, которые фокусируют излучение, локально повышая напряженность электрического поля, создаваемую излучением. В результате, при заданной электропрочности (минимальной напряженности поля для создания плазмы) возможно получить плазму, имея излучение с недостаточной напряженностью до фокусировки. Для разряда нужны затравочные электроны. Они могут образовываться, например, при воздействии солнечной радиации. Затравочные электроны разгоняются в электромагнитном поле и накапливают энергию для реакций ионизации. Важная особенность разработанной модели состоит в зависимости плазменных эффектов от параметров СВЧ-системы (нагрева и дополнительного набора реакций), достигаемой за счет дополнительных предварительных этапов моделирования. Подкритический СВЧ-разряд характеризуется объемным воздействием, достигая высоких концентраций электронов [22]. Использование пропана объяснено наличием надежных подробных схем горения (используется Grimech 3.0).

Условия среды, при которых исследуются особенности горения пропано-воздушных смесей, определены экспериментами из работы [3], используемые для верификации модели. Параметры выполнения работы: давление 13 кПа и температура 150 К окружающей среды, которая движется со скоростью до 500 м/с, обтекая специальную полую антенну-инициатор; через полость подается топливная смесь с заданным давлением избытка и составом (стехиометрическая смесь с давлением 100 кПа, чистый пропан — 20 кПа); на антенну падает СВЧ-излучение с длиной волны 12,5 см и мощностью 3 кВт.

Математическая модель

Модель для расчета плазменного поджига газобразных смесей углеводородов с воздухом за счет импульсного СВЧ-разряда с неравновесной плазмой вблизи инициатора построена на основе схемы расщепления по физическим процессам и состоит из трех этапов (рис. 1).

Этап 1. Расчет функции распределения электронной энергии (ФРЭЭ), зависимостей коэффициентов реакций с участием электронов как функций амплитуды внешнего электрического поля при заданных частоте, количественной плотности среды, составе среды и зависимостей сечений таких реакций от энергии электрона.

Этап 2. Расчет в трехмерной постановке распространения СВЧ-излучения в присутствии фокусирующих систем, антенны, заданной проводящей области «плазма» вблизи антенны. Решены уравнения Гельмгольца для монохромного излучения. Основная искомая величина — распределение электрического поля в плазме, по которому вычислены локальные значения коэффициентов реакций, а также мощность джоулевого нагрева. Локальные значения усреднены по области «плазма» и использованы на Этапе 3. Плазменные эффекты зависят от конкретной мощности СВЧ-излучения, геометрии и состояния среды.

Этап 3. Расчет газодинамики горения в присутствии области «плазма» с дополнительным набором реакций и заданной удельной мощностью нагрева. Решены уравнения газодинамики для сжимаемой среды и законы сохранения сортов частиц.

Разделение модели на разные этапы позволяет существенно сократить потребление вычислительных ресурсов, времени вычислений и происходит за счет использования специальных приближений.

Допущения и приближения. Плазма СВЧ-разряда — неравновесна. Большинство реакций ионизации с повышением проводимости происходят за счет горячего электронного газа низкой плотности. Внешний источник питания — импульсный, соответственно плазма не успевает перейти в равновесную стадию.

Плазма образуется и приобретает устойчивое значение проводимости и форму в течение нескольких микросекунд и далее ее поддерживает. Благодаря этому допускается задание плазмы извне с параметрами, основанными на экспериментальной статистике.

Состояние электронного газа определено локальным значением амплитуды электрического поля. При

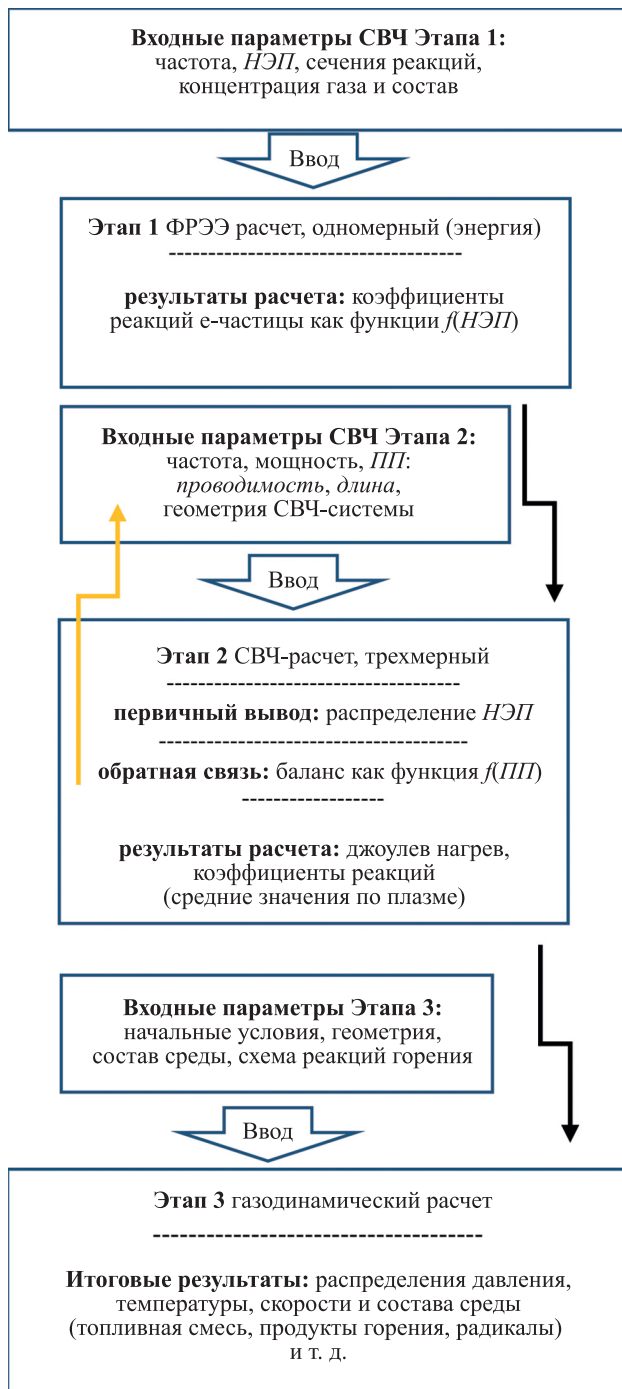


Рис. 1. Модель с расщеплением по физическим процессам.

Жирным шрифтом обозначены: входные параметры; результаты расчета этапов для использования на следующих этапах и итоговые результаты расчетов.

Курсивом — параметры, которые необходимо перебирать в ходе решения этапов.

НЭП — напряженность электрического поля; *ПП* — плазменные параметры

Fig. 1. The setup of the three-stage model used in this study.

The following are indicated in bold: input parameters; the results of the calculation of the stages to be used in the next stages and the final results of the calculations.

Italics — parameters that need to be sorted out during the solution stages.

НЭП — electric field strength; *ПП* — plasma parameters

этом состав среды считается постоянным. Небольшие изменения состава среды (менее 1 %, до поджига) пренебрежимо слабо меняют ФРЭЭ и значения коэффициентов плазменных реакций.

На данном этапе развития модели не учтены реакции фотоионизации. Плазменные эффекты без фотоионизации ожидаются менее интенсивными.

Этап 1. Решим стационарные кинетические уравнения Больцмана для однородного электронного газа в среде во внешнем осциллирующем электрическом поле с заданной амплитудой для поиска равновесной ФРЭЭ. Используем метод конечных элементов — построим сетку базисных функций в одномерном пространстве электронной энергии.

Уравнения Больцмана для функции распределения электронов по энергиям имеют вид:

$$\frac{\partial}{\partial \epsilon} \left(W F_e - D \frac{\partial F_e}{\partial \epsilon} \right) = \Sigma S_{in}, \quad (1)$$

$$\gamma = \sqrt{\frac{2|e|}{m_e}},$$

$$W = -\gamma \epsilon^2 \Sigma \sigma(\epsilon),$$

$$D = \frac{\gamma}{3} E_N^2 + \frac{\gamma k_B T_g}{|e|} \epsilon^2 \sigma(\epsilon),$$

где E_N — приведенный модуль амплитуды электрического поля; ϵ — энергия электрона; F_e — ФРЭЭ; W и D — коэффициенты подвижности и диффузии электронной энергии в пространстве энергий; S_{in} — мощности охлаждения электронного газа неупругими реакциями с участием электронов; γ — электронный параметр; $\sigma(\epsilon)$ — зависимости сечений реакций от кинетической энергии электронов из базы данных Ixcat, а также работы [6]; T_g — температура среды; k_B — постоянная Больцмана; $|e|$ — элементарный заряд, m_e — масса электрона.

Электронный газ нагревается внешним осциллирующим электрическим полем, электроны соударяются с частицами среды. В правой части уравнения (1) учитываются потери энергии электронов на реакции в среде. Определим коэффициент подвижности электронной энергии с помощью сечений соударений, а диффузии — нагревом внешнего электрического поля и взаимодействием со средой.

Используем естественные граничные условия для вычисления ФРЭЭ при следующих правилах: не существует электронов с отрицательной кинетической энергией и, следовательно, невозможно их взаимодействие с электронами, у которых энергия больше нуля, также не существует электронов с бесконечной энергией

$$\left. \frac{\partial F_e}{\partial \epsilon} \right|_{\epsilon=0} = 0, \quad F_e|_{\epsilon=\epsilon_{\max}} = 0$$

В численном расчете считается, что нет электронов с достаточно большой энергией ϵ_{\max} (выбрано значение 150 эВ).

При расчете ФРЭЭ применим приведенные параметры внешнего электрического поля, нормированные на

количественную плотность среды N_g , вычисленную из уравнения состояния идеального газа:

$$N_g = \frac{p_g}{k_B T_g}, E_N = \frac{E}{N_g}, f_N = \frac{f_{MW}}{N_g},$$

где p_g — давление среды; E и f_{MW} — модуль амплитуды и частота электрического поля; f_N — приведенный модуль частота электрического поля.

Рассмотрим на Этапе 1 реакции с участием электронов: упругие соударения, возбуждения, ионизации, диссоциации и прилипания. Исследуемые частицы среды: азот, кислород, пропан.

Для численного решения уравнения (1) построим сетку из 3000 линейных элементов в пространстве энергии от 0 до 150 эВ с коэффициентом увеличения размера элемента 10. Итеративно подберем стационарное распределение электронов по энергии таким образом, чтобы полный источник электронной энергии и потоки в пространстве энергии были равны нулю. Итеративный процесс начнем с максвелловского распределения со средней энергией 2 эВ. Последовательность приближений выполним по методу Ньютона–Рафсона до тех пор, пока не будет получен заданный уровень невязки 10^{-8} .

Рассчитаем коэффициенты реакций k_{ei} по заданным сечениям и ФРЭЭ

$$k_{ei} = \gamma \int_0^\infty \epsilon \sigma_e(\epsilon) F_e(\epsilon) d\epsilon. \quad (2)$$

Коэффициенты (2) вычислим для ряда значений амплитуды электрического поля, которые оказываются табличными зависимостями от этого поля. Данные коэффициенты необходимы для постобработки решения на Этапе 2.

Этап 2. Решим уравнения Гельмгольца для монохромного СВЧ-излучения при наличии проводящей области:

$$\nabla \times (\nabla \times \mathbf{E}) - k_0^2 \left(1 - \frac{i\sigma_{cond}}{2\pi f_{MW} \epsilon_0} \right) \mathbf{E} = 0, \quad (3)$$

$$\mathbf{B} = -\frac{1}{2\pi i f_{MW}} [\nabla \times \mathbf{E}],$$

где ∇ — оператор набла; $\mathbf{E}(x, y, z)$ — вектор напряженности электрического поля; $\mathbf{B}(x, y, z)$ — вектор магнитной индукции; k_0 — волновой вектор в открытом пространстве; i — мнимая единица; σ_{cond} — проводимость среды (не равна нулю в плазме); ϵ_0 — электрическая постоянная.

Зададим область «плазма» исходя из экспериментальной статистики, при этом четко видно, где возникает плазма, какая ее форма и примерное значение проводимости или концентрации зарядов. В модели учтен инструмент проверки плазмы на возможность ее существования с точки зрения сравнения реакций, повышающих и понижающих проводимость, а также уточняющий свойства плазмы. При этом целесообразно сделать перебор нескольких параметров плазмы, например длину и концентрацию электронов в ней.

По заданной концентрации электронов в плазме, а также их подвижности при известной количественной плотности среды и известном составе плазмы определим ее проводимость для уравнения (3):

$$\sigma_{cond} \equiv |e| (b_e n_e + \sum_i |z_i|^2 b_i n_i) \approx |e| b_e n_e,$$

где b_e и b_i — подвижности электронов и ионов в электрическом поле в среде; z_i — зарядовое число иона. Ионы на два порядка менее подвижны, имеют сопоставимую концентрацию (плазма квазинейтральная), поэтому их вкладом можно пренебречь.

Для расчета распространения СВЧ-излучения (3) в фокусирующей системе в модели учтены следующие элементы: инициатор и зеркальная поверхность под ним, плазма с заданными свойствами и зона входа для СВЧ-излучения. Геометрия для расчета СВЧ-поля показана на рис. 2. В данном случае плазма представляет шесть цилиндров, расположенных на конце инициатора — окружности с равномерным распределением по ней.

Уравнения (3) дополним граничными условиями. Инициатор и зеркало идеально отражают излучение, поэтому:

$$\mathbf{n} \times \mathbf{E} = 0,$$

где \mathbf{n} — вектор нормали поверхности.

Для экономии вычислительных ресурсов решим половину модели. На плоскости симметрии зададим условие нулевой касательной компоненты электрического поля, которое имеет вид:

$$\mathbf{n} \times \mathbf{B} = 0.$$

Над инициатором установим условие входа для излучения через площадь 100 см^2 (для половины модели 50 см^2) с заданной поляризацией и мощностью. Электрическое поле направлено вдоль инициатора и имеет амплитуду E_0 , следовательно:

$$\mathbf{E} = E_0 \mathbf{e}_x.$$

На остальных границах определено условие рассеяния второго порядка

$$\mathbf{n} \times (\nabla \times (\mathbf{E})) - i k_0 \mathbf{n} (\mathbf{E} \times \mathbf{n}) - \frac{1}{2 i k_0} \nabla \times (\mathbf{n} \nabla \times \mathbf{E}) = 0.$$

Для решения уравнений (3) Этапа 2 построим сетку со следующими параметрами: на расстоянии от малых объектов выбраны размеры элементов с характерной длиной до 1/5 от длины волны. Вблизи малых объектов сетка сгущается так, чтобы не менее шести элементов были расположены в узких промежутках, таких как диаметр плазмоида. Выберем квадратичные элементы — 600 тыс. элементов. Используем геометрический многосеточный метод решения разностных уравнений на основе V-цикла. Создадим дополнительно сетку с уменьшенным в два раза количеством элементов.

Вычислим средние значения коэффициентов реакций k_{je}^n с участием частиц среды j и электронов, удель-

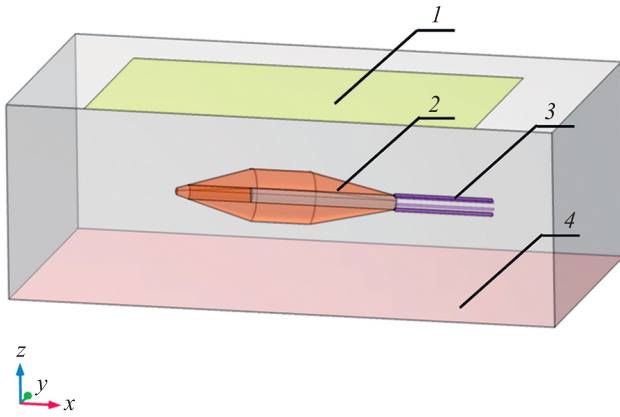


Рис. 2. Используемая геометрия для расчета СВЧ-поля.
1 — зона входа СВЧ-излучения; 2 — антенна; 3 — область «плазма»; 4 — отражающая плоскость; xz — плоскость симметрии (условие отсутствия касательной компоненты магнитной индукции); остальные границы (условие свободного прохождения излучения второго порядка)

Fig. 2. The studied geometry for MW-field calculation (1 — MW-radiation input zone, 2 — initiator, 3 — plasma domain, 4 — mirror-plane, xz — plane of symmetry — a condition of zero tangential magnetic field; the rest boundaries — a condition of second order open boundary)

ную мощность джоулева нагрева W_{joule} по области «плазма» Ω_{plas} и оператор интегрирования $\hat{I}(A)$:

$$\begin{aligned} \hat{I}(A) &= \int_{\Omega_{plas}}^{def} A dx, \\ W_{joule} &= \frac{\hat{I}(\mathbf{j}, \mathbf{E})}{\hat{I}(1)}, \\ k_{je}^n &= \frac{\hat{I}(k_{je}^n(|\mathbf{E}|))}{\hat{I}(1)}, \end{aligned} \quad (4)$$

где A — некоторая полевая функция, к которой применяется оператор интегрирования; dx — элементарный объем; \mathbf{j} — плотность тока проводимости.

Выполним проверку плазмы на возможность ее существования. Рассмотрим различные плазмиды, и введем критерий баланса электронов — разницу между всеми скоростями реакций ионизации, рекомбинации и прилипания. Если при заданных условиях баланс электронов B меньше нуля, то плазма не существует, существенно больше нуля — плазма развивается. Если плазма с положительным балансом, близким к нулю, то она максимально развита, и ее необходимо выбрать. Введем скорость реакции частиц i, j с заданными молярными долями w :

$$S_{ij}^n = k_{ij}^n n_i n_j, n_i = N_e w_i.$$

Тогда получим суммарные скорости всех реакций ионизации S^+ с участием электронов и прилипания:

$$S^+ = \sum_{n \in \{e^+\}} i, j; S_{ie}^n, S^e = \sum_{n \in \{e^-\}} i, j; S_{ie}^n,$$

где $\{e^+\}$ и $\{e^-\}$ — множества реакций типа ионизации и прилипания соответственно. Вычислим реакцию

рекомбинации S^- по известному коэффициенту рекомбинации C^- в приближении квазинейтральности. Типичное значение коэффициента рекомбинации равно $10^{-13} \text{ м}^3/\text{с}$.

$$S^- = C^- n_e n_i = [n_e = n_i] = C_{rec} n_e^2.$$

В результате имеем:

$$B = S^+ - S^e - S^-.$$

Этап 3. Решим уравнения газодинамики в пакете Ansys FLUENT для сжимаемой среды в осесимметричной постановке. Для проверки модели поджига выполним сравнение с экспериментальными данными. В среде присутствуют частицы воздуха, пропана, промежуточных и конечных продуктов горения (схемы реакций Grimech 3.0). Используем модель турбулентности $k-\epsilon$ со стандартными для пакета Ansys FLUENT параметрами. При этом средняя мощность джоулева нагрева (4) задана равномерно в области «плазма». Примем, что плазма неподвижна, значит в плазме известно количество электронов и ионов, ранее полученное на Этапе 2 сохранится. Чтобы не вводить новые частицы (электроны и ионы) и не переделывать значительно схему реакций, а также не настраивать их источники, они не рассматриваются явно. Вместо этого коэффициенты реакций умножаются на концентрации таких частиц, а реакции, начиная с этого этапа считаются одночастичными, а не двухчастичными и происходят только в домене плазма. С технической точки зрения это упрощенный аналог «замороженных» частиц в среде.

Рассмотренные реакции с участием ионов — реакции обмена заряда, сопровождающиеся диссоциацией частиц, при этом коэффициенты реакций взяты из работы [23].

Используем сетку с характерным размером элемента 0,1 мм и элементами второго порядка, и неявную разностную схему первого порядка. Решим переходную задачу до получения стационарного решения. Шаг по времени составляет 2 мкс.

Геометрия и граничные условия показаны на рис. 3. По инициатору-форсунке подается топливная смесь пропана с воздухом, сама форсунка обдувается чистым воздухом.

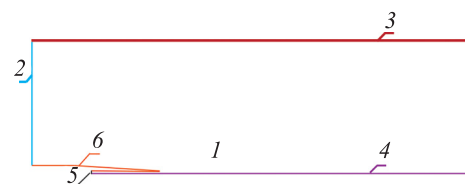


Рис. 3. Геометрия для расчета газодинамики горения в присутствии области «плазма» (1).

Граничные условия: 2 и 3 — вход и выход для внешнего потока воздуха; 4 — ось симметрии; 5 — вход для топливной смеси; 6 — стенка (антенна)

Fig. 3. Geometry to study gas dynamics of propane-air flame in case of plasma domain (1).

Boundary conditions: 2 — inlet, exterior flux of air, 3 — outlet, 4 — symmetry axis, 5 — inlet, fuel mixture, 6 — wall condition (initiator)

Анализ результатов

Проверим трехэтапную приближенную модель СВЧ-поджига газообразных смесей с воздухом по имеющимся экспериментальным данным.

Анализ результатов Этапов 1 и 2. На Этапе 1 выполнены расчеты кинетических уравнений (1) для электронов в пространстве энергии. Электроны нагреты внешним СВЧ-полем, их амплитуда изменяется параметрически, и они могут совершать упругие столкновения, а также реакции возбуждения, диссоциации и ионизации частиц среды.

Отметим важный результат: основные плазменные реакции — диссоциация кислорода и пропана с образованием отдельных атомов, молекулы водорода, целой молекулы метана или его радикала. При этом возможна ионизация частиц одновременно с диссоциацией. Подобные реакции приводят к заметной плазменной конверсии пропана, к плазменному поджигу (вместе с диссоциацией молекулы кислорода), и к поддержке проводимости плазмы. На Этапе 2 вычислено распространение СВЧ-излучения при наличии антенны. После получения результатов для разных проводимостей плазмы и ее длины (l) выполнена проверка баланса реакций (рис. 4). В результате баланс уменьшился и имеет отрицательное значение при увеличении длины и проводимости плазмы (не показаны на рис. 4), так как сложнее ее поддерживать при заданной мощности СВЧ.

Выберем плазму с параметрами: длина 6 мм, концентрация электронов 10^{20} $1/\text{м}^3$. Проведем оценку коэффициентов реакций при суммарной по всем сортам частиц удельной мощности нагрева 220 ГВт/ м^3 . Отметим, что только 67 % мощности джоулева нагрева ионов и электронного газа будет приложено к среде, так как часть энергии рассеивается.

Анализ результатов Этапа 3. На Этапе 3 осуществлено сравнение результатов расчета газодинамики в присутствии области «плазма» с данными экспериментов из работы [3]. В результате эксперимента получены фотографии пламени при следующих условиях: давление среды 13 кПа, температура 150 К, внешний поток воздуха 85–500 м/с, который подается через ан-

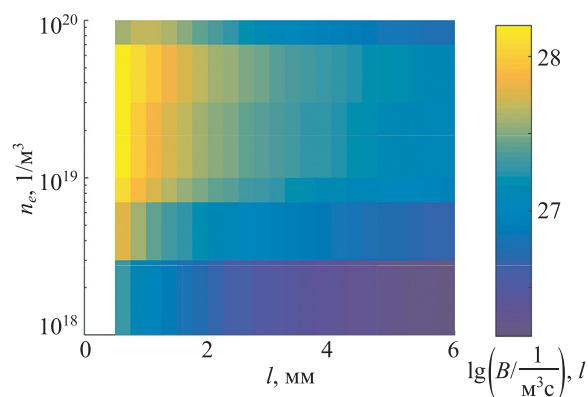


Рис. 4. Зависимость десятичного логарифма функции баланса от длины плазмоида и концентрации электронов

Fig. 4. The Balance Decimal logarithm values at different plasma lengths and electron concentrations

тенну. На выходе получена стехиометрическая смесь пропан-воздух с избыточным давлением 101 кПа или чистый пропан — 20,2 кПа. Фотографии пламени при данных условиях были сопоставлены с результатами численного моделирования, выполненного при тех же условиях, и добавлены комментарии о наблюдаемых особенностях пламени.

Рассмотрим три экспериментальных случая с различными параметрами.

Случай 1: скорость внешнего потока воздуха 85 м/с, через инициатор подается стехиометрическая смесь. Фотография пламени и результаты расчета (поля температуры и модуля скорости движения) показаны в таблице. На фотографии видно, что имеется область свечения, обусловленная наличием возбужденных частиц. Свечение имеет яркий фиолетовый оттенок — излучают частицы с высоким уровнем возбуждения, характерным для неравновесной плазмы. Такие частицы могут образовываться в плазме, затем диффузионно распространяться. Также наблюдается рассеивание свечения в среде. Иными словами, область фиолетового свечения, видимая в эксперименте, должна быть не меньше, чем область действия плазмы, и предположительно пламени. По фотографии можно сказать, что ее длина составляет примерно половину высоты конусной задней части инициатора. Пламени пропана (голубое свечение) не видно, значит, область горения находится внутри области фиолетового свечения.

В результате расчета получены поля температуры и скорости движения среды с подходящими под эксперимент особенностями: на выходе инициатора происходит горение с повышением температуры до 20 кК и с расширением среды со скоростью до 1,2 км/с. Указанные области имеют размер около 1/3, 1/2 от высоты конусной части инициатора, а их форма близка к овальной. Заметим, что результаты расчета для описываемых условий не противоречат эксперименту. Аналогично проанализируем случаи использования стехиометрической смеси и чистого пропана таблицу при скорости внешнего потока 500 м/с.


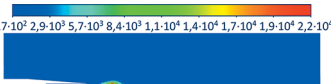
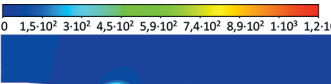

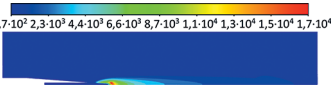
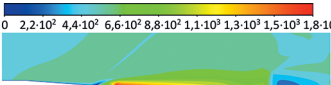

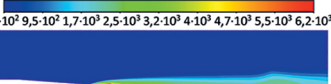
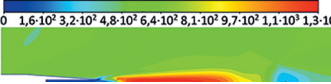
Случай 2: в таблице видно, что в сверхзвуковом потоке область свечения смещена от инициатора, в остальном особенности аналогичны случаю скорости потока 85 м/с. Результаты расчетов согласуются с экспериментом: имеется пламя, область повышенной температуры меньше области яркого свечения.

Отметим случай 3 — горение при подаче чистого пропана при скорости внешнего потока 500 м/с. Пропана достаточно, чтобы было четко видно пламя в большом объеме вне области фиолетового свечения. Длина пламени сопоставима с высотой конуса инициатора. Распределение температуры по данным моделирования: имеется длинная струя с высокой температурой порядка 2 кК, которая снижается при движении газа от плазмы в направлении выхода. В эксперименте в этом же направлении снижается интенсивность свечения пламени.

На рис. 5 приведены зависимости интегральных потоков ряда продуктов f сгорания и активных радикалов через самую правую границу модели от скорости внешнего потока V . Представлены данные для стехио-

Таблица. Фотографии из эксперимента для верификации модели и результаты расчета для случаев стехиометрической смеси и чистого пропана

Table. Photos from the experiment for model verification and calculation results for the cases of stoichiometric mixture and pure propane

Скорость внешнего потока, м/с	Фотография из эксперимента для верификации модели	Результаты расчета	
		распределение температуры, К	распределение модуля скорости, м/с
Случай 1 (подаётся стехиометрическая смесь)			
85			
Случай 2 (подаётся стехиометрическая смесь)			
500			
Случай 3 (подаётся чистый пропан)			
500			

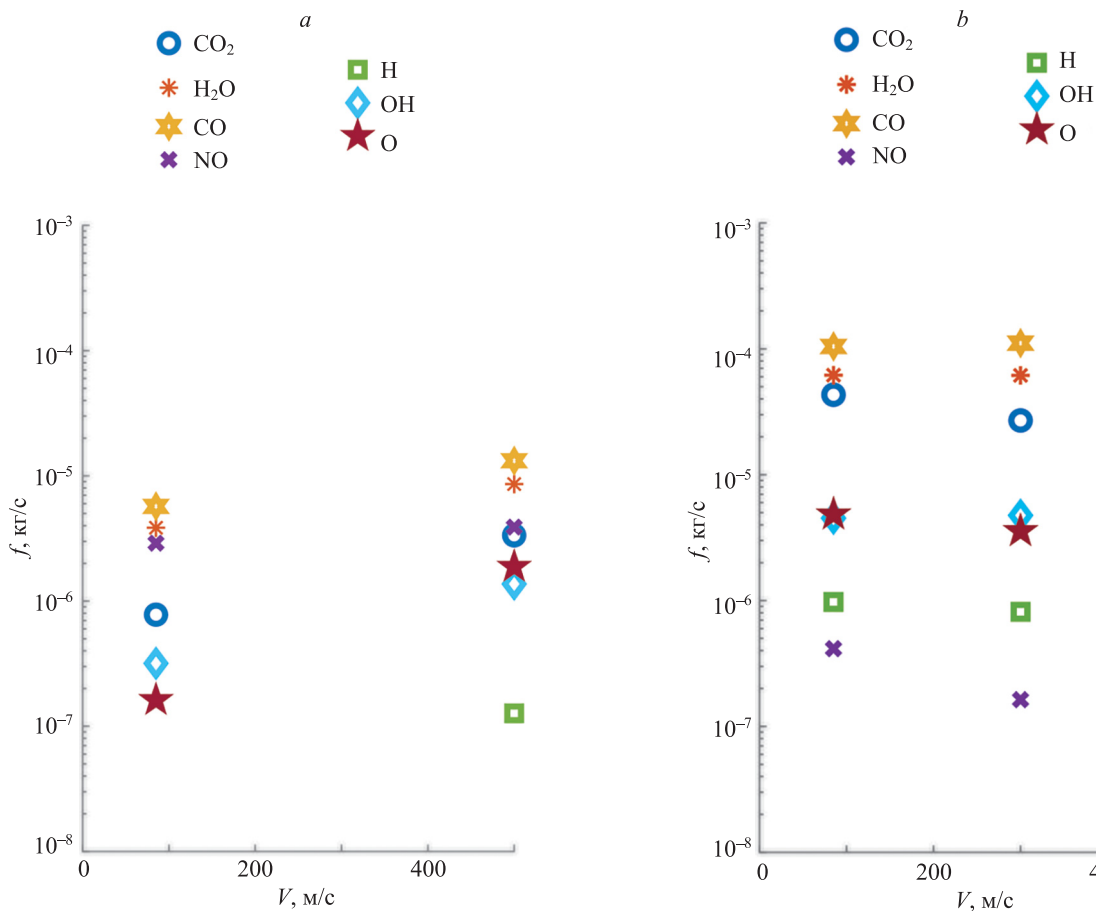


Рис. 5. Зависимость потока ряда продуктов сгорания через самую правую границу модели «выход» от скорости внешнего потока воздуха: стехиометрическая смесь (а) и чистый пропан (б)

Fig. 5. The dependence of the mass flux of a number of combustion products through the most right boundary “output” of the model vs. the speed of the external air flow: a stoichiometric mixture (a), pure propane (b)

метрической смеси и чистого пропана с избыточными давлениями 100 и 20 кПа соответственно.

В случае стехиометрической смеси при увеличении скорости внешнего потока увеличиваются и потоки продуктов сгорания и радикалов. Однако их концентрации в среде падают: увеличение потоков в 2–5 раз при увеличении скорости почти в 6 раз. Заметим, что имеется эффект эжекции, на большей скорости при заданном давлении подачи смеси, ее во внешнюю среду должно попадать больше. Следовательно, смесь не успевает прогореть, однако наблюдается факт горения.

Отметим, что при увеличении скорости внешнего обтекания получим увеличение потока атомарных кислорода и водорода на порядок. Данная особенность может быть полезна для применения системы как «плазменно-пламенного» факела для разогрева и «плазменной», тепловой конверсии сложных топливных смесей.

Если заменить стехиометрическую смесь на чистый пропан, то снова получим горение, а также на порядок больше атомарного водорода в потоке, но в меньшей степени увеличится количество атомарного кислорода и гидроксильной группы. Такое увеличение потоков связано с тем, что в расчетную область попадает больше пропана, содержащего водород. Плазма же воздействует лишь на него, подогревая и осуществляя конверсию при сильно пониженном содержании кислорода в потоке топливной струи. «Активированный» пропан (радикалы пропана, водород) на границе двух потоков (струя топлива, внешний поток воздуха) реагируют с кислородом. Происходят реакции, подобные водородной цепи, атомизируется кислород, образуется гидроксо-группа, но для этого нужно время, поэтому у таких двух веществ потоки не возрасли более, чем на порядок относительно случая стехиометрической смеси.

У воды, угарного и углекислого газов кратно увеличился поток при изменении смеси, поскольку в расчетной области существенно больше поступает пропана. При этом оксид азота образуется в меньшем количестве: плазма более не действует на внешний поток воздуха, не образует атомарные кислород, азот, которые затем соединятся в оксид. С точки зрения создания наиболее горячего пламени в значительном объеме с уменьшением выброса оксида азота, с увеличением количества активных радикалов горение чистого пропана во внешнем потоке воздуха перспективнее, чем стехиометрической смеси пропан-воздух.

В результате горения струи чистого пропана в потоке воздуха с дефицитом кислорода внутри струи пропана в продуктах сгорания образован сильно разогретый активизированный синтез-газ с высоким содержанием активных радикалов.

Имеет смысл дальнейшее исследование горения чистых газообразных топлив в потоке с окислителем для поджига смесей с жидким топливом (керосином) неравновесной плазмой через «посредника». Прямой

поджиг затруднителен — электроны быстро прилипают к каплям топлива.

Можно сделать вывод, что результаты моделирования приемлемо соотносятся с экспериментальными данными. Отсутствующая область фиолетового свечения в результате эксперимента объясняется неучтенными в расчете фотонами, остальные видимые особенности пламени и плазмы согласуются.

Заключение

Выполнены расчеты на основе трехэтапной приближенной модели поджига пропана в воздухе с помощью подкритического сверхвысокочастотного разряда. Результаты моделирования близки к экспериментальным данным. Модель верифицирована за счет сравнения полей температуры и концентраций различных сортов частиц с фотографиями из эксперимента других авторов. Для всех рассмотренных случаев выполнен поджиг топливной смеси.

Использована СВЧ-система со следующими параметрами: мощность излучения 3 кВт, длина волны излучения 12,5 см. При этом параметры среды: давление 13 кПа, температура 150 К, она движется с заданной скоростью движения относительно антенны-инициатора до 500 м/с. Через инициатор подается либо стехиометрическая смесь пропан-воздух, либо чистый пропан с избыточными давлениями 101 и 20 кПа.

Модель позволила осуществить анализ продуктов сгорания топливной смеси, сделать первичные оценки перед тем, как ставить эксперимент или численное моделирование, где будет более строго и согласованно решаться полная система уравнений плазмогодинамики и горения. Например, один из интересных результатов — сгорание чистого пропана во внешнем потоке воздуха с образованием горячего сингаза со значительным количеством радикалов: атомы водорода, кислорода, гидроксильной группы. Такую топливную смесь, полагаем, можно использовать для «плазменно-газового» поджига смесей с жидким топливом, и это требует дальнейших исследований.

Отметим, что в работе использовано ключевое приближение: параметры и геометрия плазмы считаются известными. Это приводит к существенному снижению ресурсозатратности модели и времени расчета, поскольку не рассматриваются многочисленные уравнения переноса частиц на этапе расчета СВЧ-полей, требующие соблюдения критериев устойчивости. Допускается повторное использование результатов предложенных Этапов 1 и 2 моделирования для новых задач. В дальнейшем планируется добавить дополнительный этап оценки эффектов взаимодействия фотонов со средой, добавить модель роста стримеров, использовать упрощенную модель для предварительных оценок возможности создания детонационной волны в круглой трубе за счет интенсивного плазменного поджига.

Литература

References

- Feng R., Wang Z., Sun M., Wang H., Huang Y., Yang Y., Liu X., Wang C., Tian Y., Luo T., Zhu J. Multi-channel gliding arc plasma-assisted ignition in a kerosene-fueled model scramjet engine // *Aerospace Science and Technology*. 2022. V. 126. P. 107606. <https://doi.org/10.1016/j.ast.2022.107606>
- Starikovskaia S.M. Plasma assisted ignition and combustion // *Journal of Physics D: Applied Physics*. 2006. V. 39. P. R265–R299. <https://doi.org/10.1088/0022-3727/39/16/R01>
- Bulat M.P., Bulat P.V., Denissenko P.V., Esakov I.I., Grachev L.P., Lavrov P.V., Volkov K.N., Volobuev I.A. Plasma-assisted ignition and combustion of lean and rich air/fuel mixtures in low- and high-speed flows // *Acta Astronautica*. 2020. V. 176. P. 700–709. <https://doi.org/10.1016/j.actaastro.2020.04.028>
- Chen Q., Ge J., Zheng T., Che X., Nie W. The role of non-equilibrium plasma kinetic effect on GCH_4/GOX rocket engine combustion performance // *Journal of Physics: Conference Series*. 2020. V. 1707. P. 012015. <https://doi.org/10.1088/1742-6596/1707/1/012015>
- Kotel'nikov V.A., Kotel'nikov M.V., Filippov G.S. Electrical and physical parameters of plasma fluxes in exhaust from a liquid-propellant rocket engine // *Journal of Machinery Manufacture and Reliability*. 2018. V. 47. N 6. P. 488–494. <https://doi.org/10.3103/S1052618818060067>
- Janev R.K., Reiter D. Collision processes of C_2H_y and C_2H_y^+ hydrocarbons with electrons and protons // *Physics of Plasmas*. 2004. V. 11. N 2. P. 780–829. <https://doi.org/10.1063/1.1630794>
- Zhou S., Nie W., Tian Y. High frequency combustion instability control by discharge plasma in a model rocket engine combustor // *Acta Astronautica*. 2021. V. 179. P. 391–406. <https://doi.org/10.1016/j.actaastro.2020.11.010>
- Bulat M.P., Bulat P.V., Denissenko P.V., Esakov I.I., Grachev L.P., Volkov K.N., Volobuev I.A. Ignition of lean and stoichiometric air-propane mixture with a subcritical microwave streamer discharge // *Acta Astronautica*. 2018. V. 150. P. 153–161. <https://doi.org/10.1016/j.actaastro.2017.11.030>
- Kim W., Cohen J. Plasma-assisted combustor dynamics control at ambient and realistic gas turbine conditions // *Proceedings of the ASME Turbo Expo*. 2017. V. 4A. P. V04AT04A037. <https://doi.org/10.1115/GT2017-63477>
- Bulat M., Bulat P., Denissenko P., Esakov I., Grachev L., Volkov K., Volobuev I. Numerical Simulation of ignition of premixed air/fuel mixtures by microwave streamer discharge // *IEEE Transactions on Plasma Science*. 2019. V. 47. N 1. P. 62–68. <https://doi.org/10.1109/TPS.2018.2869259>
- Sharma A., Subramaniam V., Solmaz E., Raja L. Fully coupled modeling of nanosecond pulsed plasma assisted combustion ignition // *Journal of Physics D: Applied Physics*. 2019. V. 52. N 9. P. 095204. <https://doi.org/10.1088/1361-6463/aaf690>
- Saifutdinov A.I., Kustova E.V. Dynamics of plasma formation and gas heating in a focused-microwave discharge in nitrogen // *Journal of Applied Physics*. 2021. V. 129. N 2. P. 023301. <https://doi.org/10.1063/5.0031020>
- Biturin V.A., Bocharov A.N., Dobrovolskaya A.S., Kuznetsova T.N., Popov N.A., Filimonova E.A. Numerical modeling of pulse-periodic nanosecond discharges // *Journal of Physics: Conference Series*. 2021. V. 2100. N 1. P. 012032. <https://doi.org/10.1088/1742-6596/2100/1/012032>
- Popov N.A., Starikovskaia S.M. Relaxation of electronic excitation in nitrogen/oxygen and fuel/air mixtures: fast gas heating in plasma-assisted ignition and flame stabilization // *Progress in Energy and Combustion Science*. 2022. V. 91. P. 100928. <https://doi.org/10.1016/j.pecs.2021.100928>
- Zheng T., Che X., Li L., Chen C., Nie W., Li X. Numerical study of plasma assisted combustion for a rocket combustor using GCH_4/GOX as propellants // *Journal of Physics: Conference Series*. 2018. V. 1064. P. 012013. <https://doi.org/10.1088/1742-6596/1064/1/012013>
- Zheng Z., Nie W., Zhou S., Tian Y., Zhu Y., Shi T., Tong Y. Characterization of the effects of a plasma injector driven by AC dielectric barrier discharge on ethylene-air diffusion flame structure // *Open Physics*. 2020. V. 18. N 1. P. 58–73. <https://doi.org/10.1515/phys-2020-0008>
- Deng J., He L., Liu X., Chen Y. Numerical simulation of plasma-assisted combustion of methane-air mixtures in combustion chamber // *Plasma Science and Technology*. 2018. V. 20. N 12. P. 125502. <https://doi.org/10.1088/2058-6272/aacdef>
- Feng R., Wang Z., Sun M., Wang H., Huang Y., Yang Y., Liu X., Wang C., Tian Y., Luo T., Zhu J. Multi-channel gliding arc plasma-assisted ignition in a kerosene-fueled model scramjet engine. *Aerospace Science and Technology*, 2022, vol. 126, pp. 107606. <https://doi.org/10.1016/j.ast.2022.107606>
- Starikovskaia S.M. Plasma assisted ignition and combustion. *Journal of Physics D: Applied Physics*, 2006, vol. 39, no. 16, pp. R265–R299. <https://doi.org/10.1088/0022-3727/39/16/R01>
- Bulat M.P., Bulat P.V., Denissenko P.V., Esakov I.I., Grachev L.P., Lavrov P.V., Volkov K.N., Volobuev I.A. Plasma-assisted ignition and combustion of lean and rich air/fuel mixtures in low- and high-speed flows. *Acta Astronautica*, 2020, vol. 176, pp. 700–709. <https://doi.org/10.1016/j.actaastro.2020.04.028>
- Chen Q., Ge J., Zheng T., Che X., Nie W. The role of non-equilibrium plasma kinetic effect on GCH_4/GOX rocket engine combustion performance. *Journal of Physics: Conference Series*, 2020, vol. 1707, pp. 012015. <https://doi.org/10.1088/1742-6596/1707/1/012015>
- Kotel'nikov V.A., Kotel'nikov M.V., Filippov G.S. Electrical and physical parameters of plasma fluxes in exhaust from a liquid-propellant rocket engine. *Journal of Machinery Manufacture and Reliability*, 2018, vol. 47, no. 6, pp. 488–494. <https://doi.org/10.3103/S1052618818060067>
- Janev R.K., Reiter D. Collision processes of C_2H_y and C_2H_y^+ hydrocarbons with electrons and protons. *Physics of Plasmas*, 2004, vol. 11, no. 2, pp. 780–829. <https://doi.org/10.1063/1.1630794>
- Zhou S., Nie W., Tian Y. High frequency combustion instability control by discharge plasma in a model rocket engine combustor. *Acta Astronautica*, 2021, vol. 179, pp. 391–406. <https://doi.org/10.1016/j.actaastro.2020.11.010>
- Bulat M.P., Bulat P.V., Denissenko P.V., Esakov I.I., Grachev L.P., Volkov K.N., Volobuev I.A. Ignition of lean and stoichiometric air-propane mixture with a subcritical microwave streamer discharge. *Acta Astronautica*, 2018, vol. 150, pp. 153–161. <https://doi.org/10.1016/j.actaastro.2017.11.030>
- Kim W., Cohen J. Plasma-assisted combustor dynamics control at ambient and realistic gas turbine conditions. *Proceedings of the ASME Turbo Expo*, 2017, vol. 4A, pp. V04AT04A037. <https://doi.org/10.1115/GT2017-63477>
- Bulat M., Bulat P., Denissenko P., Esakov I., Grachev L., Volkov K., Volobuev I. Numerical Simulation of ignition of premixed air/fuel mixtures by microwave streamer discharge. *IEEE Transactions on Plasma Science*, 2019, vol. 47, no. 1, pp. 62–68. <https://doi.org/10.1109/TPS.2018.2869259>
- Sharma A., Subramaniam V., Solmaz E., Raja L. Fully coupled modeling of nanosecond pulsed plasma assisted combustion ignition. *Journal of Physics D: Applied Physics*, 2019, vol. 52, no. 9, pp. 095204. <https://doi.org/10.1088/1361-6463/aaf690>
- Saifutdinov A.I., Kustova E.V. Dynamics of plasma formation and gas heating in a focused-microwave discharge in nitrogen. *Journal of Applied Physics*, 2021, vol. 129, no. 2, pp. 023301. <https://doi.org/10.1063/5.0031020>
- Biturin V.A., Bocharov A.N., Dobrovolskaya A.S., Kuznetsova T.N., Popov N.A., Filimonova E.A. Numerical modeling of pulse-periodic nanosecond discharges. *Journal of Physics: Conference Series*, 2021, vol. 2100, no. 1, pp. 012032. <https://doi.org/10.1088/1742-6596/2100/1/012032>
- Popov N.A., Starikovskaia S.M. Relaxation of electronic excitation in nitrogen/oxygen and fuel/air mixtures: fast gas heating in plasma-assisted ignition and flame stabilization. *Progress in Energy and Combustion Science*, 2022, vol. 91, pp. 100928. <https://doi.org/10.1016/j.pecs.2021.100928>
- Zheng T., Che X., Li L., Chen C., Nie W., Li X. Numerical study of plasma assisted combustion for a rocket combustor using GCH_4/GOX as propellants. *Journal of Physics: Conference Series*, 2018, vol. 1064, pp. 012013. <https://doi.org/10.1088/1742-6596/1064/1/012013>
- Zheng Z., Nie W., Zhou S., Tian Y., Zhu Y., Shi T., Tong Y. Characterization of the effects of a plasma injector driven by AC dielectric barrier discharge on ethylene-air diffusion flame structure. *Open Physics*, 2020, vol. 18, no. 1, pp. 58–73. <https://doi.org/10.1515/phys-2020-0008>
- Deng J., He L., Liu X., Chen Y. Numerical simulation of plasma-assisted combustion of methane-air mixtures in combustion chamber. *Plasma Science and Technology*, 2018, vol. 20, no. 12, pp. 125502. <https://doi.org/10.1088/2058-6272/aacdef>

18. Zettervall N., Fureby C., Nilsson E.J.K. A reduced chemical kinetic reaction mechanism for kerosene-air combustion // *Fuel*. 2020. V. 269. P. 117446. <https://doi.org/10.1016/j.fuel.2020.117446>
19. Ma J.Z., Luan M.Y., Xia Z.-J., Wang J.-P., Zhang S.-J., Yao S.-B., Wang B. Recent progress, development trends, and consideration of continuous detonation engines // *AIAA Journal*. 2020. V. 58. N 12. P. 4976–5035. <https://doi.org/10.2514/1.J058157>
20. Bulat P.V., Chernyshov P., Esakov I.I., Grachev L., Lavrov P., Melnikova A.I., Volkov K.N., Volobuev I.A. Multi-point ignition of air/fuel mixture by the initiated subcritical streamer discharge // *Acta Astronautica*. 2022. V. 194. P. 504–513. <https://doi.org/10.1016/j.actaastro.2021.09.043>
21. Dobrov Y.V., Lashkov V.A., Mashek I.Ch., Khoronzuk R.S. Investigation of heat flux on aerodynamic body in supersonic gas flow with local energy deposition // *AIP Conference Proceedings*. 2018. V. 1959. P. 050009. <https://doi.org/10.1063/1.5034637>
22. Булат П.В., Есаков И.И., Грачев Л.П., Денисенко П.В., Булат М.П., Волобуев И.А. Математическое и компьютерное моделирование горения и детонации подкритическим стримерным разрядом // Научно-технический вестник информационных технологий, механики и оптики. 2017. Т. 17. № 4. С. 569–592. <https://doi.org/10.17586/2226-1494-2017-17-4-569-592>
23. Kossyi I.A., Kostinsky A.Yu., Matveyev A.A., Silakov V.P. Kinetic scheme of the non-equilibrium discharge in nitrogen-oxygen mixtures // *Plasma Sources Science and Technology*. 1992. V. 1. N 3. P. 207–220. <https://doi.org/10.1088/0963-0252/1/3/011>
18. Zettervall N., Fureby C., Nilsson E.J.K. A reduced chemical kinetic reaction mechanism for kerosene-air combustion. *Fuel*, 2020, vol. 269, pp. 117446. <https://doi.org/10.1016/j.fuel.2020.117446>
19. Ma J.Z., Luan M.Y., Xia Z.-J., Wang J.-P., Zhang S.-J., Yao S.-B., Wang B. Recent progress, development trends, and consideration of continuous detonation engines. *AIAA Journal*, 2020, vol. 58, no. 12, pp. 4976–5035. <https://doi.org/10.2514/1.J058157>
20. Bulat P.V., Chernyshov P., Esakov I.I., Grachev L., Lavrov P., Melnikova A.I., Volkov K.N., Volobuev I.A. Multi-point ignition of air/fuel mixture by the initiated subcritical streamer discharge. *Acta Astronautica*, 2022, vol. 194, pp. 504–513. <https://doi.org/10.1016/j.actaastro.2021.09.043>
21. Dobrov Y.V., Lashkov V.A., Mashek I.Ch., Khoronzuk R.S. Investigation of heat flux on aerodynamic body in supersonic gas flow with local energy deposition. *AIP Conference Proceedings*, 2018, vol. 1959, pp. 050009. <https://doi.org/10.1063/1.5034637>
22. Bulat P.V., Esakov I.I., Grachev L.P., Denissenko P.V., Bulat M.P., Volobuev I.A. Modeling and simulation of combustion and detonation by subcritical streamer discharge. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2017, vol. 17, no. 4, pp. 569–592. (in Russian). <https://doi.org/10.17586/2226-1494-2017-17-4-569-592>
23. Kossyi I.A., Kostinsky A.Yu., Matveyev A.A., Silakov V.P. Kinetic scheme of the non-equilibrium discharge in nitrogen-oxygen mixtures. *Plasma Sources Science and Technology*, 1992, vol. 1, no. 3, pp. 207–220. <https://doi.org/10.1088/0963-0252/1/3/011>

Авторы

Булат Павел Викторович — доктор физико-математических наук, кандидат экономических наук, главный научный сотрудник, Балтийский государственный технический университет «ВОЕНМЕХ» им. Д.Ф. Устинова, Санкт-Петербург, 190005, Российская Федерация, [sc 55969578400](https://orcid.org/0000-0003-0099-9953), <https://orcid.org/0000-0003-0099-9953>, pavelbulat@mail.ru

Волков Константин Николаевич — доктор физико-математических наук, лектор Университет Кингстона, Лондон, SW15 3DW, Великобритания; ведущий научный сотрудник, Балтийский государственный технический университет «ВОЕНМЕХ» им. Д.Ф. Устинова, Санкт-Петербург, 190005, Российская Федерация, [sc 8663950000](https://orcid.org/0000-0001-6055-2323), <https://orcid.org/0000-0001-6055-2323>, k.volkov@kingston.ac.uk

Мельникова Анжелика Игоревна — младший научный сотрудник, Балтийский государственный технический университет «ВОЕНМЕХ» им. Д.Ф. Устинова, Санкт-Петербург, 190005, Российская Федерация, [sc 57222383655](https://orcid.org/0000-0001-7229-0965), <https://orcid.org/0000-0001-7229-0965>, angelika044@gmail.com

Ренев Максим Евгеньевич — младший научный сотрудник, Балтийский государственный технический университет «ВОЕНМЕХ» им. Д.Ф. Устинова, Санкт-Петербург, 190005, Российская Федерация; аспирант, Санкт-Петербургский государственный университет, Санкт-Петербург, 199034, Российская Федерация, [sc 57211271545](https://orcid.org/0000-0002-0583-0607), <https://orcid.org/0000-0002-0583-0607>, renewme@mail.ru

Authors

Pavel V. Bulat — D. Sc. (Physics and Mathematics), PhD (Economics), Chief Researcher, Baltic State Technical University «VOENMEH» named after D.F. Ustinov, Saint Petersburg, 190005, Russian Federation, [sc 55969578400](https://orcid.org/0000-0003-0099-9953), <https://orcid.org/0000-0003-0099-9953>, pavelbulat@mail.ru

Konstantin N. Volkov — D. Sc. (Physics and Mathematics), Senior Lecturer, Kingston University, London, SW15 3DW, United Kingdom; Leading Researcher, Baltic State Technical University «VOENMEH» named after D.F. Ustinov, Saint Petersburg, 190005, Russian Federation, [sc 8663950000](https://orcid.org/0000-0001-6055-2323), <https://orcid.org/0000-0001-6055-2323>, k.volkov@kingston.ac.uk

Anzhelika I. Melnikova — Junior Researcher, Baltic State Technical University «VOENMEH» named after D.F. Ustinov, Saint Petersburg, 190005, Russian Federation, [sc 57222383655](https://orcid.org/0000-0001-7229-0965), <https://orcid.org/0000-0001-7229-0965>, angelika044@gmail.com

Maksim E. Renev — Junior Researcher, Baltic State Technical University «VOENMEH» named after D.F. Ustinov, Saint Petersburg, 190005, Russian Federation; PhD Student, St. Petersburg State University (SPbSU), Saint Petersburg, 199034, Russian Federation, [sc 57211271545](https://orcid.org/0000-0002-0583-0607), <https://orcid.org/0000-0002-0583-0607>, renewme@mail.ru

Статья поступила в редакцию 15.05.2022
Одобрена после рецензирования 17.06.2022
Принята к печати 28.07.2022

Received 15.05.2022
Approved after reviewing 17.06.2022
Accepted 28.07.2022



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

doi: 10.17586/2226-1494-2022-22-4-804-811

Numerical study on the straight, helical and spiral capillary tube for the CO₂ refrigerant

Pravin Jadhav¹, Anjan Kumar Sahu², Sunita Ballal³

^{1,3} Karamveer Bhaurao Patil College of Engineering, Satara, 415001, India

² Majhighariani Institute of Technology and Science, Rayagada, Odisha, 765017, India

¹ pravin.jadhav@kbpcoes.edu.in, <https://orcid.org/0000-0002-8043-8883>

² anjansahu111@gmail.com, <https://orcid.org/0000-0003-4385-1326>

³ Sunita.ballal@kbpcoes.edu.in, <https://orcid.org/0000-0002-8392-8095>

Abstract

A numerical study has been carried out for straight, spiral and helical capillary tubes and their performance has been compared with CO₂ refrigerant. The numerical models are developed based on the fundamental conservation principles of mass, momentum, and energy. Within this, outer loop, the ordinary differential equations are solved from the inlet to the exit of the capillary tube. The study has been carried out to calculate the mass flow rate by bisection method where the mass is iteratively calculated at the specified capillary length or vice versa. In-house coding programming employs the finite difference approach for numerical solutions. The characterization of the capillary tube has been done by calculating the length for the given mass or by calculating mass for the given length. The comparison of the straight capillary with helical capillary tube (50 mm coil diameter) and spiral capillary tube (50 mm pitch) has been reported. For a change in tube diameter, surface roughness, and length, the percentage reduction in mass flow rate in capillary tubes (straight, helical, and spiral) is calculated. The percentage reduction in mass in a helical capillary tube compared to the straight capillary tube is about 7–9 %. The percentage reduction in mass in a spiral tube compared to the straight capillary tube is nearly 23–26 %. Additionally, the percentage reduction in mass in a spiral tube compared to the helical capillary tube is almost 17–19 %. Additionally, the percentage reduction in length in a spiral tube compared to straight capillary tube ranges from 37 % to 43 %. Similarly, the percentage reduction in length in a spiral tube compared to helical capillary tube is ranging from 25 % to 32 %.

Keywords

straight tube, helical tube, spiral tube, capillary tube, adiabatic, mass flow rate

For citation: Jadhav P., Sahu A., Ballal S. Numerical study on the straight, helical and spiral capillary tube for the CO₂ refrigerant. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2022, vol. 22, no. 4, pp. 804–811. doi: 10.17586/2226-1494-2022-22-4-804-811

УДК 621.564

Численное исследование прямой, винтовой и спиральной капиллярных трубок для хладагента CO₂

Правин Джадхав¹, Анджан Кумар Саху², Сунита Баллал³

^{1,3} Инженерный колледж Карамвир Бхаурао Патил, Сатара, 415001, Индия

² Институт технологии и науки Маджхигариани, Раягада, Одиша, 765017, Индия

¹ pravin.jadhav@kbpcoes.edu.in, <https://orcid.org/0000-0002-8043-8883>

² anjansahu111@gmail.com, <https://orcid.org/0000-0003-4385-1326>

³ Sunita.ballal@kbpcoes.edu.in, <https://orcid.org/0000-0002-8392-8095>

Аннотация

Выполнено численное исследование прямых, спиральных и винтовых капиллярных трубок с хладагентом CO₂ и проведено сравнение их характеристик. Численные модели разработаны на основе фундаментальных принципов сохранения массы, импульса и энергии. Решены обыкновенные дифференциальные уравнения во

© Jadhav P., Sahu A., Ballal S., 2022

внешнем контуре от входа до выхода капиллярной трубки. Проведены расчеты массового расхода капиллярной трубки и хладагента методом деления отрезка пополам, где масса итеративно рассчитана при заданной длине капилляра или наоборот. Использовано собственное программирование численным методом конечных разностей. Определение характеристик капиллярной трубки выполнено путем расчета ее длины для заданной массы или расчета массы для заданной длины. Выполнено сравнение прямой капиллярной трубки с винтовой (внутренний диаметр 50 мм) и спиральной (шаг 50 мм). Для изменения диаметра трубки, шероховатости поверхности и длины рассчитан процент снижения массового расхода в прямых, винтовых и спиральных капиллярных трубках. В результате сравнения с прямой капиллярной трубкой получено процентное уменьшение массы трубок: для винтовой трубки — 7–9 %, для спиральной — 23–26 %. Для спиральной трубки также получено процентное уменьшение показателей: массы по сравнению с винтовой трубкой — 17–19 %; длины по сравнению с прямой трубкой — 37–43 % и винтовой — 25–32 %.

Ключевые слова

прямая трубка, винтовая трубка, спиральная трубка, капиллярная трубка, адиабатическая, массовый расход

Ссылка для цитирования: Джадхав П., Саху А.К., Баллал С. Численное исследование прямой, винтовой и спиральной капиллярных трубок для хладагента CO₂ // Научно-технический вестник информационных технологий, механики и оптики. 2022. Т. 22, № 4. С. 804–811 (на англ. яз.). doi: 10.17586/2226-1494-2022-22-4-804-811

Introduction

The use of heat pumps, refrigeration systems increased by almost 10 % in terms of total and interest income. Among these systems, a large area is covered by small capacity applications, and a capillary tube is used as an expansion device. The capillary tube is a relatively low-cost and simple narrow tube with different cross-sections: straight, helical, and spiral. The coiled capillary tube is preferable compared to the straight capillary tube. Along with the economic system, an environmentally friendly system is the need of the day. CO₂ refrigerant is a favorite among all-natural refrigerants. The sustainable solution for small capacity HVAC applications is the transcritical CO₂ system and capillary tube [1–2]. Numerous numerical studies of a straight capillary tube with various refrigerants were carried out. Stoecker and Jones [3] designed the numerical model of the straight capillary tube. The numerical model of the straight capillary tube was designed using basic equations of conservation: mass, energy, momentum. Jabaraj et al. [4] mentioned the influencing parameter in the capillary tube. The study was carried out for R22 and M20 refrigerants. Among all the geometric parameters, the tube diameter and tube length were more influential. Rasti and Jeong [5] formed a correlation for calculating the mass flow rate of R-12, R-22, R-134a, R-152a, R-404A, R-407C, R-410A, R-507A, and R-600a for the straight capillary tube. The flow characterization of the adiabatic capillary tubes was carried out. The developed correlations had been in comparatively good agreement with different application range. Choi et al. [6] framed an experimental model for straight capillary tube with R-290, R-22, R-407C refrigerant. The effect of numerous geometric parameters and operating parameters was calculated, and an empirical correlation was developed. Chingulpitak and Wongwises [7] revealed the numerical study for the R-22, R-407C, and R-410A on the helically coiled capillary tube employing the adiabatic and homogenous two-phase condition. After completing rigorous research, they suggest that the Mori and Nakamaya friction factor equation was the best among various correlations. Jadhav and Agrawal [8–10] carried out a numerical study for the helical and spiral capillary tube with CO₂ and R22 refrigerant. They suggest the coil diameter influenced the helical capillary tube and pitch

in the spiral capillary tube. Additionally, the gas cooler temperature in the CO₂ system is a crucial parameter to attain the optimum pressure. Mittal et al. [11] developed a numerical model for spiral capillary tubes for R22, R407C, and R410a refrigerants. The model was developed by considering adiabatic and metastable flow conditions. Flow behavior was reported at various operating and geometric conditions. Khan et al. [12] carried out an experimental study on an adiabatic spiral capillary tube with an R-134a refrigerant. The influence of geometrical parameters on the flow behavior of the capillary tube and operating parameters such as subcooling was investigated. A reduction of 5–15 % in the mass flow rate of the spiral capillary tube compared to the straight capillary tube has been reported. An empirical correlation was also developed to predict the mass flow rate of the refrigerant.

In the past, the literature reveals the studies on straight, helical, and spiral capillary tubes individually. However, capillary tubes comparative studies have not been done at the mature level for CO₂ refrigerant. Additionally, one must compare the straight, helically, and spirally coiled capillary tubes for a compact design at one glance. A compact and environmentally sustainable HVAC system is the need of today's world, and that can be addressed with various tube configurations of capillary tube and CO₂ refrigerants. The objective of the present study is to exhaustively compare the straight, helical, and spiral capillary tubes with carbon dioxide refrigerant, which helps to design the straight and coiled capillary tube. The comparison was carried out for more practical tube dimensions and operating conditions of the transcritical CO₂ system.

Mathematical Modeling

Mathematical modeling of the capillary tube is carried out for the straight, helical and spiral capillary tube. The capillary tube is divided into small discretized elements as shown in Fig. 1 for straight, helical and spiral capillary tube. In CO₂ system, two different flow regions are observed in the capillary tube, viz. single phase and two-phase region. For simplicity, the mathematical modeling of capillary tube is divided into single phase and two-phase region.

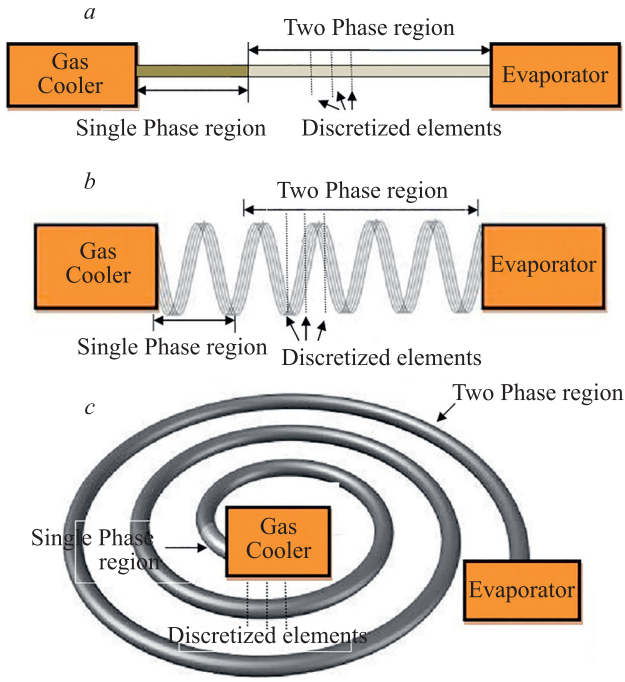


Fig. 1. Straight (a), helical (b) and spiral (c) capillary tube with discretized elements

Single Phase region

After the refrigerant cooling in the gas cooler, the single phase gaseous refrigerant pass through the capillary tube. To calculate flow characteristics of single phase region, basic equations of conservation of mass, momentum and energy are applied for straight, helical, and spiral capillary tube. The conservation of mass is applied for the elemental area, then it may be written as

$$m = m_1 = m_2$$

$$\frac{AV}{v} = \frac{A_1 V_1}{v_1} = \frac{A_2 V_2}{v_2}$$

Here m , m_1 , m_2 are the mass of CO₂ refrigerant flowing through of the elemental area of the tube, kg/s; A is the cross-sectional area tube, m²; V is the velocity of the fluid, m/s; and v is the specific volume, m³/kg. The flow in the capillary tube is steady and unidirectional. Applying Steady Flow Energy Equation (SFEE) to elemental area we get

$$h_1 + \frac{V_1^2}{2} + z_1 + Q = h_2 + \frac{V_2^2}{2} + z_2 + W.$$

The flow of the refrigerant is passing through the discretized session 1 and 2. Here h is the specific enthalpy, kJ/kg; V is the velocity of the fluid, m/s; z is the head across tube, m; Q is the heat transfer; and W is work done, kW. Assuming that capillary tube is adiabatic ($Q = 0$) with no any work done ($W = 0$) and the tube is horizontal ($z_1 = z_2$), above equation can be written as

$$h_1 + \frac{V_1^2}{2} = h_2 + \frac{V_2^2}{2}.$$

Using above equations

$$dh + \frac{dv^2 G^2}{2} = 0.$$

Here G is the mass flux, kg/(m²·s). Additionally, conservation of momentum is applied to the elemental area. The summation of the forces on the elemental area is equal to the acceleration of the refrigerant. The net momentum is given as

$$\Sigma f = ma$$

$$-dP - \frac{f_{sp} V G dL}{2d} = G dV$$

$$dL = \frac{2d}{f_{sp}} \left(\frac{v}{dv} - \frac{dp}{vG^2} \right).$$

Here f is force, N; P is the pressure, N/m; a is the acceleration of the fluid through tube, m/s²; L and dL are the total length and elemental length, respectively, m; f_{sp} is the friction factor in single capillary tube. In addition to the basic equations, friction factor equations are calculated by using following friction factor correlation for straight, helical and spiral capillary tube as follows.

1) For straight tube, Churchill [13], friction factor correlations may be expressed as

$$f_{st} = 8 \left[\left(\frac{8}{Re} \right)^{12} + (A^{16} + B^{16})^{-\frac{3}{12}} \right]^{\frac{1}{12}}.$$

Here f_{st} is friction factor in the straight capillary tube, Re is the Reynolds number, A and B may be calculated as

$$A = 2.457 \ln \left(\frac{1}{\left(\frac{7}{Re} \right)^{0.9} + \frac{0.27\varepsilon}{d}} \right), B = \frac{37530}{Re}, Re = \frac{Gd}{\mu},$$

$$dL = \frac{2d}{f_{sp}} \left(\frac{v}{dv} - \frac{dp}{vG^2} \right),$$

where ε is the surface roughness, mm; and μ is the dynamic viscosity, Pa·s.

2) For helical capillary tube, the elemental length is calculated using Mori and Nakamaya [14] friction factor relation

$$f_{hsp} = \frac{C_1 \left(\frac{d}{D_c} \right)^{0.5}}{\left[Re \left(\frac{d}{D_c} \right)^{\frac{n}{2}} \right]^{\frac{1}{n+1}}} \left[1 + \frac{C_2}{\left[Re \left(\frac{d}{D_c} \right)^{\frac{n}{2}} \right]^{\frac{1}{n+1}}} \right].$$

Here d is tube diameter; D_c is coil diameter of the helical capillary tube; f_{hsp} is the friction factor of the helical capillary tube. The value of n is considered as 5 because the Reynolds Number (Re) is larger than 10⁵.

$$\ln C_1 = \frac{1}{n+1} \times$$

$$\times \left\{ \frac{1}{4} [-3 \ln(2n+1) + (16n-7) \ln(2n-1) - (8n-3) \times \right.$$

$$\times [\ln n + \ln(4n-1)] + (6n-1)] + n \ln \alpha + 9 \ln 2 \left. \right\},$$

where α is the linearity constant, and it is calculated by employing the general friction factor equation

$$\ln C_2 = \frac{1}{n+1} \times \left\{ \begin{aligned} &\frac{1}{4} [3 \ln(2n+1) - (15n+4) \ln n + (19n-4) \times \\ &\times \ln(2n-1) - (7n-4) \ln(4n-1)] + n \ln \alpha - \\ &- n \ln(6n-1) - 9n \ln 2 \end{aligned} \right\}$$

$$dL = \frac{2d}{f_{hsp}} \left(\frac{v}{dv} - \frac{dp}{vG^2} \right).$$

- 3) Similarly, for spiral capillary tube elemental length is calculated using Ju et al. [15] friction factor relation

$$f_{csp} = f_{st} (1 + 0.11 \text{Re}^{0.23} (d/D_c))^{0.14},$$

f_{csp} is the friction factor of the spiral capillary tube and f_{st} is the friction factor in straight capillary tube

$$f_{st} = 0.1 (1.46 \varepsilon/d + 100/\text{Re})^{0.25}$$

$$dL = \frac{2d}{f_{csp}} \left(\frac{v}{dv} - \frac{dp}{vG^2} \right) \left(\frac{\cos\left(\frac{d\theta}{2}\right)}{\cos\left(\frac{d\theta}{4}\right)} \right).$$

Here, ε is the surface roughness, mm; and θ is angular displacement, rad.

Two-Phase region

While passing the refrigerant through subcritical region, two-phase region is existed. In CO₂ refrigerant the density ratio is low, so the two phase flow is considered as a homogenous flow. The modeling of the capillary tube is similar to the single-phase region except the dryness fraction (x) in the subcritical region. Using the dryness fraction, the enthalpy, entropy and specific entropy of the refrigerants are calculated as

$$h = h_f + xh_{fg},$$

$$v = v_f + xv_{fg},$$

$$s = s_f + xs_{fg}.$$

Here h , h_f , h_g , are specific enthalpy, specific enthalpy at saturated liquid, and specific enthalpy at saturated gas, respectively, and latent heat of evaporation is given as $h_{fg} = h_g - h_f$. Similarly, v , v_f , v_g , are specific volume,

specific volume at saturated liquid, and specific volume at saturated gas, respectively, and difference of specific volume of saturated gas and saturated liquid is termed as $v_{fg} = v_g - v_f$. Similarly, s , s_f , s_g , are specific entropy, specific entropy at saturated liquid, and specific entropy at saturated gas, respectively, and $s_{fg} = s_g - s_f$. The mean values of the friction factors (similar to section “Single phase region”) for straight, helical, and spiral capillary tube and specific volume are used to calculate the length of the respective capillary tube.

Solution Technique

The characterization of the capillary tube has been done by calculating the length for the given mass or calculating mass for the given length. The CO₂ refrigerant properties are given for this calculation. The calculation of mass is termed as simulation of the capillary tube and determination of length is called as designing of the capillary tube.

In the designing of the capillary tube, initially, for a given tube geometry and mass flow rate, evaporator temperature (T_{ev}) gas cooler temperature (T_{gc}) and pressure (P_{gc}) used in calculating elemental length. Adding the elemental length gives the total tube length. Likewise, total length is calculated for every flow regions, viz. supercritical, transcritical and subcritical region (as shown in Fig. 2). The pressure is considered as a marching parameter to identify the flow regions in the capillary tube. While the pressure is above the critical value (71 bar), the flow region is super-critical region. If the pressure is in between critical pressure and saturation pressure, the flow region is the transcritical region. At last, the flow region is subcritical region if it is between the saturation pressure and the evaporator pressure.

In the simulation study, as shown in Fig. 3, mass of the refrigerant is calculated for the given length. Initially mass is assumed and calculated for the total length of the tube. Comparing the calculated length (c) with the desired length (d) and if deviation is found in it, then the mass is to be changed, and again calculated length of tube. The cycle is repeated until the desired length gets equal to the calculated length and at that time assumed mass is the exact mass of the tube. The better approximate value may be found by bisection method. For numerical solutions, FORTRAN programming is used with in-house coding using the finite difference approach.

Result and Discussion

The straight, helical and spiral capillary tube has been compared. In the comparison considered, T_{gc} and

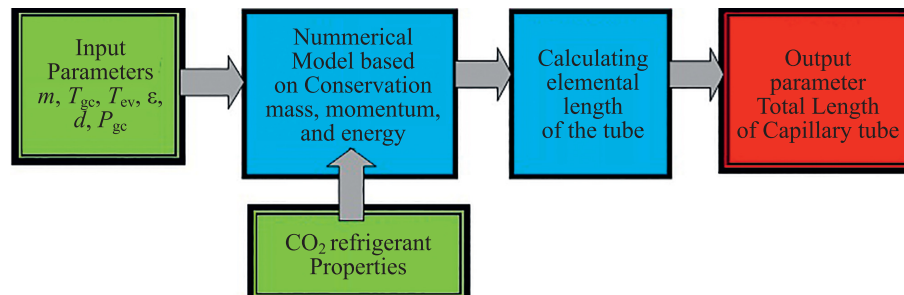


Fig. 2. Block diagram of designing of the capillary tube

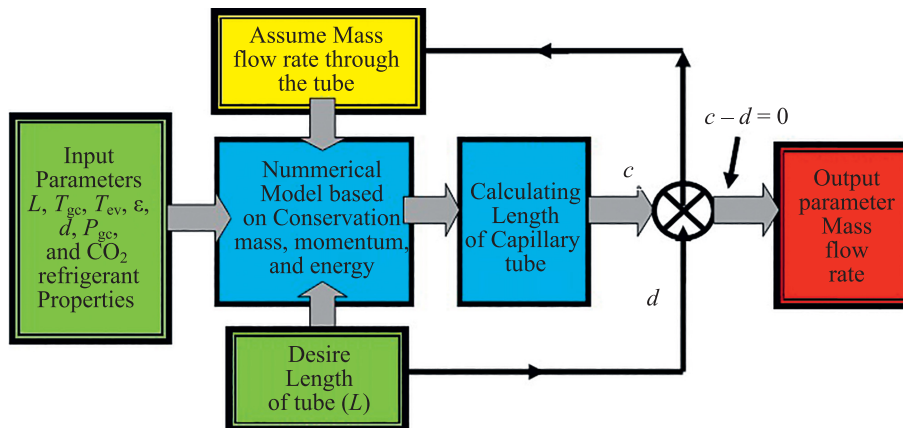


Fig. 3. Block diagram of simulation of the capillary tube

P_{gc} are gas cooler temperature and gas-cooler pressure, respectively. Similarly, T_{ev} and P_{ev} are evaporator temperature and evaporator pressure, respectively. Fig. 4 shows validation of present model result with the test result of Agrawal and Battacharyya for straight capillary tube. The present model results agree with the experimental results.

Fig. 5 indicates the validation of present model result with Wang et al. for helical capillary tube. The results agree with the test results of Wang et al.

Fig. 6 indicates the validation of present model result with Jadhav and Agrawal for spiral capillary tube. The results agree with the test results of Jadhav and Agrawal.

The comparison of straight capillary tube with helical tube (50 mm coil diameter) and spiral capillary tube (50 mm pitch) has been reported. For change in tube diameter, surface roughness and length of tube, the percentage reductions in mass flow rate in capillary tubes (straight, helical, and spiral) are calculated. The change in mass flow rate for straight, helical and spiral capillary tube, when changing the tube diameter from 1 mm to 1.8 mm, is indicated. The geometric and operating conditions are: tube length 1.4 m, evaporating temperature 270 K, Surface roughness 0.0005 m, gas cooler temperature 310 K, coil diameter 50 mm (helical capillary tube only), and pitch

50 mm (for spiral capillary tube only). As the tube diameter increases, the mass flow rate increases significantly in straight, helical and spiral capillary tube due to flow resistance minimizing.

Fig. 7, a indicates the percentage reduction in mass for straight helical and spiral capillary tube when changing the tube diameter from 1 mm to 1.8 mm in five steps with 0.2 mm intervals, i.e. 1 mm, 1.2 mm, 1.4 mm, 1.6 mm, 1.8 mm. Percentage reduction in mass in a helical tube compared to straight capillary tube is from 7.07 % to 8.73 %. Percentage reduction in mass in a spiral tube compared to straight capillary tube is from 23.52 % to 25.64 %. Percentage reduction in mass in a spiral tube

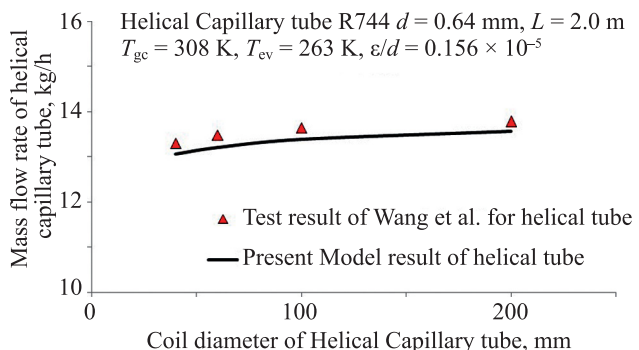


Fig. 5. Validation of the present model results of helical capillary tube with Wang et al. for helical capillary tube

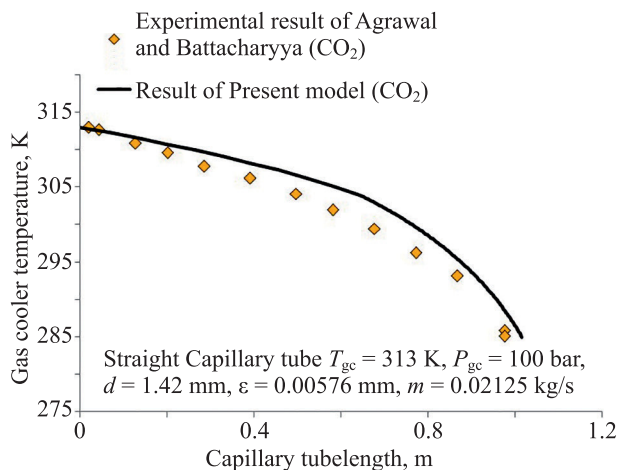


Fig. 4. Validation of present model result with the test result of Agrawal and Battacharyya for straight capillary tube

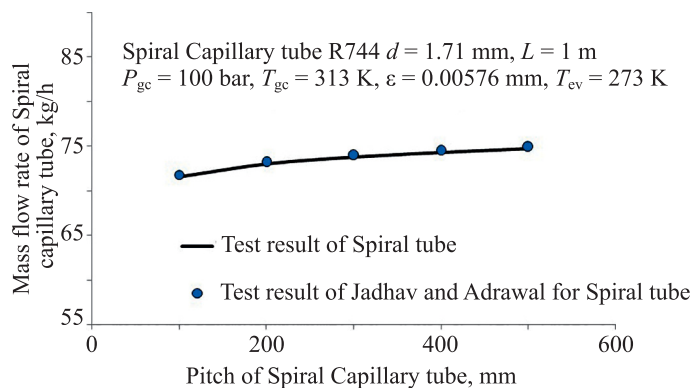


Fig. 6. Validation of present model result with Jadhav and Agrawal for spiral capillary tube

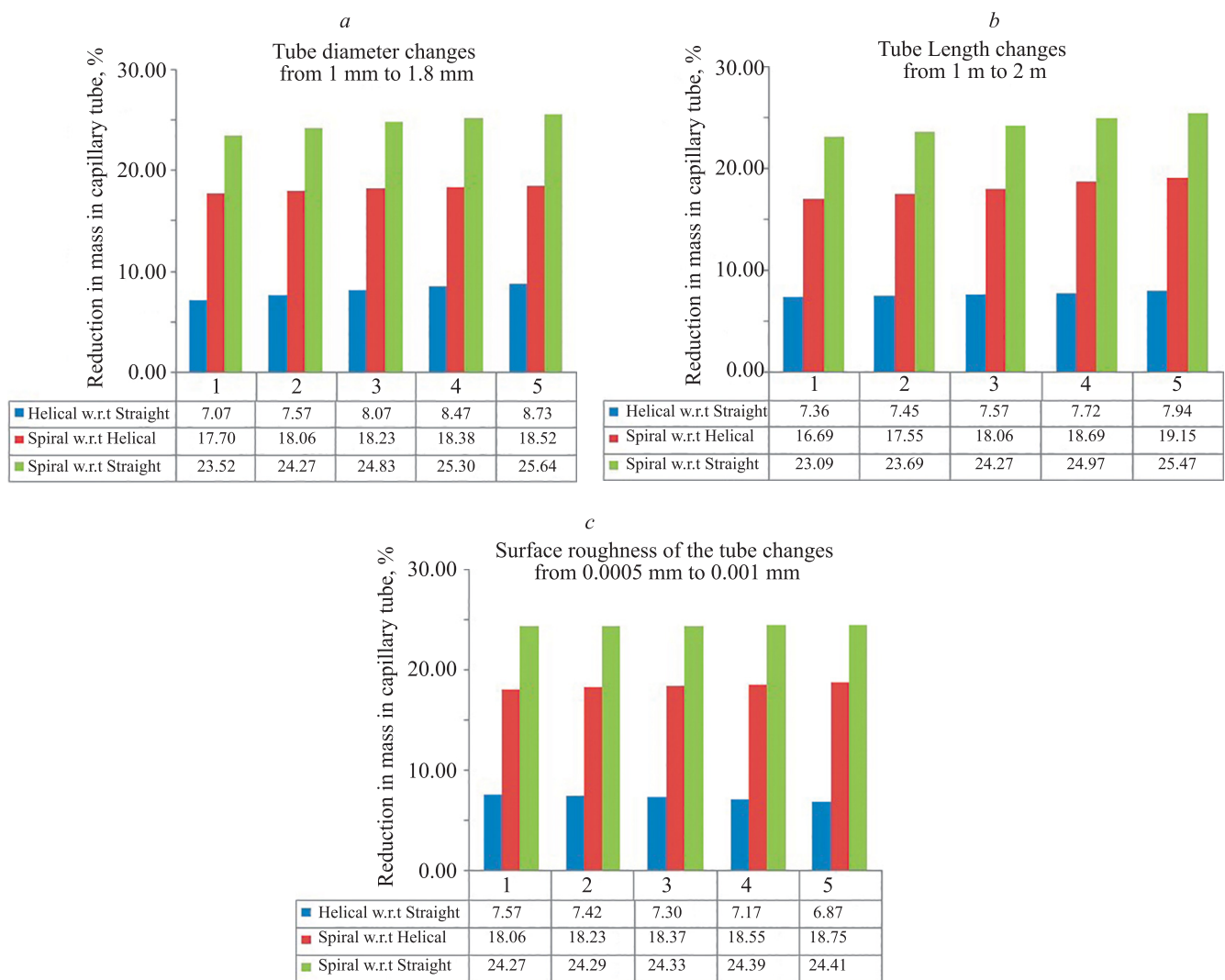


Fig. 7. Comparison of the a mass flow rate in straight helical and spiral capillary tube with change in tube diameter (a), tube length (b), surface roughness (c)

compared to helical capillary tube is from 17.70 % to 18.52 %.

The change in mass flow rate for straight, helical and spiral capillary tube, when changing the tube length from 1 m to 2 m, is presented. The geometric and operating conditions are: tube diameter 1.2 m, evaporating temperature 270 K, surface roughness 0.0005 m, gas cooler temperature 310 K, coil diameter 50 mm, and pitch 50 mm. As the tube length increases, the mass flow rate decreases in straight, helical and spiral capillary tube. The increase in length increases the friction in the tube that results in decrease in mass flow rate. Fig. 7, b indicates the percentage reduction in mass for straight helical and spiral capillary tube when changing the tube length from 1 m to 2 m in five steps with 0.25 m intervals, i.e. 1 m, 1.25 m, 1.5 m, 1.75 m, 2 m. Percentage reduction in mass in a helical tube compared to straight capillary tube is from 7.36 % to 7.94 %. Percentage reduction in mass in a spiral tube compared to straight capillary tube is from 23.09 % to 25.47 %. Percentage reduction in mass in a spiral tube compared to helical capillary tube is from 16.69 % to 19.15 %.

The change in mass flow rate for straight, helical and spiral capillary tube, when changing the surface roughness from 0.0005 mm to 0.001 mm, is reported. The geometric and operating conditions are: tube diameter 1.2 m, evaporating temperature 270 K, tube length 1.4 m, gas cooler temperature 310 K, coil diameter 50 mm, and pitch 50 mm. As the surface roughness increases, the resistance to flow increases that results in decrease in mass flow rate for straight, helical and spiral capillary tube. Fig. 7, c indicates the percentage reduction in mass for straight helical and spiral capillary tube when changing the surface roughness from 0.0005 mm to 0.001 mm in five steps with 0.000125 mm intervals, i.e., 0.0005 mm, 0.000625 mm, 0.00075 mm, 0.000875 mm, 0.001 mm. Percentage reduction in mass in a helical tube compared to straight capillary tube is from 7.57 % to 6.87 %. Percentage reduction in mass in a spiral tube compared to straight capillary tube is in the range 24.27–24.41 %. Percentage reduction in mass in a spiral tube compared to helical capillary tube is in the range 18.06–18.75 %.

Fig. 8 indicates the percentage reduction in length for straight helical and spiral capillary tube for change in the

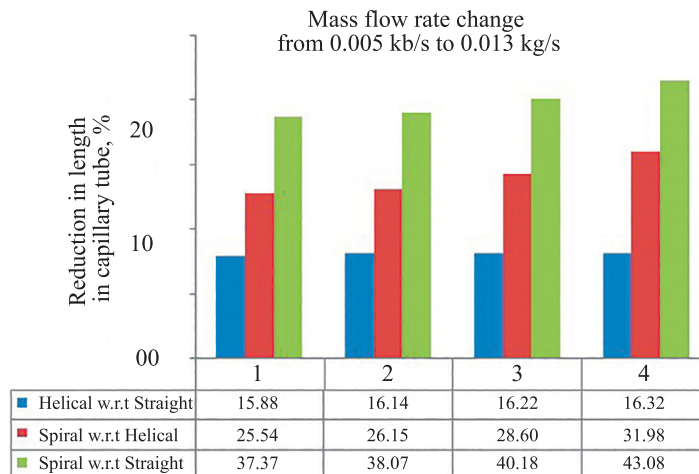


Fig. 8. Percentage reductions in length for straight helical and spiral capillary tube for change in the mass flow rate

mass flow rate from 0.005 kg/s to 0.013 kg/s. Percentage reduction in length in a helical tube compared to straight capillary tube is from 15.88 % to 16.32 %. Percentage reduction in length in a spiral tube compared to helical capillary tube is from 25.54 % to 31.98 %. Percentage reduction in length in a spiral tube compared to straight capillary tube is from 37.37 % to 43.08 %.

Conclusion

Numerical study of straight, spiral and helical capillary tube and comparison of their characteristics has been carried out. The models are developed based on the basic principles of conservation of mass, momentum, and energy. Within this outer loop, the ordinary differential equations are solved from the inlet to the exit of the capillary tube. The study has been done to find the mass flow rate by bisection method, where the mass is iteratively calculated at the specified capillary length or vice versa. For numerical solutions, FORTRAN programming is used with in-

house coding using the finite difference approach. The comparison of straight capillary with helical tube (50 mm coil diameter) and spiral capillary tube (50 mm pitch) has been reported. For changing the tube diameter, surface roughness and length of tube, the percentage reduction in mass flow rate in capillary tubes (straight, helical, and spiral) is calculated. The percentage reduction in mass in a helical tube compared to straight capillary tube is about 7–9 %. The percentage reduction in mass in a spiral tube compared to straight capillary tube is near about 23–26 %. The percentage reduction in mass in a spiral tube compared to helical capillary tube is nearly 17–19 %. The percentage reduction in length is calculated, as the mass flow rate changes from 0.005 kg/s to 0.013 kg/s. The percentage reduction in length in a helical tube compared to straight capillary tube is nearly 16 %. At the same time, percentage reduction in length in a spiral tube compared to straight capillary tube ranges from 37 % to 43 %. Similarly, percentage reduction in length in a spiral tube compared to helical capillary tube is ranging from 25 % to 32 %.

References

- Kim M., Pettersen J., Bullard C.W. Fundamental process and system design issues in CO₂ vapor compression systems. *Progress in Energy and Combustion Science*, 2004, vol. 30, no. 2, pp. 119–174. <https://doi.org/10.1016/j.pecs.2003.09.002>
- Jadhav P., Agrawal N. A review on flow characteristics of the straight and coiled capillary tubes. *International Journal of Air-Conditioning and Refrigeration*, 2021, vol. 29, no. 3, pp. 2130004. <https://doi.org/10.1142/S2010132521300044>
- Stoecker W., Jones J. *Refrigeration and Air Conditioning*. 2nd ed. McGraw-Hill, 1982, pp. 260–272.
- Jabaraj D., Vетри Kathirvel A., Mohan Lal D. Flow characteristics of HFC407C/HC600a/HC290 refrigerant mixture in adiabatic capillary tubes. *Applied Thermal Engineering*, 2006, vol. 26, no. 14–15, pp. 1621–1628. <https://doi.org/10.1016/j.applthermaleng.2005.11.017>
- Rasti M., Jeong J. A generalized continuous empirical correlation for the refrigerant mass flow rate through adiabatic straight and helically coiled capillary tubes. *Applied Thermal Engineering*, 2018, vol. 143, pp. 450–460. <https://doi.org/10.1016/j.applthermaleng.2018.07.124>
- Choi J., Kim Y., Kim H. A generalized correlation for refrigerant mass flow rate through adiabatic capillary tubes. *International Journal of Refrigeration*, 2003, vol. 26, no. 8, pp. 881–888. [https://doi.org/10.1016/S0140-7007\(03\)00079-3](https://doi.org/10.1016/S0140-7007(03)00079-3)

Литература

- Kim M., Pettersen J., Bullard C.W. Fundamental process and system design issues in CO₂ vapor compression systems // *Progress in Energy and Combustion Science*. 2004. V. 30. N 2. P. 119–174. <https://doi.org/10.1016/j.pecs.2003.09.002>
- Jadhav P., Agrawal N. A review on flow characteristics of the straight and coiled capillary tubes // *International Journal of Air-Conditioning and Refrigeration*. 2021. V. 29. N 3. P. 2130004. <https://doi.org/10.1142/S2010132521300044>
- Stoecker W., Jones J. *Refrigeration and Air Conditioning* / 2nd ed. McGraw-Hill, 1982. P. 260–272.
- Jabaraj D., Vетри Kathirvel A., Mohan Lal D. Flow characteristics of HFC407C/HC600a/HC290 refrigerant mixture in adiabatic capillary tubes // *Applied Thermal Engineering*. 2006. V. 26. N 14–15. P. 1621–1628. <https://doi.org/10.1016/j.applthermaleng.2005.11.017>
- Rasti M., Jeong J. A generalized continuous empirical correlation for the refrigerant mass flow rate through adiabatic straight and helically coiled capillary tubes // *Applied Thermal Engineering*. 2018. V. 143. P. 450–460. <https://doi.org/10.1016/j.applthermaleng.2018.07.124>
- Choi J., Kim Y., Kim H. A generalized correlation for refrigerant mass flow rate through adiabatic capillary tubes // *International Journal of Refrigeration*. 2003. V. 26. N 8. P. 881–888. [https://doi.org/10.1016/S0140-7007\(03\)00079-3](https://doi.org/10.1016/S0140-7007(03)00079-3)

7. Chingulpitak S., Wongwises S. Two-phase flow model of refrigerants flowing through helically coiled capillary tubes. *Applied Thermal Engineering*, 2010, vol. 30, no. 14-15, pp. 1927–1936. <https://doi.org/10.1016/j.applthermaleng.2010.04.026>
8. Jadhav P., Agrawal N. Comparative study on a straight and helical capillary tube for CO₂ and R22 refrigerant. *Journal of Thermal Science and Engineering Applications*, 2021, vol. 13, no. 2, pp. 024502. <https://doi.org/10.1115/1.4047822>
9. Jadhav P., Agrawal N. A comparative study of flow characteristics of adiabatic spiral and helical capillary tube in a CO₂ transcritical system. *International Journal of Ambient Energy*, 2021, in press. <https://doi.org/10.1080/01430750.2021.1913645>
10. Jadhav P., Agrawal N. Flow behavior of spiral capillary tube for CO₂ transcritical cycle. *Journal of Thermal Analysis and Calorimetry*, 2020, vol. 141, no. 6, pp. 2177–2188. <https://doi.org/10.1007/s10973-020-09536-8>
11. Mittal M., Kumar R., Gupta A. Numerical analysis of adiabatic flow of refrigerant through a spiral capillary tube. *International Journal of Thermal Sciences*, 2009, vol. 48, no. 7, pp. 1348–1354. <https://doi.org/10.1016/j.ijthermalsci.2009.01.003>
12. Khan M., Kumar R., Sahoo P. An experimental study of the flow of R-134a inside an adiabatic spirally coiled capillary tube. *International Journal of Refrigeration*, 2008, vol. 31, no. 6, pp. 970–978. <https://doi.org/10.1016/j.ijrefrig.2008.01.008>
13. Churchill S.W. Friction-factor equation spans all fluid-flow regimes. *Chemical Engineering*, 1977, vol. 84, no. 24, pp. 91–92.
14. Mori Y., Nakayama W. Study on forced convective heat transfer in curve pipes II. *International Journal of Heat and Mass Transfer*, 1967, vol. 10, no. 1, pp. 37–59. [https://doi.org/10.1016/0017-9310\(67\)90182-2](https://doi.org/10.1016/0017-9310(67)90182-2)
15. Ju H., Huang Z., Xu Y., Duan B., Yu Y. Hydraulic performance of small bending radius helical coil-pipe. *Journal of Nuclear Science and Technology*, 2001, vol. 38, no. 10, pp. 826–831. <https://doi.org/10.1080/18811248.2001.9715102>

Authors

Pravin Jadhav — PhD (Mech. Eng.), Assistant Professor, Karamveer Bhaurao Patil College of Engineering, Satara, 415001, India, <https://orcid.org/0000-0002-8043-8883>, pravin.jadhav@kbpcoes.edu.in

Anjan Kumar Sahu — PhD (Mech. Eng.), Professor, Majhighariani Institute of Technology and Science, Rayagada, Odisha, 765017, India, <https://orcid.org/0000-0003-4385-1326>, anjansahu111@gmail.com

Sunita Ballal — M.Sc (Mathematics), Head of Department, Karamveer Bhaurao Patil College of Engineering, Satara, 415001, India, <https://orcid.org/0000-0002-8392-8095>, Sunita.ballal@kbpcoes.edu.in

Авторы

Джадхав Правин — PhD, доцент, Инженерный колледж Карамвир Бхаурао Патил, Сатара, 415001, Индия, <https://orcid.org/0000-0002-8043-8883>, pravin.jadhav@kbpcoes.edu.in

Саху Анджан Кумар — PhD, профессор, Институт технологии и науки Маджхигариани, Раягада, Одиша, 765017, Индия, <https://orcid.org/0000-0003-4385-1326>, anjansahu111@gmail.com

Баллал Сунита — магистр математики, заведующий кафедрой, Инженерный колледж Карамвир Бхаурао Патил, Сатара, 415001, Индия, <https://orcid.org/0000-0002-8392-8095>, Sunita.ballal@kbpcoes.edu.in

Received 27.04.2022

Approved after reviewing 22.06.2022

Accepted 28.07.2022

Статья поступила в редакцию 27.04.2022

Одобрена после рецензирования 22.06.2022

Принята к печати 28.07.2022



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

КРАТКИЕ СООБЩЕНИЯ BRIEF PAPERS

doi: 10.17586/2226-1494-2022-22-4-812-816

УДК 535.421

Восстановление информации в зашумленном канале методом фантомной визуализации со спектральным мультиплексированием в оптическом диапазоне

Егор Николаевич Опарин¹✉, Владимир Сергеевич Шумигай²,
Азат Олфатович Исмагилов³, Антон Николаевич Цыпкин⁴

^{1,2,3,4} Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

¹ en_oparin@itmo.ru✉, <https://orcid.org/0000-0002-4009-7594>

² vshumigay@itmo.ru, <https://orcid.org/0000-0001-8597-1196>

³ ismagilov.azat@itmo.ru, <https://orcid.org/0000-0002-5844-2966>

⁴ tsyppkinan@itmo.ru, <https://orcid.org/0000-0002-9254-1116>

Аннотация

Техника фантомной визуализации позволяет получать информацию об объекте в условиях зашумленности каналов передачи, соизмеримой с интенсивностью задействованных в восстановлении спекл-структур. Одним из основных недостатков данной методики является сравнительно низкая скорость работы, что ограничивает ее применимость при работе с быстро протекающими процессами или движущимися объектами. Предложена модификация техники вычислительной фантомной визуализации, позволяющая снять это ограничение. Продемонстрировано, что спектральное мультиплексирование спекл-структур дает возможность ускорить процесс восстановления изображения. Например, при увеличении числа спектральных каналов с 4 до 10, происходит увеличение отношения сигнал-шум в шесть раз. В то же время использование классической монохромной фантомной визуализации не позволяет восстановить изображение в тех же условиях и с тем же количеством измерений. Предложенная методика привлекательна для таких требовательных к быстродействию приложений, как передача данных и дистанционное зондирование.

Ключевые слова

фантомная визуализация, спектральный суперконтинуум, пространственный модулятор света, передача информации, дистанционное зондирование

Благодарности

Работа выполнена при частичной финансовой поддержке Минобрнауки РФ в рамках Государственного задания (Паспорт № 2019-0903) и Физико-технического мегафакультета Университета ИТМО в рамках гранта НИРМА.

Ссылка для цитирования: Опарин Е.Н., Шумигай В.С., Исмагилов А.О., Цыпкин А.Н. Восстановление информации в зашумленном канале методом фантомной визуализации со спектральным мультиплексированием в оптическом диапазоне // Научно-технический вестник информационных технологий, механики и оптики. 2022. Т. 22, № 4. С. 812–816. doi: 10.17586/2226-1494-2022-22-4-812-816

Information reconstruction from noisy channel using ghost imaging method with spectral multiplexing in visible range

Egor N. Oparin¹✉, Vladimir S. Shumigay², Azat O. Ismagilov³, Anton N. Tsyppkin⁴

^{1,2,3,4} ITMO University, Saint Petersburg, 197101, Russian Federation

¹ en_oparin@itmo.ru✉, <https://orcid.org/0000-0002-4009-7594>

² vshumigay@itmo.ru, <https://orcid.org/0000-0001-8597-1196>

³ ismagilov.azat@itmo.ru, <https://orcid.org/0000-0002-5844-2966>

⁴ tsyppkinan@itmo.ru, <https://orcid.org/0000-0002-9254-1116>

© Опарин Е.Н., Шумигай В.С., Исмагилов А.О., Цыпкин А.Н., 2022

Abstract

The ghost imaging technique allows us to recover information about an object in conditions of noisy transmission channels, commensurate with the intensity of the speckle structures involved in the reconstruction. One of the main disadvantages of this technique is relatively slow reconstruction speed. This limits its applicability for study of dynamic processes or fast-moving objects. In this paper, we propose a modification of the computational ghost imaging technique that allows us to overcome this limitation. It is shown that the spectral multiplexing of the speckle patterns speeds up the image reconstruction. Increase in the number of spectral channels from 4 to 10 leads to the increase of the signal-to-noise ratio by the factor of 6. Simultaneously, under the same conditions and with the same number of measurements classical monochrome ghost imaging does not reconstruct the picture at all. This makes the proposed technique attractive for high-speed demanding applications such as communications and remote sensing.

Keywords

ghost imaging, supercontinuum, spatial light modulator, data transmission, remote sensing

Acknowledgements

The study was partially funded by the Ministry of Education and Science of the Russian Federation (Passport No. 2019-0903) and RPMA grant of School of Physics and Engineering of ITMO University.

For citation: Oparin E.N., Shumigay V.S., Ismagilov A.O., Tsyarkin A.N. Information reconstruction from noisy channel using ghost imaging method with spectral multiplexing in visible range. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2022, vol. 22, no. 4, pp. 812–816 (in Russian). doi: 10.17586/2226-1494-2022-22-4-812-816

Интерес к передаче информации через открытое пространство существует давно. Работающие в радиодиапазоне системы широко внедрены в индустрию, однако разработка аналогичных оптических систем далека от завершения. В последние годы в данной области ведутся работы по кодированию информации в вихревых пучках [1], а также разработки алгоритмов для отслеживания положения абонентов таких сетей для снижения помех и ошибок в канале [2]. Точное позиционирование абонента необходимо, так как существующая концепция оптических коммуникаций через открытое пространство подразумевает использование сравнительно узконаправленных источников и приемников [3].

Наряду с применением адаптивной оптики и алгоритмов машинного обучения [4] решению задачи дистанционного отслеживания объектов в условиях высоких шумов может способствовать техника фантомной визуализации [5]. Данная техника основана на расчете значений корреляционной функции псевдослучайного спекл-поля, освещающего поверхность объекта, и интенсивности, регистрируемой одиночным детектором после взаимодействия излучения с объектом [6]. В ряде методик, называемых вычислительной фантомной визуализацией, для восстановления изображения объекта используется целый набор псевдослучайных спекл-структур, заданных с помощью пространственного модулятора света [7]. Таким образом, информация об объекте передается с помощью шума. Внешние влияния на канал с передаваемой информацией замедляют процесс восстановления фантомного изображения, но не делают его невозможным, т. е. такая система передачи данных — помехоустойчива.

Основное ограничение техники фантомной визуализации — сравнительно большое время восстановления информации. Это накладывает дополнительные ограничения на визуализацию быстро протекающих процессов или быстро движущихся объектов. Переход от последовательного освещения объекта спекл-структурами к параллельному может позволить обойти данный недостаток. При этом не избежать взаимной интерференции спекл-структур, сформированных монохроматическим

излучением. Решением может стать использование широкополосного источника — спектрального суперконтинуума. Это позволит одновременно детектировать несколько спеклов, разнесенных по спектру. Причем количество уникальных спеклов ограничено исключительно техническими возможностями задействованного пространственного модулятора света.

В настоящей работе предложена методика ускорения восстановления информации об объекте в присутствии шумов, по интенсивности сравнимых с задействованными спекл-структурами. Техника базируется на мультиспектральной фантомной визуализации. С ее использованием экспериментально показано восстановление изображения тестового объекта в условиях, при которых классическая фантомная визуализация не позволила восстановить информацию об объекте.

Схема экспериментальной реализации предложенной методики представлена на рис. 1. В качестве источника излучения использован суперконтинуум видимого диапазона частот, полученный при лазерной филамен-

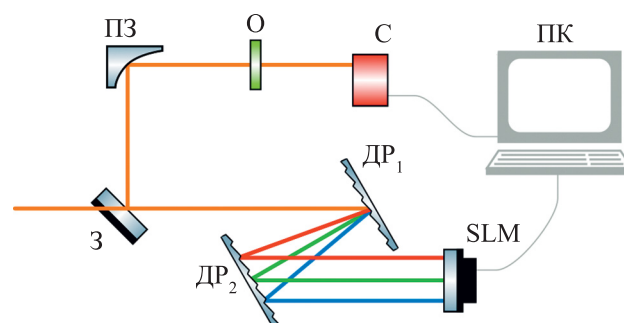


Рис. 1. Экспериментальная установка фантомной визуализации со спектральным мультиплексированием.

DP₁ и DP₂ — дифракционные решетки;
SLM — пространственный модулятор света; З — выходное зеркало; ПЗ — параболическое зеркало; О — объект;
С — спектрометр

Fig. 1. Experimental setup for ghost imaging with spectral multiplexing. Here, DP₁ and DP₂ are diffraction gratings; SLM is a spatial light modulator; З is an output mirror; ПЗ is a parabolic mirror; О is an object; and С is a spectrometer

тации в воде [8]. Псевдослучайные спекл-картины в плоскости объекта заданы с помощью пространственного модулятора света HOLOEYE LETO-2 (с разрешением 1920×1080 и размером пиксела $6,4 \times 6,4$ мкм). Детектирование излучения, прошедшего через объект, осуществлено с помощью спектрометра ASP-100 производства Avesta.

Мультиплексирование выполнено за счет разложения спектра суперконтинуума в пространстве таким образом, чтобы каждый участок SLM (Spatial Light Modulator) был освещен отдельным диапазоном. Сделано это было с помощью двух дифракционных решеток: первая раскладывала излучение в спектр, а вторая коллимировала формируемый пучок, обеспечивая тем самым освещение поверхности SLM плоским волновым фронтом. Восстановление фантомного изображения требует все спектральные диапазоны с закодированными в них псевдослучайными спекл-структурами вновь собрать в один пучок. Для этого излучение было пропущено через пару решеток в обратном направлении, после чего сформированный пучок направлялся на выходное зеркало, расположенное ниже входного пучка суперконтинуума.

Количество спектральных каналов, по которым возможно мультиплексирование, в предлагаемой методике

ограничено в первую очередь разрешением SLM. Чем больше используется поддиапазонов, тем меньшее количество пикселей отводится на модуляцию каждого из них, и тем ниже качество и разрешение итогового изображения. При этом нижнюю границу ограничивает строгая зависимость вносимого SLM фазового сдвига от длины волны излучения [9]. При модуляции широкополосного излучения меньшее число спектральных каналов говорит о большем диапазоне длин волн, охватываемом каждым участком SLM, что также приводит к некорректному восстановлению изображения. В качестве компромиссного варианта, а также для демонстрации работоспособности методики, выполним мультиплексирование по десяти спектральным каналам. Выбранное количество является условным и зависит от параметров конкретного эксперимента.

Сгенерированный при лазерной филаментации суперконтинуум охватывает диапазон от 450 до 1200 нм, но в настоящей работе он был ограничен 550–650 нм. Данное сужение диапазона связано с тем, что при разделении изначального спектра суперконтинуума на десять каналов, изменение фазовых сдвигов в пределах одного канала должно достигать 16 %. Сужение диапазона позволило ограничить фазовый сдвиг одним процентом. При большем уменьшении диапазона фа-

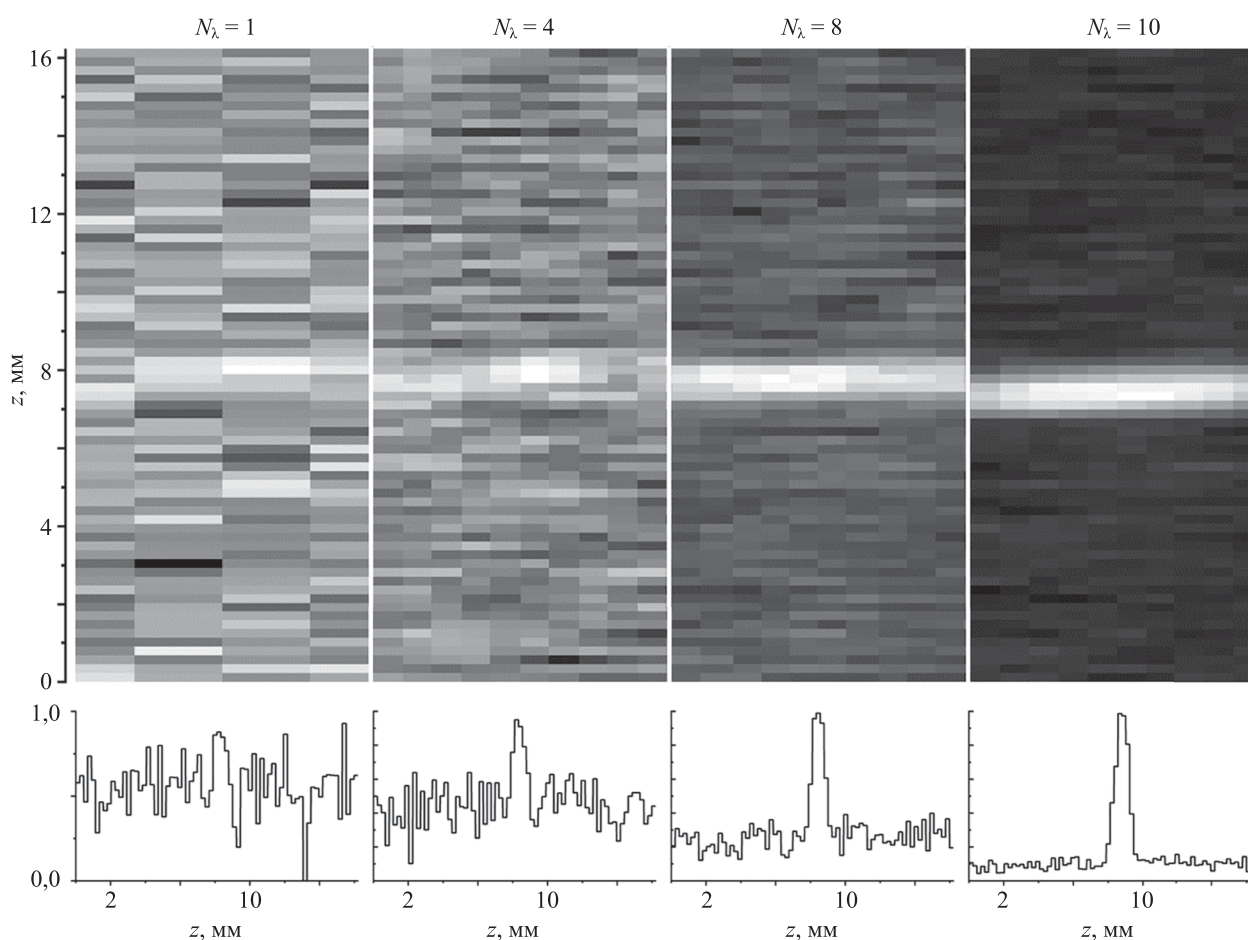


Рис. 2. Фантомные изображения щели, восстановленные с использованием различного количества спектральных каналов N_λ . Вертикальные сечения полученных изображений показаны на вставке ниже, где z — высота изображения

Fig. 2. Ghost images of the slit reconstructed using a different number of spectral channels N_λ . Vertical cross-sections of the obtained images are shown on the insert below with z is image height

зовый сдвиг еще больше уменьшился, но вместе с тем стало бы невозможным надежное разделение каналов на спектрометре.

Для подтверждения работоспособности предлагаемой методики и возможности восстановления информации об объекте в условиях высоких шумов, выполнено восстановление фантомного изображения щели шириной 1,5 мм. Для наглядности эксперимента выбраны такие условия, чтобы фантомная визуализация на основе монохроматического излучения не позволила наблюдать изображение объекта. Во всех проведенных экспериментах задействовано 5000 псевдослучайных спекл-структур. На рис. 2 видно, что с ростом количества задействованных спектральных каналов N_λ растет и качество изображения.

Предложена модификация методики вычислительной фантомной визуализации, основанная на

спектральном мультиплексировании задействованных в восстановлении фантомных изображений спекл-структур. Результат восстановления достигнут за счет использования широкополосного источника излучения — суперконтинуума видимого диапазона. Проведен эксперимент с использованием предложенной методики по восстановлению изображения щели. Продемонстрировано, что в условиях, при которых восстановление информации об объекте методом классической фантомной визуализации невозможно, предложенная методика позволяет получить изображение объекта. Показан рост качества изображения с увеличением числа спектральных каналов. Например, при увеличении количества каналов с 4 до 10, соотношение сигнал-шум выросло в шесть раз.

Литература

1. Willner A.E., Zhao Z., Liu C., Zhang R., Song H., Pang K., Manukyan K., Song H., Su X., Xie G., Ren Y., Yan Y., Tur M., Molisch A.F., Boyd R.W., Zhou H., Hu N., Minoofar A., Huang H. Perspectives on advances in high-capacity, free-space communications using multiplexing of orbital-angular-momentum beams // *APL Photonics*. 2021. V. 6. N 3. P. 030901. <https://doi.org/10.1063/5.0031230>
2. Kaymak Y., Rojas-Cessa R., Feng J., Ansari N., Zhou M., Zhang T. A survey on acquisition, tracking, and pointing mechanisms for mobile free-space optical communications // *IEEE Communications Surveys & Tutorials*. 2018. V. 20. N 2. P. 1104–1123. <https://doi.org/10.1109/COMST.2018.2804323>
3. Chan V.W.S. Free-space optical communications // *Journal of Lightwave Technology*. 2006. V. 24. N 12. P. 4750–4762. <https://doi.org/10.1109/JLT.2006.885252>
4. Na Y., Ko D.K. Deep-learning-based high-resolution recognition of fractional-spatial-mode-encoded data for free-space optical communications // *Scientific Reports*. 2021. V. 11. N 1. P. 2678. <https://doi.org/10.1038/s41598-021-82239-8>
5. Clemente P., Durán V., Torres-Company V., Tajahuerce E., Lancis J. Optical encryption based on computational ghost imaging // *Optics Letters*. 2010. V. 35. N 14. P. 2391–2393. <https://doi.org/10.1364/OL.35.002391>
6. Shapiro J.H. Computational ghost imaging // *Physical Review A*. 2008. V. 78. N 6. P. 061802. <https://doi.org/10.1103/PhysRevA.78.061802>
7. Zhang D.-J., Li H.-G., Zhao Q.-L., Wang S., Wang H.-B., Xiong J., Wang K. Wavelength-multiplexing ghost imaging // *Physical Review A*. 2015. V. 92. N 1. P. 013823. <https://doi.org/10.1103/PhysRevA.92.013823>
8. Тсупкин А.Н., Путилин С.Е., Мельник М.В., Макаров Е.А., Беспалов В.Г., Козлов С.А. Generation of high-intensity spectral supercontinuum of more than two octaves in a water jet // *Applied Optics*. 2016. V. 55. N 29. P. 8390–8394. <https://doi.org/10.1364/AO.55.008390>
9. Harm W., Jesacher A., Thalhammer G., Bernet S., Ritsch-Marte M. How to use a phase-only spatial light modulator as a color display // *Optics Letters*. 2015. V. 40. N 4. P. 581–584. <https://doi.org/10.1364/OL.40.000581>

Авторы

Опарин Егор Николаевич — лаборант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, orcid.org/0000-0002-4009-7594, en_oparin@itmo.ru

Шумигой Владимир Сергеевич — лаборант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, <https://orcid.org/0000-0001-8597-1196>, vshumigay@itmo.ru


References


1. Willner A.E., Zhao Z., Liu C., Zhang R., Song H., Pang K., Manukyan K., Song H., Su X., Xie G., Ren Y., Yan Y., Tur M., Molisch A.F., Boyd R.W., Zhou H., Hu N., Minoofar A., Huang H. Perspectives on advances in high-capacity, free-space communications using multiplexing of orbital-angular-momentum beams. *APL Photonics*, 2021, vol. 6, no. 3, pp. 030901. <https://doi.org/10.1063/5.0031230>
2. Kaymak Y., Rojas-Cessa R., Feng J., Ansari N., Zhou M., Zhang T. A survey on acquisition, tracking, and pointing mechanisms for mobile free-space optical communications. *IEEE Communications Surveys & Tutorials*, 2018, vol. 20, no. 2, pp. 1104–1123. <https://doi.org/10.1109/COMST.2018.2804323>
3. Chan V.W.S. Free-space optical communications. *Journal of Lightwave Technology*, 2006, vol. 24, no. 12, pp. 4750–4762. <https://doi.org/10.1109/JLT.2006.885252>
4. Na Y., Ko D.K. Deep-learning-based high-resolution recognition of fractional-spatial-mode-encoded data for free-space optical communications. *Scientific Reports*, 2021, vol. 11, no. 1, pp. 2678. <https://doi.org/10.1038/s41598-021-82239-8>
5. Clemente P., Durán V., Torres-Company V., Tajahuerce E., Lancis J. Optical encryption based on computational ghost imaging. *Optics Letters*, 2010, vol. 35, no. 14, pp. 2391–2393. <https://doi.org/10.1364/OL.35.002391>
6. Shapiro J.H. Computational ghost imaging. *Physical Review A*, 2008, vol. 78, no. 6, pp. 061802. <https://doi.org/10.1103/PhysRevA.78.061802>
7. Zhang D.-J., Li H.-G., Zhao Q.-L., Wang S., Wang H.-B., Xiong J., Wang K. Wavelength-multiplexing ghost imaging. *Physical Review A*, 2015, vol. 92, no. 1, pp. 013823. <https://doi.org/10.1103/PhysRevA.92.013823>
8. Тсупкин А.Н., Путилин С.Е., Мельник М.В., Макаров Е.А., Беспалов В.Г., Козлов С.А. Generation of high-intensity spectral supercontinuum of more than two octaves in a water jet. *Applied Optics*, 2016, vol. 55, no. 29, pp. 8390–8394. <https://doi.org/10.1364/AO.55.008390>
9. Harm W., Jesacher A., Thalhammer G., Bernet S., Ritsch-Marte M. How to use a phase-only spatial light modulator as a color display. *Optics Letters*, 2015, vol. 40, no. 4, pp. 581–584. <https://doi.org/10.1364/OL.40.000581>


Authors


Egor N. Oparin — Laboratory Assistant, ITMO University, Saint Petersburg, 197101, Russian Federation, orcid.org/0000-0002-4009-7594, en_oparin@itmo.ru

Vladimir S. Shumigay — Laboratory Assistant, ITMO University, Saint Petersburg, 197101, Russian Federation, <https://orcid.org/0000-0001-8597-1196>, vshumigay@itmo.ru

Исмагилов Азат Олфатович — научный сотрудник, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация,  57195673891, <https://orcid.org/0000-0002-5844-2966>, ismagilov.azat@itmo.ru

Цыпкин Антон Николаевич — доктор физико-математических наук, научный сотрудник, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация,  56366230300, <https://orcid.org/0000-0002-9254-1116>, tsyppkinan@itmo.ru

Azat O. Ismagilov — Scientific Researcher, ITMO University, Saint Petersburg, 197101, Russian Federation,  57195673891, <https://orcid.org/0000-0002-5844-2966>, ismagilov.azat@itmo.ru

Anton N. Tsyppkin — D. Sc. (Physics & Mathematics), Scientific Researcher, ITMO University, Saint Petersburg, 197101, Russian Federation,  56366230300, <https://orcid.org/0000-0002-9254-1116>, tsyppkinan@itmo.ru

Статья поступила в редакцию 26.05.2022

Одобрена после рецензирования 25.06.2022

Принята к печати 30.07.2022

Received 26.05.2022

Approved after reviewing 25.06.2022

Accepted 30.07.2022



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

ТРЕБОВАНИЯ К ОФОРМЛЕНИЮ СТАТЕЙ

Журнал публикует научные обзоры, полнотекстовые статьи и краткие сообщения.

Полнотекстовая статья должна иметь четкую структуру, включающую в себя аннотацию, ключевые слова, введение, несколько содержательных разделов и заключение.

В аннотации, рассчитанной на самый широкий круг читателей, необходимо в объеме 200–400 слов в форме краткого реферата (Abstract) изложить научное содержание статьи: предмет, цель работы, метод или методология проведения работы, краткое описание эксперимента, полученные результаты, рекомендации по их применению.

Ключевые слова должны отражать информацию, представленную в статье. Основное ключевое слово указывается первым.

Во введении необходимо представить содержательную постановку рассматриваемого вопроса, провести краткий анализ известных из научной литературы решений (со ссылками на источники), дать критику их недостатков, показать научную новизну и преимущество (особенности) предлагаемого подхода.

В основном тексте статьи должна быть представлена строгая постановка решаемой задачи, изложены и обстоятельно разъяснены (доказаны) полученные утверждения и выводы, приведены результаты экспериментальных исследований или математического моделирования, иллюстрирующие сделанные утверждения. Основной текст статьи должен быть разбит на содержательные разделы (2–3 раздела).

В заключении необходимо кратко сформулировать основные результаты, прокомментировать их и, если возможно, указать на направления дальнейших исследований и области применения.

Пристатейный список литературы (рекомендуется): для обзорной статьи — не менее 50, для полнотекстовой статьи — не менее 15, для краткого сообщения — не менее 8 литературных источников.

Ссылки на стандарты и иные нормативные документы, а также на неперсонифицированные интернет-ресурсы в список литературы не вносятся, а оформляются в виде сносок.

Объем обзорной статьи предварительно согласовывается с редакцией.

Объем полнотекстовой статьи, включая иллюстрации, таблицы и список литературы, не должен превышать 8 страниц машинописного текста (с рисунками и таблицами), шрифт 12 pt, один интервал.

Объем краткого сообщения — до 3 страниц, шрифт 12 pt, один интервал. Рубрикация текста не требуется. Статьи принимаются в электронном виде (ntvitmo@itmo.ru). В распечатанном виде с подписями авторов материалы представляются в редакцию по запросу редакции после прохождения процедуры рецензирования.

Комплект документов должен включать:

- текст статьи с заверстанными рисунками и таблицами;
- формы сведений об авторах (на каждого автора); допускается указание нескольких мест работы автора;
- файлы с рисунками к статье в оригинальном формате (предпочтительно JPEG) с максимальным разрешением; допускается представление цветных рисунков, если в черно-белом варианте теряется полезная информация;
- лицензионное соглашение;
- согласие на обработку персональных данных.

REQUIREMENTS FOR EXECUTION OF PAPERS

Scientific reviews, full-text and brief papers are published.

A full-text paper should have a well-defined structure, including an abstract, keywords, introduction, several substantive chapters and a conclusion.

An abstract is intended for general public. It is necessary to set forth the scientific content of the paper limited by 200–400 words in the form of a brief abstract: the subject, the work objective, the method or methodology of work, a brief description of the experiment, obtained results, recommendations for their application.

Keywords should be connected with the information presented in the paper. The main keyword is given first.

The introduction should contain a meaningful statement of the issue in question, a brief analysis of the solutions known from scientific literature (with references to sources), a criticism of their shortcomings, scientific novelty and advantages (features) of the proposed approach.

The main body of the paper should represent exact statement of the problem being solved, obtained assertions and conclusions are to be set forth and explained in detail (proved), the results of experimental studies or mathematical modeling should be given to illustrate the statements made. The main text of the paper needs to be split up into meaningful sections (2–3 sections).

In conclusion it is necessary to summarize the main results, comment on them and, if possible, to indicate the areas of future research and application.

A list of references is recommended: for a review paper — at least 50 items, for full-text paper — not less than 15, for a brief paper — at least 8 references. References to standards and other regulatory documents, as well as to Internet resources, are not included in a reference list and are documented as footnotes.

The length of a review paper is coordinated with the editorial board beforehand.

The length of a full-text paper, including illustrations, tables and references, should not exceed 8 pages of typewritten text (figures and tables included), font 12 pt, single-spaced.

The length of a brief paper is up to 3 pages, font 12 pt, single-spaced.

No rubrication is required.

Papers are accepted in electronic form (ntvitmo@itmo.ru).

Materials are submitted to the editors on request after the procedure of reviewing in a printed form signed by the authors.

A set of documents should include:

- The text of the paper with make-up figures and tables;
- The forms with information about the authors (for each author), several employers can be specified by the author;
- Files with pictures to the paper in their original format (preferably JPEG) with the maximum resolution; should useful information is lost in the black and white variant, color drawings are permitted;
- Publishers license agreement;
- Consent to the processing of personal data.

PHOTONICS AND OPTOINFORMATICS

Investigation of congruent lithium niobate crystal dispersion properties in the terahertz frequency range

Shumigai V.S., Oparin E.N., Nabilkova A.O., Melnik M.V., Tsyarkin A.N., Kozlov S.A. 635

OPTICAL ENGINEERING

Polarization extinction ratio in polarization maintaining fiber sealed with glass solder

Kalugin E.E., Mukhtubaev A.B., Meshkovskiy I.K. 643

Method for remote control of radiation parameters of spacecraft based on X-ray fluorescence analysis

Lukyanova L.A., Svitnev I.V., Kharitonova E.A., Gavrilov I.E. 650

Fiber-optic amplitude bend direction and magnitude sensor

Dmitriev A.A., Grebnev K.V., Smirnov D.S., Varzhel S.V. 659

AUTOMATIC CONTROL AND ROBOTICS

Compensation of external disturbances for MIMO systems with control delay

Nguyen K.T., Vlasov S.M., Pyrkin A.A., Skobeleva A.V. 666

COMPUTER SCIENCE

Building cryptographic schemes based on elliptic curves over rational numbers

Davydov V.V., Dakuo J.-M.N., Ioganson I.D., Khutsaeva A.F. 674

An algorithm for generating design solutions for data and design-production procedures management at the stages of the lifecycle of an electronic product

Donetskaya Ju.V. 681

A method for improving the efficiency of integrated processing of Earth remote sensing data in solving problems of spatial objects monitoring

Karin S.A., Karin A.I. 691

Development of a model for detecting network traffic anomalies in distributed wireless ad hoc networks

Legashev L.V., Grishina L.S., Parfenov D.I., Zhigalov A.Yu. 699

Applying the FN-corrector to improve the quality of audio event classification

Golubkov A.M., Shuranov E.V. 708

Strengthening the role of microarchitectural stages of embedded systems design

Kolchurin M.V., Pinkevich V.Yu., Platonov A.E. 716

A multivariate binary decision tree classifier based on shallow neural network (in English)

Marakhimov A.R., Kudaybergenov J.K., Khudaybergenov K.K., Ohundadaev U.R. 725

Improvement and comparison the performance of fuzzing testing algorithms for applications in Google Thread Sanitizer

Doronin O.V. 734

A method for protecting neural networks from computer backdoor attacks based on the trigger identification

Menisov A.B., Lomako A.G., Dudkin A.S. 742

Software development system for creation adaptive user interfaces

Tagirova L.F., Subbotin A.V., Zubkova T.M. 751

A method of detecting information security incidents based on anomalies in the user's biometric behavioral characteristics

Esipov D.A., Aslanova N., Shabala E.E., Shchetinin D.S., Popov I.Yu. 760

Light weight recommendation system for social networking analysis using a hybrid BERT-SVM classifier algorithm (in English)

Kiruthika N.S., Thailambal G. 769

MODELING AND SIMULATION

Modeling of random processes based on Karhunen-Loeve decomposition

Efimov A.S. 779

Numerical dissipation control of a hybrid large-particle method in vortex instability problems

Sadin D.V. 785

Numerical model of a pulsed subcritical streamer microwave discharge for problems of plasma ignition of fuel mixtures in the gas phase

Bulat P.V., Volkov K.N., Melnikova A.I., Renev M.E. 792

Numerical study on the straight, helical and spiral capillary tube for the CO₂ refrigerant (in English)

Jadhav P., Sahu A., Ballal S. 804

BRIEF PAPERS

Information reconstruction from noisy channel using ghost imaging method with spectral multiplexing in visible range

Oparin E.N., Shumigay V.S., Ismagilov A.O., Tsyarkin A.N. 812