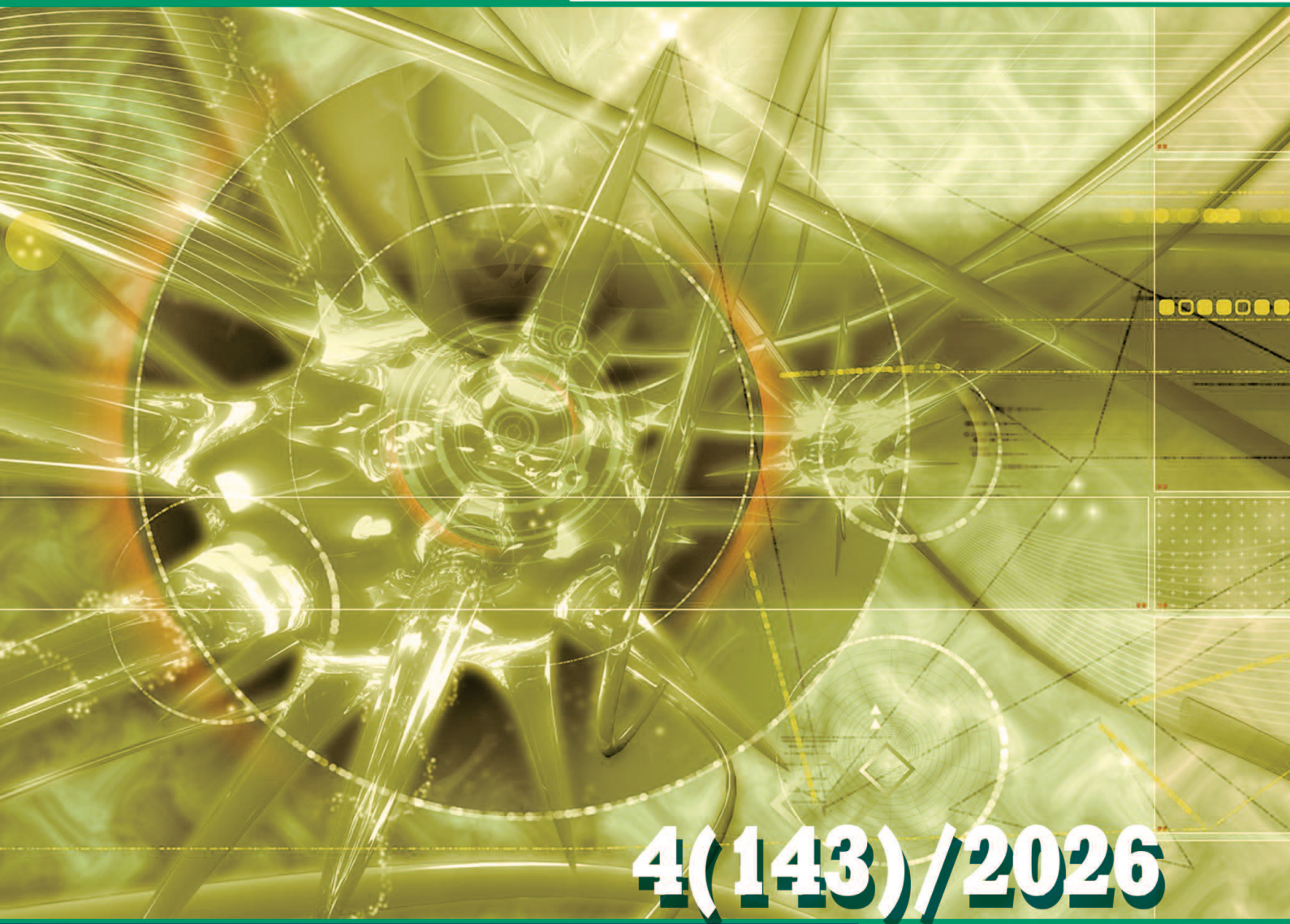
ISSN 1684-8853 (print); ISSN 2541-8610 (online)

ИНФОРМАЦИОННО-УПРАВЛЯЮЩИЕ СИСТЕМЫ

НАУЧНЫЙ ЖУРНАЛ



4(143)/2026

doi:10.31799/1684-8853-2026-4

ISSN 1684-8853 (print); ISSN 2541-8610 (online)

SCIENTIFIC JOURNAL

4(143)/2026

PEER REVIEWED JOURNAL

INFORMATSIONNO-UPRAVLIAIUSHCHIE SISTEMY (INFORMATION AND CONTROL SYSTEMS)

Contact information
The Editorial and Publishing Center, SUAI
67A, Bol'shaya Morskaia, 190000, Saint Petersburg, Russia
Website: http://i-us.ru/en, e-mail: ius.spb@gmail.com
Tel.: +7 - 812 494 70 02

The Journal was registered in the Ministry of Press, Broadcasting and Mass Media of the Russian Federation.
Registration Certificate JD № 77-12412 from April, 19, 2002.
Re-registration in the Federal Service for Supervision in the Sphere of Telecom, Information Technologies and Mass Communications (ROSKOMNADZOR) due to change of the founder A. Vostrikov,
PI № FS 77-82226 from November, 23, 2021

Submitted for publication 03.07.26. Passed for printing 26.08.26. Signed for printing: 28.08.26.
Format 60×84$_{1/8}$. Phototype CentSchbkCyrill BT. Digital printing.

Layout original is made at the Editorial and Publishing Center, SUAI.
67A, Bol'shaya Morskaia, 190000, Saint Petersburg, Russia
Printed from slides at the Editorial and Publishing Center, SUAI.
67A, Bol'shaya Morskaia, 190000, Saint Petersburg, Russia

The journal is indexed in Scopus.

Free distribution.

doi:10.31799/1684-8853-2026-4

ISSN 1684-8853 (print); ISSN 2541-8610 (online)

НАУЧНЫЙ ЖУРНАЛ

4(143)/2026

ИНФОРМАЦИОННО-УПРАВЛЯЮЩИЕ СИСТЕМЫ

РЕЦЕНЗИРУЕМОЕ ИЗДАНИЕ

Адрес редакции: 190000, г. Санкт-Петербург,
ул. Большая Морская, д. 67, лит. А, ГУАП, РИЦ
Тел.: (812) 494-70-02, эл. адрес: ius.spb@gmail.com,
сайт: http://i-us.ru

Журнал входит в БД Scopus и в Перечень рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук.

Сдано в набор 03.07.26. Подписано в печать 26.08.26. Дата выхода в свет: 28.08.26.
Формат 60×841/8. Гарнитура CentSchbkCyrill BT. Печать цифровая.
Усл. печ. л. 9,2. Уч.-изд. л. 12,5. Тираж 1000 экз (1-й завод 50 экз.). Заказ № 269.
Оригинал-макет изготовлен в редакционно-издательском центре ГУАП.
190000, г. Санкт-Петербург, ул. Большая Морская, д. 67, лит. А.
Отпечатано в редакционно-издательском центре ГУАП.
190000, г. Санкт-Петербург, ул. Большая Морская, д. 67, лит. А.
Распространяется бесплатно.

Журнал зарегистрирован в Министерстве РФ по делам печати, телерадиовещания и средств массовых коммуникаций.
Свидетельство о регистрации ПИ № 77-12412 от 19 апреля 2002 г.
Перерегистрирован в Роскомнадзоре.
Свидетельство о регистрации ПИ № ФС77-82226 от 23 ноября 2021 г.

УДК 004.89:004.822:004.048
doi:10.31799/1684-8853-2026-4-2-16
EDN: GUSQBR

Научные статьи
Articles

Обобщенный сценарий поддержки принятия решений на основе использования больших языковых моделей в информационно-управляющих системах

А. В. Смирнов[a], *доктор техн. наук, профессор, orcid.org/0000-0001-8364-073X, smir@iias.spb.su*
Т. В. Левашова[a], *канд. техн. наук, старший научный сотрудник, orcid.org/0000-0002-1962-7044*
А. В. Пономарев[a], *канд. техн. наук, доцент, orcid.org/0000-0002-9380-5064*
[a]*Санкт-Петербургский Федеральный исследовательский центр Российской академии наук, 14-я линия В. О., 39, Санкт-Петербург, 199178, РФ*

Введение: *большие языковые модели повышают эффективность принятия решений пользователями информационно-управляющих систем. Разные сценарии поддержки принятия решений сталкиваются с повторяющимися (типовыми) задачами.* ***Цель:*** *разработать обобщенный сценарий поддержки принятия решений на основе больших языковых моделей, который соответствует нормативным требованиям к таким системам и обеспечивает реализацию процесса принятия решений как выполнение процесса решения типовых задач.* ***Результаты:*** *путем анализа сценариев поддержки принятия решений на основе больших языковых моделей выявлены типовые задачи, которые выполняются в этих сценариях, установлены используемые виды информационных ресурсов и разработана спецификация требований к системам поддержки принятия решений на основе больших языковых моделей. В соответствии с этой спецификацией сформулированы принципы поддержки принятия решений на основе больших языковых моделей и предложена концепция организации системы поддержки принятия решений, воплощающая эти принципы. Главным результатом является обобщенный сценарий поддержки принятия решений на основе больших языковых моделей в нотации Business Process Model Notation, опирающийся на предложенную концепцию и реализующий стандартный процесс принятия решений как выполнение последовательности типовых задач.* ***Практическая значимость:*** *предложенный обобщенный сценарий представляет практический интерес в качестве прототипа сценария проблемно-независимой системы поддержки принятия решений, в котором большая языковая модель используется в качестве основной и (или) вспомогательной модели при поддержке принятия решений.*

Ключевые слова *— поддержка принятия решений, большие языковые модели, онтология, сценарий.*

Для цитирования: Смирнов А. В., Левашова Т. В., Пономарев А. В. Обобщенный сценарий поддержки принятия решений на основе использования больших языковых моделей в информационно-управляющих системах. *Информационно-управляющие системы*, 2026, № 4, с. 2–16. doi:10.31799/1684-8853-2026-4-2-16, EDN: GUSQBR

For citation: Smirnov A. V., Levashova T. V., Ponomarev A. V. Generalized scenario of LLM-powered decision support in information and control systems. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2026, no. 4, pp. 2–16 (In Russian). doi:10.31799/1684-8853-2026-4-2-16, EDN: GUSQBR

Введение

Системы поддержки принятия решений (СППР) в информационно-управляющих системах — это интерактивные средства автоматизации, помогающие руководителям и менеджерам принимать обоснованные решения в сложных ситуациях. Они собирают, обрабатывают и анализируют большие объемы данных и информации, строят формальные модели, оценивают риски и предлагают оптимальные решения, повышая эффективность принятых решений и бизнес-процессов.

В последние годы большое внимание в области поддержки принятия решений (ППР) уделяется большим языковым моделям (БЯМ), поскольку они достаточно удобны в использовании благодаря их адаптации к индивидуальным предпочтениям и поведению пользователей [1–3]. Анализ исследований по ППР на основе БЯМ показал, что к настоящему времени предложен целый ряд общих и достаточно универсальных подходов к этой проблеме. Также было замечено, что в сценариях ППР, реализованных в этих подходах, выполняются похожие (типовые) задачи. Это послужило мотивацией для разработки концепции организации СППР на основе БЯМ, которая позволяет реализовать обобщенный сценарий принятия решений — от анализа данных до выработки рекомендаций — как процесс выполнения БЯМ-ориентированных типовых задач.

В данной работе объектом исследования являются СППР, основанные на БЯМ, в том числе на БЯМ-агентах. В качестве предмета исследования выступает организация СППР указанного класса и методы разработки таких систем. Пространство решений охватывает множество вариантов организации СППР на основе БЯМ с соответствующими сценариями.

Целью работы является разработка обобщенного сценария ППР, обеспечивающего реализацию стандартного процесса ППР в виде последовательности выполнения БЯМ-ориентированных типовых задач.

Для достижения цели были рассмотрены сценарии ППР, реализуемые в различных подходах; определены однотипные задачи, которые повторяются в разных сценариях; сформулированы спецификация требований к разрабатываемой СППР и принципы, лежащие в основе такой СППР; предложена концептуальная организация СППР и сформирован порядок выполнения типовых задач в обобщенном сценарии ППР. Предлагаемая СППР представляет собой пример искусственного интеллекта (ИИ), когда компетенции лица, принимающего решения (ЛПР), и его понимание задачи принятия решений дополняются информационными и аналитическими возможностями ИИ.

Большие языковые модели в сценариях СППР

В данном разделе приводится обзор сценариев ППР на основе БЯМ, реализованных в различных прикладных областях, и на основании данного обзора специфицируются требования к таким СППР. Под сценарием ППР понимается формализованная структура, определяющая последовательность шагов, алгоритмов и аналитических действий, которую использует СППР для помощи ЛПР в выборе варианта действий или решения задачи из нескольких возможных в конкретной ситуации, которые наилучшим образом соответствуют его предпочтениям. Сценарии, в которых для реализации ППР используются БЯМ-агенты, также включены в данный обзор, поскольку при выборе сценариев для анализа основной акцент делался именно на использование БЯМ вне зависимости от архитектуры СППР.

Рассмотренные сценарии охватывают ряд прикладных областей: медицину [4–12], управление чрезвычайными ситуациями [13], логистику [14–16], планирование и составление расписаний [17, 18], управление умным городом [19], государственное управление [20], информационную безопасность [21], оценку климатических рисков [22]. Разнообразие анализируемых прикладных областей объясняется попыткой как можно полнее охватить пространство возможных сценариев встраивания БЯМ в процессы ППР. Большинство сценариев взято из области медицины, поскольку именно там БЯМ показали себя единственным инструментом, способным эффективно работать с основным ресурсом здравоохранения — медицинскими текстами и данными, освобождая врачей от рутинной обработки таких ресурсов и предоставляя им время для выполнения действительно сложных клинических задач и повышая доступность знаний. Как следует из анализа, организация рассмотренных сценариев не сильно зависит от области их применения. Сценарии отличаются, например, сервисами обработки информации, наличием и объемом руководящих документов, строгостью критериев оценки, практиками оценки. При этом их организации похожи.

Обзор сценариев поддержки принятия решений на основе БЯМ

В результате анализа сценариев ППР, в которых используются БЯМ, было установлено, что в различных сценариях решаются похожие (однотипные) задачи. Выделено четыре типа задач: уточнение задачи принятия решений, учет предпочтений пользователя, использование знаний проблемной области (ПрО), генерация и оценка альтернатив. Рассматриваемая в работе концепция типовой задачи не предполагает, что такая задача решается в каждом сценарии. Типовые задачи иллюстрируются на примере сценариев, в которых есть акцент на решение задач определенного типа или которые наилучшим образом подходят для иллюстрации, вне зависимости от ПрО сценариев. Для каждого сценария указано, содержимое каких информационных ресурсов задействовано в процессе ППР.

Уточнение задачи принятия решений

Под задачей принятия решений понимается задача, которую пользователь сформулировал в своем запросе к СППР. Уточнение может потребоваться как пользователю, так и СППР. Пользователю требуется уточнение, если он не согласен с решением, которое ему предложила СППР, или сомневается в нем. СППР может прибегнуть к уточнению задачи, когда она не смогла верно интерпретировать запрос, ей недостает информации (например, входных данных) для решения задачи и когда область поиска решения видится слишком обширной.

Уточнение задачи принятия решений, когда пользователь сомневается в правильности предложенного решения, используется в части сценария разработки плана пожарной эвакуации, связанной с уточнением предлагаемого БЯМ плана [13]. Задача уточняется в процессе дискуссии между БЯМ и пользователем, если у него возникают сомнения в правильности предложенного плана. Используемые в сценарии информационные ресурсы: онтология (явная спецификация концептуализации [23]) ПрО «пожарная безопасность» и граф знаний, в котором представлены конкретные сценарии для выработки стратегии и анализа конкретных ситуаций (тушение пожара или планирование путей эвакуации).

Задача принятия решений при недостатке информации для ее решения уточняется в сценарии, где БЯМ ставит диагноз и рекомендует план лечения пациента [4]. В этом сценарии постановка задачи уточняется итеративно в процессе выработки моделью рекомендации по мере обнаружения БЯМ недостающей информации (например, результатов медицинских исследований, которые не сделаны, но нужны для постановки диагноза). Используемые в сценарии информационные ресурсы: электронные медицинские карты и клинические рекомендации по диагностике и лечению.

Уточнение задачи принятия решений в целях сужения области поиска решения является частью сценария поиска клинических рекомендаций [5]. В этом сценарии на основе введенных пользователем данных (симптомы, жалобы, история болезни) выбирается код заболевания по Международному классификатору болезней (МКБ), который позволяет сузить область поиска рекомендаций. Используемые информационные ресурсы: МКБ, клинические рекомендации, память БЯМ (факты, «запомненные» БЯМ во время обучения [24]), в которой хранятся ранее найденные рекомендации.

Область поиска также сужается за счет учета предпочтений пользователя. Но данный тип задачи отнесен в самостоятельную категорию и рассматривается ниже.

Учет предпочтений пользователя

В сценариях СППР на основе БЯМ явные и неявные предпочтения пользователей устанавливаются как непосредственно БЯМ, так и другими компонентами СППР.

Большие языковые модели, как правило, отвечают за выявление предпочтений, когда они вовлечены в процесс принятия решений, что характерно для многоагентных систем. Например, в сценарии составления расписания онлайн-совещания агенты договариваются о времени проведения совещания [17]; в процессе дискуссии обозначаются предпочтения пользователей относительно времени проведения конкретного совещания, а также неявные предпочтения относительно того, какой временной интервал (например, утро, после обеда, в конце рабочего дня и т. п.) более удобен тому или иному пользователю для участия в онлайн-мероприятиях. В другом многоагентном сценарии — сценарии выбора места групповой поездки [18] — организуется дискуссия, в которой агенты предлагают места для поездки, стараясь учесть явные предпочтения всех участников поездки (культурные интересы, развлечения, стоимость и т. п.). Если некоторый агент соглашается с предложенным местом поездки, означает, что у данного участника изменились предпочтения. В этом случае можно говорить о выявлении неявных предпочтений. В обоих сценариях — организация совещания и планирование поездки — БЯМ обрабатывает текстовые сообщения, которые в данной статье рассматриваются как информационный ресурс.

Примером выяснения предпочтений пользователей при поддержке БЯМ является логистический сценарий доставки на «последней миле» [14]. В нем решение, найденное СППР, предлагается заинтересованным пользователям (например, клиентам, руководителям компаний, местным сообществам и т. п.), которые в текстовой форме делают качественные комментарии к этому решению. Фактически эти комментарии отражают предпочтения заинтересованных сторон, определением которых занимается специальный модуль. Обязанности БЯМ в этом сценарии — интерпретация текстовых комментариев и обновление постановки логистической задачи с учетом выявленных предпочтений. Комментарии рассматриваются как информационный ресурс.

Использование знаний проблемной области

В данной работе под знанием понимается формализованная и структурированная модель ПрО, предназначенная для автоматического рассуждения. В анализируемых сценариях знания, представленные при помощи таких моделей, являются информационными ресурсами. При этом в некоторых сценариях БЯМ является участником разработки подобного ресурса. С точки зрения способа представления знаний в рассмотренных сценариях используются онтологии, графы знаний и специализированные модели представления знаний.

Онтология ПрО и граф знаний. Онтологией называется формализованная система аксиом и определений на языке логики, которая предназначена для описания знаний ПрО в целях последующего автоматического логического вывода. Онтология предоставляет структуру, которая определяет понятия, отношения и сущности в графе знаний. Граф знаний — это конкретное применение онтологии, представляющее реальные данные в виде графовой структуры, где узлы — это сущности, а ребра — отношения. Примерами сценариев, в которых используются формальные модели онтологии и графа знаний, являются:

— сценарии разработки плана пожарной эвакуации и тушения пожара [13], в которых онтология представляет экспертные знания из области пожарной безопасности, а граф знаний — сценарии, для которых формируются рекомендации-планы;

— сценарий оптимизации интермодальных грузовых перевозок [15], где OWL-онтология (OWL, Web Ontology Language — язык описания онтологий) объединяет тематические группы знаний: а) знания предметной области (знания

системы анализа грузоперевозок и мультимодальной транспортной сети); б) знания, описывающие постановку задачи принятия решений; в) модели оптимизации, извлеченные из различных академических статей. Граф знаний формируется из онтологии для описания конкретного сценария;

— сценарий поддержки принятия клинических решений на основе объяснимого ИИ [6], в котором используется только граф знаний (без онтологии). Этот граф организует и связывает знания из различных источников и служит хранилищем постоянно обновляемых медицинских рекомендаций, которые в случае, когда запрос на предоставление рекомендации не может быть обработан из-за недостатка знаний в графе, ищутся во внешних документах и включаются в граф;

— сценарий ППР для вывода рекомендаций пациентам с набором заболеваний, в котором БЯМ транслирует текстовые регламентирующие документы в компоненты онтологической модели [7]. Онтология объединяет в себе знания о медикаментозной терапии для различных заболеваний.

Специализированные модели представления знаний — это методы описания фактов, объектов и закономерностей конкретной ПрО. Примерами сценариев, в которых знания представляются при помощи специализированных моделей, являются:

— сценарий поиска клинических рекомендаций в справочнике МКБ-10 [5];

— сценарий системы оповещения врачей о рекомендуемых мероприятиях и процедурах (вакцинациях, исследованиях и т. п.), систематизирующий причины отказа от предложенных мер в базе знаний конкретной СППР [8];

— сценарий СППР для клинических исследований на основе БЯМ [9] как представитель множества сценариев из области медицины, в которых задействована база знаний OpenEHR [25] — открытый стандарт управления, хранения и обмена электронными историями болезней;

— сценарий автоматического управления кооперативным вождением [16], в котором для накопления знаний используется внутренний модуль памяти. Модуль хранит: а) обработанные системой сценарии, б) рекомендованные для этих сценариев решения и в) принятые решения. Структурирование знаний происходит за счет использования векторной базы данных.

Генерация и оценка альтернатив

В рассмотренных сценариях на основании критериев количества генерируемых альтернатив и наличия зависимости от контекстных изменений можно выделить следующие множества альтернатив, подлежащих оценке: одна альтернатива, фиксированное множество альтернатив, динамическое множество альтернатив.

Одна альтернатива. Фактически одна альтернатива — это рекомендация, предложенная СППР, или оптимальное решение.

В случае альтернативы как рекомендации ЛПР может принять ее в том виде, в котором ее предложила СППР. Может подкорректировать, т. е. породить новую альтернативу. Может отклонить и принять собственное решение, отличное от рекомендованного. Сценарии, подпадающие под рассматриваемую категорию, включают в себя: сценарий ППР при постановке диагноза и разработке рекомендации плана лечения пациента [4], сценарий многоагентной СППР для сортировки и планирования лечения в отделениях неотложной помощи [10], сценарий интеллектуальной системы проактивного мониторинга центров информационной безопасности [21], сценарий оценки климатических рисков [22], сценарий поддержки принятия административных решений [20]. В последнем сценарии БЯМ, обученная на законодательных документах и примерах решений, помимо самой рекомендации, готовит черновик заключения о положительном или отрицательном решении. Информационными ресурсами в приведенных сценариях служат текстовые документы и запросы пользователей (врачей, администраторов, заявителей и т. п.).

Примерами сценариев, предполагающих решение задачи оптимизации, являются сценарии 1) интермодальных грузовых перевозок [15] и 2) СППР для управления умным городом [19]. В первом сценарии решением является оптимальный маршрут транспортировки из пункта отправления в пункт назначения, критериями оптимизации — максимизация топливной экономичности грузоперевозок, минимизация выбросов парниковых газов, минимизация общих эксплуатационных расходов. Во втором сценарии оптимизация применяется для решения различных задач принятия решений, связанных с муниципальным управлением (например, где лучше построить школу, разбить парк и т. п.). В этом сценарии множество альтернатив генерирует БЯМ, и впоследствии решается оптимизационная задача. Критерии оптимизации зависят от конкретной задачи принятия решений, одним из критериев является минимизация затрат на реализацию решения. Информационными ресурсами в рассмотренных сценариях служат запросы пользователей, онтология ПрО, граф знаний, документы (в частности, документы по развитию города в сценарии управления умным городом), знания БЯМ.

Фиксированное множество альтернатив. Такие множества используются в СППР, которые ориентированы на экспертную оценку. Среди рассмотренных можно указать сценарии, в которых множество альтернатив состоит из ответов разных

БЯМ на один и тот же запрос. Рекомендация, предложенная каждой БЯМ, рассматривается как альтернатива. К данной категории относится сценарий интраоперационной ППР в пластической хирургии [11], который используется для сравнения разных БЯМ, оцениваемых по пятибалльной шкале в отношении умения планировать операцию, владения знаниями по общей анатомии и хирургических процедур, а также способности находить решения и альтернативы в случае возможных осложнений. Еще один из этой категории — сценарий ППР в персонализированной онкологии [12], где четыре БЯМ генерируют ответ на запрос, сформулированный при помощи типизированного промпта. Этот сценарий предназначен для принятия решений в ситуациях, когда для ответа нет стандартизированных рекомендаций, а есть множество источников с разными рекомендациями. Решения, предложенные несколькими БЯМ, выносятся на рассмотрение экспертным советом. Решения, предложенные только одной БЯМ, отбрасываются. Информационным ресурсом в указанных сценариях являются «собственные» знания БЯМ.

Динамическое множество альтернатив. Это сценарий, когда альтернативы создаются по мере развития текущей ситуации. Примерами являются вышеупомянутые сценарии разработки плана эвакуации при пожаре [13], организации онлайн-совещания [17] и планирования совместной поездки [18]. В сценарии по разработке плана эвакуации динамически развивающейся является ситуация с распространением пожара, а альтернативами — планы эвакуации. В двух других сценариях динамически развивающаяся ситуация — ситуация, сложившаяся в процессе принятия решений, а альтернативы — время проведения совещания и место групповой поездки с учетом предпочтений участников совещания и поездки соответственно. Информационные ресурсы в приведенных сценариях — это онтология ПрО, граф знаний, текстовые сообщения (включая запросы пользователей), электронные календари (в сценарии организации расписания) и «собственные» знания БЯМ.

Виды информационных ресурсов, задействованных в рассмотренных в данном обзоре сценариях, обобщены в табл. 1. Резюмируя, такими ресурсами являются информационные системы и сервисы, пользователи, их запросы и текстовые сообщения, промпты, структурированные и неструктурированные источники знаний, в том числе доступные в веб, формализованные знания ПрО, знания БЯМ.

■ ***Таблица 1.*** Информационные ресурсы, задействованные в сценариях ППР на основе БЯМ
■ ***Table 1.*** Information sources involved in LLM-based decision support scenarios

Используемый контент	Информационный объект	Вид информационного ресурса
Постановка задачи принятия решений	Запрос пользователя, текстовые сообщения, промпт	Пользователь
Знания ПрО	Онтология, граф знаний, память СППР, запрос пользователя, содержимое текстовых документов, шаблоны формальных документов, промпт	Пользователь, веб-ресурс, СППР, информационная система, информационный сервис, неструктурированные и структурированные документы
Структура знаний ПрО	Хранилище знаний	Веб-ресурс, СППР, структурированный документ
Альтернатива, рекомендация	Вычислительный сервис, чат, знания БЯМ	Пользователь, СППР, БЯМ, программное обеспечение
Критерии оптимизации	Запрос пользователя, знания БЯМ, содержимое документов	Пользователь, веб-ресурс, БЯМ, неструктурированные и структурированные документы
Текст запроса, текст сообщений	Чат	Пользователь
Принятое решение	Чат, формальный документ	Пользователь

Спецификация требований к ППР на основе БЯМ

В соответствии с рассмотренными сценариями разработана спецификация требований к ППР на базе БЯМ, структурированная по четырем ключевым группам.

A. *Требования к СППР на основе БЯМ*:

A.Req-01: в СППР должна быть предусмотрена реализация взаимодействия между пользователем и системой через чат или (и) интерфейс;

A.Req-02: в сценарии функционирования СППР должна быть предусмотрена возможность инициации взаимодействия с пользователем со стороны БЯМ;

A.Req-03: СППР должна предоставлять пользователю решения в понятном, пригодном для экспертного анализа и оценки виде;

A.Req-04: предлагаемые системой решения должны учитывать предпочтения пользователей и личностные качества ЛПР (например, компетентность, инициативность, напористость, неконфликтность, коммуникабельность и др.);

A.Req-05: СППР должна быть контекстно-ориентированной в части динамически меняющихся постановки задачи, предпочтений пользователей и текущей ситуации;

A.Req-06: СППР должна учитывать новейшие достижения, интегрируя современные технологии для разработки альтернатив, такие как дополненный интеллект, машинное обучение, нейронные сети и др.;

A.Req-07: в СППР должна быть предусмотрена возможность построения модели знаний ПрО при помощи БЯМ или иных средств;

A.Req-08: СППР должна обладать «памятью» и позволять накапливать знания об обработанных сценариях, предложенных рекомендациях и принятых решениях в виде, подходящем для анализа и повторного использования.

B. *Требования к БЯМ* (некоторые требования, как, например, обработка информации, предполагают интеграцию БЯМ со специальными механизмами):

B.Req-01: в многоагентных СППР БЯМ-агенты должны уметь взаимодействовать в чате;

B.Req-02: БЯМ должна уметь поддерживать многомодальное взаимодействие с пользователем, интерпретировать его запросы и формализовывать постановку задачи принятия решений на основе этих запросов и формальной модели, используемой в СППР;

B.Req-03: БЯМ должна уметь динамически обновлять постановку задачи пользователя в соответствии с изменениями текущей ситуации;

B.Req-04: БЯМ должна уметь обрабатывать и интегрировать разнородную информацию (знания), в том числе из веб-ресурсов и от пользователя, и представлять их в требуемом виде, а в особо оговариваемых случаях и оценивать их;

B.Req-05: если БЯМ отвечает за разработку модели знаний ПрО, то она должна уметь систематизировать знания в соответствии с логической системой, принятой в разрабатываемой модели;

B.Req-06: если БЯМ отвечает за предоставление рекомендаций и разработку альтернатив, то она должна уметь:

B.Req-06.1: предлагать рекомендации (альтернативы) на основе постановки задачи пользователя, знаний из информационных ресурсов и из модели знаний ПрО, а также на основе оценки собранной информации;

B.Req-06.2: при разработке альтернатив учитывать новейшие достижения, в том числе опубликованные в веб-ресурсах;

B.Req-06.3: корректировать сгенерированные альтернативы или динамически порождать новые по мере развития текущей ситуации;

B.Req-07: если БЯМ отвечает за оценку и выбор альтернатив, то она должна уметь оценивать и выбирать альтернативы на основе явно заданных критериев, а также критериев, рекомендуемых источниками знаний ПрО, в том числе доступных в веб, и собственных знаний;

B.Req-08: если БЯМ отвечает за выявление предпочтений пользователей, то она должна уметь извлекать явные и неявные предпочтения пользователя на основании исходящей от пользователя информации в процессе разработки альтернатив и принятия решений;

B.Req-09: БЯМ должна уметь преобразовывать предназначенные пользователю результаты в вид, понятный ему и пригодный для экспертного анализа и оценки;

B.Req-10: БЯМ должна сопровождать объяснениями свои ответы, рекомендации, а также результаты разработки, оценки и выбора альтернатив, в том числе в определенных случаях ссылками на источники знаний, на основании которых она пришла к этим ответам/рекомендациям/результатам;

B.Req-11: если в обязанности БЯМ входит документирование принятого решения, то она должна уметь составлять формальные документы на основе имеющихся образцов.

C. *Требования к организации процесса ППР*:

C.Req-01: при необходимости должна быть предусмотрена возможность использования:

C.Req-01.1: препромптинга — метода, который предоставляет контекст, ограничения и инструкции БЯМ перед выполнением основной задачи, для получения более точных ответов;

C.Req-01.2: промптинга в соответствии с методологией Prompt-Driven Development (PDD) и концепцией Prompt Operations (PromptOps — операционная дисциплина управления жизненным циклом промптов для БЯМ).

D. *Требования к экспертам проблемной области*:

D.Req-01: эксперты-ЛПР должны быть компетентны в части оценки решений, которые предлагает СППР;

D.Req-02: инженеры RAG (Retrieval-Augmented Generation) и промпт-инженеры должны соответствовать квалификационным требованиям, предъявляемым к инженерам по промпту.

Способом проверки групп требований A–C является тестирование, требований группы D — статический анализ.

Для определения применимости требований к задаче разработки СППР на основе БЯМ использовался метод MoSCoW, устанавливающий приоритеты для требований в зависимости от их важности и срочности (табл. 2). Заложенные в метод приоритеты: Must (обязательно), Should (желательно, но не критично — важные задачи, не влияющие на работу текущей версии продукта), Could (можно сделать — задачи, направленные на улучшение существующей версии продукта), Won't (не делать сейчас — наименее приоритетные задачи).

■ ***Таблица 2.*** Приоритетность требований

■ ***Table 2.*** Requirements prioritization

Требование	Описание	Обоснование
Must		
A.Req-01 A.Req-03 A.Req-07 B.Req-01 B.Req-02 B.Req-09	Чат/интерфейс, понятный для эксперта вид решений, базовая модель знаний, формальная постановка задачи на основе запроса	Формируют минимальный рабочий каркас СППР
B.Req-04	Обеспечение данными для анализа	Внешняя информация необходима для связи с реальностью и устранения галлюцинаций БЯМ
D.Req-01	Компетентность экспертов	Если ЛПР некомпетентен в ПрО, СППР теряет смысл
Should		
A.Req-05 A.Req-08 B.Req-03	Динамический контекст, память системы	Необходимы для полностью готовой СППР, но минимально жизнеспособная СППР может работать в статике без самообучения
B.Req-06.1 B.Req-07	Генерация, оценка и выбор альтернатив	Ключевые задачи СППР
B.Req-10	Предоставление объяснений	Обеспечение доверия к СППР со стороны экспертов
C.Req-01	Использование промптинга	Промптинг требуется по мере усложнения системы
Could		
D.Req-02	Квалифицированные инженеры	На этапе разработки минимально жизнеспособной СППР возможно привлечение смежных специалистов
A.Req-02 A.Req-04	Инициатива общения со стороны БЯМ, учет личностных качеств	На начальном этапе разработки СППР достаточно реактивного режима (запрос — ответ), личностные качества учитываются во вторую очередь
B.Req-05	Автоматизированная разработка базы знаний ПрО	Трудоемко, на начальном этапе можно использовать готовую базу знаний
B.Req-06.2 B.Req-06.3	Учет важнейших достижений и динамическая корректировка альтернатив	Базовая СППР должна заработать в статике
B.Req-08	Выявление предпочтений	На начальном этапе разработки СППР достаточно работать с явными запросами
Won't		
A.Req-06	Интеграция современных технологий для разработки альтернатив	На текущем этапе ядром выступает БЯМ, дополнительные технологии могут быть встроены в расширенную версию СППР
B.Req-11	Автоматическое составление формальных документов	Рутинная задача, которую легко можно встроить в финальную версию СППР

Концептуальная организация СППР на основе БЯМ

Ключевые принципы, на которые опирается предлагаемая концепция, сформулированы в соответствии со спецификацией требований к ППР на основе БЯМ, описанными в одноименном подразделе. Принципы структурированы по трем направлениям.

Направление 1. Область применения и функциональность СППР.

1. СППР предназначена для помощи пользователю-ЛПР в детальном анализе проблемной ситуации, формулировке задачи принятия решений и принятии решений.

2. СППР выполняет функции по а) сбору и поиску информации; б) генерации рекомендуемых решений; в) оценке и выбору решений, предложенных ЛПР или системой, оставляя при этом принятие окончательного решения за человеком; г) запоминанию сценариев, результатов их выполнения и принятых решений для формирования рекомендаций в будущем.

Направление 2. Взаимодействие с пользователем.

1. СППР интерпретирует задачу принятия решений, сформулированную на естественном языке, использование которого в качестве основного средства коммуникации устраняет необходимость создания специализированных интерфейсов и упрощает внедрение СППР. Формулировка на естественном языке может быть дополнена графическими схемами и голосовым общением, но текстовое сообщение играет ведущую роль, что обусловлено тем, что над текстом человек, как правило, думает гораздо больше, чем над спонтанной речью.

2. СППР взаимодействует с пользователем, когда это необходимо, например для уточнения его запроса, согласования модели задачи принятия решений и т. д.

3. БЯМ предоставляет объяснения рекомендуемым решениям и оценкам, обеспечивая прослеживаемую аргументацию, что позволяет проверить обоснованность результатов.

4. СППР выявляет, анализирует и учитывает при предоставлении рекомендаций явные и неявные предпочтения пользователей.

Направление 3. Представление и обработка задачи принятия решений.

1. СППР учитывает контекст. Она выявляет и принимает во внимание окружение задачи принятия решений, точку зрения и предпочтения ЛПР, а также другие временные данные, связанные с этой задачей.

2. СППР строит модель задачи принятия решений (или несколько семантически совместимых моделей). Специфика моделей может варьироваться в зависимости от ПрО и самой задачи. Построение модели позволяет прояснить сложность текущей задачи и ее связи с другими элементами ПрО и дает возможность применять эффективные и интерпретируемые методы, что соответствует принципу объяснимости.

3. СППР декомпозирует сложные задачи на более простые. Такая декомпозиция часто основывается на модели задачи принятия решений и имеющихся представлениях задач, схожих с задачей принятия решений, в виде процессной модели (последовательности шагов, операций и исполнителей).

4. Для выполнения задачи принятия решений СППР использует доступные знания, которые могут быть представлены в структурированном (на основе модели представления знаний) и неструктурированном видах (документы, мультимедиа и др.).

Предлагаемая концепция учитывает базовые принципы ППР на основе БЯМ и исходит из того, что СППР, использующая БЯМ, должна быть способна: 1) уточнять формулировку и постановку задачи принятия решений; 2) генерировать контекстно-зависимые рекомендации с учетом предпочтений пользователя; 3) оценивать сгенерированные рекомендации; 4) поддерживать ЛПР при принятии окончательного решения, сопровождая объяснениями ответы БЯМ, предлагаемые рекомендации, а также результаты разработки, оценки и выбора альтернатив.

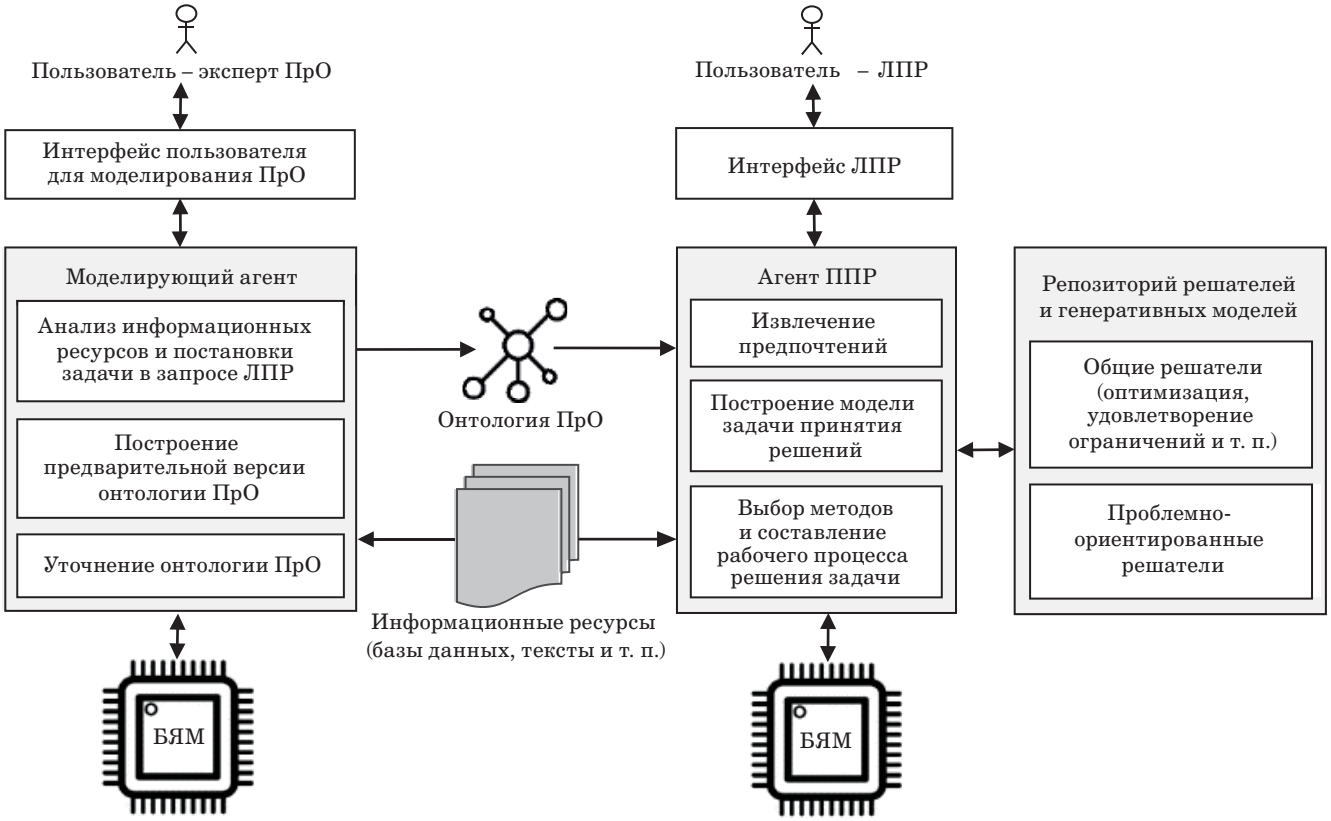
Основными компонентами СППР, использующей БЯМ, являются информационные ресурсы, онтология ПрО, БЯМ, БЯМ-агенты и репозиторий решателей и генеративных моделей (рис. 1).

Информационные ресурсы — источники данных, информации и знаний, которые внедрены в онтологию ПрО или в которых можно найти информацию, требующуюся для выполнения задачи принятия решений.

Онтология ПрО — специализированная база знаний ПрО, предоставляющая знания для выполнения задачи принятия решений.

Концептуальная организация предполагает наличие двух БЯМ-агентов: моделирующего агента и агента ППР. Под БЯМ-агентом понимается автономная интеллектуальная система, которая использует БЯМ в качестве центрального когнитивного компонента («мозга») для восприятия среды, планирования, рассуждения, распознавания сложных целей и выполнения сложных многошаговых задач [26].

Моделирующий агент взаимодействует с экспертом для создания онтологии ПрО, которая соответствует представлению и потребностям эксперта.



■ ***Рис. 1.*** Концептуальные компоненты БЯМ-ориентированной СППР
■ ***Fig. 1.*** Conceptual components of LLM-powered DSS

Агент ППР играет центральную роль. Он отвечает за поддержание диалога с ЛПР. В ходе этого диалога агент ППР решает следующие задачи: 1) уточнение описания задачи принятия решений и выявление предпочтений пользователя; 2) построение модели задачи принятия решений — онтологической модели этой задачи с учетом предпочтений пользователя — посредством извлечения из онтологии знаний, релевантных запросу пользователя, в случае необходимости — при поддержке пользователя; 3) выбор соответствующего решателя и (или) генеративной модели из репозитория. Для обогащения диалога может использоваться дополнительная информация из онтологии ПрО и доступных информационных ресурсов.

Большая языковая модель используется для разработки онтологии ПрО, интерпретации высказываний пользователей и их связи с моделью задачи принятия решений. Она также может применяться в качестве резервного решателя (для поиска оптимального или эффективного решения), однако предпочтение отдается специализированным решателям из репозитория, так как они могут гарантировать строгое (или, по крайней мере, контролируемое) удовлетворение требований пользователя и обладают большей объяснимостью. Кроме того, БЯМ может использоваться для унификации представления данных и (или) формирования консолидированного представления решения.

Репозиторий решателей и генеративных моделей — хранилище вычислительных моделей, из которого выбираются модели, предлагающие методы для решения конкретной задачи принятия решений.

Предложенная концептуальная организация СППР может быть реализована на основе существующих технологий и программных библиотек построения систем, основанных на БЯМ. Так, агенты — и моделирующий, и агент ППР — могут быть построены на базе платформы Lang* (множество взаимосвязанных библиотек, включающее, в частности, LangChain и LangGraph). Например, LangGraph предлагает адекватный набор абстракций для реализации различных сценариев обработки информации, необходимых как для моделирования ПрО в ходе взаимодействия пользователя и БЯМ, так и для извлечения предпочтений конечного пользователя и ППР.

Интеграция репозитория решателей и генеративных моделей может осуществляться различными способами: посредством концепции инструментов (tools) и, соответственно, использования возможностей обращения БЯМ к инструментам; посредством протокола MCP или же иным способом на основе более традиционных технологий обнаружения сервисов, опирающихся, например, на OWL-S.

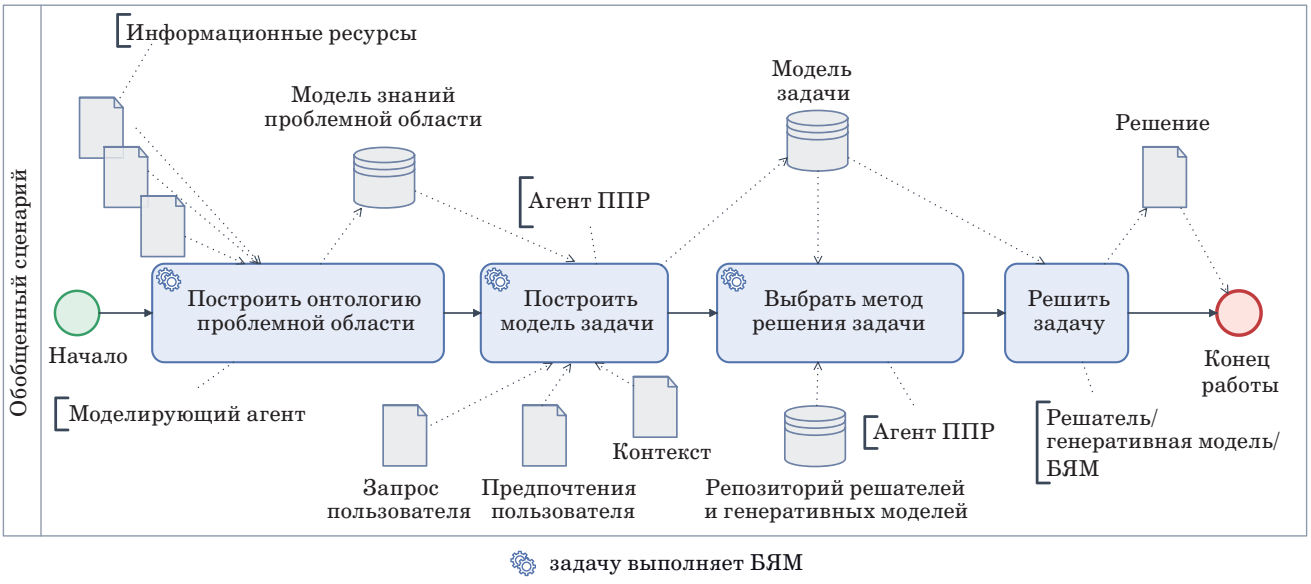
Обобщенный сценарий поддержки принятия решений на основе БЯМ

Обобщенный сценарий ППР на основе БЯМ (рис. 2) охватывает два этапа жизненного цикла СППР – разработки (анализ предметной области и сбор требований, проектирование, разработка и тестирование системы) и эксплуатации.

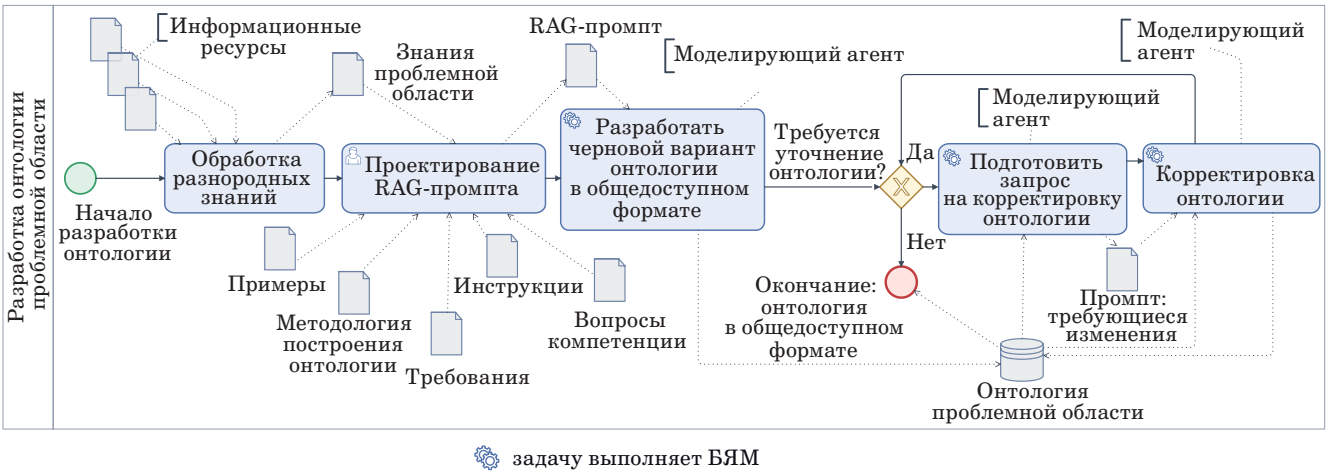
На этапе разработки осуществляются два тесно взаимосвязанных параллельных процесса: 1) анализ неструктурированных (текстовых) и структурированных (базы данных) информационных ресурсов ПрО; 2) построение и уточнение модели ПрО, включающей онтологию ПрО, модели задач, возможную декомпозицию задач на подзадачи и предполагаемые сценарии решений. Оба процесса выполняются при взаимодействии эксперта ПрО и моделирующего агента.

На этапе эксплуатации СППР идентифицируется цель ЛПР и происходит построение и последовательное уточнение модели задачи принятия решений, которая затем используется для генерации и оценки решений. Модель задачи принятия решений строит агент ППР, используя онтологию ПрО. Модель включает в себя знания ПрО, релевантные данной задаче, и предпочтения пользователя-ЛПР. Агент ППР наполняет модель контекстной информацией и выбирает компонент для решения задачи из репозитория решателей и генеративных моделей. Этот компонент решает задачу и предоставляет рекомендуемое решение.

Хотя вероятность получения полностью согласованной и детализированной онтологии непосредственно от БЯМ невысока, в то же время многие исследования (например, [28–30] и др.) обоснованно утверждают, что автоматизированное построение онтологий возможно и вполне перспективно. В обобщенном сценарии разработка онтологии реализована как итеративный процесс (рис. 3). Моделирующий агент формирует черновой вариант онтологии, который выглядит как множество RDF (Resource Description Framework) троек, релевантных рассматриваемой ПрО. RDF совместим со всеми стандартами Semantic Web (семантической паутины), Linked Data (связанных данных), а также баз знаний уровня предприятий. Другие форматы представления онтологий (например, концептуальные графы), как правило, поддерживают возможности перевода между собственными форматами и форматом RDF. Черновой вариант онтологии затем уточняется посредством взаимодействия с экспертом до тех пор, пока эксперт не будет удовлетворен результатом, т. е. созданной онтологией ПрО в желаемом формате.



■ ***Рис. 2.*** Обобщенный сценарий поддержки принятия решений на основе БЯМ (стандарт BPMN [27])
■ ***Fig. 2.*** Generalized scenario of LLM-based decision support (BPMN representation [27])

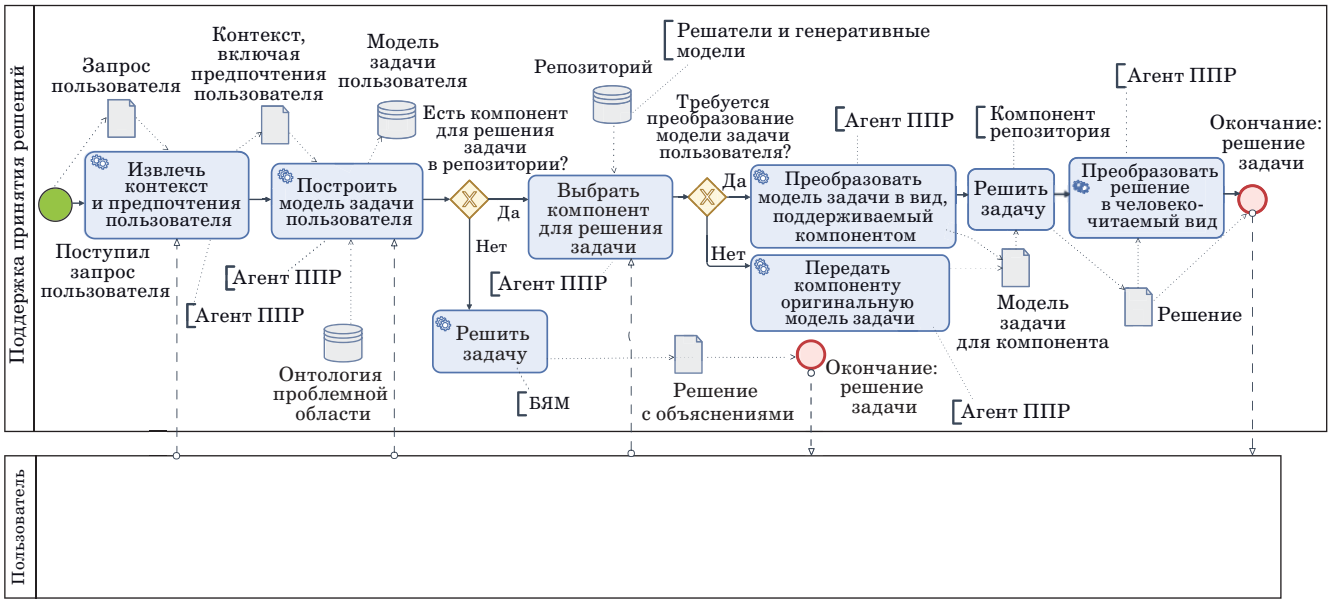


■ ***Рис. 3.*** Разработка онтологии проблемной области
■ ***Fig. 3.*** Development of domain ontology

При разработке онтологии используется RAG-промптинг. Вначале определяются информационные ресурсы, в которых содержатся знания ПрО (базы данных, документы, модели знаний и т. п.). Если найденные ресурсы содержат знания в виде диаграмм или графических материалов, то знания, представленные таким образом, преобразовываются в текстовый формат с помощью дополнительных инструментов. После чего посредством RAG-промптинга содержимое ресурсов передается в БЯМ.

Помимо передачи содержимого ресурсов в БЯМ, промпты описывают схему разработки онтологии в соответствии с выбранной методологией (например, METHONTOLOGY [31], On-to-Knowledge [32], AMOD (Agile Methodology for Ontology Development) [33] и др.), специфицируют требования к онтологии, указывают формат представления знаний, формулируют вопросы компетенции, а в случае применения техники few-shot-промптинг содержат примеры ожидаемых результатов.

В рамках обобщенного сценария рабочий процесс ППР (рис. 4) реализует классическую модель принятия решений — от идентификации задачи до рекомендации решения. Агент ППР поддерживает взаимодействие с ЛПР на всех этапах данного процесса. БЯМ выполняет функции,



■ ***Рис. 4.*** Рабочий процесс поддержки принятия решений на основе БЯМ
■ ***Fig. 4.*** Workflow of LLM-based decision support

связанные с преобразованием информации и знаний и предоставлением пользователю объяснений. Также БЯМ решает задачу принятия решений, если в хранилище решателей и генеративных моделей не найдено компонента, который мог бы ее решить.

Заключение

В работе представлен комплексный подход к интеграции больших языковых моделей в системы поддержки принятия решений. Проведенный анализ существующих сценариев использования БЯМ в СППР позволил систематизировать опыт их применения в информационно-управляющих системах из различных прикладных областей и выявить ключевые закономерности. Основным результатом анализа стало выделение четырех типовых задач, повторяющихся в различных ПрО: уточнение постановки задачи принятия решений, выявление и учет предпочтений пользователя, использование знаний ПрО, генерация и оценка альтернатив.

На основе выявленных типовых задач разработана спецификация требований к СППР на основе БЯМ, которая легла в основу базовых принципов функционирования таких систем.

Одним из результатов исследований является концепция организации СППР, отличительной особенностью которой выступает двунаправленное совмещение возможностей БЯМ и онтологического моделирования. С одной стороны, онтология ПрО и модель задачи принятия решений используются для повышения качества и интерпретируемости рекомендаций, формируемых с участием БЯМ. С другой стороны, сама БЯМ применяется как инструмент для построения и актуализации онтологии, что особенно важно в динамически меняющихся ПрО. Такой подход позволяет компенсировать ограничения языковых моделей (галлюцинации, неполнота знаний) за счет структурированных знаний и одновременно преодолеть высокую трудоемкость ручного построения онтологий за счет генеративных возможностей ИИ.

Разработанный обобщенный сценарий ППР объединяет все типовые задачи в единый рабочий процесс. Особенностью сценария является четкое разделение функций между агентами и внешними специализированными решателями. Такой подход позволяет использовать БЯМ как для непосредственного решения задач (в случаях отсутствия специализированных решателей), так и в качестве «интеллектуального посредника», обеспечивающего интерпретацию запросов, преобразование данных и объяснение полученных результатов.

Финансовая поддержка

Работа выполнена при финансовой поддержке Российского научного фонда (проект № 25-11-00127 «Модели и методы дополненного интеллекта для контекстно-зависимой поддержки принятия решений с использованием больших языковых моделей»).

Литература

1. **Abolnejadian M., Amirshahi S., Brehmer M., Crisan A.** AInsight: Augmenting expert decision-making with on-the-fly insights grounded in historical data. *Proceedings of the 7th ACM Conference on Conversational User Interfaces, CUI 2025*, Waterloo, ON, Canada, July 8–10, ACM, 2025, pp. 13:1–13:7. doi:10.1145/3719160.3737633
2. **Handler A., Larsen K. R., Hackathorn R.** Large language models present new questions for decision support. *International Journal of Information Management*, 2024, vol. 79, Article 102811. doi:10.1016/j.ijinfomgt.2024.102811
3. **Deng R., Martin G., Wang T., Zhang G., Liu Y., Weng C., Wang Y., Rousseau J. F., Peng Y.** CPG-Prompt: Translating clinical guidelines into large language model-executable decision support. *Journal of the American Medical Informatics Association*, 2026, vol. 33, iss. 4, pp. 855–862. doi:10.1093/jamia/ocag026
4. **Hager P., Jungmann F., Holland R., Bhagat K., Hubrecht I., Knauer M., Vielhauer J., Makowski M., Braren R., Kaissis G., Rueckert D.** Evaluation and mitigation of the limitations of large language models in clinical decision-making. *Nature Medicine*, 2024, vol. 30, iss. 9, pp. 2613–2622. doi:10.1038/s41591-024-03097-1
5. **Евлахова М. А., Зубков А. В.** Обзор применения больших языковых моделей в информационных задачах систем поддержки принятия решения на примере здравоохранения. *Инженерный вестник Дона*, 2025, вып. 123, № 3, с. 920–945. EDN: DEFDEH. https://ivdon.ru/ru/magazine/archive/n3y2025/9956 (дата обращения: 13.05.2026).
6. **Park C., Lee H., Lee S., Jeong O**. Synergistic joint model of knowledge graph and LLM for enhancing XAI-based clinical decision support systems. *Mathematics*, 2025, vol. 13, iss. 6, Article 949. doi:10.3390/math13060949
7. **Грибова В. В., Шалфеева Е. А.** Смарт-стандарты в реабилитации: онтологическая модель и методы

формализации. *Информатика и системы управления*, 2026, вып. 87, № 1, с. 61–73. doi:10.22250/18142400_2026_87_1_61, EDN: QZAWZY

8. **Liu S., McCoy A. B., Wright A. P., Nelson S. D., Huang S. S., Ahmad H. B., Carro S. E., Franklin J., Brogan J., Wright A.** Why do users override alerts? Utilizing large language model to summarize comments and optimize clinical decision support. *Journal of the American Medical Informatics Association*, 2024, vol. 31, iss. 6, pp. 1388–1396. doi:10.1093/jamia/ocae041
9. **Kaiser P., Yang S., Bach M., Breit C., Mertz K., Stieltjes B., Ebbing J., Wetterauer C., Henkel M.** The interaction of structured data using openEHR and large language models for clinical decision support in prostate cancer. *World Journal of Urology*, 2025, vol. 43, iss. 1, Article 67. doi:10.1007/s00345-024-05423-1
10. **Han S., Choi W.** Development of a large language model-based multi-agent clinical decision support system for Korean Triage and Acuity Scale (KTAS)-based triage and treatment planning in emergency departments. *Advances in Artificial Intelligence and Machine Learning*, 2025, vol. 5, iss. 1, pp. 3261–3275. doi:10.54364/AAIML.2025.51187
11. **Gomez-Cabello C. A., Borna S., Pressman S. M., Haider S. A., Forte A. J.** Large language models for intraoperative decision support in plastic surgery: A comparison between ChatGPT-4 and Gemini. *Medicina*, 2024, vol. 60, iss. 6, Article 957. doi:10.3390/medicina60060957
12. **Benary M., Wang X. D., Schmidt M., Soll D., Hilfenhaus G., Nassir M., Sigler C., Knödler M., Keller U., Beule D., Keilholz U., Leser U., Rieke D. T.** Leveraging large language models for decision support in personalized oncology. *JAMA Network Open*, 2023, vol. 6, iss. 11, Article e2343689. doi:10.1001/jamanetworkopen.2023.43689
13. **Durmus D., Giretti A., Ashkenazi O., Carbonari A., Isaac S.** The role of large language models for decision support in fire safety planning. *Proceedings of the 41st International Symposium on Automation and Robotics in Construction (ISARC 2024)*, 2024, pp. 339–346. doi:10.22260/ISARC2024/0045
14. **Ghiani G., Solazzo G., Elia G.** Integrating large language models and optimization in semi-structured decision making: methodology and a case study. *Algorithms*, 2024, vol. 17, iss. 12, Article 582. doi:10.3390/a17120582
15. **Tupayachi J., Xu H., Omitaomu O. A., Camur M. C., Sharmin A., Li X.** Towards next-generation urban decision support systems through AI-powered construction of scientific ontology using large language models — A case in optimizing intermodal freight transportation. *Smart Cities*, 2024, vol. 7, iss. 5, pp. 2392–2421. doi:10.3390/smartcities7050094
16. **Fang S., Liu J., Ding M., Cui Y., Lv C., Hang P., Sun J.** Towards interactive and learnable cooperative driving automation: A large language model-driven decision-making framework. *IEEE Transactions on Vehicular Technology*, 2025, vol. 74, iss. 8. pp. 11894–11905. doi:10.1109/TVT.2025.3552922
17. **Lawless C., Lawless C., Schoeffer J., Le L., Rowan K., Sen S., St. Hill C., Suh J., Sarrafzadeh B.** "I want it that way": Enabling interactive decision support using large language models and constraint programming. *ACM Transactions on Interactive Intelligent Systems*, 2024, vol. 14, iss. 3, pp. 1–33. doi:10.1145/3685053
18. **Shiiku S., Takeuchi Y.** Exploring the mutual adaptation of large language models and emergent decision-making in simulated small group interactions. *Proceedings of the 12th International Conference on Human-Agent Interaction*, ACM, 2024, pp. 270–277. doi:10.1145/3687272.3688317
19. **Kalyuzhnaya A., Mityagin S., Lutsenko E., Getmanov A., Aksenkin Y., Fatkhiev K., Fedorin K., Nikitin N. O., Chichkova N., Vorona V., Boukhanovsky A.** LLM agents for smart city management: Enhancing decision support through multi-agent AI systems. *Smart Cities*, 2025, vol. 8, iss. 1, Article 19. doi:10.3390/smartcities8010019
20. **Pesch P. J.** Potentials and challenges of Large Language Models (LLMs) in the context of administrative decision-making. *European Journal of Risk Regulation*, 2025, vol. 16, iss. 1, pp. 76–95. doi:10.1017/err.2024.99
21. **Прохорец И. О., Шматок В. А., Попов М. А.** Использование больших языковых моделей и машинного обучения в составе интеллектуальной системы проактивного мониторинга центров информационной безопасности. *Безопасность информационных технологий,* 2026, вып. 33, № 2, с. 82–93. doi:10.26583/bit.2026.2.09, EDN: QWWSON
22. **Изосимова К. С., Попов В. В., Соловьева А. Ю., Краснова В. С., Чернышева Т. Ю.** Разработка интеллектуальной СППР для составления рекомендаций по адаптации к климатическим рискам. *Математическое и информационное моделирование: материалы Всероссийской конференции молодых ученых*, Тюмень, ТюмГУ-Press, 2025, с. 82–85.
23. **Gruber T. R.** A translation approach to portable ontology specifications. *Knowledge Acquisition*, 1993, vol. 5, iss. 2, pp. 199–220. doi:10.1006/knac.1993.1008
24. **Manning C. D.** Human language understanding & reasoning. *Daedalus*, 2022, vol. 151, iss. 2, pp. 127–138. doi:10.1162/daed_a_01905
25. *openEHR Architecture Overview*. The openEHR Foundation. https://specifications.openehr.org/releases/BASE/latest/architecture_overview.html#_architecture_overview (дата обращения: 13.05.2026).
26. **Cheng Y., Zhang C., Zhang Z., Meng X., Hong S., Li W., Wang Z., Wang Z., Yin F., Zhao J., He X.** *Exploring Large Language Model Based Intelligent Agents: Definitions, Methods, and Prospects*. 2024. http://arxiv.org/abs/2401.03428 (дата обращения: 13.05.2026).

27. **Briol P.** *BPMN, the Business Process Modeling Notation*. 1st ed. Lulu.com, 2008. 124 p.
28. **Neuhaus F.** Ontologies in the era of large language models — a perspective. *Applied Ontology*, 2023, vol. 18, iss. 4, pp. 399–407. doi:10.3233/AO-230072
29. **Chen K., Yang Y., Chen B., Hernández López J. A., Mussbacher G., Varró D.** Automated domain modeling with large language models: A comparative study. *Proceedings of the 26th International Conference on Model Driven Engineering Languages and Systems (MODELS)*, IEEE, 2023, pp. 162–172. doi:10.1109/MODELS58315.2023.00037
30. **Saeedizade M. J., Blomqvist E.** Navigating ontology development with large language models. *Proceedings of the 21st European Semantic Web Conference ESWC 2024*, A. Meroño Peñuela, A. Dimou, R. Troncy, O. Hartig, M. Acosta, M. Alam, H. Paulheim, P. Lisena (Eds). Cham, Springer, 2024, vol. 14664, pp. 143–161. doi:10.1007/978-3-031-60626-7_8
31. **Fernández-López M., Gómez-Pérez A., Juristo N., Fernandez M., Gomes-Perez A., Juristo N.** METHONTOLOGY: From ontological art towards ontological engineering. *AAAI Proceedings of the Symposium on Ontological Engineering*, AAAI, 1997, pp. 33–40.
32. **Sure Y., Staab S., Studer R.** On-To-Knowledge Methodology (OTKM). *Handbook on Ontologies*. Berlin, Heidelberg, Springer, 2004, pp. 117–132. doi:10.1007/978-3-540-24750-0_6
33. **Abdelghany A., Darwish N., Hefni H.** An agile methodology for ontology development. *International Journal of Intelligent Engineering and Systems*, 2019, vol. 12, iss. 2, pp. 170–181. doi:10.22266/ijies2019.0430.17

UDC 004.89:004.822:004.048
doi:10.31799/1684-8853-2026-4-2-16
EDN: GUSQBR

Generalized scenario of LLM-powered decision support in information and control systems

A. V. Smirnov[a], Dr. Sc., Tech., Professor, Principal Researcher, orcid.org/0000-0001-8364-073X, smir@iias.spb.su
T. V. Levashova[a], PhD, Tech., Senior Researcher, orcid.org/0000-0002-1962-7044
A. V. Ponomarev[a], PhD, Tech., Associate Professor, Senior Researcher, orcid.org/0000-0002-9380-5064
[a]St. Petersburg Federal Research Center of the RAS, 39, 14th Line, 199178, Saint-Petersburg, Russian Federation

Introduction: Large language models improve the efficiency of decision-making for users of information and control systems. Different decision support scenarios involve recurring (typical) tasks. **Purpose:** To develop a generalized decision support scenario powered by large language models that meets the regulatory requirements for such systems and ensures the implementation of the decision-making process as a process of solving typical tasks. **Results:** Based on the analysis of decision support scenarios powered by large language models, this study identifies typical tasks, establishes the kinds of information resources used, and develops a specification of requirements for decision support powered by large language models. Following this specification, the paper proposes principles of decision support powered by large language models and a conceptual organization of a decision support system that embodies these principles. The main research result is a generalized scenario of decision support powered by large language models in Business Process Model Notation, grounded in the developed conceptual organization and combining the typical tasks. **Practical relevance:** The proposed generalized scenario is of practical interest as a prototype scenario for a problem-independent decision support system, in which a large language model is used as a primary and/or auxiliary model in decision support.

Keywords — decision support, large language model, ontology, scenario.

For citation: Smirnov A. V., Levashova T. V., Ponomarev A. V. Generalized scenario of LLM-powered decision support in information and control systems. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2026, no. 4, pp. 2–16 (In Russian). doi:10.31799/1684-8853-2026-4-2-16, EDN: GUSQBR

Financial support
The research is due to the Russian Science Foundation Project No. 25-11-00127 “Models and methods of LLM-assisted augmented intelligence for context-aware decision support”.

References

1. Abolnejadian M., Amirshahi S., Brehmer M., Crisan A. AInsight: Augmenting expert decision-making with on-the-fly insights grounded in historical data. *Proceedings of the 7th ACM Conference on Conversational User Interfaces, CUI 2025*, ACM, 2025, pp. 13:1–13:7. doi:10.1145/3719160.3737633
2. Handler A., Larsen K. R., Hackathorn R. Large language models present new questions for decision support. *International Journal of Information Management*, 2024, vol. 79, Article 102811. doi:10.1016/j.ijinfomgt.2024.102811
3. Deng R., Martin G., Wang T., Zhang G., Liu Y., Weng C., Wang Y., Rousseau J. F., Peng Y. CPGPrompt: Translating clinical guidelines into large language model-executable decision support. *Journal of the American Medical Informatics Association*, 2026, vol. 33, no. 4, pp. 855–862. doi:10.1093/jamia/ocag026
4. Hager P., Jungmann F., Holland R., Bhagat K., Hubrecht I., Knauer M., Vielhauer J., Makowski M., Braren R., Kaissis G., Rueckert D. Evaluation and mitigation of the limitations of large language models in clinical decision-making. *Nature Medicine*, 2024, vol. 30, no. 9, pp. 2613–2622. doi:10.1038/s41591-024-03097-1
5. Evlakhova M. A., Zubkov A. V. An overview of the use of large language models in information tasks of decision support systems using the example of healthcare. *Ingineering Journal of Don*, 2025, vol. 123, no. 3, pp. 920–945 (In Russian). EDN: DEFDEH. Available at: https://ivdon.ru/ru/magazine/archive/n3y2025/9956 (accessed 13 May 2026).
6. Park C., Lee H., Lee S., Jeong O. Synergistic joint model of knowledge graph and LLM for enhancing XAI-based clinical decision support systems. *Mathematics*, 2025, vol. 13, no. 6, Article 949. doi:10.3390/math13060949

7. Gribova V. V., Shalfeeva E. A. Smart standards in rehabilitation: An ontological model and formalization methods. *Information Science and Control Systems*, 2026, vol. 87, no. 1, pp. 61–73 (In Russian). doi:10.22250/18142400_2026_87_1_61, EDN: QZAWZY
8. Liu S., McCoy A. B., Wright A. P., Nelson S. D., Huang S. S., Ahmad H. B., Carro S. E., Franklin J., Brogan J., Wright A. Why do users override alerts? Utilizing large language model to summarize comments and optimize clinical decision support. *Journal of the American Medical Informatics Association*, 2024, vol. 31, no. 6, pp. 1388–1396. doi:10.1093/jamia/ocae041
9. Kaiser P., Yang S., Bach M., Breit C., Mertz K., Stieltjes B., Ebbing J., Wetterauer C., Henkel M. The interaction of structured data using openEHR and large language models for clinical decision support in prostate cancer. *World Journal of Urology*, 2025, vol. 43, no. 1, Article 67. doi:10.1007/s00345-024-05423-1
10. Han S., Choi W. Development of a large language model-based multi-agent clinical decision support system for Korean Triage and Acuity Scale (KTAS)-based triage and treatment planning in emergency departments. *Advances in Artificial Intelligence and Machine Learning*, 2025, vol. 5, no. 1, pp. 3261–3275. doi:10.54364/AAIML.2025.51187
11. Gomez-Cabello C. A., Borna S., Pressman S. M., Haider S. A., Forte A. J. Large language models for intraoperative decision support in plastic surgery: A comparison between ChatGPT-4 and Gemini. *Medicina*, 2024, vol. 60, no. 6, Article 957. doi:10.3390/medicina60060957
12. Benary M., Wang X. D., Schmidt M., Soll D., Hilfenhaus G., Nassir M., Sigler C., Knödler M., Keller U., Beule D., Keilholz U., Leser U., Rieke D. T. Leveraging large language models for decision support in personalized oncology. *JAMA Network Open*, 2023, vol. 6, no. 11, Article e2343689. doi:10.1001/jamanetworkopen.2023.43689
13. Durmus D., Giretti A., Ashkenazi O., Carbonari A., Isaac S. The role of large language models for decision support in fire safety planning. *Proceedings of the 41st International Symposium on Automation and Robotics in Construction (ISARC 2024)*, 2024, pp. 339–346. doi:10.22260/ISARC2024/0045
14. Ghiani G., Solazzo G., Elia G. Integrating large language models and optimization in semi-structured decision making: Methodology and a case study. *Algorithms*, 2024, vol. 17, no. 12, Article 582. doi:10.3390/a17120582
15. Tupayachi J., Xu H., Omitaomu O. A., Camur M. C., Sharmin A., Li X. Towards next-generation urban decision support systems through AI-powered construction of scientific ontology using large language models — A case in optimizing intermodal freight transportation. *Smart Cities*, 2024, vol. 7, no. 5, pp. 2392–2421. doi:10.3390/smartcities7050094
16. Fang S., Liu J., Ding M., Cui Y., Lv C., Hang P., Sun J. Towards interactive and learnable cooperative driving automation: A large language model-driven decision-making framework. *IEEE Transactions on Vehicular Technology*, 2025, vol. 74, no. 8. pp. 11894–11905. doi:10.1109/TVT.2025.3552922
17. Lawless C., Lawless C., Schoeffer J., Le L., Rowan K., Sen S., St. Hill C., Suh J., Sarrafzadeh B. "I want it that way": Enabling interactive decision support using large language models and constraint programming. *ACM Transactions on Interactive Intelligent Systems*, 2024, vol. 14, no. 3, pp. 1–33. doi:10.1145/3685053
18. Shiiku S., Takeuchi Y. Exploring the mutual adaptation of large language models and emergent decision-making in simulated small group interactions. *Proceedings of the 12th International Conference on Human-Agent Interaction*, ACM, 2024, pp. 270–277. doi:10.1145/3687272.3688317
19. Kalyuzhnaya A., Mityagin S., Lutsenko E., Getmanov A., Aksenkin Y., Fatkhiev K., Fedorin K., Nikitin N. O., Chichkova N., Vorona V., Boukhanovsky A. LLM agents for smart city management: Enhancing decision support through multi-agent AI systems. *Smart Cities*, 2025, vol. 8, no. 1, Article 19. doi:10.3390/smartcities8010019
20. Pesch P. J. Potentials and challenges of Large Language Models (LLMs) in the context of administrative decision-making. *European Journal of Risk Regulation*, 2025, vol. 16, no. 1, pp. 76–95. doi:10.1017/err.2024.99
21. Prokhorets I., Shmatok V., Popov M. The use of large language models and machine learning as part of an intelligent proactive information security monitoring system. *IT Security*, 2026, vol. 33, no. 2, pp. 82–93 (In Russian). doi:10.26583/bit.2026.2.09, EDN: QWWSON
22. Izosimova K., Popov V., Solovieva A., Krasnova V., Chernysheva T. Development of intelligent DSS for generating recommendations on adaptation to climate risks. *Materialy Vserossijskoj konferencii molodyh uchenyh "Matematicheskoe i informacionnoe modelirovanie"* [Proc. of the All-Russian Conf. of Young Scientists "Mathematical and information modeling"]. Tyumen', Tyumenskij gosudarstvennyj universitet Press Publ., 2025, pp. 82–85 (In Russian).
23. Gruber T. R. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 1993, vol. 5, no. 2, pp. 199–220. doi:10.1006/knac.1993.1008
24. Manning C. D. Human language understanding & reasoning. *Daedalus*, 2022, vol. 151, no. 2, pp. 127–138. doi:10.1162/daed_a_01905
25. *openEHR Architecture Overview*. The openEHR Foundation. Available at: https://specifications.openehr.org/releases/BASE/latest/architecture_overview.html#_architecture_overview (accessed 13 May 2026).
26. Cheng Y., Zhang C., Zhang Z., Meng X., Hong S., Li W., Wang Z., Wang Z., Yin F., Zhao J., He X. *Exploring Large Language Model Based Intelligent Agents: Definitions, Methods, and Prospects*. 2024. Available at: http://arxiv.org/abs/2401.03428 (accessed 13 May 2026).
27. Briol P. *BPMN, the Business Process Modeling Notation*. 1st ed. Lulu.com, 2008. 124 p.
28. Neuhaus F. Ontologies in the era of large language models — a perspective. *Applied Ontology*, 2023, vol. 18, no. 4, pp. 399–407. doi:10.3233/AO-230072
29. Chen K., Yang Y., Chen B., Hernández López J.A., Mussbacher G., Varró D. Automated domain modeling with large language models: A comparative study. *Proceedings of the 26th International Conference on Model Driven Engineering Languages and Systems (MODELS)*, IEEE, 2023, pp. 162–172. doi:10.1109/MODELS58315.2023.00037
30. Saeedizade M. J., Blomqvist E. Navigating ontology development with large language models. *Procceedings of the 21st European Semantic Web Conference ESWC 2024*. A. Meroño Peñuela, A. Dimou, R. Troncy, O. Hartig, M. Acosta, M. Alam, H. Paulheim, P. Lisena (Eds), Cham, Springer, 2024, vol. 14664, pp. 143–161. doi:10.1007/978-3-031-60626-7_8
31. Fernández-López M., Gómez-Pérez A., Juristo N., Fernandez M., Gomes-Perez A., Juristo N. METHONTOLOGY: From ontological art towards ontological engineering. *AAAI Proceedings of the Symposium on Ontological Engineering*, AAAI, 1997, pp. 33–40.
32. Sure Y., Staab S., Studer R. *On-To-Knowledge Methodology (OTKM)*. In: *Handbook on Ontologies*. Berlin, Heidelberg, Springer, 2004, pp. 117–132. doi:10.1007/978-3-540-24750-0_6
33. Abdelghany A., Darwish N., Hefni H. An agile methodology for ontology development. *International Journal of Intelligent Engineering and Systems*, 2019, vol. 12, no. 2, pp. 170–181. doi:10.22266/ijies2019.0430.17

УДК 003.26
doi:10.31799/1684-8853-2026-4-17-27
EDN: GJXNHN

Научные статьи
Articles

Постквантовый алгоритм цифровой подписи с новым механизмом задания вспомогательных скрытых групп

Н. А. Молдовян[a], *доктор техн. наук, профессор, главный научный сотрудник, orcid.org/0000-0002-4483-5048, moldovyan.na@talantiuspeh.ru*
А. А. Молдовян[б], *доктор техн. наук, профессор, главный научный сотрудник, orcid.org/0000-0001-5480-6016, maa1305@yandex.ru*
[a]*Научный центр информационных технологий и искусственного интеллекта Научно-технологического университета «Сириус», Олимпийский пр., 1, Краснодарский край, федеральная территория «Сириус», 354340, РФ*
[б]*Санкт-Петербургский Федеральный исследовательский центр Российской академии наук, 14-я линия В. О., 39, Санкт-Петербург, 199178, РФ*

Введение: *с точки зрения практического применения известные постквантовые алгоритмы электронной цифровой подписи имеют недостаток, состоящий в их сравнительно большом общем размере открытого ключа и подписи.* ***Цель:*** *разработать постквантовый алгоритм электронной цифровой подписи с малым общим размером подписи и открытого ключа при заданном уровне стойкости.* ***Метод:*** *применение скрытых коммутативных групп в конечной алгебре квадратных матриц, заданных над простым конечным полем GF(p), и формирование открытого ключа в виде набора матриц, которые связаны с секретными матрицами через степенные выражения.* ***Результаты:*** *разработан практичный постквантовый алгебраический алгоритм цифровой подписи со скрытыми коммутативными группами, стойкость которого основана на вычислительной трудности решения больших систем степенных уравнений. Применен новый способ задания вспомогательных скрытых групп и новый механизм верификации подписи по двум проверочным уравнениям с вхождением всех элементов подписи в каждое из них. В качестве алгебраического носителя использованы конечные алгебры матриц μ × μ при значениях μ = 3, 5 и 7, задающих три версии предложенного алгоритма. Для всех версий алгоритма получены оценки стойкости к прямой атаке, к атаке на основе известных подписей и к подделке подписи.* ***Практическая значимость:*** *разработанный алгоритм обладает малыми размерами открытого ключа и подписи и является кандидатом на практичную постквантовую схему подписи.*

Ключевые слова *— постквантовая криптография, компьютерная безопасность, электронная цифровая подпись, система степенных уравнений, вычислительно трудная задача, конечная алгебра матриц.*

Для цитирования: Молдовян Н. А., Молдовян А. А. Постквантовый алгоритм цифровой подписи с новым механизмом задания вспомогательных скрытых групп. *Информационно-управляющие системы*, 2026, № 4, с. 17–27. doi:10.31799/1684-8853-2026-4-17–27, EDN: GJXNHN

For citation: Moldovyan N. A., Moldovyan A. A. Post-quantum digital signature algorithm with a new mechanism for specifying auxiliary hidden groups. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2026, no. 4, pp. 17–27 (In Russian). doi:10.31799/1684-8853-2026-4-17-27, EDN: GJXNHN

Введение

Значительное внимание мирового криптографического сообщества уделяется разработке постквантовых криптоалгоритмов с открытым ключом (ОК). Одним из перспективных направлений постквантовой криптографии является использование вычислительной трудности решения больших систем степенных уравнений (БССУ) [1, 2]. На протяжении более тридцати пяти лет в рамках этого направления алгоритмы открытого шифрования и электронной цифровой подписи (ЭЦП) строились на трудно обратимых отображениях с секретной лазейкой [3–5]. Алгоритмы ЭЦП такого типа обладают малым размером подписи, однако длина ОК является чрезвычайно большой (от десятков килобайт при уровне стойкости 2^{80} до нескольких мегабайт при стойкости 2^{256}) [1], что с точки зрения практического применения является существенным недостатком.

Попыткой устранения этого недостатка явилась новая парадигма [6, 7] задания трудно обратимых отображений в виде набора многочленов, описывающих операции возведения в небольшую степень n в векторном конечном поле с секретной модификацией, а потайной лазейки — в виде операции извлечения корня степени n. Этот подход позволяет уменьшить размер ОК до нескольких десятков раз, однако и в этом случае упомянутый недостаток не устраняется в полной мере.

Использование вычислительной трудности синдромного декодирования для построения постквантовых алгоритмов ЭЦП [8, 9] позволяет обеспечить достаточно малый размер ОК (около 200 байт), однако размер подписи составляет сотни килобайт. Разработка алгоритмов ЭЦП с использованием синдромного декодирования ограниченных ошибок [10] привела к сокращению размера ЭЦП до 17 Кбайт [11], ≈10 Кбайт [12, 13] и 7 Кбайт [14].

Представляет интерес сравнительно новая концепция построения алгебраических алгоритмов ЭЦП, стойкость которых основана на вычислительной сложности решения БССУ [15, 16]. В рамках этой концепции в качестве алгебраического носителя используются конечные некоммутативные ассоциативные алгебры, а цифровая подпись включает подгоночный элемент в виде многомерного вектора **S**, формируемого путем выбора случайного элемента **G** из секретной коммутативной группы и маскирования вектора **G** путем умножения его слева и справа на фиксированные секретные векторы **D** и **F** : **S** = **DGF**. В работах [17, 18] показана ограниченность рандомизации, задаваемой последней формулой, которая создает уязвимость к атакам на основе известных подписей. Для устранения этой уязвимости предложены механизмы усиления рандомизации подгоночного элемента подписи путем вычисления вектора **S** 1) по двум случайным векторам, выбираемым из скрытой коммутативной группы [19]; 2) по двум случайным взаимно некоммутативным векторам, выбираемым из двух скрытых коммутативных групп [20, 21], и 3) по трем случайным векторам, выбираемым из трех скрытых коммутативных групп [22].

Алгебраическая структура, в рамках которой выполняются вычисления, связанные с формированием секретного и открытого ключей и с генерацией и верификацией подписи, называется носителем алгоритма ЭЦП. В качестве алгебраического носителя алгебраических алгоритмов ЭЦП, использующих вычислительную трудность решения БССУ, перспективно применение конечных алгебр квадратных матриц $\mu \times \mu$, заданных над простым конечным полем $GF(p)$, размер порядка которого выбирается в зависимости от размерности матриц [19, 23]. В целом постквантовые алгебраические алгоритмы ЭЦП рассматриваемого типа имеют общий размер ОК и подписи менее 1 Кбайт, что обусловливает значительный практический интерес к ним. Однако их базовая концепция является сравнительно новой, и требуется расширение исследований по анализу их стойкости, включая рассмотрение специфических атак, использующих особенности их построения.

Проверочные уравнения известных алгебраических алгоритмов ЭЦП относятся к трем различным типам, отличающимся механизмом обеспечения стойкости к подделке подписи путем решения проверочного уравнения относительно подгоночного элемента подписи **S** как неизвестного алгебраического элемента [21]. Известны следующие типы упомянутого механизма:

1) двукратное вхождение элемента **S** в проверочное уравнение, включая возведение **S** в степень большого размера;

2) использование в проверочном уравнении операции возведения в степень, вычисляемую как значение однонаправленной функции, вычисляемое от **S**;

3) комбинирование первых двух механизмов.

В данной работе предлагается четвертый тип проверочных соотношений, отличающийся заданием двух проверочных уравнений первой степени относительно **S**, задающих при попытках подделки подписи переопределенную несовместную систему линейных уравнений. На основе нового механизма задания стойкости к подделке подписи разработан практичный постквантовый алгоритм ЭЦП с двумя базовыми и двумя вспомогательными скрытыми коммутативными группами.

Алгебраический носитель

В разработанном алгоритме ЭЦП в качестве алгебраического носителя используются конечные алгебры квадратных матриц, заданных над простым конечным полем $GF(p)$ с числом строк, равным простым числам $\mu = 3, 5$ и 7. Эти значения μ задают три версии разработанного алгоритма с общими формулами для 1) вычисления открытого ключа по секретному ключу, 2) вычисления элементов цифровой подписи и 3) выполнения процедуры проверки подлинности подписи. Свойства указанных матричных алгебр, используемых при разработке алгоритмов ЭЦП, стойкость которых базируется на вычислительной трудности решения БССУ, детально представлены в работах [19, 23].

В данной работе используются следующие известные факты [19–24].

1. Для простых значений μ конечная алгебра матриц $\mu \times \mu$, заданная над полем $GF(p)$ нечетного простого порядка p, содержит достаточно большое число ($\gg p^{\mu}$) коммутативных циклических групп порядка $p^{\mu} - 1$.

2. Достаточно легко могут быть сгенерированы простые значения r такие, что числа $p = 2r - 1$ и

$$q = p^{\mu-1} + p^{\mu-2} + \ldots + p + 1 \qquad (1)$$

являются простыми. Легко видеть, что в этом случае q является делителем числа $p^{\mu} - 1$,

и матрица **G** порядка q генерирует циклическую группу <**G**> простого порядка q. Под термином «порядок матрицы» в данной статье понимается ее порядок как элемента мультипликативной группы конечной алгебры матриц.

В разработанном алгоритме ЭЦП используются простые значения p, для которых r и q также являются простыми в связи с тем, что при генерации подписи требуется вычислять обратные значения для случайных чисел по модулю q и по модулю rq. При таком выборе практически устраняется проблема появления случаев, когда обратные значения не существуют. Разрядность чисел r, p и q для различных версий разработанного алгоритма представлена в табл. 1.

3. Если задана невырожденная нескалярная матрица **P**, содержащаяся в некоторой некоммутативной группе Γ, то элементы любой матрицы, содержащейся в Γ, могут быть представлены в виде линейных комбинаций некоторого уникального набора из μ скалярных значений $a_i \in GF(p)$, где $i = 0, 1, \ldots, \mu - 1$, в которых коэффициенты определяются значениями элементов матрицы **P**, которую будем называть представителем группы Γ. Заметим, что коммутативная группа Γ содержится в некотором коммутативном подкольце порядка p^{μ} конечного кольца всех возможных матриц размера $\mu \times \mu$. При фиксированной матрице-представителе **P** каждому уникальному набору из μ скалярных значений соответствует уникальная матрица из упомянутого подкольца. Очевидно, что каждой матрице из группы Γ соответствует уникальный набор указанных скалярных значений.

Последнее означает, что выбор η матриц из некоторой скрытой коммутативной циклической группы <**G**> задает μ^2 скалярных неизвестных (элементы первой выбранной неизвестной матрицы **G**′), соответствующих выбору первой неизвестной матрицы, и $\mu(\eta - 1)$ скалярных неизвестных, соответствующих выбору остальных матриц, координаты каждой из которых могут быть описаны через уникальный набор из μ скалярных неизвестных при рассмотрении матрицы **G**′ в качестве представителя группы <**G**>.

Генерация секретного и открытого ключей

Секретный ключ формируется по следующему алгоритму.

1. Генерируются четыре случайные попарно некоммутирующие обратимые матрицы **G**, **J**, **Q** и **U** такие, что порядок **G** и **J** равен q, а порядок **Q** и **U** равен rq.

2. Генерируются случайные невырожденные попарно некоммутирующие матрицы **A**, **B**, **D** и **F**, каждая из которых является некоммутирующей с матрицами **G**, **J**, **Q** и **U**, и случайные натуральные числа α, β, γ, ε, λ и τ, не превышающие значение q.

3. Генерируются четыре случайных натуральных числа x, z, u и v ($< q$) и вычисляются матрицы $\mathbf{K}_1 = \mathbf{Q}\mathbf{J}^{-z}\mathbf{G}^{-x}$ и $\mathbf{K}_2 = \mathbf{J}^{-v}\mathbf{G}^{-u}\mathbf{U}$.

В результате выполнения этого алгоритма получаем секретный ключ в виде набора из восьми матриц (**A**, **B**, **D**, **F**, **G**, **J**, **Q**, **U**) и набора из 10 чисел (α, β, γ, ε, λ, τ, x, z, u, v). Открытый ключ представляет собой набор из шести матриц (**Y**, **Z**, **W**, **N**, $\mathbf{T}_1$, $\mathbf{T}_2$), которые вычисляются в зависимости от секретного по следующим формулам:

$$\mathbf{Y} = \mathbf{A}\mathbf{Q}^{\alpha}\mathbf{A}^{-1},\ \mathbf{Z} = \mathbf{D}\mathbf{G}\mathbf{D}^{-1},$$

$$\mathbf{W} = \mathbf{F}^{-1}\mathbf{J}\mathbf{F},\ \mathbf{N} = \mathbf{B}\mathbf{U}^{\tau}\mathbf{B}^{-1}; \quad (2)$$

$$\mathbf{T}_1 = \mathbf{A}\mathbf{Q}^{\beta}\mathbf{K}_1\mathbf{G}^{\gamma}\mathbf{D}^{-1},$$

$$\mathbf{T}_2 = \mathbf{F}^{-1}\mathbf{J}^{\varepsilon}\mathbf{K}_2\mathbf{U}^{\lambda}\mathbf{B}^{-1}. \quad (3)$$

Размеры открытого и секретного ключей представлены в табл. 2.

Формулы, используемые на шаге 3, определяют следующую связь базовых скрытых коммутативных групп <**G**> и <**J**> с вспомогательными скрытыми группами <**Q**> и <**U**>:

$$\mathbf{Q} = \mathbf{K}_1\mathbf{G}^{x}\mathbf{J}^{z}; \quad (4)$$

$$\mathbf{U} = \mathbf{G}^{u}\mathbf{J}^{v}\mathbf{K}_2. \quad (5)$$

■ ***Таблица 1.*** Размер используемых простых чисел
■ ***Table 1.*** The size of used primes

Версия алгоритма	Разрядность, бит		
	r	p	q
μ = 3	63	64	128
μ = 5	31	32	128
μ = 7	23	24	144

■ ***Таблица 2.*** Размеры параметров предложенного алгоритма ЭЦП
■ ***Table 2.*** The size of used primes

Версия алгоритма	Размер, байт		
	открытого ключа	секретного ключа	ЭЦП
μ = 3	432	896	216
μ = 5	600	1168	236
μ = 7	882	1656	291

Формулы (4) и (5) отличаются от формул задания связи вспомогательных и базовых скрытых групп из работы [21] наличием множителей в виде невырожденных матриц $\mathbf{K}_1$ и $\mathbf{K}_2$. Это отличие упрощает процедуру генерации секретного ключа, что служит сопутствующим эффектом. Главным назначением этого модифицирования является предотвращение возможности использовать альтернативные формулы для вычисления ЭЦП, что обусловливало бы необходимость оценивания стойкости с учетом альтернативной процедуры генерации подписи.

Процедура генерации и верификации ЭЦП. Доказательство корректности разработанного алгоритма ЭЦП

В разработанном алгоритме ЭЦП предполагается использование некоторой 512-битной хеш-функции $\Phi(\cdot)$, стойкой к коллизиям. Процесс генерации подписи требует использования секретного ключа и включает следующие шаги.

1. Сгенерировать четыре случайных натуральных числа $k < rq$, $t < q$, $b < q$, $g < rq$ и вычислить две рандомизирующие матрицы $\mathbf{R}_1$ и $\mathbf{R}_2$ по следующим формулам:

$$\mathbf{R}_1 = \mathbf{A}\mathbf{Q}^k\mathbf{J}^t\mathbf{F}; \tag{6}$$

$$\mathbf{R}_2 = \mathbf{D}\mathbf{G}^b\mathbf{U}^g\mathbf{B}^{-1}. \tag{7}$$

2. Вычислить четыре 128-битных рандомизирующих элемента ЭЦП e_1, e_2, e_3, e_4, конкатенация которых представляет собой значение 512-битной хеш-функции $\Phi(\cdot)$, вычисленной от документа M с присоединенными к нему рандомизирующими матрицами: $e_1||e_2||e_3||e_4 = \Phi(M||\mathbf{R}_1||\mathbf{R}_2)$.

3. Вычислить элементы подписи ξ и ψ:

$$\xi = (e_3\alpha)^{-1}(k - \beta - 1) \bmod rq; \tag{8}$$

$$\psi = (e_4\tau)^{-1}(g - \lambda - 1) \bmod rq. \tag{9}$$

4. Найти значения s и n путем решения следующей системы из двух уравнений первой степени:

$$\begin{cases} \gamma + e_1 s + n = x \bmod q, \\ e_3 s + n = u + b \bmod q, \end{cases} \tag{10}$$

что соответствует выполнению вычислений по формулам

$$s = (e_3 - e_1)^{-1}(u + b + \gamma - x) \bmod q;$$

$$n = x - \gamma - e_1 s \bmod q. \tag{11}$$

5. Найти значения σ и d путем решения следующей системы из двух уравнений первой степени:

$$\begin{cases} d + e_2\sigma = z + t \bmod q, \\ d + e_4\sigma + \varepsilon = v \bmod q, \end{cases} \tag{12}$$

что соответствует выполнению вычислений по формулам

$$\sigma = (e_4 - e_2)^{-1}(v - \varepsilon - z - t) \bmod q;$$

$$d = z + t - e_2\sigma \bmod q. \tag{13}$$

6. Вычислить значение подгоночной матрицы $\mathbf{S}$ по формуле

$$\mathbf{S} = \mathbf{D}\mathbf{G}^n\mathbf{J}^d\mathbf{F}. \tag{14}$$

Подпись представляет собой набор элементов ($e_1, e_2, e_3, e_4, s, \sigma, \xi, \psi, \mathbf{S}$). Размер подписи для различных версий разработанного алгоритма представлен в табл. 2. Вычислительная сложность процедуры генерации ЭЦП определяется, главным образом, шестью операциями возведения в степень (вычисление следующих матриц $\mathbf{Q}^k$, $\mathbf{J}^t$, $\mathbf{G}^b$, $\mathbf{U}^g$, $\mathbf{G}^n$ и $\mathbf{J}^d$) и равна примерно $2 \cdot 10^8$ (для версий алгоритма со значениями параметра $\mu = 3, 5$) и $3 \cdot 10^8$ битовые операции (для версии с параметром $\mu = 7$).

Алгоритм проверки подлинности ЭЦП ($e_1, e_2, e_3, e_4, s, \sigma, \xi, \psi, \mathbf{S}$) к электронному документу M осуществляется по ОК ($\mathbf{Y}, \mathbf{Z}, \mathbf{W}, \mathbf{N}, \mathbf{T}_1, \mathbf{T}_2$) и описывается следующим образом.

1. Вычислить проверочные матрицы $\mathbf{R}'_1$ и $\mathbf{R}'_2$ по следующим формулам:

$$\mathbf{R}'_1 = \mathbf{Y}^{e_3\xi}\mathbf{T}_1\mathbf{Z}^{e_1 s}\mathbf{S}\mathbf{W}^{e_2\sigma}; \tag{15}$$

$$\mathbf{R}'_2 = \mathbf{Z}^{e_3 s}\mathbf{S}\mathbf{W}^{e_4\sigma}\mathbf{T}_2\mathbf{N}^{e_4\psi}. \tag{16}$$

2. Вычислить 512-битное хеш-значение, представленное в виде конкатенации четырех 128-битных чисел: $e'_1||e'_2||e'_3||e'_4 = \Phi(M||\mathbf{R}'_1||\mathbf{R}'_2)$.

3. Если $e'_1||e'_2||e'_3||e'_4 = e_1||e_2||e_3||e_4$, то подпись подлинная, иначе — ложная.

Вычислительная сложность процедуры верификации ЭЦП примерно равна сложности процедуры генерации подписи и определяется в основном шестью операциями экспоненцирования матриц в степень большой разрядности. При использовании алгоритма быстрого возведения в степень, основанного на процессе последовательного возведения в квадрат, сложность верификации ЭЦП составляет примерно $2 \cdot 10^8$ (для случаев $\mu = 3, 5$) и $3 \cdot 10^8$ битовые операции (для случая $\mu = 7$).

Для подписи (e_1, e_2, e_3, e_4, s, σ, ξ, ψ, **S**), вычисленной в соответствии с процедурой генерации ЭЦП, вычисление проверочных матриц $\mathbf{R}'_1$ и $\mathbf{R}'_2$ по формулам (15) и (16) дает следующее:

$$\mathbf{R}'_1 = \left(\mathbf{A}\mathbf{Q}^{\alpha}\mathbf{A}^{-1}\right)^{e_3\xi}\left(\mathbf{A}\mathbf{Q}^{\beta}\mathbf{K}_1\mathbf{G}^{\gamma}\mathbf{D}^{-1}\right)\left(\mathbf{D}\mathbf{G}\mathbf{D}^{-1}\right)^{e_1 s} \times$$

$$\times\left(\mathbf{D}\mathbf{G}^{n}\mathbf{J}^{d}\mathbf{F}\right)\left(\mathbf{F}^{-1}\mathbf{J}\mathbf{F}\right)^{e_2\sigma} =$$

$$= \mathbf{A}\mathbf{Q}^{\alpha e_3\xi+\beta}\mathbf{K}_1\mathbf{G}^{\gamma+e_1 s+n}\mathbf{J}^{d+e_2\sigma}\mathbf{F} =$$

[учитывая формулы (11) и (13), далее имеем]

$$= \mathbf{A}\mathbf{Q}^{\alpha e_3\xi+\beta}\mathbf{K}_1\mathbf{G}^{x}\mathbf{J}^{z+t}\mathbf{F} = \mathbf{A}\mathbf{Q}^{\alpha e_3\xi+\beta}\left(\mathbf{K}_1\mathbf{G}^{x}\mathbf{J}^{z}\right)\mathbf{J}^{t}\mathbf{F} =$$

[учитывая (2), (6) и (8), далее имеем]

$$= \mathbf{A}\mathbf{Q}^{\alpha e_3\xi+\beta+1}\mathbf{J}^{t}\mathbf{F} = \mathbf{A}\mathbf{Q}^{k}\mathbf{J}^{t}\mathbf{F} = \mathbf{R}_1;$$

$$\mathbf{R}'_2 = \left(\mathbf{D}\mathbf{G}\mathbf{D}^{-1}\right)^{e_3 s}\left(\mathbf{D}\mathbf{G}^{n}\mathbf{J}^{d}\mathbf{F}\right)\left(\mathbf{F}^{-1}\mathbf{J}\mathbf{F}\right)^{e_4\sigma} \times$$

$$\times\left(\mathbf{F}^{-1}\mathbf{J}^{\varepsilon}\mathbf{K}_2\mathbf{U}^{\lambda}\mathbf{B}^{-1}\right)\left(\mathbf{B}\mathbf{U}^{\tau}\mathbf{B}^{-1}\right)^{e_4\psi} =$$

$$= \mathbf{D}\mathbf{G}^{e_3 s+n}\mathbf{J}^{d+e_4\sigma+\varepsilon}\mathbf{K}_2\mathbf{U}^{\lambda+\tau e_4\psi}\mathbf{B}^{-1} =$$

[с учетом второго уравнения в (10) и второго уравнения в (12) далее имеем]

$$= \mathbf{D}\mathbf{G}^{u+b}\mathbf{J}^{v}\mathbf{K}_2\mathbf{U}^{\lambda+\tau e_4\psi}\mathbf{B}^{-1} =$$

$$= \mathbf{D}\mathbf{G}^{-b}\left(\mathbf{G}^{u}\mathbf{J}^{v}\mathbf{K}_2\right)\mathbf{U}^{\lambda+\tau e_4\psi}\mathbf{B}^{-1} =$$

[учитывая выражения (3), (7) и (9), далее имеем]

$$= \mathbf{D}\mathbf{G}^{b}\left(\mathbf{G}^{u}\mathbf{J}^{v}\mathbf{K}_2\right)\mathbf{U}^{\lambda+\tau e_4\psi}\mathbf{B}^{-1} = \mathbf{D}\mathbf{G}^{b}\mathbf{U}^{1+\lambda+\tau e_4\psi}\mathbf{B}^{-1} =$$

$$= \mathbf{D}\mathbf{G}^{b}\mathbf{U}^{g}\mathbf{B}^{-1} = \mathbf{R}_2.$$

Поскольку для подлинной подписи в ходе процесса верификации имеют место равенства $\mathbf{R}'_1 = \mathbf{R}_1$ и $\mathbf{R}'_2 = \mathbf{R}_2$, то выполняется и равенство $e'_1||e'_2||e'_3||e'_4 = \Phi(M||\mathbf{R}'_1||\mathbf{R}'_2) = \Phi(M||\mathbf{R}_1||\mathbf{R}_2) = e_1||e_2||e_3||e_4$. Последнее означает корректность работы разработанного алгоритма ЭЦП.

Стойкость

Обоснованием использования множителей $\mathbf{K}_1$ и $\mathbf{K}_2$ в формулах (2) и (3), по которым задаются вспомогательные скрытые группы <**Q**> и <**U**>, является устранение возможности генерации ЭЦП с использованием модифицированной процедуры генерации ЭЦП, свободной от связи матриц **Q** и **U** с матрицами **G** и **J**. Действительно, в случае задания вычисления матриц **Q** и **U** порядка rq по формулам $\mathbf{Q} = \mathbf{G}^{x}\mathbf{J}^{z}$ и $\mathbf{U} = \mathbf{G}^{u}\mathbf{J}^{v}$ описанный выше алгоритм генерации ЭЦП имеет силу, однако имеется возможность генерации ЭЦП по альтернативному алгоритму, отличающемуся шагами 3–5, которые имеют следующий модифицированный вид.

3. Вычислить элементы подписи ξ и ψ:

$$\xi = (e_3\alpha)^{-1}(k - \beta) \bmod rq;$$

$$\psi = (e_4\tau)^{-1}(g - \lambda) \bmod rq.$$

4. Найти значения s и n путем решения системы из следующих двух уравнений первой степени:

$$\gamma + e_1 s + n = 0 \bmod q;$$

$$e_3 s + n = b \bmod q,$$

что соответствует выполнению вычислений по формулам

$$s = (e_3 - e_1)^{-1}(b + \gamma) \bmod q;$$

$$n = b - e_3 s \bmod q.$$

5. Найти значения σ и d путем решения системы из следующих двух уравнений первой степени:

$$d + e_2\sigma = t \bmod q;$$

$$d + e_4\sigma + \varepsilon = 0 \bmod q,$$

что соответствует вычислению σ и d по следующим двум формулам:

$$\sigma = (e_2 - e_4)^{-1}(\varepsilon + t) \bmod q;$$

$$d = t - e_2\sigma \bmod q.$$

Подпись, сформированная по альтернативной процедуре генерации ЭЦП, проходит процедуру верификации с проверочными уравнениями (15) и (16) как подлинная подпись. Легко видеть, что в альтернативной процедуре генерации ЭЦП матрицы **Q** и **U** могут быть произвольными, т. е. формулы не накладывают на матрицы **Q** и **U** связь с матрицами **G** и **J**. Это означает, что альтернативная процедура генерации ЭЦП соответствует модифицированной схеме подписи, в которой сохраняются проверочные уравнения (15) и (16) и связь элементов открытого ключа с элементами секретного ключа, задаваемая только формулами (4) и (5), тогда как в исходной схеме ЭЦП уравнения $\mathbf{Q} = \mathbf{G}^{x}\mathbf{J}^{z}$ и $\mathbf{U} = \mathbf{G}^{u}\mathbf{J}^{v}$ также должны учитываться.

Уменьшение числа матричных уравнений с восьми до шести, которые должны быть совместно решены для вычисления секретного ключа, соответствует существенному снижению уровня стойкости (например, с уровня 2^{192} до $\approx 2^{128}$ для случая версии алгоритма ЭЦП, соответствующей значению $\mu = 3$). Таким образом, механизм задания вспомогательных скрытых групп по формулам (2) и (3), в которых $\mathbf{K}_1$ и $\mathbf{K}_2$ являются нескалярными невырожденными матрицами, имеет принципиальное значение для устранения атак, использующих модифицирование процедуры генерации ЭЦП в целях снижения уровня стойкости.

Прямолинейная атака на разработанный алгоритм состоит в вычислении секретного ключа по открытому, что предполагает решение системы из восьми матричных нелинейных уравнений с неизвестными матрицами **A**, **B**, **D**, **F**, **G**, **J**, **Q**, **U**, $\mathbf{K}_1$, $\mathbf{K}_2$ и неизвестными натуральными числами α, β, γ, ε, λ, τ, x, z, u, v, входящими в матричные уравнения в качестве неизвестных степеней. Указанные неизвестные степени имеют разрядность, равную или превышающую 128 бит, что делает практически невыполнимым прямолинейное сведение системы матричных уравнений к системе скалярных уравнений.

Для осуществления такого сведения следует рассмотреть матрицы $\mathbf{Q}^{\alpha}$, $\mathbf{Q}^{\beta}$, $\mathbf{U}^{\tau}$, $\mathbf{U}^{\lambda}$, $\mathbf{G}^{x}$, $\mathbf{G}^{u}$, $\mathbf{U}^{\lambda}$, $\mathbf{U}^{\lambda}$, $\mathbf{J}^{z}$, $\mathbf{J}^{v}$ как матрицы, выбираемые из соответствующих скрытых коммутативных групп <**Q**>, <**U**>, <**G**> и <**J**> и описываемые по элементам матриц-представителей **Q**, **U**, **G** и **J** и уникальным наборам из μ скалярных неизвестных. Элементы матриц **Q**, **U**, **G** и **J** являются скалярными неизвестными, причем каждая из этих неизвестных матриц задает μ^2 независимых скалярных неизвестных. Каждая матрица, задаваемая как некоторая степень одной из матриц-представителей, вносит только μ независимых скалярных неизвестных (см. раздел «Алгебраический носитель»). С учетом этих замечаний получаем некоторую исходную скалярную БССУ, в которой число скалярных неизвестных $\eta_1 = 10\mu^2 + 10\mu$, а число скалярных уравнений $\eta_2 = 8\mu^2$. Предположим, что нам удается зафиксировать значения $2\mu^2 + 10\mu$ скалярных неизвестных таким образом, что получаемая после этого БССУ имеет решения. Тогда вычислительная трудоемкость ее решения, т. е. стойкость W разработанного алгоритма к прямолинейной атаке, может быть оценена по данным работы [4], показывающим зависимость трудоемкости от числа уравнений в случае равенства числа уравнений и неизвестных (см. табл. 1 в работе [4]). С учетом того, что при заданном числе уравнений с ростом порядка поля, в котором задана БССУ, сложность ее решения возрастает, получены оценки стойкости трех версий разработанного алгоритма к прямолинейной атаке (табл. 3).

■ ***Таблица 3.*** Оценка стойкости W к прямолинейной атаке

■ ***Table 3.*** Security evaluation W to the direct attack

Версия алгоритма	η_1	η_2	W
$\mu = 3$	120	72	2^{192}
$\mu = 5$	300	200	$> 2^{256}$
$\mu = 7$	560	392	$> 2^{256}$

Подделка подписи представляет собой атаку, состоящую в вычислении подписи к некоторому документу M. Например, попытку подделки подписи можно осуществить по следующему алгоритму.

1. Сгенерировать произвольные невырожденные матрицы $\mathbf{R}'_1$ и $\mathbf{R}'_2$.

2. Вычислить хеш-значение $(e_1, e_2, e_3, e_4) = \Phi(M||\mathbf{R}'_1||\mathbf{R}'_2)$ и сгенерировать случайные натуральные числа $s < q$, $\sigma < q$, $\xi < rq$ и $\psi < rq$.

3. Систему из двух матричных уравнений, задаваемых формулами (15) и (16), представить в виде системы из $2\mu^2$ линейных скалярных уравнений, в которых неизвестными являются μ^2 элементов матрицы **S**.

4. Найти решение системы уравнений, составленной на шаге 3.

Легко видеть, что система линейных уравнений, формируемая на шаге 3, с пренебрежимо малой вероятностью ($\ll 2^{256}$) является совместной. Действительно, каждая из систем, включающих μ^2 линейных скалярных уравнений, составленных по матричным уравнениям (15) и (16), может быть легко решена. Однако каждое из двух получаемых решений включает $V = \mu^2$ скалярных значений и является случайным ввиду того, что коэффициенты в этих двух системах линейных уравнений являются случайными. Вероятность совпадения рассматриваемых двух решений может быть оценена значением $p^{-V} \approx 2^{-V|p|}$, где $|p|$ — разрядность числа p в битах. Последнее значение определяет значение трудоемкости Λ описанной процедуры подделки подписи: $\Lambda \approx 2^{V|p|} \gg 2^{256}$. Таким образом, по сравнению с алгоритмами ЭЦП, описанными в работах [19, 21, 23], в разработанном алгоритме защищенность от подделки подписи обеспечивается заданием несовместности системы линейных скалярных уравнений, к которым сводятся матричные проверочные уравнения с неизвестной матрицей **S**.

Более эффективным способом подделки подписи является модифицирование известной подписи $(e_1, e_2, e_3, e_4, s, \sigma, \xi, \psi, \mathbf{S})$ таким образом, что модифицированная подпись окажется подлинной для заданного документа M. Это может быть осуществлено по следующему алгоритму.

1. В соответствии с формулами (15) и (16) по известной подписи вычислить матрицы $\mathbf{R}'_1$ и $\mathbf{R}'_2$.

2. По заданному документу и вычисленным матрицам $\mathbf{R}'_1$ и $\mathbf{R}'_2$ вычислить хеш-значение $(e'_1, e'_2, e'_3, e'_4) = \Phi(M||\mathbf{R}'_1||\mathbf{R}'_2)$.

3. Вычислить элементы ξ' и ψ' модифицированной подписи: $\xi' = e_3\xi/e'_3 \bmod rq$ и $\psi' = e_4\psi/e'_4 \bmod rq$.

4. Вычислить элементы s' и σ' модифицированной подписи: $s' = e_1 s/e'_1 \bmod q$ и $\sigma' = e_2\sigma/e'_2 \bmod q$.

5. Если $e'_4 = e_4\sigma/\sigma' \bmod q$ и $e'_3 = e_3 s/s' \bmod q$, то выдать набор значений $(e'_1, e'_2, e'_3, e'_4, s', \sigma', \xi', \psi', \mathbf{S})$ как подлинную подпись к документу M.

В противном случае взять другую известную подпись и перейти к шагу 1 или модифицировать документ M и перейти к шагу 2.

Вероятность выполнения на шаге 5 каждого из двух равенств может быть оценена как $\text{Prob}(e'_4 = e_4\sigma/\sigma' \bmod q) \approx \text{Prob}(e'_3 = e_3 s/s' \bmod q) \approx \approx 2^{-|q|}$. Вероятность одновременного выполнения обоих равенств равна $\approx 2^{-2|q|} \le 2^{256}$. Последнее значение определяет трудоемкость последнего способа подделки подписи, равную $\approx 2^{256}$.

Заметим, что на шаге 4 последнего алгоритма можно вычислить элементы s' и σ' модифицированной подписи по формулам $s' = e_3 s/e'_3 \bmod q$ и $\sigma' = e_4\sigma/e'_4 \bmod q$. Тогда успешная подделка имеет вероятность, равную произведению $\text{Prob}(e'_2 = e_2\sigma/\sigma' \bmod q) \times \text{Prob}(e'_1 = e_1 s/s' \bmod q) \approx 2^{-2|q|}$.

Атака на основе известных подписей. Для разработанного в настоящей статье алгоритма и известных алгоритмов [21–24], в которых матрицы используются в качестве секретных элементов, актуальным является рассмотрение атаки на основе известных подписей. Такую атаку можно отнести к структурным атакам, которые используют особенности построения алгоритма. Данная структурная атака ориентирована на вычисление только части секретного ключа. После того, как часть секретного ключа становится известной, вычислительная трудоемкость решения БССУ, рассматриваемой в рамках прямой атаки, существенно снижается. Для того чтобы атака на основе известных подписей не снижала оценку стойкости к прямой атаке, при разработке алгоритма следует обеспечить стойкость к упомянутой структурной атаке, равную или более высокую по сравнению со стойкостью к прямой атаке.

При практическом применении алгоритмов ЭЦП потенциальный нарушитель имеет доступ к многим различным известным подписям. В структурной атаке может использоваться одна из следующих случайных матриц, связанных с ЭЦП: $\mathbf{S}$, $\mathbf{R}_1$ и $\mathbf{R}_2$. Также можно рассмотреть версии атаки с использованием двух и трех матриц из указанной тройки. Однако рассмотрение всех версий такой структурной атаки приводит к одинаковым оценкам стойкости. Матрица $\mathbf{S}$ является элементом подписи, а матрицы $\mathbf{R}_1$ и $\mathbf{R}_2$ могут быть вычислены с использованием всех элементов подписи по формулам (15) и (16) соответственно. Рассмотрим вариант с использованием матрицы $\mathbf{S}$.

В наборе из m известных подписей выделим подгоночные матрицы $\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_\eta$, каждая из которых вычислена по формуле (14) при фиксированных значениях секретных матриц $\mathbf{D}$ и $\mathbf{F}$ и уникальных неизвестных значениях матриц $(\mathbf{G}_i^n)$ и $(\mathbf{J}_i^d)$, где $i = 1, 2, \ldots, \eta$, выбираемых из коммутативных скрытых групп $<\mathbf{G}>$ и $<\mathbf{J}>$.

Матрицу $(\mathbf{G}_1^n)$ можно взять в качестве матрицы-представителя группы $<\mathbf{G}>$, а матрицу $(\mathbf{J}_1^d)$ – в качестве представителя группы $<\mathbf{J}>$. Тогда выбор каждой из случайных матриц $(\mathbf{G}_i^n)$ и $(\mathbf{J}_i^d)$ для $i = 2, 3, \ldots, \eta$ можно описать как выбор набора из μ случайных скалярных значений. Система степенных уравнений, записанных по формуле (14), может быть сведена к скалярной БССУ, включающей $\eta\mu^2$ уравнений, $4\mu^2$ фиксированных неизвестных (элементы матриц $\mathbf{D}$, $\mathbf{F}$, $(\mathbf{G}_1^n)$ и $(\mathbf{J}_i^n)$) и $2\mu(\eta - 1)$ уникальных скалярных неизвестных. Для получения БССУ с равным числом уравнений и неизвестных требуется некоторое число известных подписей η_0, которое может быть получено как решение следующего уравнения относительно η:

$$\eta\mu^2 = 4\mu^2 + 2\mu(\eta - 1). \qquad (17)$$

Из (17) получаем $\eta_0 = (\mu - 2)^{-1}(4\mu - 2)$. Последнее значение задает скалярную БССУ, содержащую $\mu^2(\mu - 2)^{-1}(4\mu - 2)$ скалярных уравнений. Сложность решения этой БССУ определяет значение стойкости Λ к атаке на основе известных подписей. Для значения $\mu = 3$ число степенных уравнений в БССУ равно $\eta_2 = 90$, а для значений $\mu = 5$ и 7 $\eta_2 = 150$, что с учетом данных [4] дает оценки стойкости $\Lambda > 2^{192}$ (для случая $\mu = 3$) и $\Lambda > 2^{256}$ (для случаев $\mu = 5$ и 7), которые не меняют оценки стойкости W, полученные из рассмотрения прямолинейной атаки и приведенные в табл. 3.

При использовании значений матриц $\mathbf{R}_1$ и $\mathbf{S}$ в рамках атаки на основе известных подписей она реализуется по следующему алгоритму.

1. Для каждой i-й известной подписи ($i = 1, 2, \ldots, \eta$) по проверочному уравнению (15) вычисляется матрица $(\mathbf{R}'_1)_i$, которая в соответствии с формулой (6) равна рандомизирующей матрице $(\mathbf{R}_1)_i$. Индекс i указывает принадлежность соответствующей матрицы i-й известной подписи.

2. С использованием формул (6) и (14) составляется система из 2η матричных уравнений с фиксированными матричными неизвестными $\mathbf{A}$, $\mathbf{D}$, $\mathbf{F}$, $(\mathbf{G}')_1$ и $(\mathbf{J}')_1$ и уникальными матричными неизвестными $\mathbf{G}_1$, $\mathbf{J}_1$, $\mathbf{G}_i$, $\mathbf{J}_i$, $(\mathbf{G}')_i$ и $(\mathbf{J}')_i$, где $i = 2, 3, \ldots, \eta$ и штрихом выделены неизвестные матрицы, относящиеся к рандомизирующей матрице $(\mathbf{R}_1)_i$, т. е. $(\mathbf{R}_1)_i = \mathbf{A}(\mathbf{G}')^k(\mathbf{J}')^t\mathbf{F}$.

3. По системе матричных уравнений, составленной на шаге 2, записывается скалярная БССУ, включающая $2\mu^2\eta$ уравнений в поле $GF(p)$ и $5\mu^2 + 2\mu\eta + 2\mu(\eta - 1)$ скалярных неизвестных, где последние два слагаемых относятся к уникальным матричным неизвестным, каждая из которых выбирается из скрытых циклических групп $<\mathbf{G}>$ и $<\mathbf{J}>$ и описываемых через координаты представителей $(\mathbf{G}')_1$ и $(\mathbf{J}')_1$ этих групп и μ уникальных скалярных неизвестных.

■ ***Таблица 4.*** Сравнение с постквантовыми стандартами
■ ***Table 4.*** Comparison with post-quantum standards

Алгоритм	Уровень стойкости, бит (номер стандарта)	Размер, байт		
		ЭЦП	открытого ключа	секретного ключа
Версия $\mu = 3$	192 (–)	216	432	896
Версия $\mu = 5$	256 (–)	236	600	1168
Версия $\mu = 7$	256 (–)	291	882	1656
ML-DSA-65	192 (FIPS 204)	3293	1952	4032
ML-DSA-87	256 (FIPS 204)	4627	2592	4896
SLH-DSA-192f	192 (FIPS 205)	35 664	48	–
SLH-DSA-256f	256 (FIPS 205)	49 856	64	–

4. Решается БССУ, составленная на шаге 4.

Легко установить, что в решаемой БССУ равенство числа уравнений и неизвестных достигается при числе известных подписей, равном $\eta_0 = (2\mu - 4)^{-1}(5\mu - 2)$. При $\mu = 3$ имеем число степенных уравнений $\eta_2 > 100$, а при $\mu = 5$ и 7 $\eta_2 > 150$, т. е. имеем вычислительную сложность последнего варианта атаки на основе известных подписей не меньше, чем для первого варианта.

Сравнение размеров подписи и открытого ключа в разработанном алгоритме и в постквантовых стандартах НИСТ FIPS 204 и FIPS 205 [25] при уровнях 192-битной и 256-битной стойкости представлено в табл. 4, из которой видно, что первый имеет существенно меньший суммарный размер ЭЦП и открытого ключа.

Заключение

Предложен новый механизм задания вспомогательных скрытых групп в алгебраических схемах ЭЦП, основанных на вычислительной трудности решения БССУ, на основе которого разработан практичный постквантовый алгоритм ЭЦП с использованием алгебр матриц $\mu \times \mu$ в качестве алгебраического носителя. При обеспечении высокого уровня стойкости три версии разработанного алгоритма, соответствующие значениям $\mu = 3, 5, 7$, обладают малыми размерами подписи и открытого ключа, существенно меньшими по сравнению с известными постквантовыми алгоритмами ЭЦП. Предложенный алгоритм может быть использован в качестве прототипа для разработки практичного постквантового стандарта ЭЦП.

Финансовая поддержка

Результаты получены при финансовой поддержке проекта «Технологии противодействия ранее неизвестным квантовым киберугрозам», реализуемого в рамках Государственной программы федеральной территории «Сириус» «Научно-технологическое развитие федеральной территории «Сириус» (соглашение № 23-03 от 27.09.2024 г.).

Благодарность

Авторы признательны анонимному рецензенту за замечания и рекомендации, которые позволили существенно улучшить представление полученных результатов.

Литература

1. **Ding J., Petzoldt A., Schmidt D. S.** *Multivariate Cryptography*. In: *Multivariate Public Key Cryptosystems*. Ser.: *Advances in Information Security*, 2020, vol. 80, pp. 7–23. Springer, New York, NY. doi:10.1007/978-1-0716-0987-3_2
2. **Ikematsu Y., Nakamura S., Takagi T.** Recent progress in the security evaluation of multivariate public-key cryptography. *IET Information Security*, 2022, vol. 17, pp. 1–17. doi:10.1049/ise2.12092
3. **Shuaiting Q., Wenbao H., Yifa L., Luyao J.** Construction of extended multivariate public key cryptosystems. *International Journal of Network Security*, 2016, vol. 18, no. 1, pp. 60–67.
4. **Ding J., Petzoldt A.** Current state of multivariate cryptography. *IEEE Security and Privacy Magazine*, 2017, vol. 15, no. 4, pp. 28–36.
5. **Sumit Debnath, Dheerendra Mishra.** Post-quantum digital signature scheme based on multivariate cubic problem. *Journal of Information Security and Applications*, 2020, vol. 53, no. 1, pp. 1–8. doi:10.1016/j.jisa.2020.102512

6. **Молдовян Д. Н.** Альтернативный способ построения алгоритмов многомерной криптографии. *Вопросы защиты информации*, 2022, № 3, с. 13–21. doi:10.52190/2073-2600_2022_3_13, EDN: VKCAOV
7. **Костина А. А., Морозова Е. В., Молдовян Д. Н.** Параметризуемые способы задания векторных конечных полей для криптоалгоритмов на нелинейных отображениях. *Вопросы защиты информации*, 2024, № 1, с. 3–10. doi:10.52190/2073-2600_2024_1_3, EDN: QAYIUE
8. **Vysotskaya V. V., Chizhov I. V.** The security of the code-based signature scheme based on the Stern identification protocol. *Прикладная дискретная математика*, 2022, № 57, с. 67–90. doi:10.17223/20710410/57/5, EDN: FFRFUH
9. **Ниткин И. С.** Применение метода компактного описания подстановки для модификации схемы цифровой подписи на основе протокола аутентификации Штерна. *Информационно-управляющие системы*, 2025, № 6, с. 51–63. doi:10.31799/1684-8853-2025-6-51-63, EDN: AEXSDC
10. **Baldi M., Battaglioni M., Chiaraluce F., Horlemann A.-L., Persichetti E., Santini P., Weger V.** A new path to code-based signatures via identification schemes with restricted errors. *Advances in Mathematics of Communications*, 2025, no. 19(5), pp. 1360–1381. doi:10.3934/amc.2024058
11. **Feneuil T., Joux A., Rivain M**. Shared permutation for syndrome decoding: new zero-knowledge protocol and code-based signature. *Designs, Codes and Cryptography*, 2023, vol. 91, pp. 563–608. doi:10.1007/s10623-022-01116-1
12. **Feneuil T., Joux A., Rivain M.** Syndrome Decoding in the Head: Shorter Signatures from Zero-Knowledge Proofs. *Lecture Notes in Computer Science*, 2022, vol. 13508, pp. 541–572. Springer, Cham. doi:10.1007/978-3-031-15979-4_19
13. **Carozza E., Couteau G., Joux A**. Short Signatures from Regular Syndrome Decoding in the Head. *Lecture Notes in Computer Science*, 2023, vol. 14008. Springer, Cham. doi:10.1007/978-3-031-30589-4_19
14. **Baldi M., Bitzer S., Pavoni A., Santini P., Wachter-Zeh A., Weger V.** Zero Knowledge Protocols and Signatures from the Restricted Syndrome Decoding Problem. *Lecture Notes in Computer Science*, 2024, vol. 14602, pp. 243–274. Springer, Cham. doi:0.1007/978-3-031-57722-2_8
15. **May D. T., Bac D. T., Minh N. H., Kurysheva A. A., Kostina A. A., Moldovyan D. N.** Signature algorithms on non-commutative algebras over finite fields of characteristic two. *Future Data and Security Engineering. Big Data, Security and Privacy, Smart City and Industry 4.0 Applications (FDSE 2022)*, T. K. Dang, J. Küng, T. M. Chung (eds). *Communications in Computer and Information Science*, 2022, vol. 1688, pp. 273–284. Springer, Singapore. doi:10.1007/978-981-19-8069-5_18
16. **Moldovyan D. N.** A new type of digital signature algorithms with a hidden group. *Computer Science Journal of Moldova*, 2023, vol. 31, no. 1(91), pp. 111–124. doi:10.56415/csjm.v31.06
17. **Курышева А. А.** Способ задания полной рандомизации подписи и алгебраический алгоритм на его основе. *Вопросы защиты информации*, 2024, № 1, с. 38–46. doi:10.52190/2073-2600_2024_1_38, EDN: RPLELR
18. **Молдовян Д. Н., Костина А. А.** Способ усиления рандомизации подписи в алгоритмах ЭЦП на некоммутативных алгебрах. *Вопросы кибербезопасности*, 2024, № 4(62), с. 71–81. doi:10.21681/2311-3456-2024-4-71-81, EDN: RPQEUZ
19. **Молдовян А. А., Молдовян Д. Н., Костина А. А.** Способ усиления рандомизации в постквантовых алгоритмах ЭЦП со скрытой группой. *Информатика и автоматизация*, 2025, т. 24, № 6, с. 1810–1835. doi:10.15622/ia.24.6.9, EDN: WESESL
20. **Duong M. T., Moldovyan A. A., Moldovyan D. N., Nguyen M. H., Do B. T.** Structure of quaternion-type algebras and a post-quantum signature algorithm. *International Journal of Electrical and Computer Engineering (IJECE)*, 2025, vol. 15, no. 3, pp. 2965–2976. doi:10.11591//ijece.v15i3.pp2965-2976
21. **Dinh K. L., Do T. B., Moldovyan N. A., Petrenko A. S.** Typical algebraic signature schemes with two hidden groups. *Future Data and Security Engineering (FDSE 2025)*, T. K. Dang, J. Küng, T. M. Chung (eds). *Communications in Computer and Information Science*, 2026, vol. 2709, pp. 83–99. Springer, Singapore. doi:10.1007/978-981-95-4724-1_7
22. **Dinh K.-L., Nguyen L.-G., Nguyen H.-M., Do T.-B., Moldovyan A. A., Moldovyan D. N., Kostina A. A.** Post-quantum signature algorithm on finite matrix algebras, using three hidden groups. *2025 2nd International Conference on Cryptography and Information Security (VCRIS) Proceedings*, Hanoi, Vietnam, October 30–31, 2025, pp. 1–8. doi:10.1109/VCRIS68011.2025.11250573
23. **Захаров Д. В., Костина А. А., Морозова Е. В., Молдовян Д. Н.** Алгоритм ЭЦП на алгебре матриц 3×3, использующий две скрытые группы. *Вопросы кибербезопасности*, 2025, № 3(67), с. 45–54. doi:10.21681/2311-3456-2025-3-45-54, EDN:QRNBDD
24. **Молдовян А. А., Молдовян Н. А., Молдовян Д. Н., Костина А. А.** Постквантовый алгоритм ЭЦП на основе трудности решения степенных уравнений. *Информатика и автоматизация*, 2026, т. 25, № 3, с. 958–985. doi:10.15622/ia.25.3.11
25. *ML-DSA vs SLH-DSA: Which to Choose*. https://dev.to/quantumsequrity/ml-dsa-vs-slh-dsa-which-to-choose-1c8c?ysclid=mqrutxllbb16437002 (дата обращения: 24.06.2024).

UDC 003.26
doi:10.31799/1684-8853-2026-4-17-27
EDN: GJXNHN

Post-quantum digital signature algorithm with a new mechanism for specifying auxiliary hidden groups

N. A. Moldovyan[a], Dr. Sc., Tech., Professor, Chief Researcher, orcid.org/0000-0002-4483-5048, moldovyan.na@talantiuspeh.ru
A. A. Moldovyan[b], Dr. Sc., Tech., Professor, Chief Researcher, orcid.org/0000-0001-5480-6016, maa1305@yandex.ru
[a]Scientific Center of Information Technologies and Artificial Intelligence of Sirius University, 1, Olimpiyskiy Av., 354340, Krasnodar region, federal territory «Sirius», Russian Federation
[b]St. Petersburg Federal Research Center of the RAS, 39, 14th Line, 199178, Saint-Petersburg, Russian Federation

Introduction: From the point of view of practical application, the known post-quantum digital signature algorithms have a disadvantage of having a relatively large total size of the public key and signature. **Purpose:** To develop a post-quantum signature algorithm with a small total size of the signature and public key at a given level of security. **Method:** The application of hidden commutative groups in the finite-type algebra of square matrices that are defined over a ground finite field GF(p), and the formation of a public key as a set of matrices that are related to secret matrices through power expressions. **Results:** We develop a practical post-quantum algebraic digital signature algorithm with hidden commutative groups. The security of the algorithm is based on the computational difficulty of solving large systems of power equations. We employ a new method for defining auxiliary hidden groups and a new signature verification mechanism using two verification equations, each involving all signature elements. The finite algebras of $\mu \times \mu$ matrices with $\mu = 3$, 5, and 7 are used as the algebraic support, defining three versions of the proposed algorithm. Security estimates against direct attacks, attacks based on known signatures, and signature forgery have been obtained for all versions of the algorithm. **Practical relevance:** The developed algorithm has small public key size and signature size and is a candidate for a practical post-quantum signature scheme.

Keywords — post-quantum cryptography, computer security, digital signature, system of power equations, computationally difficult problem, finite matrix algebra.

For citation: Moldovyan N. A., Moldovyan A. A. Post-quantum digital signature algorithm with a new mechanism for specifying auxiliary hidden groups. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2026, no. 4, pp. 17–27 (In Russian). doi:10.31799/1684-8853-2026-4-17-27, EDN: GJXNHN

References

1. Ding J., Petzoldt A., Schmidt D. S. *Multivariate Cryptography*. In: *Multivariate Public Key Cryptosystems*. Ser.: *Advances in Information Security*, 2020, vol. 80, pp. 7–23. Springer, New York, NY. doi:10.1007/978-1-0716-0987-3_2
2. Ikematsu Y., Nakamura S., Takagi T. Recent progress in the security evaluation of multivariate public-key cryptography. *IET Information Security*, 2022, vol. 17, pp. 1–17. doi:10.1049/ise2.12092
3. Shuaiting Q., Wenbao H., Yifa L., Luyao J. Construction of extended multivariate public key cryptosystems. *International Journal of Network Security*, 2016, vol. 18, no. 1, pp. 60–67.
4. Ding J., Petzoldt A. Current state of multivariate cryptography. *IEEE Security and Privacy Magazine*, 2017, vol. 15, no. 4, pp. 28–36.
5. Sumit Debnath, Dheerendra Mishra. Post-quantum digital signature scheme based on multivariate cubic problem. *Journal of Information Security and Applications*, 2020, vol. 53, no. 1, pp. 1–8. doi:10.1016/j.jisa.2020.102512
6. Moldovyan D. N. Alternative method for developing algorithms of multivariate cryptography. *Information Security Questions*, 2022, no. 3, pp. 13–21 (In Russian). doi:10.52190/2073-2600_2022_3_13, EDN: VKCAOV
7. Kostina A. A., Morozova E. V., Moldovyan D. N. Parameterizable methods for specifying vector finite fields for cryptoalgorithms on nonlinear mappings. *Information Security Questions*, 2024, no. 1, pp. 3–10 (In Russian). doi:10.52190/2073-2600_2024_1_3, EDN: QAYIUE
8. Vysotskaya V. V., Chizhov I. V. The security of the code-based signature scheme based on the Stern identification protocol. *Applied Discrete Mathematics*, 2022, no. 57, pp. 67–90. doi:10.17223/20710410/57/5, EDN: FFRFUH
9. Nitkin I. S. Permutation compact description method and its application for Stern-based digital signature modification. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2025, no. 6, pp. 51–63 (In Russian). doi:10.31799/1684-8853-2025-6-51-63, EDN: AEXSDC
10. Baldi M., Battaglioni M., Chiaraluce F., Horlemann A.-L., Persichetti E., Santini P., Weger V. A new path to code-based signatures via identification schemes with restricted errors. *Advances in Mathematics of Communications*, 2025, no. 19(5), pp. 1360–1381. doi:10.3934/amc.2024058
11. Feneuil T., Joux A., Rivain M. Shared permutation for syndrome decoding: new zero-knowledge protocol and code-based signature. *Designs, Codes and Cryptography*, 2023, vol. 91, pp. 563–608. doi:10.1007/s10623-022-01116-1
12. Feneuil T., Joux A., Rivain M. *Syndrome Decoding in the Head: Shorter Signatures from Zero-Knowledge Proofs*. In: *Lecture Notes in Computer Science*, 2022, vol. 13508, pp. 541–572. Springer, Cham. doi:10.1007/978-3-031-15979-4_19
13. Carozza E., Couteau G., Joux A. *Short Signatures from Regular Syndrome Decoding in the Head*. In: *Lecture Notes in Computer Science*, 2023, vol. 14008. Springer, Cham. doi: 10.1007/978-3-031-30589-4_19
14. Baldi M., Bitzer S., Pavoni A., Santini P., Wachter-Zeh A., Weger V. Zero Knowledge Protocols and Signatures from the Restricted Syndrome Decoding Problem. In: *Lecture Notes in Computer Science*, 2024, vol. 14602, pp. 243–274. Springer, Cham. doi:0.1007/978-3-031-57722-2_8
15. May D. T., Bac D. T., Minh N. H., Kurysheva A. A., Kostina A. A., Moldovyan D. N. Signature algorithms on non-commutative algebras over finite fields of characteristic two. *Future Data and Security Engineering. Big Data, Security and Privacy, Smart City and Industry 4.0 Applications (FDSE 2022)*, T. K. Dang, J. Küng, T. M. Chung (eds). *Communications in Computer and Information Science*, 2022, vol. 1688, pp. 273–284. Springer, Singapore. doi:10.1007/978-981-19-8069-5_18
16. Moldovyan D. N. A new type of digital signature algorithms with a hidden group. *Computer Science Journal of Moldova*, 2023, vol. 31, no. 1(91), pp. 111–124. doi:10.56415/csjm.v31.06
17. Kurysheva A. A. A method for specifying complete randomization of a signature and an algebraic algorithm based on it. *Information Security Questions*, 2024, no. 1, pp. 38–46 (In Russian). doi:10.52190/2073-2600_2024_1_38, EDN: RPLELR
18. Moldovyan D. N., Kostina A. A. A method for strengthening signature randomization in signature algorithms on non-commutative algebras. *Voprosy kiberbezopasnosti*, 2024, no. 4(62), pp. 71–81 (In Russian). doi:10.21681/2311-3456-2024-4-71-81, EDN: RPQEUZ
19. Moldovyan A. A., Moldovyan D. N., Kostina A. A. Randomization in post-quantum digital signature algorithms with a secret group. *Informatics and Automation*, 2025, vol. 24, no. 6, pp. 1810–1835 (In Russian). doi:10.15622/ia.24.6.9, EDN: WESESL

20. Duong M. T., Moldovyan A. A., Moldovyan D. N., Nguyen M. H., Do B. T. Structure of quaternion-type algebras and a post-quantum signature algorithm. *International Journal of Electrical and Computer Engineering (IJECE)*, 2025, vol. 15, no. 3, pp. 2965–2976. doi:10.11591//ijece.v15i3.pp2965-2976
21. Dinh K. L., Do T. B., Moldovyan N. A., Petrenko A. S. Typical algebraic signature schemes with two hidden groups. *Future Data and Security Engineering (FDSE 2025)*, T. K. Dang, J. Küng, T. M. Chung (eds). *Communications in Computer and Information Science*, 2026, vol. 2709, pp. 83–99. Springer, Singapore. doi:10.1007/978-981-95-4724-1_7
22. Dinh K.-L., Nguyen L.-G., Nguyen H.-M., Do T.-B., Moldovyan A. A., Moldovyan D. N., Kostina A. A. Post-quantum signature algorithm on finite matrix algebras, using three hidden groups. *2025 2nd International Conference on Cryptography and Information Security (VCRIS) Proceedings*, Hanoi, Vietnam, October 30–31, 2025, pp. 1–8. doi:10.1109/VCRIS68011.2025.11250573
23. Zakharov D. V., Moldovyan D. N., Kostina A. A., Morozova E. V. A digital signature algorithm on the algebra of 3×3 matrices, which uses two hidden groups. *Voprosy kiberbezopasnosti*, 2025, no. 3(67), pp. 45–54 (In Russian). doi:10.21681/2311-3456-2025-3-45-54, EDN: QRNBDD
24. Moldovyan A. A., Moldovyan N. A., Moldovyan D. N., Kostina A. A. Post-quantum digital signature algorithm based on the difficulty of solving power equations. *Informatics and Automation*, 2026, vol. 25, no. 3, pp. 958–983 (In Russian). doi:10.15622/ia.25.3.11
25. *ML-DSA vs SLH-DSA: Which to Choose*. Available at: https://dev.to/quantumsequrity/ml-dsa-vs-slh-dsa-which-to-choose-1c8c?ysclid=mqrutxllbb16437002 (accesed 24 June 2024).

УВАЖАЕМЫЕ АВТОРЫ!

Научные базы данных, включая Scopus и Web of Science, обрабатывают данные автоматически. С одной стороны, это ускоряет процесс обработки данных, с другой — различия в транслитерации ФИО, неточные данные о месте работы, области научного знания и т. д. приводят к тому, что в базах оказывается несколько авторских страниц для одного и того же человека. В результате для всех по отдельности считаются индексы цитирования, что снижает рейтинг ученого.

Для идентификации авторов в сетях Thomson Reuters проводит регистрацию с присвоением уникального индекса (ID) для каждого из авторов научных публикаций.

Процедура получения ID бесплатна и очень проста, есть возможность провести регистрацию на 12 языках, включая русский (чтобы выбрать язык, кликните на зеленое поле вверху справа на стартовой странице): https://orcid.org

УДК 004.056
doi:10.31799/1684-8853-2026-4-28-40
EDN: HRWBEL

Научные статьи
Articles

Теоретико-игровая модель адаптивного управления защищенностью распределенных IoT/IoRT-систем на основе обнаружения аномалий и прогнозирования уязвимости безопасности

В. А. Липатников[a], *доктор техн. наук, профессор, orcid.org/0000-0002-3736-4743, lipatnikovanl@mail.ru*
В. А. Задбоев[a], *младший научный сотрудник, orcid.org/0009-0003-9362-1307*
Д. В. Макаренко[a], *адъюнкт, orcid.org/0009-0002-4732-2990*
И. А. Андреев[a], *оператор роты (научной), orcid.org/0009-0007-3617-3439*
[a]*Военная академия связи им. Маршала Советского Союза С. М. Буденного, Тихорецкий пр., 3, Санкт-Петербург, 193232, РФ*

Введение: *современные распределенные IoT/IoRT-системы непрерывно функционируют в условиях нестационарного трафика, ограниченных ресурсов периферийных узлов и изменяющихся многоэтапных киберугроз, что требует перехода от реактивных схем контроля к превентивному управлению защищенностью. Особую сложность представляет необходимость сохранять устойчивость функционирования при одновременном воздействии всех указанных факторов. При этом традиционные механизмы контроля не всегда позволяют своевременно учитывать динамику атаки и изменение состояния системы.* ***Цель:*** *разработать теоретико-игровую модель адаптивного управления защищенностью распределенных IoT/IoRT-систем, объединяющую обнаружение аномалий, прогнозирование краткосрочной уязвимости безопасности и превентивный выбор режима контроля при ограниченных ресурсах.* ***Результаты:*** *построена модель конфликта «Злоумышленник–Защитник», в которой решение о выборе режима контроля принимается по минимаксному критерию с учетом текущей аномальности, накопленной аномальности, прогноза уязвимости и остатка ресурса безопасности. Предложенный подход обеспечивает повышение качества управленческих решений в условиях неопределенности действий нарушителя и позволяет рассматривать защиту как непрерывный процесс адаптации к изменяющимся условиям функционирования системы. Экспериментальная проверка на данных N-BaIoT показала, что теоретико-игровая стратегия обеспечивает меньший средний и накопленный ущерб по сравнению со статической, пороговой и риск-ориентированной стратегиями, что подтверждает целесообразность использования адаптивного выбора режима контроля для снижения последствий многоэтапных кибервоздействий.* ***Практическая значимость:*** *возможно построение вычислительно реализуемых средств превентивного управления защищенностью распределенных IoT/IoRT-систем на стандартной аппаратной платформе.*

Ключевые слова *— IoT, IoRT, защищенность, обнаружение аномалий, прогнозирование уязвимости, теоретико-игровая модель, превентивное управление, киберугрозы, многоэтапные атаки, ресурсные ограничения.*

Для цитирования: Липатников В. А., Задбоев В. А., Макаренко Д. В., Андреев И. А. Теоретико-игровая модель адаптивного управления защищенностью распределенных IoT/IoRT-систем на основе обнаружения аномалий и прогнозирования уязвимости безопасности. *Информационно-управляющие системы*, 2026, № 4, с. 28–40. doi:10.31799/1684-8853-2026-4-28-40, EDN: HRWBEL
For citation: Lipatnikov V. A., Zadboev V. A., Makarenko D. V., Andreev I. A. Game-theoretic model of adaptive security management for distributed IoT/IoRT systems based on anomaly detection and security vulnerability forecasting. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2026, no. 4, pp. 28–40 (In Russian). doi:10.31799/1684-8853-2026-4-28-40, EDN: HRWBEL

Введение

Распределенные системы интернета вещей (IoT) и интернета вещей робототехнических комплексов (IoRT) широко применяются в задачах мониторинга, управления и обмена данными в реальном времени. Рост числа периферийных узлов, неоднородность каналов связи и изменение характера сетевого взаимодействия усложняет обеспечение защищенности таких систем, особенно в условиях многоэтапных и изменяющихся атакующих воздействий [1–4].

Для IoT/IoRT-сред характерно сочетание трех факторов: нестационарного трафика, ограниченных вычислительных и энергетических ресурсов периферийных узлов и высокой изменчивости киберугроз. Поэтому традиционные средства контроля безопасности, основанные на статических правилах, сигнатурах и фиксированных порогах, нередко оказываются недостаточно гибкими: они хуже адаптируются к новым типам атак, чувствительны к несбалансированности данных и не всегда обеспечивают своевременную реакцию при развитии сложных сценариев вторжения [5–9].

Особую сложность представляют многоэтапные и коррелированные во времени атаки, при которых вредоносное воздействие формируется не одним событием, а последовательностью взаимосвязанных действий [10–13]. В таких условиях недостаточно фиксировать только текущее отклонение от нормы. Необходимо учитывать накопление признаков угрозы во времени, прогнозировать

изменение состояния безопасности и выбирать режим контроля с упреждением относительно следующей фазы атаки.

Однако трудность состоит в том, что усиление контроля безопасности чаще всего сопровождается более жесткими режимами мониторинга и реагирования, которые в свою очередь повышают вычислительные, коммуникационные и организационные затраты. В то же время попытки ослабить контроль увеличивают вероятность успешной реализации атаки [14–20]. Следовательно, задача выбора режима безопасности имеет конфликтный характер: защитная сторона стремится уменьшить ущерб с минимальными затратами ресурсов, в то время как атакующая сторона стремится увеличить ущерб и перевести систему в неблагоприятное состояние.

Именно в этом смысле рассматриваемая в статье модель является теоретико-игровой. Ее центральный элемент состоит не просто в вычислении показателей аномальности или прогноза уязвимости, а в том, что выбор режима контроля формализуется как задача двух игроков в виде Злоумышленника и Защитника с противоположными целями. Защитник выбирает подходящие режимы контроля безопасности, минимизируя потери в наихудшем допустимом сценарии атакующего воздействия, тогда как Злоумышленник рассматривается как сторона, стремящаяся провести атаку с максимально возможным ущербом. Тем самым обнаружение аномалий и прогнозирование уязвимости в данной работе выступают не самостоятельной целью, а информационной основой для игрового выбора защитного действия. Работы [21–23] используются не как источник вводимых далее формул, а как предметные аналоги активной защиты и сетевого контроля, относительно которых уточняется предлагаемая игровая постановка.

Такой подход позволяет перейти от реактивной схемы к превентивной. Если пороговые и статические методы усиливают контроль уже после заметного роста аномальности, то в предлагаемой модели решение принимается с учетом не только текущего отклонения, но и накопленной аномальности, прогноза краткосрочной уязвимости и ресурсного состояния системы, что в свою очередь позволяет интерпретировать действия Защитника как опережающие в конфликте «Злоумышленник – Защитник».

В связи с этим в статье рассматривается задача построения теоретико-игровой модели адаптивного управления защищенностью, в которой показатели аномальности и прогноза уязвимости используются как информационная основа для выработки оптимальной стратегии контроля безопасности в конфликте «Злоумышленник – Защитник».

Постановка задачи

В настоящей статье вводится самостоятельная система обозначений и соотношений теоретико-игровой модели адаптивного управления защищенностью распределенной IoT/IoRT-системы. Работы [21–23] используются как предметная основа для сопоставления с активной защитой, сетевым контролем и многоэтапными атаками, но математическая постановка, обозначения (табл. 1) и правила выбора режима контроля формируются непосредственно в данной работе.

С учетом конфликтного характера взаимодействия и интегральной оценки риска выбор режима контроля формулируется как минимаксная задача.

Минимаксная постановка используется как робастный вариант выбора режима контроля в условиях неполной информации о текущем намерении нарушителя. При этом Злоумышленник не предполагается обладателем всего возможного арсенала атак на каждом шаге: множество допустимых стратегий формируется как сценарное множество, согласованное с наблюдаемыми признаками, уровнем аномальности, накоплением угрозы и ресурсным состоянием системы:

$$r_t^* = \arg\min_{r_t \in R} \max_{h_t \in H_t} L\left(r_t, h_t, R_t, B_t\right). \qquad (1)$$

В соответствии с выражением (1) на каждом шаге выбирается такой режим контроля, который минимизирует потери Защитника при наиболее неблагоприятном допустимом сценарии воздействия. Такая запись не исключает использования статистической информации: если апостериорные вероятности сценариев атак доступны и достаточно устойчивы, минимаксное правило заменяется байесовской формой [см. (15)]. В базовом варианте статьи выбран минимаксный критерий, поскольку для распределенных IoT/IoRT-систем в переходных эпизодах «нормальный трафик → атакующий трафик» априорное распределение действий нарушителя, как правило, неполно, нестационарно и зависит от текущей конфигурации системы.

Теоретико-игровая модель адаптивного управления защищенностью распределенных IoT/IoRT-систем

Теоретико-игровая модель строится как контур, объединяющий формирование текущей аномальности, накопление признаков угрозы, прогнозирование краткосрочной уязвимости и минимаксный выбор режима контроля безопасности.

■ ***Таблица 1.*** Основные обозначения и базовые соотношения модели
■ ***Table 1.*** Main symbols and basic relations of the model

Обозначение	Описание	Определение/условие
t	Дискретное время функционирования системы	$t \in \{1, 2, \ldots, T\}$
$\mathbf{x}_t$	Вектор наблюдаемых признаков на шаге t	$\mathbf{x}_t = (x_{t1}, x_{t2}, \ldots, x_m)$
A_t	Показатель текущей аномальности	$A_t \in [0, 1]$
$\hat{V}_{t+1}$	Прогнозный показатель уязвимости на следующем шаге	$\hat{V}_{t+1} \in [0, 1]$
R_t	Интегральная оценка риска	$R_t = \alpha A_t + \eta S_t + (1 - \alpha - \eta)\hat{V}_{t+1}$, $\alpha, \eta \geq 0$, $\alpha + \eta \leq 1$
r_t	Режим контроля безопасности, выбираемый Защитником	$r_t \in R = \{R_1, R_2, \ldots, R_k\}$
R_1, R_2, R_3	Режимы контроля безопасности	R_1 – минимальный, R_2 – усиленный, R_3 – критический
B_t	Остаток ресурса безопасности на шаге t	$B_{t+1} = \min(B_{\max}, \max(0, B_t - c(r_t) + u_t))$
$c(r_t)$	Расход ресурса при выбранном режиме	$c(r_t) \geq 0$
u_t	Восстановление или пополнение ресурса	$u_t \geq 0$
B_{crit}	Критический порог ресурса безопасности	При $B_t < B_{crit}$ система переходит в зону ресурсного дефицита
h_t	Сценарий/стратегия атакующего воздействия Злоумышленника	$h_t \in H_t$
H_t	Множество допустимых сценариев/стратегий нарушителя на шаге t	Включает низкоинтенсивные, всплесковые и многоэтапные воздействия
$D_t(r_t, h_t, B_t)$	Ущерб (потери Защитника) на шаге t	Зависит от режима контроля, сценария атаки и остатка ресурса
J	Целевой функционал защитной стороны	$J = \sum_{t=1}^{T} \left(D_t(r_t, h_t, B_t) + \lambda c(r_t) \right)$, $\lambda \geq 0$
$L(\cdot)$	Локальная функция потерь	Учитывает риск, ущерб, стоимость режима и текущее ресурсное состояние системы

Формирование показателя аномальности наблюдаемого поведения

Для описания нормального режима работы системы вводится профиль нормального состояния, т. е. параметризованное описание типичных значений наблюдаемых признаков и допустимых отклонений от них:

$$\mu_t = \beta \mu_{t-1} + (1 - \beta) x_t, \ \beta \in [0, 1]. \quad (2)$$

Для оценки отклонения текущего состояния от профиля нормы используется нормированная ошибка.

Под профилем нормального состояния в работе понимается векторная характеристика штатного поведения системы, включающая сглаженные оценки средних значений признаков и их допустимой изменчивости. Параметр β является коэффициентом памяти профиля: при $\beta > 1$ профиль изменяется медленнее и сильнее учитывает предыдущее состояние, при $\beta > 0$ профиль быстрее адаптируется к новым штатным наблюдениям. Символ μ обозначает оценку среднего значения признака в профиле нормы. Обновление профиля выполняется только для интервалов, не отнесенных к аномальным, что снижает риск включения атакующего поведения в норму:

$$e_t = \frac{1}{m} \sum_{j=1}^{m} \left| x_{tj} - \mu_{tj} \right|. \quad (3)$$

Тогда показатель текущей аномальности определяется как

$$A_t = \min\left(1, \frac{e_t}{\theta_t}\right), \quad (4)$$

где динамический порог нормализации задает допустимый уровень отклонения текущего признакового вектора от профиля нормального состояния.

Значение порога адаптируется по истории нормальных отклонений и используется для приведения ошибки к сопоставимой шкале аномальности.

Для адаптации порога к изменяющимся условиям среды используется правило:

Коэффициент адаптации порога определяет скорость реакции на изменение штатного трафика: малые значения обеспечивают устойчивость к выбросам, большие значения ускоряют перестройку порога при изменении нормального режима. Порог не обновляется по интервалам, для которых текущая аномальность превышает допустимую границу:

$$\theta_t = \gamma\theta_{t-1} + (1-\gamma)e_t,\ \gamma \in [0,1]. \quad (5)$$

Поскольку единичные выбросы не всегда соответствуют реальной атаке, вводится накопленный показатель аномальности:

Коэффициент памяти накопленного показателя задает вклад предшествующих интервалов наблюдения. Поэтому единичный выброс оказывает ограниченное влияние, а последовательность слабых отклонений приводит к росту накопленной аномальности и отражает развитие многоэтапного воздействия:

$$S_t = \rho S_{t-1} + (1-\rho)A_t,\ \ \rho \in [0,1]. \quad (6)$$

В отличие от мгновенной оценки текущей аномальности накопленный показатель аномальности позволяет учитывать последовательные и многоэтапные воздействия, развивающиеся во времени [20–23].

Прогнозирование уязвимости безопасности распределенной IoT/IoRT-системы

Для перехода к превентивному управлению вводится показатель прогнозной уязвимости безопасности $\hat{V}_{t+1}$, формируемый по краткой истории аномальности и ресурсного состояния системы:

$$\hat{V}_{t+1} = \sigma\left(w_1 S_t + w_2 \Delta S_t + w_3\left(1 - \frac{B_t}{B_{\max}}\right) + b\right), \quad (7)$$

где $\sigma(\cdot)$ — логистическая функция; B_t — остаток ресурса безопасности; $B_{\max}$ — максимальный доступный ресурс; w_1, w_2, w_3 — весовые коэффициенты вклада накопленной аномальности, текущего изменения аномальности и ресурсного дефицита; b — свободный параметр модели. Весовые коэффициенты задаются неотрицательными и нормируются при настройке модели, а b используется для смещения границы перехода к высокому прогнозному риску. Здесь $\Delta S_t = S_t - S_{t-1}$ характеризует изменение накопленного уровня признаков угрозы между соседними шагами наблюдения.

В соответствии с выражением (7) прогноз уязвимости возрастает при увеличении накопленной аномальности и с приближением системы к ресурсному дефициту.

Интегральная оценка риска учитывает мгновенное отклонение, накопленную аномальность и краткосрочный прогноз угрозы.

Форма интегральной оценки риска выбрана как агрегирующая функция трех независимых по смыслу компонент: мгновенного отклонения, накопления признаков угрозы и прогноза уязвимости. Такая последовательность введения показателей необходима для того, чтобы затем построить матрицу потерь игры: риск не заменяет игру, а входит в локальную функцию потерь Защитника как один из факторов ожидаемого ущерба:

$$R_t = \alpha A_t + \eta S_t + (1-\alpha-\eta)\hat{V}_{t+1},$$
$$\alpha, \eta \ge 0,\ \ \alpha + \eta \le 1. \quad (8)$$

Теоретико-игровая модель конфликта «Злоумышленник – Защитник»

На каждом шаге t модель Злоумышленника задается сценарным множеством H_t допустимых стратегий воздействия. Это множество включает не весь теоретически возможный арсенал атак, а те сценарии, которые согласуются с текущим состоянием наблюдений: низкоинтенсивное воздействие, всплесковую активность, сканирование, развитие многоэтапной атаки или интенсивное ботнет-воздействие. Защитник выбирает режим контроля безопасности из множества допустимых режимов. Для каждого сочетания «режим контроля — стратегия нарушителя» вводится функция локального ущерба:

$$D_t(r_t, h_t, B_t) = q(h_t)\big(1 - e(r_t)\big)\big(1 + \Psi(B_t)\big), \quad (9)$$

где $q(h_t)$ — базовая интенсивность ущерба для стратегии нарушителя h_t; $e(r_t)$ — защитный эффект выбранного режима контроля; $\Psi(B_t)$ — множитель деградации защиты при снижении доступного ресурса.

Множитель деградации определяется как

$$\Psi_t(B_t) = \begin{cases} 0, & B_t \ge B_{crit}, \\ \delta\left(1 - \dfrac{B_t}{B_{crit}}\right), & B_t < B_{crit}, \end{cases} \quad \delta \ge 0. \quad (10)$$

Стоимость применения режима контроля задается функцией

$$C_t = c(r_t)(1 + \chi R_t),\ \ \chi \ge 0, \quad (11)$$

где $c(r_t)$ — базовая стоимость режима.

Локальная функция потерь защищающейся стороны определяется как

$$L_t(r_t, h_t) = \omega_1 D_t + \omega_2 C_t + \omega_3 M_t,$$
$$\omega_i \geq 0, \sum_{i=1}^{3} \omega_i = 1, \quad (12)$$

где M_t — штраф за частое переключение режимов контроля, отражающий дополнительные организационные и вычислительные потери при необоснованной смене защитного режима.

В простейшем случае

$$M_t = I(r_t \neq r_{t-1}), \quad (13)$$

где $I(\cdot)$ — индикаторная функция.

Правило выбора режима контроля безопасности

При заданных R_t и B_t Защитник выбирает режим по минимаксному критерию:

$$r_t^* = \arg\min_{r_t \in R} \max_{h_t \in H_t} L_t(r_t, h_t). \quad (14)$$

Если для рассматриваемой системы доступна устойчивая статистическая оценка распределения сценариев атак, правило (14) может быть записано в байесовской форме:

$$r_t^* = \arg\min_{r_t \in R} \sum_{h_t \in H_t} \pi(h_t \mid \mathbf{z}_t) L_t(r_t, h_t), \quad (15)$$

где $\pi(h_t|\mathbf{z}_t)$ — апостериорная вероятность сценария атаки h_t при текущем состоянии $\mathbf{z}_t$.

Для практической части используется правило (14), поскольку переходные эпизоды атаки характеризуются неполной и нестационарной информацией о намерении нарушителя. Байесовская форма (15) рассматривается как расширение модели для случая, когда распределение сценариев атак может быть надежно оценено по данным эксплуатации.

После выбора режима ресурс безопасности обновляется по выражению

$$B_{t+1} = \min\left(B_{\max}, \max(0, B_t - c(r_t^*) + u_t)\right), \quad (16)$$

где u_t — частичное восстановление ресурса за шаг. Восстановление выполняется при снижении уровня угрозы и отсутствии необходимости удерживать усиленный или критический режим; пополнение ресурса описывает приток доступного вычислительного, энергетического или коммуникационного бюджета. Верхняя граница задается исходным ресурсом $B_0 = B_{\max}$, поэтому после обновления выполняется ограничение $B_{t+1} \leq B_0$. Нижняя граница исключает отрицательные значения ресурса безопасности.

Выражения (2)–(16) образуют единый контур обнаружения, прогноза, оценки риска и выбора режима контроля безопасности.

Алгоритм реализации и программная схема экспериментального исследования конфликта «Злоумышленник – Защитник»

Общая логика алгоритма

Предлагаемая модель реализуется как дискретный адаптивный цикл, в котором по вектору наблюдений $\mathbf{x}_t$ формируются показатели A_t, S_t, $\hat{V}_{t+1}$, R_t, после чего по минимаксному правилу (14) выбирается режим контроля r_t^* и обновляется ресурс безопасности по формуле (16). Для компактной записи общий алгоритм представим отображением

$$Y_t = \Psi(\mathbf{x}_t, \mathbf{z}_{t-1}, B_t), \quad (17)$$

где $Y_t = \left\{H_t, \hat{V}_{t+1}, R_t, r_t^*, B_{t+1}\right\}$ — множество вычисляемых на шаге t выходных величин; $\mathbf{z}_{t-1} = \{\mu_{t-1}, \theta_{t-1}, S_{t-1}, B_{t-1}, r_{t-1}\}$ — вектор внутреннего состояния подсистемы безопасности.

Пошаговая алгоритмическая процедура

На шаге t вычисления выполняются в следующем порядке. Сначала по входному вектору наблюдений $\mathbf{x}_t$ обновляется адаптивный профиль нормы μ_t по формуле (2). Затем по выражению (3) рассчитывается усредненная ошибка отклонения e_t, на основе которой динамический порог θ_t уточняется согласно формуле (5). После этого текущая аномальность A_t определяется по выражению (4), а накопленный показатель аномальности S_t — по формуле (6). На следующем этапе рассчитывается прогнозный показатель уязвимости $\hat{V}_{t+1}$ в соответствии с выражением (7), а затем интегральный риск R_t формируется по формуле (8). Далее для каждого допустимого режима контроля $r_t \in R$ вычисляются локальный ущерб D_t, стоимость C_t и функция потерь L_t по выражениям (9)–(13). На основе полученных значений выбирается оптимальный режим r_t^* согласно минимаксному критерию (14), после чего ресурс безопасности обновляется по формуле (16).

С учетом указанной последовательности шагов алгоритм выбора режима можно записать в виде

$$r_t^* = \arg\min_{r_t \in R} \max_{h_t \in H_t} \left[\omega_1 D_t(r_t, h_t, B_t) + \omega_2 C_t(r_t, R_t) + \omega_3 M_t(r_t, r_{t-1})\right]. \quad (18)$$

Формула (18) является алгоритмической реализацией теоретико-игрового выбора режима и непосредственно используется в программной модели при пошаговом переборе множества режимов R и сценарного множества допустимых стратегий нарушителя H_t.

Программная схема реализации модели

Для практической проверки предложенного подхода использована модульная программная схема на языке Python, включающая модуль подготовки данных, модуль обнаружения аномалий M_{det}, модуль прогнозирования уязвимости M_{pred}, игровой модуль выбора режима M_{game} и модуль оценки результатов M_{eval}. Общая структура реализации задается последовательностью

$$x_t \to M_{det} \to M_{pred} \to M_{game} \to M_{eval}. \quad (19)$$

Программная реализация имеет линейно-модульную структуру, удобную для поэтапной отладки и для замены отдельных блоков без изменения общей логики метода. Такая схема обеспечивает воспроизводимую реализацию метода на стандартной аппаратной платформе без использования специализированных вычислителей.

Сценарий вычислительного эксперимента

Экспериментальное исследование предлагается проводить в дискретном времени на горизонте T шагов. Для каждого шага используются наблюдаемые данные $\mathbf{x}_t$, значения внутреннего состояния системы и остаток ресурса безопасности B_t. В качестве начальных условий задаются: начальный профиль нормы μ_0, начальный порог θ_0, начальная накопленная аномальность S_0, исходный режим контроля r_0, полный доступный ресурс $B_0 = B_{\max}$, а также параметры α, η, ρ, β, γ, w_1, w_2, w_3, b, χ, ω_1, ω_2, ω_3 и δ. Коэффициенты w_1, w_2, w_3 относятся к прогнозному показателю уязвимости, тогда как ω_1, ω_2, ω_3 задают веса компонент локальной функции потерь.

Полный вычислительный прогон эксперимента можно представить как последовательное выполнение отображения

$$\{x_t, B_t, z_{t-1}\}_{t=1}^{\mathrm{T}} \to \left\{H_t, \hat{V}_{t+1}, R_t, r_t^*, D_t, C_t, B_{t+1}\right\}_{t=1}^{\mathrm{T}}. \quad (20)$$

Из выражения (20) следует, что на каждом шаге сохраняются не только текущие управляющие решения, но и промежуточные величины, необходимые для последующего анализа качества обнаружения, прогноза и управления защищенностью.

Метрики оценки качества обнаружения и управления защищенностью

Для подтверждения заявленной цели статьи метрики должны отражать не только качество обнаружения, но и влияние метода на защищенность и ресурсное состояние системы. В рамках вычислительного эксперимента предлагается использовать следующие показатели.

Средняя величина ущерба, приходящаяся на один шаг функционирования системы:

$$\bar{D} = \frac{1}{T}\sum_{t=1}^{T} D_t. \quad (21)$$

Накопленный ущерб рассматривается как суммарная величина потерь Защитника на всем горизонте наблюдения и непосредственно связан с уровнем поддерживаемой защищенности:

$$D_{\Sigma} = \sum_{t=1}^{T} D_t. \quad (22)$$

Суммарный расход ресурса безопасности, характеризующий общую стоимость реализованной защитной стратегии:

$$C_{\Sigma} = \sum_{t=1}^{T} c\left(r_t^*\right). \quad (23)$$

Доля времени функционирования системы в зоне ресурсного дефицита, которая отражает, насколько часто система переходит в состояние пониженной защищенности из-за истощения доступного ресурса:

$$P_{def} = \frac{1}{T}\sum_{t=1}^{T} I(B_t < B_{crit}). \quad (24)$$

В практической части статьи основное внимание уделяется метрикам, непосредственно отражающим влияние метода на защищенность и ресурсное состояние системы: среднему ущербу, накопленному ущербу, суммарному расходу ресурса и доле времени в зоне ресурсного дефицита.

Экспериментальная часть моделирования адаптивного управления защищенностью распределенных IoT/IoRT-систем

Постановка вычислительного эксперимента

Экспериментальная проверка предложенной модели выполнена на реальных данных сетевого трафика IoT-устройств. В качестве основного источника данных использован датасет N-BaIoT: Data for Network Based Detection of IoT Botnet Attacks, содержащий записи нормального и вредоносного трафика для ряда IoT-устройств (N-BaIoT Dataset to Detect IoT Botnet Attacks. https://www.kaggle.com/datasets/mkashifn/nbaiot-dataset), в том числе при атаках семейства Mirai. В базовом сценарии исследования выбрано устройство Danmini_Doorbell, для которого доступны как записи штатного функционирования, так и данные о сетевой активности при нескольких типах ботнет-атак.

Исходный набор данных был сформирован из файлов атакующего трафика семейства Mirai: ack, scan, syn, udp, udpplain. После объединения подмножеств получен массив объемом 701 648 наблюдений, включающий 49 548 “benign”-наблюдений и 652 100 “attack”-наблюдений.

На этапе предварительной обработки были выделены числовые признаки, удалены константные столбцы и выполнено заполнение пропущенных значений медианами. В дальнейшем использовалось 115 числовых признаков, формирующих вектор наблюдений $\mathbf{x}_t$, введенный в первом разделе.

В работе исходные данные использовались не для прямой бинарной классификации, а для построения последовательного контура обнаружения, прогноза и управления. Сначала по “benign”-подмножеству формировался модуль обнаружения аномалий, вычисляющий показатель текущей аномальности A_t. Далее по формулам (6)–(8) определялись накопленная аномальность S_t, прогноз краткосрочной уязвимости $\hat{V}_{t+1}$ и интегральный риск R_t. Тем самым реальный сетевой трафик преобразовывался в последовательность сигналов, непосредственно используемых в теоретико-игровом модуле выбора режима контроля безопасности.

Для проверки превентивных свойств модели из общего временного ряда были сформированы переходные эпизоды типа «нормальный трафик → атакующий трафик». Такая постановка позволяет анализировать не только поведение системы в установившемся атакующем режиме, но и момент перехода от нормального состояния к развитию угрозы, когда Защитник должен выполнить опережающий выбор режима контроля. Длина каждого эпизода составляла 2000 шагов наблюдения, а всего было сформировано 20 эпизодов, различающихся положением точки начала атаки внутри окна.

В качестве сравниваемых стратегий управления использовались четыре варианта.

1. Статическая стратегия, в которой режим контроля выбирается заранее и не изменяется во времени. Такая стратегия служит нижней базовой линией для оценки адаптивных решений.

2. Пороговая стратегия, в которой режим контроля определяется только текущим значением аномальности и заранее заданных порогов.

Именно такой тип сравнения уже применялся в работах, посвященных активной защите сетей, обработке результатов сетевого контроля и анализу поведения системы при многоэтапных атаках.

3. Риск-ориентированная стратегия, использующая прогнозный показатель риска $\hat{V}_{t+1}$, но не включающая теоретико-игровой механизм выбора. Такая стратегия позволяет отделить вклад самого прогностического блока от вклада теоретико-игрового механизма выбора.

4. Теоретико-игровая стратегия, в которой выбор режима контроля выполняется на основе теоретико-игрового критерия с учетом текущей аномальности, накопленной аномальности, прогноза уязвимости и ресурсного состояния системы.

Во всех вариантах использовалось одно и то же множество режимов контроля

$$r_t \in R = \{R_1, R_2, R_3\}, \tag{25}$$

где R_1 соответствует минимальному режиму контроля, R_2 — усиленному, а R_3 — критическому режиму с максимальной интенсивностью мониторинга и реагирования.

В базовом эксперименте начальный ресурс безопасности B_0 принимался 5000, критический порог дефицита B_{crit} — 1000, а восстановление ресурса на каждом шаге задавалось постоянной величиной. Дополнительно выполнялся стресс-анализ, в котором отношение B_{crit}/B_0 варьировалось в диапазоне от 0,20 до 0,45. В качестве основных критериев эффективности использовались показатели: средний ущерб $\bar{D}$ (21), накопленный ущерб D_Σ (22), суммарный расход ресурса C_Σ (23), доля времени в зоне дефицита P_{def} (24) и число переключений режимов контроля.

Таким образом, вычислительный эксперимент обеспечивает сопоставление реактивных и превентивных стратегий управления защищенностью в конфликте «Злоумышленник – Защитник».

Результаты основного эксперимента на переходных эпизодах типа «нормальный трафик → атакующий трафик»

Эксперимент на переходных эпизодах «нормальный трафик → атакующий трафик» ориентирован на анализ поведения модели в момент перехода от нормального функционирования к развитию вредоносной активности. В такой постановке важно не только зафиксировать отклонение от нормы, но и своевременно усилить режим контроля до полной реализации следующей фазы атаки.

Результаты основного вычислительного эксперимента получены на 20 переходных эпизодах длиной 2000 шагов. На каждом шаге по наблюдаемому трафику вычислялись текущая аномальность A_t, накопленная аномальность S_t и прогноз краткосрочной уязвимости $\hat{V}_{t+1}$. Далее на основе интегрального риска R_t выполнялся выбор режима контроля безопасности по правилу (14) для теоретико-игровой стратегии и по упрощенным правилам для базовых подходов. Эффективность сравнивалась по показателям $\bar{D}$, D_Σ, C_Σ, P_{def} и числу переключений режима.

Сводные результаты основного эксперимента представлены в табл. 2.

Теоретико-игровая стратегия показала наименьший накопленный и средний ущерб среди всех сравниваемых подходов, что указывает на устойчивое преимущество теоретико-игрового механизма выбора режима контроля.

■ ***Таблица 2.*** Результаты основного эксперимента на переходных эпизодах «нормальный трафик → атакующий трафик»
■ ***Table 2.*** Results of the main experiment on benign → attack transition episodes

Стратегия	$\bar{D}$	D_Σ	Остаточный ресурс B_{last}	Минимальный ресурс $B_{\min}$	Число переключений	C_Σ	P_{def}
Теоретико-игровая	0,233	9325,758	4875,227	4427,2	6840	79 434,4	0,0
Риск-ориентированная	0,275	10 989,231	4936,109	4520,0	2357	70 878,4	0,0
Статическая	0,398	15 939,417	5000,000	5000,0	20	56 000,0	0,0
Пороговая	0,401	16 045,683	4942,233	4522,0	2562	53 576,4	0,0

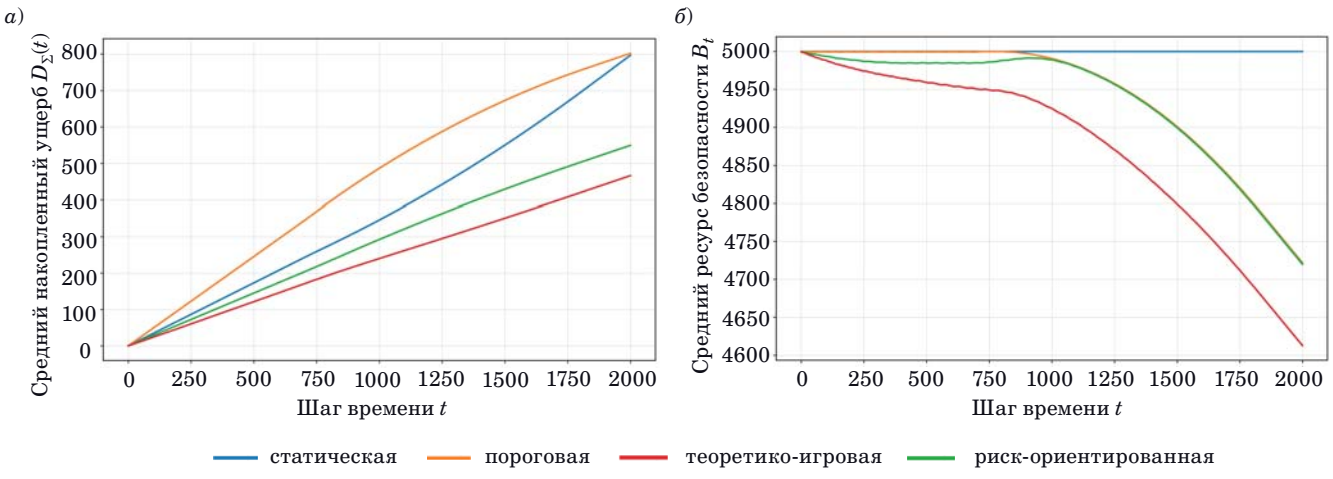
Важно, что теоретико-игровая стратегия превосходит не только статическую и пороговую схемы, но и риск-ориентированную стратегию, уже использующую прогноз $\hat{V}_{t+1}$. Следовательно, выигрыш объясняется не только наличием прогностического блока, но и тем, что прогноз используется в конфликтной постановке «Злоумышленник – Защитник» для превентивного выбора режима контроля.

Для наглядного сопоставления динамики стратегий используется рис. 1, *а*, на котором видно, что кривая теоретико-игровой стратегии располагается ниже остальных на всем протяжении эпизода. Различие с риск-ориентированной стратегией становится устойчивым после начала атакующего участка и сохраняется до завершения эпизода. Статическая и пороговая стратегии накапливают ущерб заметно быстрее, что указывает на их меньшую приспособленность к переходным сценариям развития атаки.

Наряду с уменьшением ущерба важно оценить затраты от этого выигрыша (рис. 1, *б*).

Теоретико-игровая стратегия использует ресурс безопасности наиболее интенсивно, что согласуется с логикой модели: при росте показателей A_t, S_t и $\hat{V}_{t+1}$ Защитник заранее переводит систему в более жесткий режим контроля, уменьшая ожидаемый ущерб ценой более высокого расхода ресурса. В базовом режиме эксперимента это не приводит к переходу в зону дефицита, поскольку для всех подходов выполняется $P_{def} = 0$, однако суммарный расход ресурса у теоретико-игровой стратегии выше, чем у пороговой и статической схем.

Таким образом, основной эксперимент на переходных эпизодах «нормальный трафик → атакующий трафик» показывает, что предлагаемая теоретико-игровая стратегия превентивного управления снижает как средний, так и накопленный ущерб по сравнению со статической, пороговой и риск-ориентированной стратегиями. При этом выигрыш относительно риск-ориентированной стратегии отражает вклад именно игрового механизма выбора режима, а не только эффект использования прогноза уязвимости.



■ ***Рис. 1.*** Усредненные накопленный ущерб $D_\Sigma(t)$ (*а*) и динамика ресурса безопасности B_t (*б*) для сравниваемых стратегий на переходных эпизодах «нормальный трафик → атакующий трафик»
■ ***Fig. 1.*** Average cumulative damage $D_\Sigma(t)$ (*а*) and dynamics of the security resource B_t (*б*) for comparing strategies during transitional episodes benign → attack

Результаты стресс-анализа при варьировании B_{crit}/B_0

Для проверки устойчивости полученных результатов в условиях ужесточения ресурсных ограничений выполнен стресс-анализ, в котором варьировалось отношение критического порога дефицита ресурса к начальному ресурсу безопасности:

$$\frac{B_{crit}}{B_0} \in \{0,20;|0,25;|0,30;|0,35;|0,40;|0,45\}. \quad (26)$$

В отличие от базового эксперимента, где для всех стратегий наблюдалось $P_{def} = 0$, в стрессовом режиме параметры расхода и восстановления ресурса были выбраны так, чтобы система действительно входила в зону дефицита. Это позволило оценить устойчивость сравниваемых стратегий при более жестких ресурсных условиях.

Сводные результаты стресс-анализа представлены в табл. 3. Для сокращения объема в таблицу включены три характерных значения отношения B_{crit}/B_0, отражающие нижнюю, промежуточную и верхнюю области исследованного диапазона.

При увеличении отношения B_{crit}/B_0 все стратегии демонстрируют ожидаемое ухудшение характеристик: возрастают средний ущерб $\bar{D}$, накопленный ущерб D_Σ и доля времени в зоне дефицита ресурса P_{def}. Это означает, что при ужесточении порога ресурсной устойчивости даже кратковременное снижение остатка ресурса начинает сильнее влиять на результат управления защищенностью.

Несмотря на общее ухудшение показателей теоретико-игровая стратегия сохраняет наименьший ущерб на всем исследованном диапазоне. При $B_{crit}/B_0 = 0,20$ средний ущерб для нее составляет около 0,3498, тогда как для риск-ориентированной стратегии — около 0,3934, для пороговой — около 0,4355, а для статической — около 0,5461. При $B_{crit}/B_0 = 0,45$ значения ущерба возрастают у всех стратегий, однако преимущество теоретико-игровой стратегии сохраняется: $\bar{D}$ увеличивается лишь до 0,3577, тогда как у риск-ориентированной — до 0,4224, у пороговой — до 0,4484, а у статической — до 0,5984.

■ ***Таблица 3.*** Результаты стресс-анализа при варьировании отношения B_{crit}/B_0

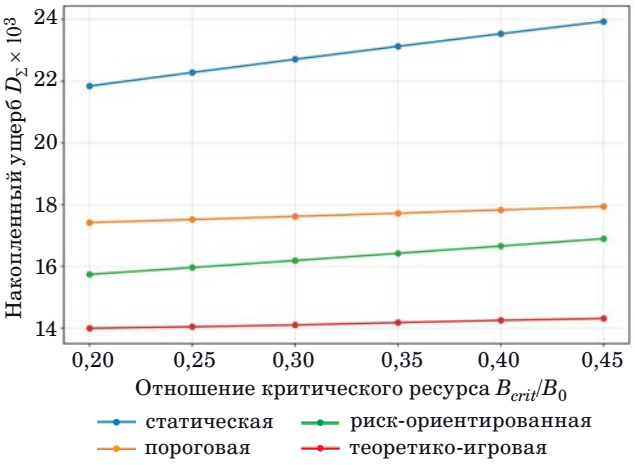
■ ***Table 3.*** Stress-analysis results for varying the ratio B_{crit}/B_0

Стратегия	$\bar{D}$	D_Σ	C_Σ	P_{def}
$B_{crit}/B_0 = 0,20$				
Теоретико-игровая	0,350	13 991,861	110 211,6	0,572
Риск-ориентированная	0,393	15 737,471	94 317,6	0,461
Пороговая	0,435	17 419,163	70 710,8	0,156
Статическая	0,546	21 842,075	72 000,0	0,333
$B_{crit}/B_0 = 0,30$				
Теоретико-игровая	0,352	14 097,891	110 950,8	0,626
Риск-ориентированная	0,405	16 187,750	94 317,6	0,528
Пороговая	0,440	17 616,008	70 710,8	0,195
Статическая	0,567	22 709,425	72 000,0	0,416
$B_{crit}/B_0 = 0,45$				
Теоретико-игровая	0,358	14 308,894	111 820,8	0,707
Риск-ориентированная	0,422	16 894,827	94 317,6	0,631
Пороговая	0,448	17 936,557	70 710,8	0,257
Статическая	0,598	23 935,354	72 000,0	0,541

Наиболее наглядно это проявляется для накопленного ущерба D_Σ (рис. 2).

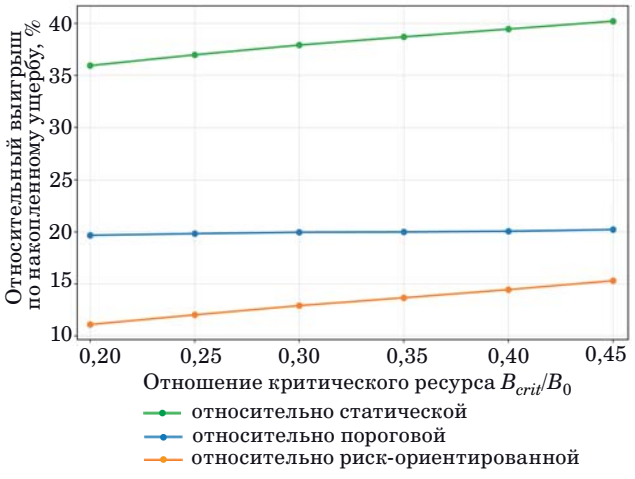
Теоретико-игровая стратегия остается лучшей по накопленному ущербу при всех рассмотренных значениях B_{crit}/B_0. Для количественной интерпретации результата на рис. 3 приведен относительный выигрыш теоретико-игровой стратегии по накопленному ущербу по сравнению с базовыми подходами.

Согласно полученным данным, выигрыш теоретико-игровой стратегии по накопленному ущербу относительно пороговой составляет примерно 19,7–20,2 % на всем исследованном диапазоне B_{crit}/B_0. Относительно риск-ориентированной стратегии выигрыш возрастает примерно от 11,1 до 15,3 %, а относительно статической — от 35,9 до 40,2 %.



■ ***Рис. 2.*** Зависимость накопленного ущерба D_Σ от отношения B_{crit}/B_0

■ ***Fig. 2.*** Dependence of accumulated damage D_Σ on the ratio B_{crit}/B_0



■ ***Рис. 3.*** Относительный выигрыш теоретико-игровой стратегии по накопленному ущербу по сравнению с базовыми подходами

■ ***Fig. 3.*** Relative gain of the game-theoretic strategy in terms of accumulated damage compared to baseline approaches

При этом уменьшение ущерба достигается ценой более интенсивного использования ресурса безопасности, например, для теоретико-игровой стратегии значения C_Σ и P_{def} выше, чем у пороговой и статической. Следовательно, предлагаемую стратегию следует интерпретировать как средство приоритетной минимизации ущерба, а не как универсальный минимум по всем метрикам одновременно.

Таким образом, стресс-анализ подтверждает, что теоретико-игровая стратегия сохраняет преимущество по ущербу в широком диапазоне ресурсных условий и остается устойчивой при ужесточении эксплуатационных ограничений.

Обсуждение результатов

Полученные результаты показывают, что предлагаемая теоретико-игровая модель превентивного управления защищенностью распределенных IoT/IoRT-систем обеспечивает снижение ущерба на переходных сценариях «нормальный трафик → атакующий трафик». В отличие от статической, пороговой и риск-ориентированной стратегий, теоретико-игровая стратегия использует текущее отклонение от профиля нормы, накопленную аномальность, прогноз краткосрочной уязвимости и ресурсное состояние системы, что позволяет Защитнику заранее усиливать режим контроля.

Существенно, что выигрыш достигается не только по сравнению с простыми базовыми подходами, но и по сравнению с риск-ориентированной стратегией, уже использующей прогноз $\hat{V}_{t+1}$. Это означает, что эффект связан не только с наличием прогностического блока, но и с включением прогноза в теоретико-игровой механизм выбора режима. Основными ограничениями исследования остаются использование одного устройства из датасета N-BaIoT, облегченная форма прогнозного блока и дискретное множество режимов контроля.

Заключение

В статье предложена теоретико-игровая модель адаптивного управления защищенностью распределенных IoT/IoRT-систем, объединяющая обнаружение аномалий, прогнозирование краткосрочной уязвимости и превентивный выбор режима контроля безопасности при ограниченных ресурсах периферийных узлов. Экспериментальная проверка на данных N-BaIoT показала, что теоретико-игровая стратегия обеспечивает меньший накопленный ущерб по сравнению со статической, пороговой и риск-ориентированной стратегиями на переходных эпизодах «нормальный трафик → атакующий трафик». Стресс-анализ при варьировании отношения B_{crit}/B_0 подтвердил сохранение этого преимущества при ужесточении ресурсных ограничений. Практическая значимость работы связана с построением воспроизводимой вычислительной схемы, пригодной для реализации на стандартной аппаратной платформе.

Литература

1. **Rose S., Borchert O., Mitchell S., Connelly S.** *Zero Trust Architecture*. Gaithersburg, National Institute of Standards and Technology, 2020. 54 p. doi:10.6028/NIST.SP.800-207
2. **Bast C., Yeh K.-H.** Emerging authentication technologies for zero trust on the Internet of Things. *Symmetry*, 2024, vol. 16, no. 8, art. 993. doi:10.3390/sym16080993
3. **Nie S., Ren J., Wu R., Han P., Han Z., Wan W.** Zero-trust access control mechanism based on blockchain and inner-product encryption in the Internet of Things in a 6G environment. *Sensors*, 2025, vol. 25, no. 2, art. 550. doi:10.3390/s25020550
4. **DeMedeiros K., Hendawi A., Alvarez M.** A survey of AI-based anomaly detection in IoT and sensor networks. *Sensors*, 2023, vol. 23, no. 3, art. 1352. doi:10.3390/s23031352
5. **Belay M. A., Blakseth S. S., Rasheed A., Salvo Rossi P.** Unsupervised anomaly detection for IoT-based multivariate time series: Existing solutions, performance analysis and future directions. *Sensors*, 2023, vol. 23, no. 5, art. 2844. doi:10.3390/s23052844
6. **Aldhaheri A., Alwahedi F., Ferrag M. A., Battah A.** Deep learning for cyber threat detection in IoT networks: A review. *Internet of Things and Cyber-Physical Systems*, 2024, vol. 4, pp. 110–128. doi:10.1016/j.iotcps.2023.09.003
7. **Fan Y., Fu T., Listopad N. I., Liu P., Garg S., Hassan M. M.** Utilizing correlation in space and time: Anomaly detection for Industrial Internet of Things (IIoT) via spatiotemporal gated graph attention network. *Alexandria Engineering Journal*, 2024, vol. 106, pp. 560–570. doi:10.1016/j.aej.2024.08.048
8. **Issa A. S. A., Sonuç E.** CLSTMNet: A Deep learning model for intrusion detection. *Journal of Physics: Conference Series*, 2021, vol. 1973, art. 012244. doi:10.1088/1742-6596/1973/1/012244
9. **Friji H., Mavromatis I., Sanchez-Mompo A., Carnelli P., Olivereau A., Khan A.** Multi-stage attack detection and prediction using graph neural networks: An IoT feasibility study. *2023 IEEE 22nd International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, 2023, pp. 614–621. doi:10.1109/TrustCom60117.2023.00095
10. **Liu W., Gao P., Zhang H., Li K., Yang W., Wei X., Shu J.** Detecting complex multi-step attacks with explainable graph neural network. *arXiv*, 2024, arXiv:2405.11335v2 [cs.CR]. doi:10.48550/arXiv.2405.11335
11. **Perales Gómez A. L., Martínez Beltrán E. T., Sánchez Sánchez P. M., Huertas Celdrán A.** TemporalFED: A software module for detecting cyberattacks in Industry 4.0 time-series data using decentralized federated learning. *arXiv*, 2025, arXiv:2308.03554v2 [cs.CR]. doi:10.48550/arXiv.2308.03554
12. **Danish M.** Enhancing cyber security through predictive analytics: Real-time threat detection and response. *International Journal of Advanced Computer Science and Applications*, 2025, vol. 16, no. 8, pp. 27–35. doi:10.14569/IJACSA.2025.0160804
13. **Jiang Y., Oo N., Meng Q., Lin L., Niyato D., Xiong Z., Lim H. W., Sikdar B.** CyGATE: Game-theoretic cyber attack-defense engine for patch strategy optimization. *arXiv*, 2025. doi:10.48550/arXiv.2508.00478
14. **Datar M., Dujardin Y.** Adaptive learning for moving target defence: Enhancing cybersecurity strategies. *arXiv*, 2025. arXiv:2508.17945v1 [cs.GT].
15. **Spyrou E. D., Kappatos V., Stylios C.** Game-theoretic secure socket transmission with a zero trust model. *Applied Sciences*, 2025, vol. 15, no. 19, art. 10535. doi:10.3390/app151910535
16. **Исаева О. С., Кулясов Н. В., Исаев С. В.** Инфраструктура сбора данных и имитации угроз безопасности сети интернета вещей. *Сибирский аэрокосмический журнал*, 2025, Т. 26, № 1, с. 8–20. doi: 10.31772/2712-8970-2025-26-1-8-20, EDN: OPICJJ
17. **Федоров Н. А., Сычев П. П.** Исследование вопросов информационной безопасности систем «Интернета вещей». *Системный анализ в науке и образовании*, 2023, № 2, с. 27–35. EDN: FGJBSV
18. **Васильев В. И., Гвоздев В. Е., Шамсутдинов Р. Р.** Обнаружение аномалий в системах промышленного Интернета вещей на основе искусственной иммунной системы. *Доклады ТУСУР*, 2021, Т. 24, № 4, с. 40–45. doi: 10.21293/1818-0442-2021-24-4-40-45, EDN: YKHKPJ
19. **Тергеуов О. С., Маликова Ф. У.** Обнаружение и устранение DDoS-атаки IoT-ботнетов на основе SIEM. *Universum: технические науки*, 2022, № 4-1 (97), с. 54–63. EDN: WJNKHL
20. **Уринов Н. Т., Мамадалиев С. С., Бахрамова М. Б., Алижонов Ш. А., Неъматжонов А. У., Акбарова М. Ш.** Определение информативных признаков и формализация базы знаний для идентификации кибератаки Stuxnet на систему Интернета вещей. *Потомки Аль-Фаргани*, 2025, т. 1, вып. 2, с. 80–85. https://doi.org/10.5281/zenodo.15580978. https://al-fargoniy.uz/ (дата обращения: 15.05.2025).
21. **Липатников В. А., Мелехов К. В., Задбоев В. А.** Способ активной защиты информационно-вычислительных сетей от многоэтапных атак. *Сборник трудов международной и межрегиональной конференции «Региональная информатика и информационная безопасность»,* Санкт-Петербург, 1–4 апреля 2024 г. СПб., с. 112–114.
22. **Липатников В. А., Мелехов К. В.** Способ обработки результатов сетевого контроля при поддержке принятия решения администратора безопасности информации. *Тр. XXVII Всерос. науч.-практ. конф. «Актуальные проблемы защиты и безопасности»*, СПб., 2024, *с.* 306–310.
23. **Липатников В. А., Мелихов И. А., Мелехов К. В.** Особенности сетевого контроля в сети передачи данных в условиях многоэтапных атак. *Технологии. Инновации. Связь: сб. матер. науч.-практ. конф.*, СПб., 2023, с. 171–174.

UDC 004.056
doi:10.31799/1684-8853-2026-4-28-40
EDN: HRWBEL

Game-theoretic model of adaptive security management for distributed IoT/IoRT systems based on anomaly detection and security vulnerability forecasting

V. A. Lipatnikov[a], Dr. Sc., Tech., Professor, orcid.org/0000-0002-3736-4743, lipatnikovanl@mail.ru
V. A. Zadboev[a], Junior Researcher, orcid.org/0009-0003-9362-1307
D. V. Makarenko[a], Post-Graduate Student, orcid.org/0009-0002-4732-2990
I. A. Andreev[a], Scientific Company Operator, orcid.org/0009-0007-3617-3439
[a]S. M. Budenny Military Academy of Communication, 3, Tikhoretskii Pr., 190064, Saint-Petersburg, Russian Federation

Introduction: Modern distributed IoT/IoRT systems operate continuously under conditions of non-stationary traffic, limited resources of edge nodes, and evolving multi-stage cyber threats, which necessitates a transition from reactive control schemes to preventive security management. A particular challenge is the need to maintain operational stability under the simultaneous influence of all the above-mentioned factors. At the same time, traditional control mechanisms do not always allow timely consideration of attack dynamics and changes in the system state. **Purpose:** To develop a game-theoretic model for adaptive security management in distributed IoT/IoRT systems that integrates anomaly detection, short-term security vulnerability prediction, and preventive selection of a control mode under resource constraints. **Results:** We construct an "Attacker–Defender" conflict model, in which the decision on selecting a control mode is made according to the minimax criterion, taking into account the current anomaly level, accumulated anomaly level, vulnerability forecast, and the remaining security resource. The proposed approach improves the quality of management decisions under uncertainty regarding the attacker's actions and makes it possible to consider security as a continuous process of adaptation to changing system operating conditions. Experimental validation on the N-BaIoT dataset shows that the game-theoretic strategy provides lower average and cumulative damage as compared to static, threshold-based, and risk-oriented strategies, which confirms the feasibility of using adaptive control mode selection to reduce the consequences of multi-stage cyber impacts. **Practical relevance:** The practical significance lies in the possibility of developing computationally feasible tools for preventive security management of distributed IoT/IoRT systems on a standard hardware platform.

Keywords – IoT, IoRT, security, anomaly detection, vulnerability forecasting, game-theoretic model, preventive control, cyber threats, multi-stage attacks, resource constraints.

For citation: Lipatnikov V. A., Zadboev V. A., Makarenko D. V., Andreev I. A. Game-theoretic model of adaptive security management for distributed IoT/IoRT systems based on anomaly detection and security vulnerability forecasting. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2026, no. 4, pp. 28–40 (In Russian). doi:10.31799/1684-8853-2026-4-28-40, EDN: HRWBEL

References

1. Rose S., Borchert O., Mitchell S., Connelly S. *Zero Trust Architecture*. Gaithersburg, National Institute of Standards and Technology, 2020. 54 p. doi:10.6028/NIST.SP.800-207
2. Bast C., Yeh K.-H. Emerging authentication technologies for zero trust on the Internet of Things. *Symmetry*, 2024, vol. 16, no. 8, art. 993. doi:10.3390/sym16080993
3. Nie S., Ren J., Wu R., Han P., Han Z., Wan W. Zero-trust access control mechanism based on blockchain and inner-product encryption in the Internet of Things in a 6G environment. *Sensors*, 2025, vol. 25, no. 2, art. 550. doi:10.3390/s25020550
4. DeMedeiros K., Hendawi A., Alvarez M. A survey of AI-based anomaly detection in IoT and sensor networks. *Sensors*, 2023, vol. 23, no. 3, art. 1352. doi:10.3390/s23031352
5. Belay M. A., Blakseth S. S., Rasheed A., Salvo Rossi P. Unsupervised anomaly detection for IoT-based multivariate time series: Existing solutions, performance analysis and future directions. *Sensors*, 2023, vol. 23, no. 5, art. 2844. doi:10.3390/s23052844
6. Aldhaheri A., Alwahedi F., Ferrag M. A., Battah A. Deep learning for cyber threat detection in IoT networks: A review. *Internet of Things and Cyber-Physical Systems*, 2024, vol. 4, pp. 110–128. doi:10.1016/j.iotcps.2023.09.003
7. Fan Y., Fu T., Listopad N. I., Liu P., Garg S., Hassan M. M. Utilizing correlation in space and time: Anomaly detection for Industrial Internet of Things (IIoT) via spatiotemporal gated graph attention network. *Alexandria Engineering Journal*, 2024, vol. 106, pp. 560–570. doi:10.1016/j.aej.2024.08.048
8. Issa A. S. A., Sonuç E. CLSTMNet: A Deep learning model for intrusion detection. *Journal of Physics: Conference Series*, 2021, vol. 1973, art. 012244. doi:10.1088/1742-6596/1973/1/012244
9. Friji H., Mavromatis I., Sanchez-Mompo A., Carnelli P., Olivereau A., Khan A. Multi-stage attack detection and prediction using graph neural networks: An IoT feasibility study. *2023 IEEE 22nd International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, 2023, pp. 614–621. doi:10.1109/TrustCom60117.2023. 00095
10. Liu W., Gao P., Zhang H., Li K., Yang W., Wei X., Shu J. Detecting complex multi-step attacks with explainable graph neural network. *arXiv*, 2024, arXiv:2405.11335v2 [cs.CR]. doi:10.48550/arXiv.2405.11335
11. Perales Gómez A. L., Martínez Beltrán E. T., Sánchez Sánchez P. M., Huertas Celdrán A. TemporalFED: A software module for detecting cyberattacks in Industry 4.0 time-series data using decentralized federated learning. *arXiv*, 2025, arXiv:2308.03554v2 [cs.CR]. doi:10.48550/arXiv.2308.03554
12. Danish M. Enhancing cyber security through predictive analytics: Real-time threat detection and response. *International Journal of Advanced Computer Science and Applications*, 2025, vol. 16, no. 8, pp. 27–35. doi:10.14569/IJACSA.2025.0160804
13. Jiang Y., Oo N., Meng Q., Lin L., Niyato D., Xiong Z., Lim H. W., Sikdar B. CyGATE: Game-theoretic cyber attack-defense engine for patch strategy optimization. *arXiv*, 2025. doi:10.48550/arXiv.2508.00478
14. Datar M., Dujardin Y. Adaptive learning for moving target defence: Enhancing cybersecurity strategies. *arXiv*, 2025. arXiv:2508.17945v1 [cs.GT].
15. Spyrou E. D., Kappatos V., Stylios C. Game-theoretic secure socket transmission with a zero trust model. *Applied Sciences*, 2025, vol. 15, no. 19, art. 10535. doi:10.3390/app151910535
16. Isaeva O. S., Kulyasov N. V., Isaev S. V. Infrastructure for collecting data and simulating security threats in the Internet of Things networks. *Siberian Aerospace Journal*, 2025, vol. 26, no. 1, pp. 8–20 (In Russian). doi:10.31772/2712-8970-2025-26-1-8-20, EDN: OPICJJ
17. Fedorov N. A., Sychov P. P. A study of security issues of Internet of Things (IoT) systems. *Sistemnyy analiz v nauke i obrazovanii*, 2023, no. 2, pp. 27–35 (In Russian). EDN: FGJBSV
18. Vasilyev V. I., Gvozdev V. E., Shamsutdinov R. R. Network anomaly detection based on artificial immune system for industrial Internet of Things. *Proceedings of the TUSUR University*, 2021, vol. 24, no. 4, pp. 40–45 (In Russian). doi:10.21293/1818-0442-2021-24-4-40-45, EDN: YKHKPJ

19. Tergeuov O. S., Malikova F. U. Detection and elimination of IoT botnets based on SIEM DDoS attacks. *Universum: tekhnicheskie nauki*, 2022, no. 4-1 (97), pp. 54–63 (In Russian). EDN: WJNKHL
20. Urinov N. T., Mamadaliev S. S., Bakhramova M. B., Alizhonov Sh. A., Ne'matzhonov A. U., Akbarova M. Sh. Definition of informative features and formalization of the knowledge base for identification of the Stuxnet cyber attack on the Internet of Things system. *Al-Farg'onyi avlodlari*, 2025, vol. 1, iss. 2, pp. 80–85. https://doi.org/10.5281/zenodo.15580978. Available at: https://al-fargoniy.uz/ (accessed 15 May 2025) (In Russian).
21. Lipatnikov V. A., Melekhov K. V., Zadboev V. A. Method of active protection of information and computing networks against multi-stage attacks. *Sbornik trudov Sankt-Peterburgskoj mezhdunarodnoj konferencii i Sankt-Peterburgskoj mezhregional'noj konferencii "Regional'naya informatika i informacionnaya bezopasnost'"* [Proc. Saint Petersburg International and Interregional Conf. "Regional Informatics and Information Security"]. Saint Petersburg, 2024, pp. 112–114 (In Russian).
22. Lipatnikov V. A., Melekhov K. V. Method for processing network control results in support of decision-making by an information security administrator. *Trudy XXVII Vserossijskoj nauchnoj-prakticheskoj konferencii "Aktual'nye problemy zashchity i bezopasnosti"* [Proc. XXVII All-Russian Sci. and Pract. Conf. "Current issues of protection and security"]. Saint Petersburg, 2024, pp. 306–310 (In Russian).
23. Lipatnikov V. A., Melikhov I. A., Melekhov K. V. Features of network control in a data transmission network under conditions of multi-stage attacks. *Materialy nauchno-prakticheskoj konferencii "Tekhnologii. Innovacii. Svyaz'"* [Mater. of the Sci. and Pract. Conf. "Technologies. Innovations. Communications"]. Saint Petersburg, 2023, pp. 171–174 (In Russian).

УДК 004.056
doi:10.31799/1684-8853-2026-4-41-51
EDN: HPKSPY

Научные статьи
Articles

Метод послойного декодирования QC-LDPC-кодов с постоянной выходной пропускной способностью (ZLayer)

П. С. Поперечный[а], *канд. техн. наук, ведущий инженер, orcid.org/0009-0008-8127-9854, yncherepop@mail.ru*
В. В. Хрусталев[б], *аспирант, orcid.org/0009-0003-8742-1878*
[а]*ООО «КНС Групп», Рочдельская ул., 15, стр. 15, Москва, 123376, РФ*
[б]*Санкт-Петербургский государственный университет аэрокосмического приборостроения, Б. Морская ул., 67, Санкт-Петербург, 190000, РФ*

Введение: *при использовании QC-LDPC-кодов требования к зависимости пропускной способности аппаратного декодера от длины кода различаются. Например, для Solid State Drive выгодно, чтобы пропускная способность декодера была постоянной для разных длин кода. Однако существующие алгоритмы декодирования QC-LDPC имеют линейную зависимость между размером кода и пропускной способностью.* ***Цель:*** *модифицировать метод декодирования QC-LDPC-кодов таким образом, чтобы он обеспечивал постоянную выходную пропускную способность декодирования.* ***Результаты:*** *предложена модификация метода послойного декодирования, заключающаяся в изменении способа формирования слоев и их количества. Каждый слой сформирован из компонент классических слоев. Количество слоев определяется коэффициентом расширения Z для QC-LDPC-кода, а размер слоя — количеством строк базового графа QC-LDPC-кода. В связи с этим мы назвали метод ZLayer. Увеличение количества слоев с увеличением длины кода позволяет сохранять постоянную выходную пропускную способность аппаратуры, реализующей метод. Предложена микроархитектура аппаратного ZLayer-декодера. Проведенный анализ корректирующей способности предлагаемого метода и декодера показал улучшение по сравнению с классическим послойным декодированием на 0,3 дБ. Представлены оценки корректирующей способности метода при использовании коэффициентов масштабирования и смещения, а также результаты подбора этих коэффициентов для различных параметров QC-LDPC-кода, применяемого в мобильных сетях пятого поколения.* ***Практическая значимость:*** *представленное решение предлагается использовать в системах с QC-LDPC-кодированием, где требуется постоянная пропускная способность. Увеличение корректирующей способности по сравнению с общепринятым решением также положительно скажется на общей эффективности системы с QC-LDPC.*

Ключевые слова *— помехоустойчивое кодирование, коды с малой плотностью проверок на четность, послойное декодирование, аппаратный декодер, QC-LDPC, постоянная пропускная способность.*

Для цитирования: Поперечный П. С., Хрусталев В. В. Метод послойного декодирования QC-LDPC-кодов с постоянной выходной пропускной способностью (ZLayer). *Информационно-управляющие системы*, 2026, № 4, с. 41–51. doi:10.31799/1684-8853-2026-4-41-51, EDN: HPKSPY

For citation: Poperechny P. S., Khrustalev V. V. Method for layer-by-layer decoding of QC-LDPC codes with constant output throughput (ZLayer). *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2026, no. 4, pp. 41–51 (In Russian). doi:10.31799/1684-8853-2026-4-41-51, EDN: HPKSPY

Введение

При передаче и хранении информации возникают ошибки, вызванные внешними помехами и несовершенством микросхем памяти. Эффективным способом борьбы с ошибками является использование помехоустойчивых кодов, которые за счет добавления к передаваемым сообщениям некоторого количества дополнительных символов, называемых контрольными, проверочными или избыточными, позволяют обнаруживать и (или) исправлять ошибки. Для добавления проверочных символов (избыточности) применяются помехоустойчивые кодеры, а для исправления ошибок — помехоустойчивые декодеры [1–5].

В теории помехоустойчивого кодирования существует большое количество различных классов кодов и методов их декодирования. Однако требования современных систем связи и хранения информации, такие как простота декодирования мягких решений и простота поддержки различных длин кодовых слов, существенно ограничивают многообразие помехоустойчивых кодов. Для высокопроизводительных систем связи и хранения информации фактически стандартом стало применение кодов LDPC (Low Density Parity Check), а точнее их подкласса, квазициклических LDPC (QC-LDPC). LDPC-коды применяются в мобильных сетях 3GPP NR 5G, WLAN и WiMAX [6–10], а также в спутниковых системах DVB-S2 и NASA GSFC [11, 12]. Кроме этого, LDPC-коды используются в современных высокопроизводительных твердотельных накопителях (Solid State Drive, SSD) [13–15].

Причинами высокой популярности QC-LDPC-кодов являются следующие их особенности: относительно простая аппаратная реализация QC-LDPC-декодера, возможность обработки мягких решений (декодирование на основании LLR — логарифма отношения правдоподобия), поддержка

одним универсальным аппаратным декодером разных кодов (с разной длиной, задаваемой при помощи коэффициента расширения Z) и возможность частично параллельного декодирования символов.

Высокая популярность QC-LDPC-кодов привела к появлению большого количества исследований и разработок, связанных с реализацией аппаратных высокопроизводительных декодеров. Предлагаемые в работах декодеры ориентированы авторами под определенный стандарт. Следует заметить, что, кроме явных характеристик кода, таких как длина и базовый граф (проверочная матрица), стандарт накладывает требования и на пропускную способность. При этом пропускная способность задается не единым числом (значением), а зависимостями ее значений от длины и скорости кода. Для разных стандартов связи и применений данная зависимость может сильно различаться. Это приводит к тому, что удачная архитектура декодера для одного применения может совершенно не годиться для другого применения. В частности, для хранения данных (SSD-дисков) удобен декодер, пропускная способность которого не зависит от размера страницы памяти, а для многих систем передачи информации требуется линейный рост пропускной способности в зависимости от длины слова.

Например, в работе [16] приводится описание частично параллельного декодера для 5G. Особенностью данной реализации является возможность одновременного декодирования нескольких коротких слов, что позволяет компенсировать падение пропускной способности при декодировании короткого слова. К недостаткам реализации следует отнести высокую сложность, в связи с чем предлагается уменьшить максимальный размер кодового слова, обрабатываемого декодером, в четыре раза по сравнению с поддерживаемым стандартом.

Аналогичный подход описан в работе [17]. Его авторы отмечают, что для коротких кодов при классическом послойном декодировании используется лишь 0,5 % аппаратной части декодера, что приводит к многократному падению пропускной способности при декодировании коротких кодов. Предлагается архитектура с возможностью переиспользовать аппаратные ресурсы для одновременного декодирования нескольких коротких слов. При этом график зависимости пропускной способности от длины кода имеет пилообразный вид (см. рис. 9 в [17]). К недостаткам данной реализации можно отнести двукратную разницу между пропускными способностями при различных длинах кода.

В работах [18, 19] представлены декодеры с послойным декодированием. Пропускная способность при такой реализации линейно зависит от длины кодового слова. Особенностью работы [18] является обработка двух уровней одновременно (когда они ортогональны), а в работе [19] предложен универсальный декодер под различные беспроводные стандарты (WLAN, WiMAX, 5G NR).

В работе [20] предложены четыре декодера с одинаковыми длинами для SSD-дисков. Изменение параметров кода декодерами не поддерживается.

Высокопроизводительная реализация QC-EG-LDPC-декодера для флеш-памяти приведена в работе [21]. Изменение параметров данного декодера также не поддерживается.

Таким образом, анализ перечисленных работ показал отсутствие архитектуры LDPC-декодера с постоянной выходной пропускной способностью, не зависящей от длины кода. Некоторые исследователи пытались реализовать подобный декодер путем переиспользования части аппаратных ресурсов для одновременного декодирования нескольких кодовых слов. Однако при таком подходе график пропускной способности является горизонтальным пилообразным вокруг некоторого среднего значения. Причем возникает двукратная разница между максимальной и минимальной пропускной способностью. Также, как показано в работах, для классического послойного QC-LDPC-декодера разница в пропускных способностях, зависящих от длины кода, может измеряться десятками раз.

В данной работе предлагается декодер QC-LDPC-кода с постоянной выходной пропускной способностью. Такое свойство декодера будет полезно в части практических применений, а именно в SSD-контроллерах, что позволит универсально использовать один контроллер с различными микросхемами памяти или вариативно менять длину кода в страницах памяти для увеличения корректирующей способности. При этом пропускная способность SSD будет сохраняться.

Послойное декодирование QC-LDPC-кодов

Исторически первый алгоритм мягкого декодирования LDPC-кодов предложил их автор Р. Галлагер. Алгоритм получил название распространения доверия (Belief Propagation, BP) [22]. Идея алгоритма заключается в последовательной передаче сообщений о надежностях между проверочными и символьными узлами в графе Таннера. Данный подход оказался возможен благодаря главному свойству LDPC-кодов — низкой плотности проверок на четность в них. Это позволило использовать каждую проверку независимо от остальных, с некоторым допущением.

В исходной версии алгоритма для вычисления апостериорных LLR в проверочных узлах требуется вычисление произведения гиперболических тангенсов и последующее сложение их при переходе к символьным вершинам. Исходя из этих двух операций алгоритм получил название суммы произведений (Sum Product, SP) [22]. Впоследствии для уменьшения вычислительной сложности была предложена аппроксимация произведения гиперболических тангенсов с использованием минимального значения среди входящих LLR в проверочный символ. Алгоритм, основанный на данной аппроксимации, получил название минимума суммы (Min Sum, MS) [23], так как при обработке проверочных вершин вычислялся минимум, а при обработке символьных вершин, как и в исходной версии, — сумма. Дальнейшие модификации алгоритма MS, такие как Normalized, Offset и различные их комбинации, позволили приблизить результат аппроксимации к исходному алгоритму BP.

На практике длины LDPC-кодов достигают тысяч или даже десятков тысяч бит (как, например, в 5G NR). Кроме этого, известно, что особенно эффективны нерегулярные LDPC-коды (те, у которых количество связей между различными проверочными и символьными вершинами многократно различается). В частности, в 5G используется нерегулярный код, что при аппаратной реализации приводит к крайне сложной схеме коммутации между вычислительными узлами и памятью. Для решения этой проблемы был предложен послойный (Layer) метод декодирования LDPC-кодов.

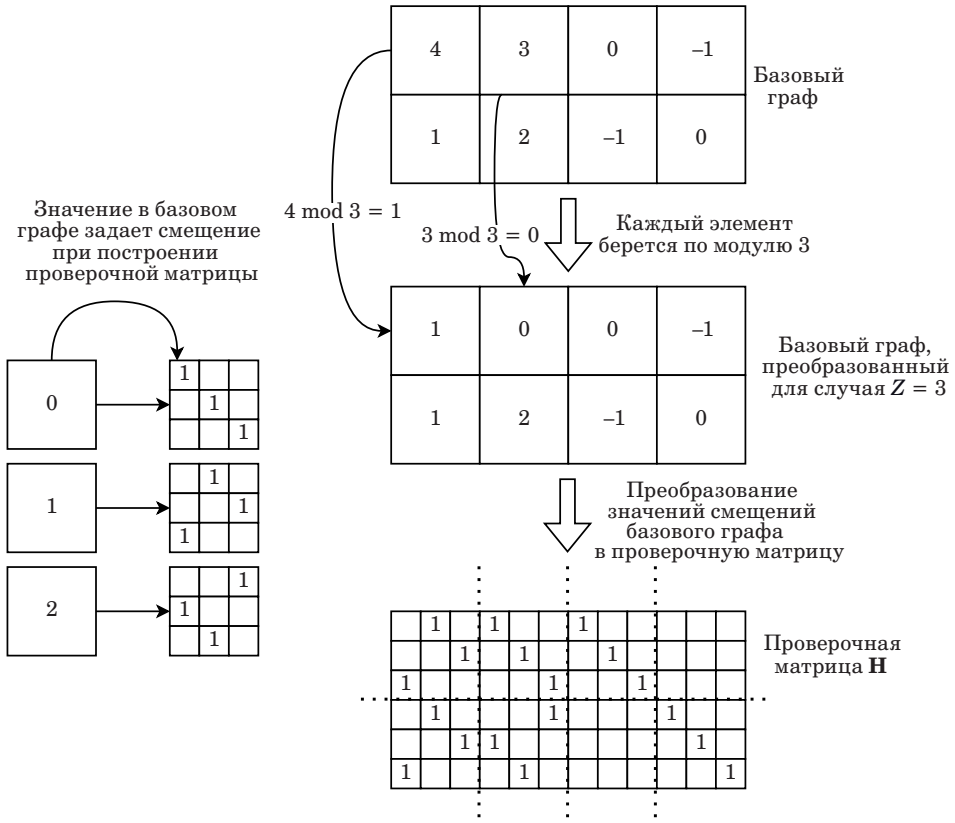
Идея послойного декодирования LDPC заключается в том, что каждая итерация классического BP-алгоритма разбивается на подытерации (слои) [24, 25]. Одной подытерацией (слоем) считается обработка одной строки проверочной матрицы (или базового графа в случае QC-LDPC). Алгоритм послойного декодирования модифицирует вектор апостериорных LLR сразу после обработки слоя, не дожидаясь обработки всей матрицы. Благодаря послойному декодированию в несколько раз увеличивается сходимость алгоритма по сравнению с классическим BP. При этом на сходимость алгоритма и его корректирующую способность начинает влиять порядок обхода слоев. Такое послойное декодирование существенно расслабляет требования к системе коммутации и упрощает аппаратную реализацию, так как достаточно обеспечить связи лишь для одного слоя, а не для всей проверочной матрицы кода целиком.

Важным требованием современных систем связи и хранения информации к помехоустойчивому декодеру является возможность гибкого (программного) изменения размера кода и его скорости. Например, в 5G NR используется 102 разных LDPC-кода, а их длина меняется от сотен до десятков тысяч бит. Для поддержки различных длин на практике применяется квазициклическая модификация LDPC-кодов (QC-LDPC). Основой QC-LDPC-кода является базовый граф (базовая проверочная матрица). Данная матрица состоит из чисел, которые задают циклические сдвиги единичной матрицы некоторого размера Z. Значение Z в QC-LDPC-кодах называется коэффициентом расширения (lifting factor). Значение –1 в базовой матрице задает нулевую матрицу размера Z. Благодаря изменению Z меняется длина кода. В случае если значение циклического сдвига в базовом графе превышает текущее значение Z, то используется остаток от деления значения сдвига на Z. Пример построения проверочной матрицы из базового графа для QC-LDPC-кода показан на рис. 1. На первом этапе базовый граф адаптируется для заданного коэффициента расширения. Для этого циркулянты матрицы (значения сдвига) из базового графа берутся по модулю коэффициента расширения Z (в примере — по модулю 3). На втором этапе выполняется подстановка единичных матриц с заданным смещением. Пустые элементы базового графа (обозначенные –1) заполняются нулевой матрицей $Z \times Z$. Как видно, для QC-LDPC-кода вес каждого столбца одного слоя не превышает 1. Это позволяет гарантированно избегать конфликтов между символами при обработке слоя.

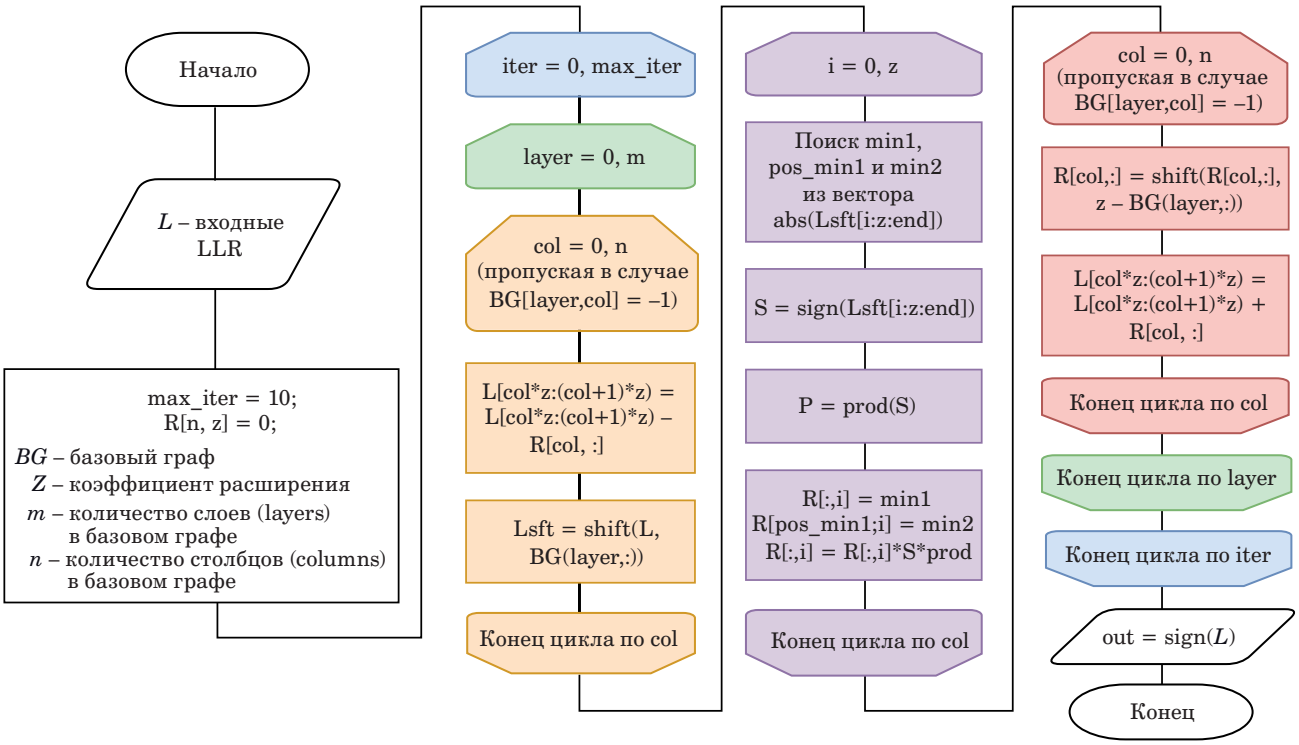
Алгоритм послойного декодирования QC-LDPC-кодов (рис. 2) состоит из двух основных вложенных циклов: внешнего цикла по итерациям (голубой цвет) и внутреннего цикла по слоям (зеленый цвет). При обработке каждого слоя внутри цикла по слоям выполняется три основных действия (каждое в своем отдельном подцикле).

Вычитание из входных в слой LLR (L) значений, на которые были изменены LLR (L) после обработки данного слоя на предыдущей итерации (желтый цвет) (разница, полученная после вычитания, обозначается R (от англ. Residual)). Для удобства дальнейшей обработки данного слоя производится циклический сдвиг элементов базового графа. Операции выполняются в цикле для каждого элемента слоя с единичной матрицей (когда значение сдвига в базовом графе не равно –1).

Выполнение операции min (фиолетовый цвет). Для каждой строки проверочной матрицы из слоя (цикл по Z) производится поиск первого и второго минимального абсолютного значения L. Вычисляется вектор знаков и результат перемножения знаков. Так как в алгоритме BP при расчете надежности символа не учитывается его значение, то первый найденный минимум подставляется для всех элементов, кроме элемента, где этот первый минимум был найден.



■ ***Рис. 1.*** Пример построения проверочной матрицы из базового графа
■ ***Fig. 1.*** An example of constructing a parity-check matrix from a base graph



■ ***Рис. 2.*** Блок-схема классического послойного алгоритма декодирования
■ ***Fig. 2.*** Block diagram of the classical layered decoding algorithm

На позицию первого минимума подставляется второй минимум. Похожее действие производится при формировании знака. Так как знак R должен содержать в себе произведение знаков L, за исключением текущего, то для этого достаточно однократно вычислить произведение всех знаков и умножать это произведение на знак текущего L.

Выполнение операции sum (красный цвет). Производится обратный сдвиг элементов слоя (цикл по элементам слоя с единичными матрицами). После выполнения сдвига значения апостериорных вероятностей для символов модифицируются на рассчитанное на предыдущем этапе значение R.

Жесткий выход декодера соответствует знакам полученных после декодирования апостериорных LLR (L). Дополнительно, на каждой итерации, после декодирования всех слоев может быть добавлен механизм досрочного завершения декодирования, например на основании синдрома.

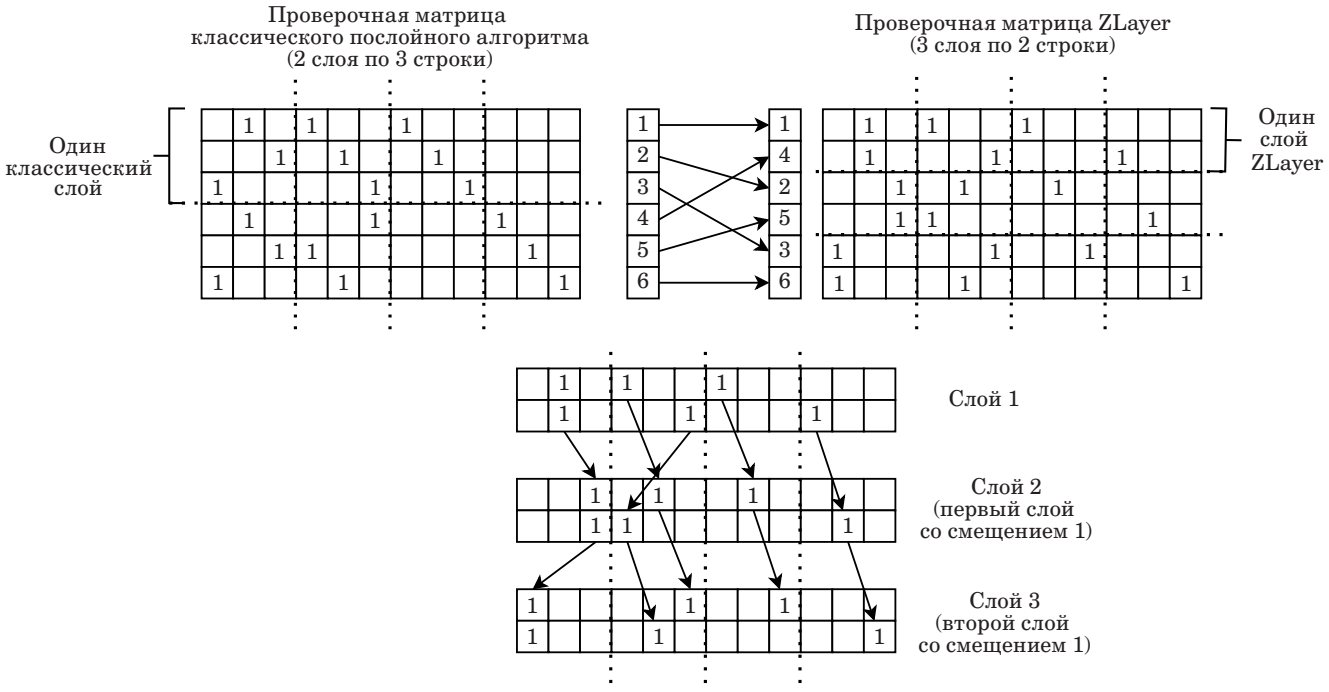
Модифицированный алгоритм послойного декодирования ZLayer

Пропускная способность классического послойного декодера QC-LDPC-кода, как правило, прямо пропорциональна Z. Это связано с тем, что аппаратный декодер параллельно обрабатывает каждую из Z строк слоя, т. е. обрабатывает слой всегда за фиксированное время вне зависимости от Z.

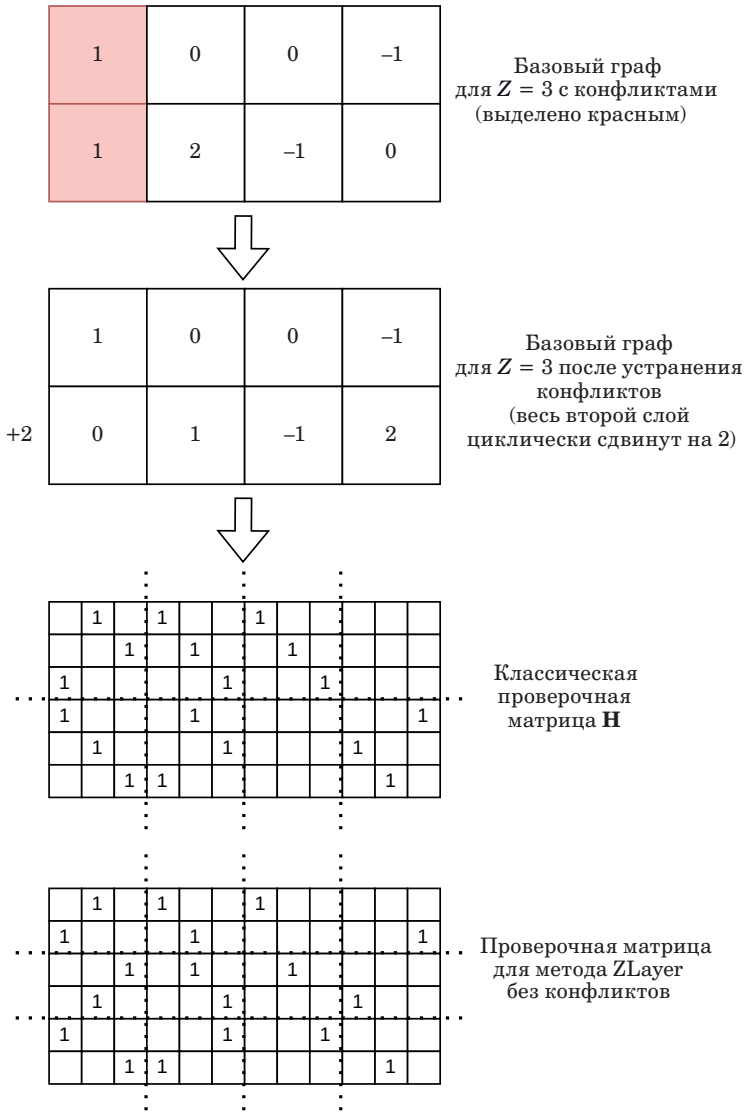
Предлагается модификация классического послойного декодера, заключающаяся в формировании не k слоев (где k — количество строк базовой матрицы) из Z элементов, а Z слоев из k элементов. В связи с тем, что количество слоев меняется и зависит от Z, а время обработки слоя фиксировано, при увеличении Z, т. е. длины кодового слова, будет увеличиваться и время его декодирования. Таким образом, пропускная способность декодера будет сохраняться вне зависимости от Z. В примере предлагаемого ZLayer переупорядочивания (рис. 3) вместо двух классических слоев из трех строк сформировано три ZLayer-слоя по две строки. Первый слой сформирован из первых строк двух исходных слоев, второй слой сформирован из вторых строк двух исходных слоев и т. д. Причем каждый следующий слой ZLayer является циклическим сдвигом предыдущего слоя в секции из Z столбцов (в примере $Z = 3$).

Вес столбца в одном слое теперь может быть больше 1 (см. рис. 3), что будет порождать конфликты при аппаратной параллельной реализации. Решение конфликтов возможно путем циклического сдвига всего исходного слоя на некоторое значение (рис. 4).

Решение конфликтов путем сдвига невозможно, если максимальное количество активных (не –1) элементов в одном из столбцов базового графа больше Z. Однако в SSD-применении коэффициент расширения сильно превышает количество слоев в базовом графе (например, размер базового графа 72×12, а $Z = 140$) [13].



■ ***Рис. 3.*** Формирование слоев в ZLayer
■ ***Fig. 3.*** Forming layers in ZLayer



■ ***Рис. 4.*** Решение конфликтов в ZLayer
■ ***Fig. 4.*** Conflict resolution in ZLayer

Микроархитектура аппаратного ZLayer-декодера

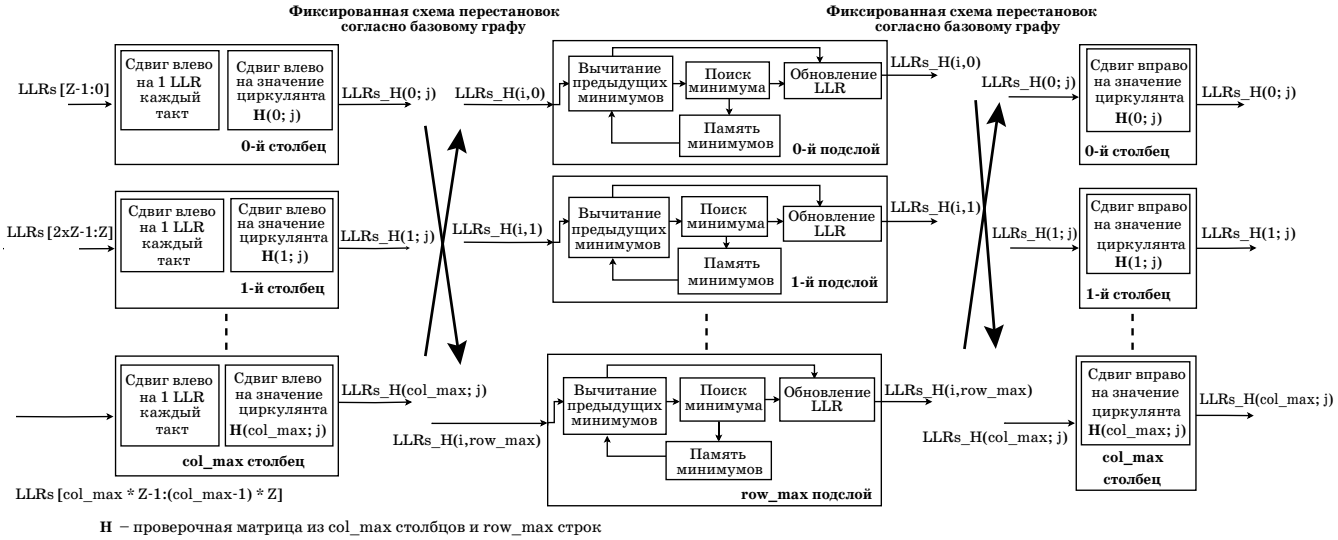
Аппаратная реализация ZLayer-алгоритма выглядит простой, так как для перехода между слоями достаточно лишь циклически сдвигать элементы. При этом для обработки одного слоя необходимо столько вычислительных элементов, сколько слоев в исходном базовом графе. Ввиду того, что число слоев в исходном базовом графе сильно меньше Z, то и вычислительных элементов требуется меньше по сравнению с классической реализацией (например, всего 12 вместо 140 для SSD-применения).

Архитектура ZLayer-декодера для SSD представлена на рис. 5. На вход схемы поступают данные в виде LLR, которые сегментируются согласно заданному Z. Количество сегментов равно количеству столбцов базовой матрицы: col_max. Схема условно делится на три части. Части разделяются между собой схемой прямой и обратной константной перестановки (схема перестановки не меняется в процессе работы). Для переупорядочивания LLR, относящихся к одному элементу базового графа, применяется блок циклического вращения. Вращение производится на величину циркулянта (значение сдвига из базовой матрицы, взятое по модулю Z).

1. Первая часть содержит входной сдвиговый регистр и блоки вращения влево:

— входной сдвиговый регистр — регистр циклического сдвига на один LLR в пределах цикла, равного заданному Z;

— блок вращения влево — регистр циклического сдвига влево на величину циркулянта, заданного в матрице **H** для текущего сегмента Z.



■ ***Рис. 5.*** Архитектура ZLayer-декодера для SSD
■ ***Fig. 5.*** ZLayer decoder architecture for SSD

2. Вторая часть содержит основные вычислительные модули в количестве row_max штук. В этой части выполняются три основные операции алгоритма послойного декодирования (см. рис. 2):

— вычитание предыдущих минимумов: из всех LLR, участвующих в Z подслоях (проверка или строка матрицы **H**), отнимается минимум, вычисленный на предыдущей итерации;

— поиск минимума: среди полученных значений находятся два минимальных по модулю значения и сохраняются в память минимумов;

— обновление LLR: к разницам прибавляются найденные минимумы согласно правилу.

3. Третья часть содержит обратные блоки вращения вправо. Далее модифицированные LLR записываются во входной сдвиговый регистр в соответствующую позицию.

Декодер реализован с использованием конвейера. В начале работы данные записываются в сдвиговые регистры, далее данные двигаются по конвейеру, каждый такт записанные данные в сдвиговом регистре сдвигаются на один LLR. В каждом сегменте данные с нулевого выхода сдвигового регистра сдвигаются влево на константу (циркулянт), заданный в базовой матрице.

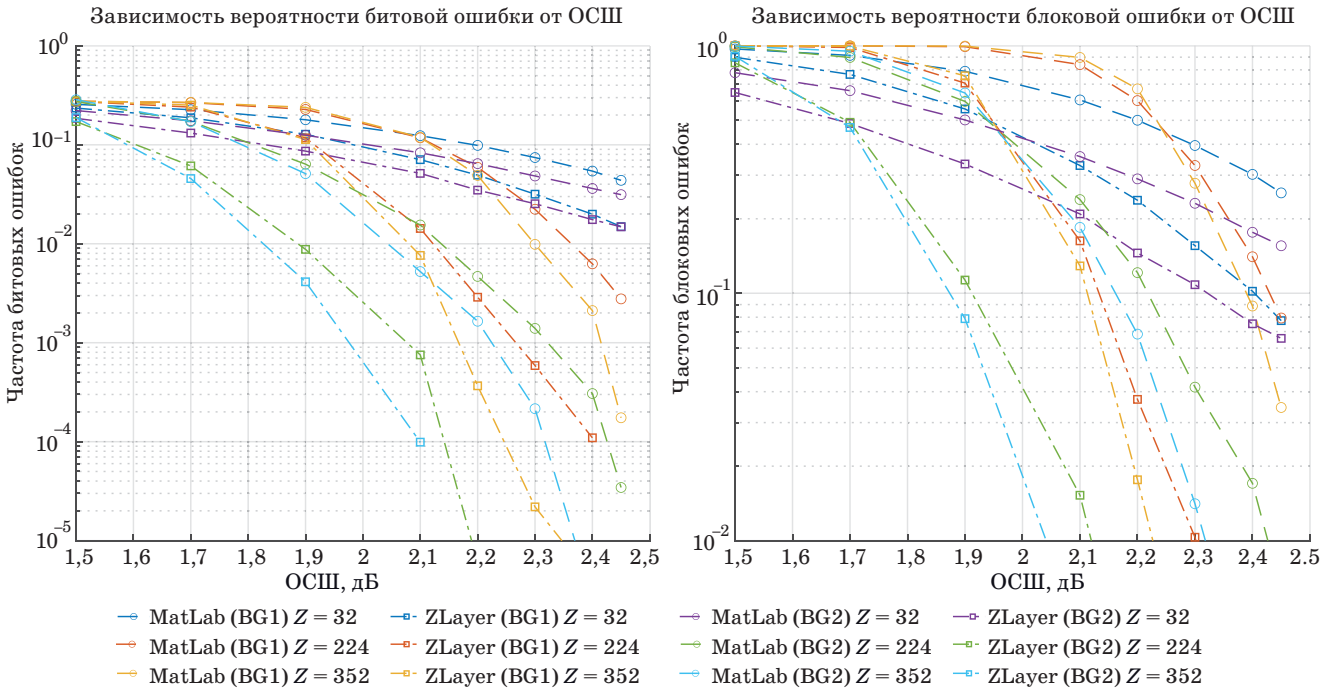
Сформированные на сдвиговых регистрах данные поступают через константную перестановку на схему «вычитание предыдущих минимумов – поиск минимума – обновление LLR». Согласно алгоритму послойного декодирования сначала из входных значений отнимаются соответствующие минимумы, найденные на предыдущем слое, далее находятся новые минимумы, которые записываются в память минимумов. Значения обновляются и поступают на выход второй части.

Результаты оценки корректирующей способности предложенного метода

Известно, что при изменении порядка прохождения по слоям в классическом декодере меняется его исправляющая способность. Для оценки корректирующей способности ZLayer была реализована его программная модель. Моделирование проводилось в среде ПО MatLab на АБГШ-канале. В качестве эталонного решения использовался встроенный в ПО MatLab LDPC-декодер. В качестве стандарта для сравнения декодеров выбран NR 5G. Коэффициенты масштабирования и смещения для обоих декодеров отключены. Графики зависимости количества битовых ошибок и количества пакетных ошибок двух декодеров для различных Z и базовых графов от отношения сигнал/шум (ОСШ) представлены на рис. 6. Эксперименты проведены для всех возможных коэффициентов расширения Z и базовых графов, поддерживаемых стандартом (всего 102 комбинации). Для ZLayer-декодера базовые графы заранее адаптировались для исключения конфликтов. Количество итераций для декодирования — 8.

Как видно на графиках, ZLayer превосходит классический послойный алгоритм декодирования LDPC-кодов. Применение ZLayer-алгоритма дает выигрыш до ~0,3 дБ.

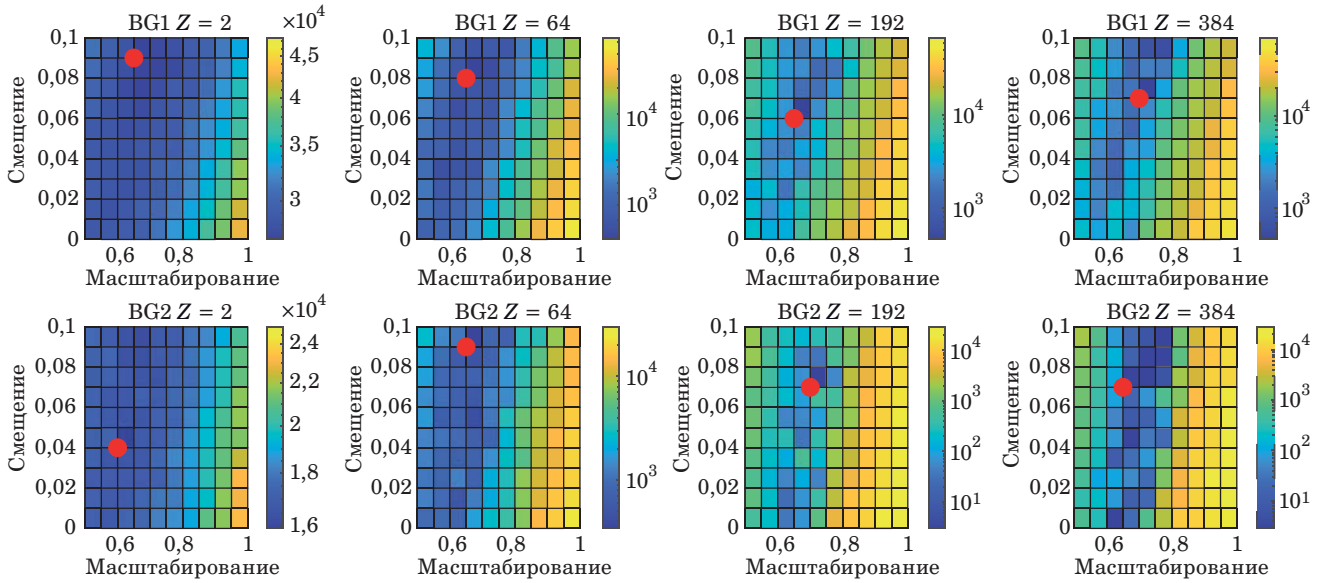
Дополнительно был проведен эксперимент, когда каждый слой разбивался еще на подслой — строки проверочной матрицы, и каждая строка проверочной матрицы обрабатывалась независимо от других и модифицировала апостериорные вероятности. Для классического алгоритма данная модификация несущественна, так как строки проверочной матрицы одного слоя ортогональны.



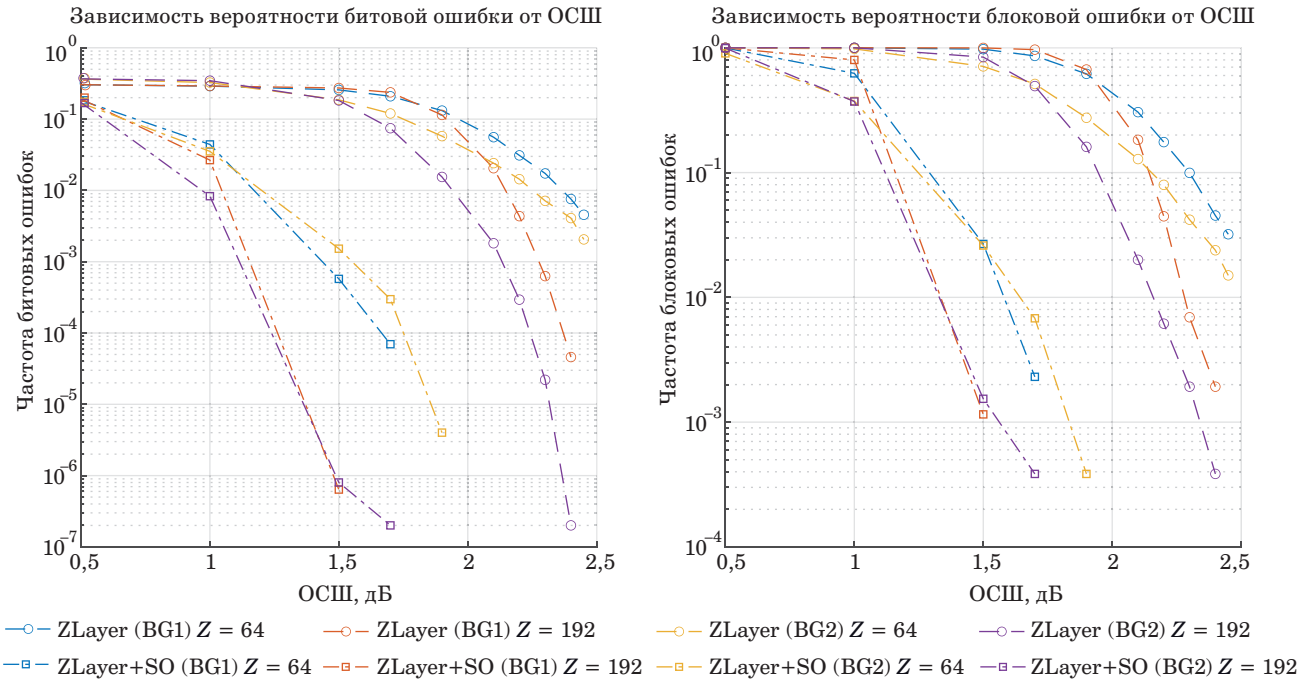
■ ***Рис. 6.*** Сравнение корректирующей способности классического послойного декодера и ZLayer
■ ***Fig. 6.*** Comparison of the correction capability of the classic layer-by-layer decoder and ZLayer

Однако для ZLayer, где могут быть конфликты, данная особенность позволяет получить дополнительный выигрыш до 0,1 дБ. Аппаратная реализация данной модификации ввиду возникших конфликтов доступа в память выглядит затратной.

Дополнительно для алгоритма ZLayer было проведено моделирование в целях подбора оптимальных коэффициентов масштабирования и смещения. Для выбранных Z и базовых графов проведены измерения зависимости количества битовых ошибок от ОСШ. Затем для каждой конфигурации было найдено общее количество ошибок, возникших при моделировании. Полученные данные объединены и представлены в виде тепловой карты (рис. 7). Красными точками отмечены минимальные значения, т. е. значения масштабирования и смещения, при которых для заданной конфигурации было получено наименьшее количество ошибок в процессе моделирования.



■ ***Рис. 7.*** Подбор коэффициентов для ZLayer
■ ***Fig. 7.*** Selecting coefficients for ZLayer



■ ***Рис. 8.*** Улучшение корректирующей способности ZLayer за счет использования коэффициентов масштабирования (0,8) и смещения (0,05)
■ ***Fig. 8.*** Improving ZLayer's correction capabilities to use Scale (0.8) and Offset (0.05)

Как видно, для различных Z и базовых графов оптимальные значения данных коэффициентов отличаются. Результаты оценки корректирующей способности ZLayer-алгоритма с использованием средних значений масштабирования и смещения представлены на рис. 8. Можно заключить, что применение даже некоторых средних значений масштабирования и смещения для ZLayer-алгоритма (ZLayer + SO) существенно улучшает его корректирующую способность.

Заключение

Представленный метод декодирования обладает постоянной пропускной способностью, что необходимо в некоторых применениях LDPC-кодов, в частности SSD-дисках. Аппаратная реализация представленного метода имеет четкую структуру, а количество вычислительных элементов в ней кратно меньше, чем в классической послойной архитектуре. Кроме этого, анализ корректирующей способности ZLayer-декодера показал улучшения по сравнению с классическим на ~0,3 дБ. Дополнительного существенного улучшения корректирующей способности предложенного метода можно добиться тонким подбором коэффициентов масштабирования и смещения для конкретных конфигураций, отношений сигнал/шум и номеров итераций.

Литература

1. **Ovchinnikov A. A., Veresova A. M., Fominykh A. A.** Decoding of linear codes for single error bursts correction based on the determination of certain events. *Информационно-управляющие системы*, 2022, № 6, с. 41–52. doi:10.31799/1684-8853-2022-6-41-52, EDN: UWXZHN
2. **Li Y.** Analysis and methodological advancements in software-defined error correction codes. *2024 4th International Signal Processing, Communications and Engineering Management Conference (ISPCEM)*, Montreal, QC, Canada, 2024, pp. 133–138. doi:10.1109/ISPCEM64498.2024.00030
3. **Ali M. M., Hashim S. J., Chaudhary M. A., Ferré G., Rokhani F. Z., Ahmad Z.** A reviewing approach to analyze the advancements of error detection and correction codes in channel coding with emphasis on LPWAN and IoT systems. *IEEE Access*, 2023, vol. 11, pp. 127077–127097. doi:10.1109/ACCESS.2023.3331417
4. **Хрусталев В. В.** Оценка характеристик открытых декодеров БЧХ и РС кодов, реализованных на современных FPGA. *Моделирование, оптимизация и информационные технологии*, 2026, т. 14, № 3. https://moitvivt.ru/ru/journal/article?id = 2174 (дата обращения: 08.04.2026). doi:10.26102/2310-6018/2026.54.3.005, EDN: CBXXSV
5. **Garani S. S., Nadkarni P. J., Raina A.** Theory behind quantum error correcting codes: An overview. *J Indian Inst Sci*, 2023, vol. 103, pp. 449–495. doi:10.1007/s41745-023-00392-7

6. **Lin C.-H., Wang C.-X., Lu C.-K.** LDPC decoder design using compensation scheme of group comparison for 5G communication systems. *Electronics*, 2021, vol. 10 (16), 2010. doi:10.3390/electronics10162010
7. **Gupta S. H., Virmani B.** LDPC for Wi-Fi and WiMAX technologies. *2009 International Conference on Emerging Trends in Electronic and Photonic Devices & Systems*, Varanasi, India, 2009, pp. 262–265. doi:10.1109/ELECTRO.2009.5441120
8. **Чжан В., Газизов Т. Р.** LDPC и полярные коды в 6G: сравнительное исследование и унифицированные фреймворки. *Труды учебных заведений связи*, 2025, т. 11, № 5, с. 42–59. doi:10.31854/1813-324X-2025-11-5-42-59, EDN: WXEIHS
9. **Portnoy S., Tikhonyuk A., Nikitin S., Voloshin A.** Performance evaluation of LDPC codes in 5G NR. *2024 Systems of Signal Synchronization, Generating and Processing in Telecommunications (SYNCHROINFO)*, Vyborg, Russian Federation, 2024, pp. 1–9. doi:10.1109/SYNCHROINFO61835.2024.10617595
10. **Ovchinnikov A. A., Fominykh A. A.** Decoding of LDPC codes for 5G standard using source distribution. *2020 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF)*, St. Petersburg, Russia, 2020, pp. 1–5. doi:10.1109/WECONF48837.2020.9131454
11. **He Y., Wang Y.** Research on Low-Density Parity-Check decoding algorithm for reliable transmission in satellite communications. *Applied Sciences*, 2024, vol. 14, iss. 2, 746. doi:10.3390/app14020746
12. **Зинченко М. Ю., Левадний А. М., Гребенко Ю. А.** Реализация LDPC декодера на ПЛИС и оптимизация потребляемой мощности. T-Comm: *Телекоммуникации и транспорт*, 2020, т. 14, № 3, с. 4–10. doi:10.36724/2072-8735-2020-14-3-4-10, EDN: XUZAWE
13. *IEEE Standard for Error Correction Coding of Flash Memory Using Low-Density Parity Check Codes*. IEEE Std 1890–2018, 2019, vol., no., pp. 1–51. doi:10.1109/IEEESTD.2019.8654228
14. **Mo J., Zhai X., Yu M., Fang Y., Han G.** SC-GC-LDPC for NAND flash memory: Construction and decoding. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2025, vol. 44, no. 11, pp. 4167–4180. doi:10.1109/TCAD.2025.3563802
15. **Song Y., Lv Y., Li W., Liu J., Shi L.** Revisiting multiple ECC on high-density NAND flash memory. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 2025, vol. 33, no. 6, pp. 1679–1692. doi:10.1109/TVLSI.2025.3551400
16. **Pourjabar S., Choi G. S.** A high-throughput multi-mode low-density parity-check decoder for 5G new radio. *Int J Circ Theor Appl*, 2022, vol. 50, iss. 4, pp. 1365–1374. doi:10.1002/cta.3208
17. **Nadal J., Baghdadi A.** Parallel and flexible 5G LDPC decoder architecture targeting FPGA. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 2021, vol. 29, iss. 6, pp. 1141–1151. doi:10.1109/TVLSI.2021.3072866
18. **Cui H., Ghaffari F., Le K., Declercq D., Lin J., Wang Z.** Design of high-performance and area-efficient decoder for 5G LDPC codes. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 2021, vol. 68, iss. 2, pp. 879–891. doi:10.1109/TCSI.2020.3038887
19. **Chen W., Li Y., Liu D.** High-throughput LDPC decoder for multiple wireless standards. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 2025, vol. 72, iss. 1, pp. 383–396. doi:10.1109/TCSI.2024.3419425
20. **Lu Q., Sham C.-W., Lau F. C. M.** On using the cyclically-coupled QC-LDPC codes in future SSDs. *2016 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)*, Jeju, Korea (South), 2016, pp. 625–628. doi:10.1109/APCCAS.2016.7804048
21. **Liu L.-W., Yuan M.-H., Liao Y.-C., Chang H.-C.** A 38.64-Gb/s Large-CPM 2-KB LDPC decoder implementation for NAND flash memories. *IEEE Open Journal of Circuits and Systems*, 2022, vol. 3, pp. 180–191. doi:10.1109/OJCAS.2022.3203849
22. **Gallager R. G.** Low-Density Parity-Check codes. *IRE Transactions on Information Theory*, 1962, vol. 8, no. 1, pp. 21–28. doi:10.1109/TIT.1962.1057683
23. **Fossorier M. P. C., Mihaljevic M., Imai H.** Reduced complexity iterative decoding of Low-Density Parity Check codes based on belief propagation. *IEEE Transactions on Communications*, 1999, vol. 47, iss. 5, pp. 673–680. doi:10.1109/26.768759
24. **Crest J. Du., Garcia-Herrero F., Mhalla M., Savin V., Valls J.** Layered decoding of quantum LDPC codes. *2023 12th International Symposium on Topics in Coding (ISTC)*, Brest, France, 2023, pp. 1–5. doi:10.1109/ISTC57237.2023.10273477
25. **Wan H., Cho J., Jang M., Zhang C. J.** Low-complexity layered decoding for LDPC codes. *2024 IEEE Globecom Workshops (GC Wkshps)*, Cape Town, South Africa, 2024, pp. 1–6. doi:10.1109/GCWkshp64532.2024.11100976

UDC 004.056
doi:10.31799/1684-8853-2026-4-41-51
EDN: HPKSPY

Method for layer-by-layer decoding of QC-LDPC codes with constant output throughput (ZLayer)

P. S. Poperechny[a], PhD, Tech., Leading Engineer, orcid.org/0009-0008-8127-9854, yncherepop@mail.ru
V. V. Khrustalev[b], Post-Graduate Student, orcid.org/0009-0003-8742-1878
[a]KNS Group LLC 15, bld. 15, Rochdelskaya St., 123376, Moscow, Russian Federation
[b]Saint-Petersburg State University of Aerospace Instrumentation, 67, B. Morskaia St., 190000, Saint-Petersburg, Russian Federation

Introduction: When using QC-LDPC codes, the requirements for the dependence of the hardware decoder throughput on the code length vary. For example, for Solid State Drive, it is advantageous for the decoder throughput to be constant for different code lengths. However, existing QC-LDPC decoding algorithms have a linear relationship between the code size and throughput. **Purpose:** To modify the QC-LDPC code decoding method so that it ensures a constant decoding output throughput. **Results:** We propose a modification of the layer-by-layer decoding method, which consists in changing the method for forming layers and their number. Each layer is formed

from components of classical layers. The number of layers is determined by the Z lifting factor for the QC-LDPC code, and the layer size is determined by the number of rows in the base graph of the QC-LDPC code. For this reason, we have decided to name it ZLayer. Increasing the number of layers with increasing code length allows maintaining a constant output throughput of the hardware that implements the method. In addition, we propose a hardware ZLayer decoder microarchitecture. The analysis of the correction capability of the proposed method and decoder reveals a 0.3 dB improvement over classical layer-by-layer decoding. We present the estimates of the method's correction capability using the Scale and Offset coefficients, as well as the results of selecting these coefficients for various parameters of the QC-LDPC code used in 5G mobile networks. **Practical relevance:** The presented solution is proposed for the use in systems with QC-LDPC coding, where constant throughput is required. An increase in the correction capability compared to the generally accepted solution will also positively impact the overall efficiency of the QC-LDPC system.

Keywords – error correction codes, low density parity check codes, layer-by-layer decoding, hardware decoder, QC-LDPC, constant throughput.

For citation: Poperechny P. S., Khrustalev V. V. Method for layer-by-layer decoding of QC-LDPC codes with constant output throughput (ZLayer). *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2026, no. 4, pp. 41–51 (In Russian). doi:10.31799/1684-8853-2026-4-41-51, EDN: HPKSPY

References

1. Ovchinnikov A. A., Veresova A. M., Fominykh A. A. Decoding of linear codes for single error bursts correction based on the determination of certain events. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2022, no. 6, pp. 41–52. doi:10.31799/1684-8853-2022-6-41-52, EDN: UWXZHN
2. Li Y. Analysis and methodological advancements in software-defined error correction codes. *2024 4th International Signal Processing, Communications and Engineering Management Conference (ISPCEM)*, Montreal, QC, Canada, 2024, pp. 133–138. doi:10.1109/ISPCEM64498.2024.00030
3. Ali M. M., Hashim S. J., Chaudhary M. A., Ferré G., Rokhani F. Z., Ahmad Z. A reviewing approach to analyze the advancements of error detection and correction codes in channel coding with emphasis on LPWAN and IoT systems. *IEEE Access*, 2023, vol. 11, pp. 127077–127097. doi:10.1109/ACCESS.2023.3331417
4. Khrustalev V. V. Evaluation of the characteristics of open BCH and RS code decoders implemented on modern FPGAs. *Modeling, Optimization and Information Technology*, 2026, vol. 14, no. 3. Available at: https://moitvivt.ru/ru/journal/article?id = 2174 (accessed 08 April 2026) (In Russian). doi:10.26102/2310-6018/2026.54.3.005, EDN: CBXXSV
5. Garani S. S., Nadkarni P. J., Raina A. Theory behind quantum error correcting codes: An overview. *J Indian Inst Sci*, 2023, vol. 103, pp. 449–495. doi:10.1007/s41745-023-00392-7
6. Lin C.-H., Wang C.-X., Lu C.-K. LDPC decoder design using compensation scheme of group comparison for 5G communication systems. *Electronics*, 2021, vol. 10 (16), 2010. doi:10.3390/electronics10162010
7. Gupta S. H., Virmani B. LDPC for Wi-Fi and WiMAX technologies. *2009 International Conference on Emerging Trends in Electronic and Photonic Devices & Systems*, Varanasi, India, 2009, pp. 262–265. doi:10.1109/ELECTRO.2009.5441120
8. Zhang W., Gazizov T. R. LDPC and polar codes in 6G: A comparative study and unified frameworks. *Proceedings of Telecommunication Universities*, 2025, vol. 11, no. 5, pp. 42–59 (In Russian). https://doi.org/10.31854/1813-324X-2025-11-5-42-59, EDN: WXEIHS
9. Portnoy S., Tikhonyuk A., Nikitin S., Voloshin A. Performance evaluation of LDPC codes in 5G NR. *2024 Systems of Signal Synchronization, Generating and Processing in Telecommunications (SYNCHROINFO)*, Vyborg, Russian Federation, 2024, pp. 1–9. doi:10.1109/SYNCHROINFO61835.2024.10617595
10. Ovchinnikov A. A., Fominykh A. A. Decoding of LDPC codes for 5G standard using source distribution. *2020 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF)*, St. Petersburg, Russia, 2020, pp. 1–5. doi:10.1109/WECONF48837.2020.9131454
11. He Y., Wang Y. Research on Low-Density Parity-Check decoding algorithm for reliable transmission in satellite communications. *Applied Sciences*, 2024, vol. 14, iss. 2, 746. doi:10.3390/app14020746
12. Zinchenko M. Y., Levadniy A. M., Grebenko Y. A. FPGA LDPC decoder implementation and optimization its power consumption. *T-Comm*, 2020, vol. 14, no. 3, pp. 4–10 (In Russian). doi:10.36724/2072-8735-2020-14-3-4-10, EDN: XUZAWE
13. *IEEE Standard for Error Correction Coding of Flash Memory Using Low-Density Parity Check Codes*. In: *IEEE Std 1890–2018*, 2019, vol., no., pp. 1–51. doi:10.1109/IEEESTD.2019.8654228
14. Mo J., Zhai X., Yu M., Fang Y., Han G. SC-GC-LDPC for NAND flash memory: Construction and decoding. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2025, vol. 44, no. 11, pp. 4167–4180. doi:10.1109/TCAD.2025.3563802
15. Song Y., Lv Y., Li W., Liu J., Shi L. Revisiting multiple ECC on high-density NAND flash memory. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 2025, vol. 33, no. 6, pp. 1679–1692. doi:10.1109/TVLSI.2025.3551400
16. Pourjabar S., Choi G. S. A high-throughput multimode low-density parity-check decoder for 5G new radio. *Int J Circ Theor Appl*, 2022, vol. 50, iss. 4, pp. 1365–1374. doi:10.1002/cta.3208
17. Nadal J., Baghdadi A. Parallel and flexible 5G LDPC decoder architecture targeting FPGA. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 2021, vol. 29, iss. 6, pp. 1141–1151. doi:10.1109/TVLSI.2021.3072866
18. Cui H., Ghaffari F., Le K., Declercq D., Lin J., Wang Z. Design of high-performance and area-efficient decoder for 5G LDPC codes. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 2021, vol. 68, iss. 2, pp. 879–891. doi:10.1109/TCSI.2020.3038887
19. Chen W., Li Y., Liu D. High-throughput LDPC decoder for multiple wireless standards. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 2025, vol. 72, iss. 1, pp. 383–396. doi:10.1109/TCSI.2024.3419425
20. Lu Q., Sham C.-W., Lau F. C. M. On using the cyclically-coupled QC-LDPC codes in future SSDs. *2016 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)*, Jeju, Korea (South), 2016, pp. 625–628. doi:10.1109/APCCAS.2016. 7804048
21. Liu L.-W., Yuan M.-H., Liao Y.-C., Chang H.-C. A 38.64-Gb/s Large-CPM 2-KB LDPC decoder implementation for NAND flash memories. *IEEE Open Journal of Circuits and Systems*, 2022, vol. 3, pp. 180–191. doi:10.1109/OJCAS.2022. 3203849
22. Gallager R. G. Low-Density Parity-Check codes. *IRE Transactions on Information Theory*, 1962, vol. 8, no. 1, pp. 21–28. doi:10.1109/TIT.1962.1057683
23. Fossorier M. P. C., Mihaljevic M., Imai H. Reduced comple-xity iterative decoding of Low-Density Parity Check codes based on belief propagation. *IEEE Transactions on Communications*, 1999, vol. 47, iss. 5, pp. 673–680. doi:10.1109/26.768759
24. Crest J. Du., Garcia-Herrero F., Mhalla M., Savin V., Valls J. Layered decoding of quantum LDPC codes. *2023 12th International Symposium on Topics in Coding (ISTC)*, Brest, France, 2023, pp. 1–5. doi:10.1109/ISTC57237.2023.10273477
25. Wan H., Cho J., Jang M., Zhang C. J. Low-complexity layered decoding for LDPC codes. *2024 IEEE Globecom Workshops (GC Wkshps)*, Cape Town, South Africa, 2024, pp. 1–6. doi:10.1109/GCWkshp64532.2024.11100976

УДК 621.391.15
doi:10.31799/1684-8853-2026-4-52-59
EDN: GVMDCL

Научные статьи
Articles

Каскадные коды для построения широкополосных сигналов: расширение на двухпередатчиковую противофазную конструкцию

Е. А. Крук[а], *доктор техн. наук, профессор, orcid.org/0000-0002-4549-2980, ekrouk@hse.ru*
В. А. Давыдов[б], *канд. техн. наук, старший научный сотрудник, orcid.org/0000-0003-1316-3346*
[а]*Научно-исследовательский институт телекоммуникаций Национального исследовательского университета «Высшая школа экономики», Москва, Таллинская ул., 34, 123458, РФ*
[б]*Институт проблем управления им. В. А. Трапезникова РАН, Профсоюзная ул., 65, Москва, 117997, РФ*

Введение: *широкополосные шумоподобные сигналы на основе псевдослучайных m-последовательностей широко применяются в системах связи с требованием энергетической скрытности. Ключевой задачей является увеличение пропускной способности без роста суммарной излучаемой мощности.* ***Цель:*** *расширение метода каскадного кодирования на случай двух независимых передатчиков, работающих в противофазном режиме с алфавитами {0, +1} и {0, −1} соответственно, при сохранении энергетической скрытности системы.* ***Методы:*** *использован аппарат теории помехоустойчивого кодирования — внешний код Рида — Соломона в каскадной схеме с внутренним кодированием m-последовательностями. Для анализа вероятности неисправимых стираний применена пуассоновская аппроксимация.* ***Результаты:*** *при нормировке амплитуды ненулевых символов до sqrt(2) · A средняя мощность каждого передатчика за период сохраняется равной $L \cdot A^2$, а суммарная мощность пары совпадает с мощностью одного исходного передатчика для стороннего энергетического обнаружителя. Коллизии при совпадении выбранных m-последовательностей обоих передатчиков интерпретируются как стирания и исправляются кодом Рида — Соломона. Среднее число стираний на кодовое слово составляет $\lambda = n_1/2^m \approx 1$.* ***Практическая значимость:*** *предложенная конструкция обеспечивает удвоение числа независимых информационных потоков при неизменной мощности каждого передатчика, сохранении скрытности по энергетическому критерию и без изменения параметров корректирующего кода.* ***Обсуждение:*** *конструкция допускает обобщение на K пар передатчиков с ортогональными наборами кодов; оптимальный выбор K определяется компромиссом между требуемой пропускной способностью и допустимым уровнем энергетической видимости.*

Ключевые слова *— широкополосные сигналы, m-последовательности, каскадные коды, код Рида — Соломона, энергетическая скрытность, стирания, противофазная конструкция.*

Для цитирования: Крук Е. А., Давыдов В. А. Каскадные коды для построения широкополосных сигналов: расширение на двухпередатчиковую противофазную конструкцию. *Информационно-управляющие системы*, 2026, № 4, с. 52–59. doi:10.31799/1684-8853-2026-4-52-59, EDN: GVMDCL
For citation: Krouk E. A., Davydov V. A. Cascade codes for wideband signal construction: extension to a two-transmitter antiphase design. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2026, no. 4, pp. 52–59 (In Russian). doi:10.31799/1684-8853-2026-4-52-59, EDN: GVMDCL

Введение

Широкополосные шумоподобные сигналы, формируемые на основе псевдослучайных m-последовательностей, находят широкое применение в системах связи с требованиями к помехозащищенности и энергетической скрытности [1–6]. Классическим подходом к повышению надежности передачи является использование каскадных кодов: внешний код Рида — Соломона (RS) совместно с внутренним кодированием m-последовательностями обеспечивает существенный выигрыш по отношению сигнал/шум при заданной вероятности ошибки [7–11].

Задача одновременного увеличения пропускной способности и сохранения энергетической скрытности системы является актуальной при проектировании специализированных систем связи [12–26]. В настоящей работе рассматривается расширение базовой каскадной конструкции с одним передатчиком на случай двух независимых передатчиков, работающих в противофазном режиме. Предложенная конструкция позволяет удвоить число независимых информационных потоков без изменения суммарной мощности для стороннего энергетического приемника и без модификации параметров используемого кода Рида — Соломона [1, 9].

Структурная схема базовой системы включает источники передачи информации, каналы связи и два типа приемников. Приемник Пр1 работает в режиме приема информации: ему известно множество S сигналов, используемых источниками. Приемник Пр2 работает в режиме обнаружения сигнала: множество S ему не известно (или известно не полностью), и он ориентируется на энергетический критерий.

Качество работы системы передачи информации определяется временем наблюдения T_0, необходимым оптимальному обнаружителю для обнаружения суммарного сигнала источников с заданными вероятностями обнаружения P_D и ложной тревоги P_F, при фиксированных вероятности ошибки P_e и отношении сигнал/шум q^2. В условиях полной априорной неопределенности оптимальным считается энергетический обнаружитель [27, 28]. Предполагается, что каналы связи в обоих направлениях эквивалентны и что радиотехнические меры повышения качества передачи уже реализованы.

Исходная конструкция: один передатчик с каскадными кодами

Предварительные замечания

Рассматривается передача двоичных сообщений, при которой в передатчике блоки из k символов сопоставляются широкополосным фазоманипулированным сигналам вида

$$s(t) = A \cdot \cos(2\pi f_0 t + \varphi(t)),$$

где A и f_0 — амплитуда и несущая частота, а $\varphi(t)$ — закон изменения фазы:

$$\varphi(t) = \varphi_i \Delta t, \ \ \varphi_i \in \{0, \pi\},$$

φ_i полностью определяется значением i-го элемента m-последовательности $\{c_j\}$, где $c_j \in \{+1, -1\}$.

Оптимальным приемником в канале с белым гауссовским шумом является корреляционный приемник [27]. Вероятность ошибочного приема

$$P_e = \Phi(\delta), \ \delta = 2A\sqrt{(T/N_0)},$$

где δ — параметр обнаружения; $q^2 = A^2T/N_0$ — отношение сигнал/шум.

Каскадные широкополосные сигналы

m-последовательность представляет собой слово двоичного линейного кода с параметрами $(L, k_0, d_{\min})$, где $L = 2^m - 1$ — длина; k_0 — число информационных символов; $d_{\min}$ — минимальное расстояние [7]. Введение дополнительного внешнего кодирования реализуется по каскадной схеме [10, 29]:

— информационная последовательность разбивается на блоки по k_1 символов, каждый рассматривается как элемент q_1-ичного алфавита;

— полученная q_1-ичная последовательность кодируется внешним кодом Рида — Соломона (n_1, k_1, t_1);

— каждый q_1-ичный символ кодируется как m-последовательность (внутренний код).

Декодирование производится в обратном порядке: сначала коррелятор по m-последовательности, затем внешний декодер Рида — Соломона.

Вероятность ошибки на бит (BER) при каскадном кодировании

$$P_b \approx P_e^{RS}(P_i),$$

где P_e^{RS} — вероятность ошибочного декодирования слова внешнего кода Рида — Соломона; P_i — вероятность ошибки внутреннего декодера (коррелятора по m-последовательности).

Вероятность ошибки после внешнего RS-декодера оценивается формулой

$$P_b \lesssim \frac{n_1}{k_1} \cdot \sum_{j=t_1+1}^{n_1} jC(n_1, j)P_i^j(1-P_i)^{n_1-j}.$$

Выигрыш от каскадного кодирования

Для оценки эффективности системы зафиксируем скорость передачи информации и ширину полосы используемых сигналов и измерим выигрыш по требуемому отношению сигнал/шум q^2 при заданной вероятности ошибки. Результаты расчетов приведены в табл. 1.

Качество передачи и скрытность

При оптимальном энергетическом обнаружении связь между вероятностью обнаружения P_D, вероятностью ложной тревоги P_F, временем наблюдения T_0 и отношением сигнал/шум q^2 определяется формулой

$$T_0 = f^{-1}(P_D, P_F, q^2),$$

где f^{-1} — функция, обратная к стандартной зависимости мощностного детектора [28].

Снижение требуемого q^2 за счет кодирования при фиксированной мощности передатчика эквивалентно увеличению T_0 для стороннего энергетического приемника, т. е. непосредственно улучшает скрытность системы [15]. Вероятность обнаружения с учетом кодирования

$$P_D^{kod} = f(P_F, q^2/\eta),$$

что при том же P_F и фиксированном T_0 означает более низкую вероятность обнаружения системой Пр2.

Расширение на две передающие стороны: противофазная конструкция

Мотивация и принцип

В исходной конструкции каждый m-ичный символ представляется элементом алфавита $\{+1, -1\}$.

■ ***Таблица 1.*** Выигрыш от каскадного кодирования (исходная конструкция с {+1, –1}-сигналами)
■ ***Table 1.*** Gain from cascade coding (original design with {+1, –1} signals)

P_e	Выигрыш η	Параметр m-последовательности	Параметры кода Рида – Соломона (n_1, k_1, t_1)	Вероятность ошибки на кодовое слово $P_{\text{слово}} \times 10^{-3}$
10^{-4}	1,1126	3	7, 5, 1	1,5
	1,2451	4	15, 11, 2	4,4
	1,3308	5	31, 25, 3	1,25
	1,3558	6	63, 53, 5	3,18
10^{-5}	1,1618	3	7, 5, 1	1,5
	1,3489	4	15, 11, 2	4,4
	1,4739	5	31, 25, 3	1,15
	1,5365	6	63, 51, 6	3,06
	1,5508	7	127, 107, 10	7,49
10^{-6}	1,1983	3	7, 5, 1	1,5
	1,4344	4	15, 11, 2	4,4
	1,6087	5	31, 23, 4	1,15
	1,7069	6	63, 51, 6	3,06
	1,7414	7	127, 105, 11	7,35
10^{-7}	1,226	3	7, 5, 1	1,5
	1,5050	4	15, 11, 2	4,4
	1,7336	5	31, 23, 4	1,15
	1,8695	6	63, 49, 7	2,94
	1,9271	7	127, 105, 11	7,35
	1,9324	8	255, 217, 19	1,74

Рассмотрим возможность одновременной работы двух независимых передатчиков на одной m-последовательности с использованием противофазного разнесения символов [4].

Введем два передатчика с алфавитами: передатчик А: символы {0, +1}; передатчик В: символы {0, –1}. В любой момент времени один передатчик передает ненулевой символ своего знака, другой — нулевой символ или наоборот. Таким образом, суммарный сигнал принимает значения из множества {–1, 0, +1}.

Нормировка мощности

В исходной конструкции при амплитуде A все $L = 2^m - 1$ чипов ненулевые, и средняя энергия за период: $E_{orig} = L \cdot A^2$.

В новой конструкции у каждого передатчика половина символов нулевые, половина ненулевые (с амплитудой B): $E_{new} = (L/2) \cdot B^2$.

Требование сохранения средней мощности каждого передатчика за период

$$E_{new} = E_{orig} \Longrightarrow B^2 = 2A^2 \Longrightarrow B = \sqrt{2}A.$$

Таким образом, мощность ненулевых символов каждого передатчика увеличивается вдвое по сравнению с исходным, но средняя мощность за период остается неизменной.

Суммарная мощность пары для стороннего приемника

При независимой работе передатчиков А и В и равновероятном выборе символов средняя мощность суммарного сигнала пары

$$E[s^2_{sum}(k)] =$$

$$= 0 \cdot (1/2) + B^2 \cdot (1/4) + B^2 \cdot (1/4) = B^2/2 = A^2.$$

Средняя энергия суммы за период L:

$$E_{sum} = L \cdot A^2 = E_{orig}.$$

Следствие: Суммарная мощность пары передатчиков А и В с нормированными амплитудами совпадает с мощностью одного исходного передатчика. Сторонний энергетический приемник Пр2 не отличает работу пары от работы одного передатчика [5, 27].

Информационные потоки и скорость передачи

Каждый из передатчиков пары независимо формирует поток символов кода Рида — Соломона на основе тех же m-последовательностей и использует тот же код Рида — Соломона (n_1, k_1, t_1), что и передатчик в исходной конструкции. Скорость каждого потока равна скорости передачи в исходной системе: $R_A = R_B = R_{orig}$.

Таким образом, суммарный объем передачи по каналу удваивается при неизменной мощности каждого передатчика и неизменной видимой мощности суммарного сигнала для стороннего приемника.

Коллизии в конструкции из двух передатчиков и их обработка кодом Рида — Соломона

В данном разделе рассматривается влияние коллизий между сигналами двух передатчиков на работу каскадной сигнально-кодовой конструкции и анализируется возможность компенсации этих коллизий с использованием уже применяемого внешнего кода Рида — Соломона [7, 9]. Особое внимание уделяется совместному учету ошибок и стираний на входе декодера Рида — Соломона и влиянию этих факторов на итоговую вероятность ошибки.

Механизм возникновения коллизий и их интерпретация как стираний

В расширенной конструкции по одному и тому же набору m-последовательностей одновременно работают два независимых передатчика, каждый из которых использует тот же внешний код Рида — Соломона с параметрами (n_1, k_1, t_1). При этом алфавит символов первого передатчика имеет вид $\{0, +1\}$, второго — $\{0, -1\}$. Ненулевые символы обоих передатчиков формируются по одной и той же m-последовательности, а нулевые символы соответствуют отсутствию излучения на данном элементарном интервале.

Рассмотрим ситуацию, когда оба передатчика в одном и том же интервале времени выбирают одну и ту же m-последовательность и передают совпадающие по модулю, но противоположные по знаку ненулевые символы. При этом суммарный сигнал на входе приемника имеет вид $s_{sum}(t) = s_A(t) + s_B(t)$, где $s_A(t) = -s_B(t) \Longrightarrow s_{sum}(t) \equiv 0$.

В отсутствие шума такая ситуация неотличима от случая, когда оба передатчика передают нулевые символы. С точки зрения легального приемника соответствующий символ внешнего кода оказывается принципиально неидентифицируемым и трактуется как стирание [9].

Вероятность стирания одного символа внешнего кода

Пусть каждый из двух передатчиков выбирает сообщения независимо и равновероятно из множества из 2^m возможных последовательностей. Вероятность того, что оба передатчика выберут одну и ту же m-последовательность, равна

$$p_e = P\{m_A = m_B\} = \frac{1}{2^m}.$$

Распределение числа стираний на кодовое слово

Случайная величина S — число стираний на кодовое слово — подчиняется биномиальному закону $S \sim \mathrm{Bin}(n_1, p_e)$. При малых значениях p_e и достаточно большом n_1 биномиальное распределение аппроксимируется распределением Пуассона с параметром $\lambda = n_1 \cdot p_e = n_1/2^m$.

В практических вариантах при $n_1 = 255$ и $m = 8$ имеем $\lambda = 255/256 \approx 1$. Таким образом, введение второй передающей стороны приводит к появлению в среднем одного стирания на кодовое слово внешнего кода для каждого передатчика.

Совместная корректировка ошибок и стираний кодом Рида — Соломона

Код Рида — Соломона с параметрами (n_1, k_1, t_1) имеет минимальное расстояние $d_{\min} = n_1 - k_1 + 1$ и способен исправлять комбинации символьных ошибок и стираний, удовлетворяющих условию [7, 9, 30]

$$2E + S \le d_{\min} - 1 = 2t_1,$$

где E — число неизвестных символьных ошибок; S — число стираний.

Оценка влияния стираний на корректирующую способность

Рассмотрим конкретные параметры внешнего кода. Пусть используется код RS(255, 239), для которого $n_1 = 255$, $k_1 = 239$, $t_1 = 8$. Максимальное число стираний при отсутствии ошибок $2t_1 = 16$. Для кода RS(255, 223) $t_1 = 16$, $2t_1 = 32$.

При использовании модели Пуассона с параметром $\lambda \approx 1$ вероятность превышения корректирующей способности

$$P\{S > 2t_1\} = 1 - \sum_{k=0}^{2t_1} \frac{e^{-\lambda}\lambda^k}{k!}.$$

Для кода RS(255, 239) $t_1 = 8$, $2t_1 = 16$:

$$P\{S > 16) \le \frac{e^{-1}}{17!} \approx 10^{-15}.$$

Для кода RS(255, 223) $t_1 = 16$, $2t_1 = 32$: аналогичная оценка дает величину порядка 10^{-30}. Эти значения существенно ниже типичных целевых значений вероятности ошибки (10^{-6}–10^{-9}), используемых при проектировании реальных систем передачи.

Совместный учет ошибок и стираний и влияние на BER

В расширенной конструкции из двух передатчиков к ошибкам E, обусловленным каналом, добавляются стирания S, связанные с редкими коллизиями. Как показано выше, математическое

ожидание числа стираний на кодовое слово составляет около 1, а вероятность того, что S достигнет значений, сопоставимых с $2t_1$, пренебрежимо мала. При этом характер шумовых ошибок не изменяется: для каждого из двух передатчиков средняя энергия на m-последовательность и соответствующее отношение сигнал/шум на входе корреляционного приемника остаются такими же, как в исходной системе.

Таким образом, условие $2E + S \le 2t_1$ продолжает выполняться с вероятностью, практически равной той же, что и в исходной системе с одним передатчиком. Итоговая вероятность ошибки на бит P_β для каждого легального приемника остается на уровне, достигавшемся в исходной системе и отраженной в табл. 1, а влияние дополнительных стираний носит характер малой поправки, существенно меньшей целевого уровня вероятности ошибки.

Заключение

Сравнение исходной и предложенной конструкций по ключевым параметрам приведено в табл. 2.

■ ***Таблица 2.*** Сравнение исходной и предложенной конструкций по ключевым параметрам

■ ***Table 2.*** Comparison of original and proposed constructions by key parameters

Параметр	Исходная конструкция	Новая конструкция (пара)
Число передатчиков	1	2
Алфавит символов	{+1, –1}	{0, +1} и {0, –1}
Амплитуда ненулевых символов	A	$\sqrt{2}A$
Средняя мощность одного передатчика	$L \cdot A^2$	$L \cdot A^2$
Видимая мощность для стороннего (Пр2)	$L \cdot A^2$	$L \cdot A^2$
Код Рида — Соломона	(n_1, k_1, t_1)	Тот же (n_1, k_1, t_1) для каждого
Скорость каждого передатчика	R_{orig}	R_{orig}
Суммарный объем передачи	R_{orig}	$2R_{orig}$
BER легального приемника	P_β	$\approx P_\beta$
Дополнительное оборудование	—	Не требуется
Вероятность неисправимых стираний	0	$\lesssim 10^{-15}$ (для $m = 8$)

Предложенная противофазная конструкция из двух передатчиков реализует следующие свойства.

1. Сохранение мощности каждого передатчика. При переходе от алфавита {+1, –1} к {0, +1} или {0, –1} с увеличением амплитуды ненулевых символов в $\sqrt{2}$ раза средняя энергия за период не изменяется.

2. Сохранение скрытности по энергетическому критерию. Суммарная мощность противофазной пары равна мощности одного исходного передатчика. Сторонний энергетический приемник не отличает работу пары от работы одиночного источника.

3. Удвоение объема передачи. Два передатчика с теми же m-последовательностями и теми же кодами Рида — Соломона передают независимые информационные потоки, каждый со скоростью исходной системы.

4. Обработка коллизий средствами уже используемого кода Рида — Соломона. При $m = 8$ и стандартных кодах RS(255, 239) или RS(255, 223) ожидаемое число стираний на кодовое слово составляет около 1, и в результате влияние дополнительных стираний носит характер малой поправки, существенно меньшей целевого уровня вероятности ошибки.

5. Отсутствие дополнительного оборудования. Для обнаружения и коррекции коллизий не требуются специальные приемники на стороне передатчиков и дополнительные протоколы обратной связи.

Основная новизна состоит в том, что предложенная конструкция реализует удвоение числа независимых информационных потоков и суммарного объема передачи при неизменной энергетической «подписи» системы для стороннего наблюдателя, неизменном BER для легальных приемников и без изменения параметров корректирующего кода.

Замечание о масштабировании. Описанная конструкция допускает обобщение на K противофазных пар, использующих K взаимно ортогональных (или квазиортогональных) наборов кодов. В этом случае суммарный объем передачи возрастает в K раз по сравнению с исходной системой, однако суммарная мощность для стороннего энергетического приемника также возрастает в K раз, что снижает энергетическую скрытность системы. Оптимальный выбор K определяется компромиссом между требуемой пропускной способностью и допустимым уровнем энергетической «видимости» [4, 15].

Литература

1. Schmidt J. H., Kochańska I., Schmidt A. M. Performance of the direct sequence spread spectrum underwater acoustic communication system with differential detection in strong multipath propagation conditions. *Archives of Acoustics*, 2024, vol. 49, no. 1, pp. 129–140. doi:10.24425/aoa.2024.148771

2. Nguyen T. P., Le H. T., Nguyen V. D. Performance analysis for DSSS with short period sequences encountering broadband interference. *IEEE Transactions on Vehicular Technology*, 2024, vol. 73, no. 11, pp. 16697–16707. doi:10.1109/TVT.2024.3414024

3. Ali A., Altaf A., Tauqeer T. Denoic extension sequences CDMA for communication systems. *Applied Science and Innovative Research*, 2025, vol. 9, no. 1, pp. 1–11. doi:10.22158/asir.v9n1p1

4. Gao C., Wu Y., Zhang Y. Secure communication using WFRFT-DSSS based on chaotic cyclic shift. *IEEE/CIC International Conference on Communications in China (ICCC)*, 2023, pp. 1–5. doi:10.1109/ICCC57788.2023.10233501

5. Andreyev Y. Analytical model of energy detector for ultra-wideband chaotic radio pulses. *Electronics*, 2023, vol. 12, no. 4, Article 954. doi:10.3390/electronics12040954

6. Li Z., Huang X., Zhao L., Chen X. Multiple access interference suppression for CDMA systems via l1 minimization. *Fundamental Research*, 2022, vol. 2, no. 5, pp. 799–806. doi:10.1016/j.fmre.2021.11.037

7. Tang Y. J., Zhang X. Fast en/decoding of Reed — Solomon codes for failure recovery. *IEEE Transactions on Computers*, 2022, vol. 71, no. 3, pp. 724–735. doi:10.1109/TC.2021.3060701

8. Ahmed K., Abdelaziz K. Enhancing erasure resilience in Reed — Solomon codes: Theory and applications in data storage systems, *Transactions on Engineering and Computing Sciences*, 2024, vol. 12, no. 5, pp. 1–15. doi:10.61173/2qsz7881

9. Hörmann F., Bartz H. Syndrome-based error-erasure decoding of interleaved linearized Reed – Solomon codes. *IEEE Transactions on Information Theory*, 2025. doi:10.48550/arXiv.2411.19101

10. Wondimu E., Annamalai P., Gared F. Performance analysis of concatenated Reed — Solomon and next generation polar codes for 6G communication systems. *Scientific Reports*, 2025, vol. 15, Article 31785. doi:10.1038/s41598-025-13672-2

11. Sukmadji A. Y., Kschischang F. R. Performance-complexity-latency trade-offs of concatenated RS-SDBCH codes. *IEEE Journal of Lightwave Technology*, 2025. doi:10.1109/JLT.2025.3530858

12. Li Z., Shi J., Si J., Lv L., Guan L., Hao B., Xing C., Quek T. Q. S. Intelligent covert communication: Recent advances and future research trends. *Engineering*, 2025, vol. 44, pp. 109–119. doi:10.1016/j.eng.2024.12.007

13. Yang Y., Ren Y., Liu W., Gong C., Huang N., Xu Z. Spread spectrum for covert communication in ultra-violet communication system. *Optics Express*, 2024, vol. 32, no. 15, pp. 25981–25994. doi:10.1364/OE.527516

14. Ning X., Yang Y., Guo K., Wang Z. Design and performance analysis of LPD communication waveform based on FrFT-FH structure. *Systems Engineering and Electronics*, 2024, vol. 46, no. 8, pp. 2857–2866. doi:10.12305/j.issn.1001-506X.2024.08.33

15. Yan S., Zhou X., Hu J., Guo S. Low probability of detection communication: Opportunities and challenges. *IEEE Wireless Communications*, 2021, vol. 28, no. 4, pp. 144–150. doi:10.1109/MWC.001.2000173

16. Aggarwal R., Kong J. S., Moore T. J., Choi J., Spasojevic P., Dagefu F. T. Covert communications with simultaneous multi-modal transmission. *Proceedings of ACM Workshop on Wireless Security and Machine Learning (WiseML)*, 2024. doi:10.1145/3643833.3656117

17. Lee J., Dinh D. T., Yeom H., Lee S.-H., Ha J. Multiuser cooperation for covert communication under quasi-static fading. *IEEE Transactions on Information Forensics and Security*, 2023, vol. 18, pp. 4625–4639. doi:10.1109/TIFS.2023.3264050

18. Doan T. B., Nguyen T.-H. Enhancing covert communication in NOMA systems with joint security and covert design. *PLOS ONE*, 2025, vol. 20, no. 1, e0317289. doi:10.1371/journal.pone.0317289

19. Si J., Liu Z., Li Z., Hu H., Guan L., Wang C., Al-Dhahir N. A cooperative deception strategy for covert communication in presence of a multi-antenna adversary. *IEEE Transactions on Communications*, 2023, vol. 71, no. 3, pp. 1528–1542. doi:10.1109/TCOMM.2022.3228637

20. Liu C., Wang C., Zhu Q., Li Y. Multi-user covert communications via intelligent spectrum control. arXiv:2601.05281 [cs.IT]. 2026. doi:10.48550/arXiv.2601.05281

21. Chen B., Gao C., Ma R., Xu D. Covert wireless communication in multichannel systems. *IEEE Wireless Communications Letters*, 2022, vol. 11, no. 9, pp. 1790–1794. doi:10.1109/LWC.2022.3180993

22. Alazab M., Al-Nemrat A., Bhatt P. Gaussian-distributed spread-spectrum for covert communications. *Sensors*, 2023, vol. 23, no. 8, Article 4081. doi:10.3390/s23084081

23. Dagefu F. T., Kong J. S., Moore T. J. Low Probability of Intercept/Detect (LPI/LPD) secure communications using antenna arrays employing rapid sidelobe time modulation. *IEEE Transactions on Antennas and Propagation*, 2024, vol. 72, no. 8, pp. 6448–6463. doi:10.1109/TAP.2024.3416756

24. Wang H., Liu Y., Zhang Y., Gao J. Robust overlapping covert communication against partial-band jamming via dynamic spread spectrum factor. *Scientific Reports*, 2025, vol. 15. doi:10.1038/s41598-025-32907-w

25. Lu X., Noneaker D. L. Physical layer covert communication in B5G wireless networks — its research, applications, and challenges. *Proceedings of the IEEE*, 2024, vol. 112, no. 1, pp. 47–82. doi:10.1109/JPROC.2023.3321768

26. Cheng Y., Lu J., Niyato D., Lyu B., Xu M., Zhu S. Performance analysis and power allocation for covert mobile edge computing with RIS-Aided NOMA. *IEEE Transactions on Mobile Computing*, 2024, vol. 23, no. 3, pp. 2046–2061. doi:10.1109/TMC.2023.3302413

27. Lorincz J., Ramljak I., Begusic D. A survey on the energy detection of OFDM signals with dynamic threshold adaptation: Open issues and future challenges. *Sensors*, 2021, vol. 21, no. 16, Article 5579. doi:10.3390/s21165579

28. Zhang Y., Liu P., Liu X., Li X. An energy detection-based spectrum-sensing method for cognitive radio. *Wireless Communications and Mobile Computing*, 2022, vol. 2022, Article 3933336. doi:10.1155/2022/3933336

29. Bac L. H., Kieu-Xuan T., Nguyen T. B. On concatenated coding scheme for high-speed Ethernet. *Proceedings of the International Conference on Advanced Communication Technology (ICACT)*, 2023, pp. 1–6. doi:10.23919/ICACT56868.2023.10023773

30. Luo G., Wan Q., Lin C. Improved error bounds for the distance distribution of Reed — Solomon codes. *IEEE Transactions on Information Theory*, 2023, vol. 69, no. 6, pp. 3602–3614. doi:10.1109/TIT.2023.3242612

UDC 621.391.15
doi:10.31799/1684-8853-2026-4-52-59
EDN: GVMDCL

Cascade codes for wideband signal construction: extension to a two-transmitter antiphase design

E. A. Krouk[a], Dr. Sc., Tech., Professor, orcid.org/0000-0002-4549-2980, ekrouk@hse.ru
V. A. Davydov[b], PhD, Tech., Senior Researcher, orcid.org/0000-0003-1316-3346
[a]Telecommunications Research Institute HSE, 34, Tallinskaya St., 123458, Moscow, Russian Federation
[b]V. A. Trapeznikov Institute of Control Sciences of RAS, 65, Profsoyuznaya St., 117997, Moscow, Russian Federation

Introduction: Wideband noise-like signals based on m-sequences are widely used in communication systems requiring energy-domain covertness. A key challenge is to increase throughput without increasing total radiated power. **Purpose:** To extend the cascade coding method to two independent transmitters operating in antiphase mode with alphabets {0, +1} and {0, –1}, preserving energy covertness. **Methods:** The error-correcting coding framework is used: an outer Reed – Solomon code combined with inner coding by m-sequences (concatenated scheme). A Poisson approximation is applied to analyze the probability of uncorrectable erasures. **Results:** Normalizing non-zero symbol amplitude to sqrt(2) · A keeps each transmitter average power at $L \cdot A^2$ per period, while the pair total power equals that of the original single transmitter for a third-party energy detector. Collisions when both transmitters select the same m-sequence are treated as erasures and corrected by the Reed – Solomon code. Expected erasures per codeword: $\lambda = n_1/2^m \approx 1$. **Practical relevance:** The proposed construction doubles independent information streams while keeping each transmitter power unchanged and preserving energy covertness without modifying code parameters. **Discussion:** The construction can be generalized to K transmitter pairs with orthogonal code sets, with optimal K being a trade-off between required throughput and admissible energy visibility.

Keywords — wideband signals, m-sequences, cascade codes, Reed – Solomon code, energy covertness, erasures, antiphase construction.

For citation: Krouk E. A., Davydov V. A. Cascade codes for wideband signal construction: extension to a two-transmitter antiphase design. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2026, no. 4, pp. 52–59 (In Russian). doi:10.31799/1684-8853-2026-4-52-59, EDN: GVMDCL

References

1. Schmidt J. H., Kochańska I., Schmidt A. M. Performance of the direct sequence spread spectrum underwater acoustic communication system with differential detection in strong multipath propagation conditions. *Archives of Acoustics*, 2024, vol. 49, no. 1, pp. 129–140. doi:10.24425/aoa.2024.148771
2. Nguyen T. P., Le H. T., Nguyen V. D. Performance analysis for DSSS with short period sequences encountering broadband interference. *IEEE Transactions on Vehicular Technology*, 2024, vol. 73, no. 11, pp. 16697–16707. doi:10.1109/TVT.2024.3414024
3. Ali A., Altaf A., Tauqeer T. Denoic extension sequences CDMA for communication systems. *Applied Science and Innovative Research*, 2025, vol. 9, no. 1, pp. 1–11. doi:10.22158/asir.v9n1p1
4. Gao C., Wu Y., Zhang Y. Secure communication using WFRFT-DSSS based on chaotic cyclic shift. *IEEE/CIC International Conference on Communications in China (ICCC)*, 2023, pp. 1–5. doi:10.1109/ICCC57788.2023.10233501
5. Andreyev Y. Analytical model of energy detector for ultra-wideband chaotic radio pulses. *Electronics*, 2023, vol. 12, no. 4, Article 954. doi:10.3390/electronics12040954
6. Li Z., Huang X., Zhao L., Chen X. Multiple access interference suppression for CDMA systems via l1 minimization. *Fundamental Research*, 2022, vol. 2, no. 5, pp. 799–806. doi:10.1016/j.fmre.2021.11.037
7. Tang Y. J., Zhang X. Fast en/decoding of Reed — Solomon codes for failure recovery. *IEEE Transactions on Computers*, 2022, vol. 71, no. 3, pp. 724–735. doi:10.1109/TC.2021.3060701
8. Ahmed K., Abdelaziz K. Enhancing erasure resilience in Reed — Solomon codes: Theory and applications in data storage systems, *Transactions on Engineering and Computing Sciences*, 2024, vol. 12, no. 5, pp. 1–15. doi:10.61173/2qsz7881

9. Hörmann F., Bartz H. Syndrome-based error-erasure decoding of interleaved linearized Reed-Solomon codes. *IEEE Transactions on Information Theory*, 2025. doi:10.48550/arXiv.2411.19101
10. Wondimu E., Annamalai P., Gared F. Performance analysis of concatenated Reed — Solomon and next generation polar codes for 6G communication systems. *Scientific Reports*, 2025, vol. 15, Article 31785. doi:10.1038/s41598-025-13672-2
11. Sukmadji A. Y., Kschischang F. R. Performance-complexity-latency trade-offs of concatenated RS-SDBCH codes. *IEEE Journal of Lightwave Technology*, 2025. doi:10.1109/JLT.2025.3530858
12. Li Z., Shi J., Si J., Lv L., Guan L., Hao B., Xing C., Quek T. Q. S. Intelligent covert communication: Recent advances and future research trends. *Engineering*, 2025, vol. 44, pp. 109–119. doi:10.1016/j.eng.2024.12.007
13. Yang Y., Ren Y., Liu W., Gong C., Huang N., Xu Z. Spread spectrum for covert communication in ultra-violet communication system. *Optics Express*, 2024, vol. 32, no. 15, pp. 25981–25994. doi:10.1364/OE.527516
14. Ning X., Yang Y., Guo K., Wang Z. Design and performance analysis of LPD communication waveform based on FrFT-FH structure. *Systems Engineering and Electronics*, 2024, vol. 46, no. 8, pp. 2857–2866. doi:10.12305/j.issn.1001-506X.2024.08.33
15. Yan S., Zhou X., Hu J., Guo S. Low probability of detection communication: Opportunities and challenges. *IEEE Wireless Communications*, 2021, vol. 28, no. 4, pp. 144–150. doi:10.1109/MWC.001.2000173
16. Aggarwal R., Kong J. S., Moore T. J., Choi J., Spasojevic P., Dagefu F. T. Covert communications with simultaneous multi-modal transmission. *Proceedings of ACM Workshop on Wireless Security and Machine Learning (WiseML)*, 2024. doi:10.1145/3643833.3656117
17. Lee J., Dinh D. T., Yeom H., Lee S.-H., Ha J. Multiuser cooperation for covert communication under quasi-static fading. *IEEE Transactions on Information Forensics and Security*, 2023, vol. 18, pp. 4625–4639. doi:10.1109/TIFS.2023.3264050
18. Doan T. B., Nguyen T.-H. Enhancing covert communication in NOMA systems with joint security and covert design. *PLOS ONE*, 2025, vol. 20, no. 1, e0317289. doi:10.1371/journal.pone.0317289
19. Si J., Liu Z., Li Z., Hu H., Guan L., Wang C., Al-Dhahir N. A cooperative deception strategy for covert communication in presence of a multi-antenna adversary. *IEEE Transactions on Communications*, 2023, vol. 71, no. 3, pp. 1528–1542. doi:10.1109/TCOMM.2022.3228637
20. Liu C., Wang C., Zhu Q., Li Y. Multi-user covert communications via intelligent spectrum control. arXiv:2601.05281 [cs.IT]. 2026. doi:10.48550/arXiv.2601.05281
21. Chen B., Gao C., Ma R., Xu D. Covert wireless communication in multichannel systems. *IEEE Wireless Communications Letters*, 2022, vol. 11, no. 9, pp. 1790–1794. doi:10.1109/LWC.2022.3180993
22. Alazab M., Al-Nemrat A., Bhatt P. Gaussian-distributed spread-spectrum for covert communications. *Sensors*, 2023, vol. 23, no. 8, Article 4081. doi:10.3390/s23084081
23. Dagefu F. T., Kong J. S., Moore T. J. Low Probability of Intercept/Detect (LPI/LPD) secure communications using antenna arrays employing rapid sidelobe time modulation. *IEEE Transactions on Antennas and Propagation*, 2024, vol. 72, no. 8, pp. 6448–6463. doi:10.1109/TAP.2024.3416756
24. Wang H., Liu Y., Zhang Y., Gao J. Robust overlapping covert communication against partial-band jamming via dynamic spread spectrum factor. *Scientific Reports*, 2025, vol. 15. doi:10.1038/s41598-025-32907-w
25. Lu X., Noneaker D. L. Physical layer covert communication in B5G wireless networks — its research, applications, and challenges. *Proceedings of the IEEE*, 2024, vol. 112, no. 1, pp. 47–82. doi:10.1109/JPROC.2023.3321768
26. Cheng Y., Lu J., Niyato D., Lyu B., Xu M., Zhu S. Performance analysis and power allocation for covert mobile edge computing with RIS-Aided NOMA. IEEE *Transactions on Mobile Computing*, 2024, vol. 23, no. 3, pp. 2046–2061. doi:10.1109/TMC.2023.3302413
27. Lorincz J., Ramljak I., Begusic D. A survey on the energy detection of OFDM signals with dynamic threshold adaptation: Open issues and future challenges. *Sensors*, 2021, vol. 21, no. 16, Article 5579. doi:10.3390/s21165579
28. Zhang Y., Liu P., Liu X., Li X. An energy detection-based spectrum-sensing method for cognitive radio. *Wireless Communications and Mobile Computing*, 2022, vol. 2022, Article 3933336. doi:10.1155/2022/3933336
29. Bac L. H., Kieu-Xuan T., Nguyen T. B. On concatenated coding scheme for high-speed Ethernet. *Proceedings of the International Conference on Advanced Communication Technology (ICACT)*, 2023, pp. 1–6. doi:10.23919/ICACT56868.2023.10023773
30. Luo G., Wan Q., Lin C. Improved error bounds for the distance distribution of Reed — Solomon codes. *IEEE Transactions on Information Theory*, 2023, vol. 69, no. 6, pp. 3602–3614. doi:10.1109/TIT.2023.3242612

УДК 004.93'1:621.396.96
doi:10.31799/1684-8853-2026-4-60-74
EDN: GGKFJQ

Научные статьи
Articles

Классификатор временных треков автомобильного радара с многомасштабной адаптацией к дисбалансу классов

Е. А. Лопухова[a], *младший научный сотрудник, orcid.org/0000-0001-6802-6010*
Г. С. Воронков[a], *канд. техн. наук, доцент, orcid.org/0000-0002-8788-2696, voronkov.gs@ugatu.su*
Е. П. Топольская[a], *канд. техн. наук, доцент, orcid.org/0000-0003-2207-2702*
[a]*Уфимский университет науки и технологий, Заки Валиди ул., 32, Уфа, 450076, РФ*

Введение: *распознавание дорожных объектов по данным автомобильного радара осложняется малым числом отражений в кадре и сильной неравномерностью классов, из-за которой редкие объекты хуже представлены в обучении.* ***Цель:*** *разработать классификатор радарных треков, устойчивый к дисбалансу классов и применимый к открытым наборам дорожного движения.* ***Методы:*** *для проведения исследований разработан и применен метод классификации радарных треков, сочетающий покадровую обработку неупорядоченных множеств радарных отражений разделяемым многослойным персептроном с инвариантной агрегацией признаков, временную трансформерную агрегацию с обучаемым классификационным токеном и прототипную классификационную голову в нормированном косинусном пространстве. Дисбаланс классов учитывался с помощью отступа, зависящего от распределения меток, классово-зависимого температурного масштабирования, весов функции потерь, рассчитанных по эффективному числу образцов, и взвешенной выборки при формировании мини-батчей; дополнительно применялось вспомогательное предсказание статистических моментов радиальной скорости.* ***Результаты:*** *эксперименты выполнены на наборах RadarScenes и Boreas-Objects-V1. RadarScenes содержит 158 последовательностей с поточечной разметкой пяти классов, а Boreas-Objects-V1 — 4699 треков с трехмерной разметкой четырех классов, включая редчайший класс с долей 1,5 %. На Boreas-Objects-V1 сбалансированная точность выросла с 0,945 до 0,970, а полнота редчайшего класса — с 0,833 до 0,917. Абляционное исследование показало, что отключение временной агрегации снижает сбалансированную точность на RadarScenes с 0,998 до 0,327, что указывает на принадлежность последовательной информации к ключевым источникам признаков, а не к вспомогательным элементам модели.* ***Практическая значимость:*** *классификатор пригоден для обучения систем радарного распознавания дорожных объектов в условиях ограниченного числа примеров редких классов.*

Ключевые слова *— автомобильный радар, радарный трек, распознавание дорожных объектов, дисбаланс классов, временная агрегация, метрическое обучение, нейронные сети.*

Для цитирования: Лопухова Е. А., Воронков Г. С., Топольская Е. П. Классификатор временных треков автомобильного радара с многомасштабной адаптацией к дисбалансу классов. *Информационно-управляющие системы*, 2026, № 4, с. 60–74. doi:10.31799/1684-8853-2026-4-60-74, EDN: GGKFJQ
For citation: Lopukhova E. A., Voronkov G. S., Topolskaya E. P. Temporal track classifier for automotive radar with multi-scale adaptation to class imbalance. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2026, no. 4, pp. 60–74 (In Russian). doi:10.31799/1684-8853-2026-4-60-74, EDN: GGKFJQ

Введение

Работоспособность функций активной безопасности транспортного средства (ТС) напрямую зависит от того, насколько надежно бортовые сенсоры распознают дорожную обстановку при переменных условиях видимости. Камеры и лидары дают высокое угловое и пространственное разрешение, но в тумане, при сильном дожде или в ночное время они деградируют ощутимо: рассеяние лидарного сигнала на аэрозоле растет с оптической глубиной трассы, а визуальное распознавание страдает от потери контраста и перекрытий [1]. В отечественных и зарубежных работах по интеллектуальным транспортным системам отмечается та же проблема: устойчивое восприятие дорожной сцены требует сенсоров с принципиально разной физической природой сигнала [2–4]. Автомобильный радар в подобных условиях держится заметно лучше — длина волны миллиметрового диапазона существенно превышает характерный размер туманных капель, что резко снижает рассеяние по сравнению с оптическими сенсорами [2, 5, 6]. Именно поэтому радар остается одним из ключевых компонентов систем помощи водителю при неблагоприятных погодных условиях [2, 3, 5].

У радарных данных есть своя принципиальная особенность, которую нельзя игнорировать. Один кадр от движущегося объекта, как правило, дает одну-три точки отражения, т. е. никакой пространственной информации, сопоставимой с плотным лидарным облаком, там нет [6]. Дискриминирующая сила сосредоточена в другом месте: динамика радиальной скорости, смещение отражений между кадрами, доплеровские сигнатуры, характерные для конкретного типа движения — все это проявляется именно во временном контексте [7, 8]. Распознавание по траекторным признакам последовательностей отражений давно

применяется в задачах классификации дорожных пользователей и в более широких задачах радарного распознавания [9]. Это означает, что агрегирование признаков по последовательности кадров — не дополнительная опция, а структурно необходимый элемент любого классификатора радарных треков.

Для обработки неупорядоченных наборов точек широко используется архитектура PointNet, обеспечивающая инвариантность к перестановке детекций за счет общего поточечного персептрона и симметричного объединения признаков; на наборе ModelNet40 она достигла 89,2 % общей точности классификации [10]. Ее расширение PointNet++ извлекает локальную структуру на нескольких масштабах и повышает точность на том же наборе до 91,9 % [11]. В российских работах по нейросетевой классификации радарных объектов также активно разрабатываются подходы к формированию информативных представлений, инвариантных к порядку поступления отражений и плотности облака [12]. В рамках радарного восприятия близкие идеи были использованы в RadarGNN, где графовая нейронная сеть была применена к данным RadarScenes. В результате макроусредненный F1, т. е. среднее гармоническое точности и полноты, одинаково усредненное по всем классам независимо от их частоты, составил 77,1 % в задаче сегментации, а средняя прецизионность, усредненная по классам и порогам перекрытия с разметкой, составила 60,2 % в задаче обнаружения объектов [13].

Для моделирования последовательностей эффективным решением является трансформер с обучаемым токеном-агрегатором Classifier Token (classification token, CLS), который через механизм самовнимания аккумулирует информацию всей последовательности и формирует компактное представление для классификации [14, 15]. Важно, что трансформерная обработка последовательностей относится не к узкоспециализированным решениям для одной предметной области, а к междисциплинарному классу методов: в отечественных исследованиях она применяется, в частности, к задаче классификации пользовательских запросов при определении профильного врача, что подтверждает переносимость механизма самовнимания и классификационного токена на данные различной природы [16]. В задачах радарного распознавания близкую роль выполняют темпоральные модули, например RadarTCN, где последовательная обработка радарных кадров используется для онлайн-классификации автомобильных радарных целей [17].

Радарный сигнал от движущегося объекта накапливается приемником в виде последовательности наборов отражений — радарного трека. Именно на уровне всего трека и выполняется классификация: пешеход, велосипедист, легковой автомобиль или тяжелое ТС. Присвоенная метка задает поведенческую модель объекта в прогнозирующем алгоритме и прямо влияет на формирование предупреждения о столкновении — ошибка здесь обходится дорого [18, 8]. Однако в реальном трафике распределение классов катастрофически неравномерно: стандартный классификатор тяготеет к доминирующей категории (прежде всего — легковым автомобилям), тогда как редкие и уязвимые классы: пешеходы, группы пешеходов, велосипедисты — получают слишком слабый градиентный сигнал [18, 19].

Отдельное направление исследований посвящено адаптации обучения к дисбалансу классов. Эта проблема имеет междисциплинарный характер: она возникает не только в задачах радарного распознавания, но и в других прикладных областях машинного обучения, где редкие, но практически значимые классы оказываются недостаточно представлены в обучающей выборке [19]. Поэтому методические выводы таких работ: необходимость контролировать качество на миноритарных классах, использовать специальные метрики и применять процедуры балансировки — могут быть перенесены на классификацию радарных треков, где редкие классы дорожных объектов также требуют отдельного контроля полноты [18, 19]. Взвешивание функции потерь через эффективное число образцов снижает ошибку на несбалансированном CIFAR-10 при коэффициенте дисбаланса 100 с 29,64 до 20,73 % [20], а метод Label-Distribution-Aware Margin (LDAM) с аддитивным отступом, зависящим от распределения меток, в схеме LDAM-DRW дает ошибку 14,82 % на том же варианте CIFAR-10 [21]. Современные обзоры по обучению на данных с длиннохвостым распределением (long-tailed learning) показывают, что основные способы работы с редкими классами включают ребалансировку выборки, модификацию функции потерь, обучение представлений и постобработку классификационной головы [22].

В метрическом обучении ArcFace применяет косинусный угловой отступ и достигает высокой точности верификации лиц за счет явного разделения классов в нормированном эмбеддинговом пространстве [23]. Близкая идея классификации по расстоянию до эталонов реализована в прототипических сетях, где каждому классу сопоставляется вектор-прототип, а решение принимается по близости объекта к этим прототипам [24]. Однако совместного применения покадровой обработки разреженных облаков точек, временной агрегации, метрической классификационной головы и адаптации к дисбалансу классов в единой модели для классификации радарных треков перечисленные подходы не рассматривали.

Для заполнения этого пробела предлагается классификатор временных радарных треков (КВРТ) с многомасштабной адаптацией к дисбалансу классов. Архитектура объединяет: покадровый кодировщик, в котором каждая точка радарного облака обрабатывается одним и тем же многослойным персептроном с разделяемыми между точками весами, а поточечные признаки затем сводятся в единый вектор кадра симметричной функцией агрегации, инвариантной к порядку точек [10, 12]; временной трансформер с CLS [14–16]; классификационную голову с косинусной метрикой [23] и обучаемыми векторами-эталонами классов [24]. Эффективность и вклад компонентов исследуются на двух открытых наборах данных:

1) RadarScenes — 158 последовательностей и пять классов: car (легковой автомобиль), pedestrian (пешеход), pedgroup (группа пешеходов), twowheeler (двухколесное ТС) и largevehicle (крупногабаритное ТС), — при этом доля редчайшего класса twowheeler не превышает 5 % треков [18];

2) Boreas-Objects-V1 — 4699 треков четырех классов: car (45,7 %), pedestrian (48,5 %), cyclist (велосипедист, 4,3 %) и misc (прочие объекты, 1,5 %, или 70 треков) [25].

Абляционное исследование количественно проверяет, является ли временная агрегация структурно необходимой, достаточен ли мощный кодировщик без адаптации к дисбалансу и какие компоненты балансировки дают измеримый прирост точности на редких классах.

Методология

Основная концепция метода заключается в том, чтобы классифицировать не отдельные радарные точки или кадры, а весь радарный трек как временную последовательность разреженных множеств отражений. Для этого применяется трехступенчатая архитектура: 1) кодирование множества точек внутри каждого кадра; 2) агрегирование покадровых признаков по времени; 3) сравнение трека с обучаемыми прототипами классов.

Адаптация к дисбалансу классов осуществляется не одним методом, а несколькими согласованными механизмами: весами функции потерь; повторной выборкой редких классов; классово-зависимым отступом; температурным масштабированием.

Наборы данных

Набор данных RadarScenes сформирован с использованием четырех серийных автомобильных радарных датчиков с рабочей частотой 77 ГГц, установленных на одном ТС; дальность обнаружения каждого датчика составляет до 100 м, угловое поле зрения — ±60° относительно оси, разрешение по дальности — 0,15 м, разрешение по радиальной скорости — 0,1 км/ч, средний период обновления — 60 мс (≈ 17 Гц). Каждая детекция описывается шестью признаками: плоскими координатами цели в системе отсчета носителя (x, y), компенсированной радиальной скоростью v_r, эффективной площадью рассеяния (мощностью отражения), дальностью и азимутальным углом; эти же шесть признаков образуют входной вектор в предложенном методе. Набор данных Boreas-Objects-V1 построен на основе датасета Boreas, записанного с использованием сканирующего кругового радара Navtech CIR204-H (диапазон — до 200 м, разрешение по дальности — 0,060 м, угловое разрешение — 0,9°, частота вращения — 4 Гц); треки сформированы из трехмерных аннотаций объектов и содержат те же шесть признаков детекций, что и при обработке RadarScenes. Минимальный порог для включения трека в обучающую выборку составлял четыре детекции и два кадра; максимальная длина трека ограничена 24 кадрами, среднее число детекций в кадре — два.

Постановка задачи

Задача, рассматриваемая в настоящей работе, обусловлена двумя структурными особенностями автомобильных радарных данных. Первая: одиночный радарный кадр содержит от одной до трех точек отражения — информации для надежной классификации там недостаточно, и распознавание опирается на всю временную последовательность кадров, а не на отдельный «снимок». Вторая: реальный трафик дает сильно несбалансированное распределение классов. В наборе Boreas-Objects-V1 редчайший класс занимает лишь 1,5 % треков — при таком соотношении стандартный классификатор просто не успевает выучить граничные условия для этой категории.

Пусть задано множество классов $\mathcal{C} = \{1, \ldots, K\}$, где K обозначает число распознаваемых категорий дорожных объектов. Радарный трек рассматривается как последовательность кадров, упорядоченных по времени:

$$\mathcal{T}_i = \left(\mathbf{X}_{i,1}, \mathbf{X}_{i,2}, \ldots, \mathbf{X}_{i,T_i}, y_i, s_i\right),$$
$$\mathbf{X}_{i,t} \in \mathbb{R}^{N_{i,t} \times d}, \; y_i \in \mathcal{C}, \qquad (1)$$

где $\mathcal{T}_i$ — i-й трек; $\mathbf{X}_{i,t}$ — матрица признаков t-го кадра этого трека; T_i — длина трека в кадрах; y_i — метка класса; s_i — идентификатор сцены записи; $N_{i,t}$ — число радарных отражений в кадре; d — размерность признакового описания одной детекции. Формула (1) задает структуру входных данных: каждый трек содержит не один вектор, а последовательность матриц переменного размера, поэтому дальнейшая модель должна уметь работать и с множеством точек внутри кадра, и с временной последовательностью кадров.

Чтобы исключить утечку данных, разбиение выполняется не по отдельным трекам, а по сценам записи [25]:

$$\forall \mathcal{T}_i \in \mathcal{D}_{train}, \forall \mathcal{T}_j \in \mathcal{D}_{test} : s_i \neq s_j,$$

где $\mathcal{D}_{train}$ и $\mathcal{D}_{test}$ — обучающая и тестовая подвыборки треков соответственно, а индексы i и j независимо пробегают треки из этих подвыборок. Это условие означает, что треки из одной и той же сцены не могут одновременно попасть в обучающую и тестовую выборки. Такое ограничение необходимо, поскольку соседние треки в пределах одной сцены могут иметь общий контекст движения, близкие условия наблюдения и схожие ошибки разметки, а значит, случайное разбиение по трекам могло бы завысить оценку качества модели.

Признаки одной детекции имеют разную физическую природу и разные численные масштабы, поэтому перед обучением они нормализуются по статистикам обучающей выборки:

$$\tilde{x}_{rj} = \frac{x_{rj} - \mu_j^{train}}{\sigma_j^{train} + \varepsilon},$$

где x_{rj} — значение j-го признака у r-й реальной детекции; μ_j^{train} и σ_j^{train} — среднее и стандартное отклонения этого признака, вычисленные только на обучающей части; ε — малая положительная константа для предотвращения деления на нуль. Нормализация делает признаки сопоставимыми по масштабу и предотвращает ситуацию, при которой признаки с большими численными значениями начинают доминировать в градиентном обучении только из-за своей размерности.

Кодировщик множества точек

Каждый радарный кадр представляет собой неупорядоченное множество отражений, поэтому его представление не должно зависеть от порядка строк во входной матрице. Для этого каждая детекция обрабатывается одним и тем же разделяемым многослойным персептроном $\phi(\cdot)$ — небольшой нейронной сетью с общими весами для всех точек, после чего точечные признаки объединяются симметричными операциями максимального и среднего агрегирования [10]:

$$\mathbf{h}_t = \psi\left(\left[\max_{n:m_{t,n}=1} \phi(\tilde{\mathbf{x}}_{t,n}); \frac{\sum_{n=1}^{N_{\max}} m_{t,n}\phi(\tilde{\mathbf{x}}_{t,n})}{\sum_{n=1}^{N_{\max}} m_{t,n}}\right]\right),$$

где $\mathbf{h}_t$ — итоговый вектор признаков кадра; $\psi(\cdot)$ — проекционный слой; $\phi(\cdot)$ — общий для всех точек персептрон; $\tilde{\mathbf{x}}_{t,n}$ — нормализованный вектор признаков n-й детекции в t-м кадре; $m_{t,n} \in \{0, 1\}$ — маска, отделяющая реальные детекции от нулевого заполнения. Максимальное агрегирование выделяет наиболее выраженные признаки среди детекций, среднее агрегирование описывает распределенную статистику всего кадра, а их конкатенация позволяет сохранить оба типа информации.

Преимуществом многослойного персептрона является инвариантность к перестановке точек внутри кадра. Если изменить порядок строк в $\mathbf{X}_t$, значения максимума и среднего по множеству реальных детекций не изменятся, поэтому и итоговый вектор $\mathbf{h}_t$ останется тем же. Это свойство важно именно для радарных данных, где порядок записей в таблице детекций не несет физического смысла [10].

Временной преобразователь с агрегирующим служебным токеном

После кодирования отдельных кадров требуется получить одно представление всего трека, поскольку класс объекта часто определяется не по форме одного кадра, а по изменению признаков во времени [7–9, 17, 18]. Для этого используется временной преобразователь на основе самовнимания с обучаемым служебным токеном CLS [14–16]:

$$\mathbf{z} = \text{Transformer}\big([\mathbf{c}_{\text{CLS}}; \mathbf{h}_1 + \mathbf{a}_1; \ldots ; \mathbf{h}_{T_{\max}} + \mathbf{a}_{T_{\max}}], \mathbf{m}^{frame}\big)_{\text{CLS}},$$

где $\mathbf{z}$ — итоговый вектор признаков всего трека; $\mathbf{c}_{\text{CLS}}$ — обучаемый токен агрегирования; $\mathbf{h}_t$ — представление t-го кадра; $\mathbf{a}_t$ — позиционный вектор, задающий номер кадра во временной последовательности; $\mathbf{m}^{frame}$ — маска реальных и дополненных кадров. Индекс CLS означает, что после обработки трансформером берется не среднее по всем выходам, а выход, соответствующий служебному токену.

Модель получает последовательность покадровых признаков и с помощью самовнимания оценивает, какие кадры важнее для классификации. Позиционные векторы необходимы, потому что без них трансформер воспринимал бы последовательность как набор элементов без временного порядка.

Классификационная голова на основе сравнения с обучаемыми прототипами классов

После получения трекового вектора $\mathbf{z}$ выполняется сравнение с обучаемыми прототипами классов. Пусть $\mathbf{P} \in \mathbb{R}^{K \times D}$ — матрица прототипов, где строка $\mathbf{p}_k$ соответствует k-му классу,

а D — размерность пространства эмбеддингов (embedding dimension). Косинусный логит k-го класса вычисляется как скалярное произведение нормированных векторов:

$$c_k = \left\langle \frac{\mathbf{z}}{\|\mathbf{z}\|_2}, \frac{\mathbf{p}_k}{\|\mathbf{p}_k\|_2} \right\rangle, \; k = 1, \ldots, K,$$

где $\mathbf{z}$ — вектор представления трека; $\mathbf{p}_k$ — вектор-прототип k-го класса, оба нормированы по евклидовой норме; c_k показывает угловое сходство между представлением трека и прототипом k-го класса. Нормирование по евклидовой норме $\|\cdot\|_2$ устраняет влияние длины векторов, поэтому решение зависит прежде всего от направления эмбеддинга, а не от его масштаба.

После перехода к угловому пространству возникает задача учесть неравномерность числа обучающих треков разных классов. Для этого в работе используется метод LDAM — подход к обучению на несбалансированных выборках, в котором к логиту истинного класса вводится аддитивный отступ, зависящий от представленности класса. В отличие от весов функции потерь, которые изменяют вклад примеров разных классов в значение ошибки, LDAM воздействует на само условие правильной классификации: на этапе обучения логит истинного класса уменьшается на величину m_{y_i}, поэтому модель должна распознать объект с дополнительным запасом по угловому сходству. Для редких классов этот запас задается больше, так как m_k выбирается пропорционально $n_k^{-1/4}$. Такая зависимость следует из анализа обобщающей способности классификатора при дисбалансе меток и позволяет аналитически связать отступ с числом обучающих примеров класса. Таким образом, к логиту истинного класса применяется классово-зависимый аддитивный отступ:

$$\tilde{c}_{ik} = c_{ik} - m_{y_i} \mathbb{I}[k = y_i],$$

где c_{ik} — косинусный логит k-го класса для i-го объекта; m_{y_i} — отступ истинного класса y_i; $\mathbb{I}[k = y_i]$ — индикатор, равный единице только для правильного класса. Отступ усложняет задачу классификации на обучении: чтобы объект был правильно распознан, логит истинного класса должен быть не просто максимальным, а достаточно большим после вычитания m_{y_i}.

После применения отступа логиты масштабируются классово-зависимой обратной температурой:

$$\ell_{ik} = \tau_k \tilde{c}_{ik}, \qquad (2)$$

где ℓ_{ik} — итоговый логит, передаваемый в softmax, а τ_k — обратная температура k-го класса. Чем больше τ_k, тем более резким становится распределение вероятностей для соответствующего класса, а меньшие значения τ_k делают решение мягче, что полезно для классов с ограниченным числом обучающих примеров.

Аналитическая настройка параметров дисбаланса

Параметры τ_k и m_k задаются через число обучающих образцов класса n_k, а не подбираются независимо для каждого класса. Сначала вычисляется эффективное число образцов k-го класса

$$E_k = \frac{1 - \beta^{n_k}}{1 - \beta},$$

где n_k — фактическое число обучающих треков этого класса; $\beta \in [0, 1)$ — параметр, задающий степень насыщения информации при добавлении похожих объектов. Если класс уже содержит много близких примеров, новый пример увеличивает E_k слабее, чем увеличивает простое количество n_k, поэтому эффективное число отражает не только объем, но и убывающий вклад дополнительных наблюдений.

На основе E_k вычисляется обратная температура:

$$\tilde{\tau}_k = E_k^{\alpha}, \; \tau_k = \bar{\tau} \frac{K \tilde{\tau}_k}{\sum_{j=1}^{K} \tilde{\tau}_j},$$

где $\tilde{\tau}_k$ — ненормированная обратная температура; α — параметр степени зависимости от представленности класса; $\bar{\tau}$ — заданное среднее значение обратной температуры; K — число классов. Нормировка во второй части формулы сохраняет средний масштаб логитов, потому что среднее значение τ_k по классам становится равным $\bar{\tau}$.

Классово-зависимый отступ задается по принципу LDAM:

$$m_k = m_{\max} \frac{n_k^{-1/4}}{\max_j n_j^{-1/4}},$$

где m_k — отступ для k-го класса; $m_{\max}$ — максимальное допустимое значение отступа, а множитель $n_k^{-1/4}$ делает отступ больше для редких классов и меньше для частых. Нормировка на $\max_j n_j^{-1/4}$ гарантирует, что самый большой отступ среди всех классов точно равен $m_{\max}$.

Функция потерь и задача оптимизации

Обучение модели выполняется по сумме основной классификационной ошибки и вспомогательной регрессионной ошибки:

$$L = L_{cls} + \lambda \cdot L_{aux}, \quad (3)$$

где L_{cls} — основная функция потерь классификации; L_{aux} — вспомогательная ошибка предсказания статистических моментов радиальной скорости; λ — коэффициент, задающий вклад вспомогательной задачи. Такая структура связывает обучение классификатора с физически значимыми характеристиками радарного сигнала, поскольку радиальная скорость и ее статистика несут информацию о типе движущегося объекта.

Основная классификационная ошибка записывается как взвешенная перекрестная энтропия (Cross-Entropy, CE):

$$L_{cls} = -\frac{1}{B}\sum_{i=1}^{B} w_{y_i} \sum_{k=1}^{K} q_{ik} \log \frac{\exp(\ell_{ik})}{\sum_{j=1}^{K} \exp(\ell_{ij})},$$

где B — размер мини-батча; w_{y_i} — вес истинного класса i-го объекта; q_{ik} — сглаженная целевая метка; ℓ_{ik} — логит после отступа и температурного масштабирования (2), а дробь внутри логарифма — вероятность класса после softmax. Веса w_{y_i} усиливают вклад редких классов, а сглаживание меток снижает переобучение за счет отказа от полностью жесткой целевой вероятности, равной единице только для истинного класса.

Вспомогательная ошибка имеет вид среднеквадратичной регрессии:

$$L_{aux} = \frac{1}{B}\sum_{i=1}^{B} \left\| \hat{\mathbf{s}}_i - \mathbf{s}_i \right\|_2^2,$$

где $\hat{\mathbf{s}}_i$ — предсказанный моделью вектор статистических моментов радиальной скорости, а $\mathbf{s}_i$ — соответствующий целевой вектор, вычисленный по данным трека. Минимизация этой ошибки заставляет скрытое представление $\mathbf{z}$ сохранять информацию о динамике объекта, что особенно важно для различения классов с похожей геометрией, но различным характером движения [18].

Классовые веса в основной функции потерь вычисляются через эффективное число образцов:

$$w_k = \frac{1/E_k}{\frac{1}{K}\sum_{j=1}^{K} 1/E_j}, \quad \frac{1}{K}\sum_{k=1}^{K} w_k = 1. \quad (4)$$

В выражении (4) редкие классы получают больший вес, потому что их эффективное число E_k меньше, а величина $1/E_k$ больше. Нормировка сохраняет средний вес равным единице, поэтому меняется относительный вклад классов, но общий масштаб функции потерь остается стабильным.

Задача оптимизации модели формулируется как минимизация средней функции потерь на обучающем множестве:

$$\theta^* = \arg\min_{\theta} \frac{1}{|\mathcal{D}_{train}|} \sum_{i:\mathcal{T}_i \in \mathcal{D}_{train}} L(\mathcal{T}_i, y_i; \theta),$$

где θ^* — значения параметров после обучения; θ — совокупность всех обучаемых параметров модели, включая параметры кодировщика, трансформера, прототипов классов и вспомогательной головы. Формула задает стандартную эмпирическую минимизацию риска, но с функцией потерь, специально адаптированной к дисбалансу классов и временной природе радарных треков.

Для выбора лучшей модели используется составная валидационная метрика:

$$Q_{val} = \frac{1}{2}\left(\mathrm{BA} + F1_{macro}\right),$$

где BA — сбалансированная точность, т. е. средняя полнота по классам; $F1_{macro}$ — макроусредненная F1-мера, одинаково учитывающая все классы независимо от их частоты. Такая метрика подходит для несбалансированных данных, потому что стандартная точность может быть высокой даже при плохом распознавании редких классов.

Контрольная базовая постановка использует линейную классификационную голову:

$$\mathbf{l}^{base} = \mathbf{W}\mathbf{z} + \mathbf{b},$$

где $\mathbf{l}^{base}$ — вектор логитов по классам; $\mathbf{W}$ — матрица весов линейного классификатора; $\mathbf{b}$ — вектор смещений. Это выражение служит базовым сравнением: в нем отсутствуют прототипы, косинусная нормализация, классово-зависимые отступы и температурное масштабирование.

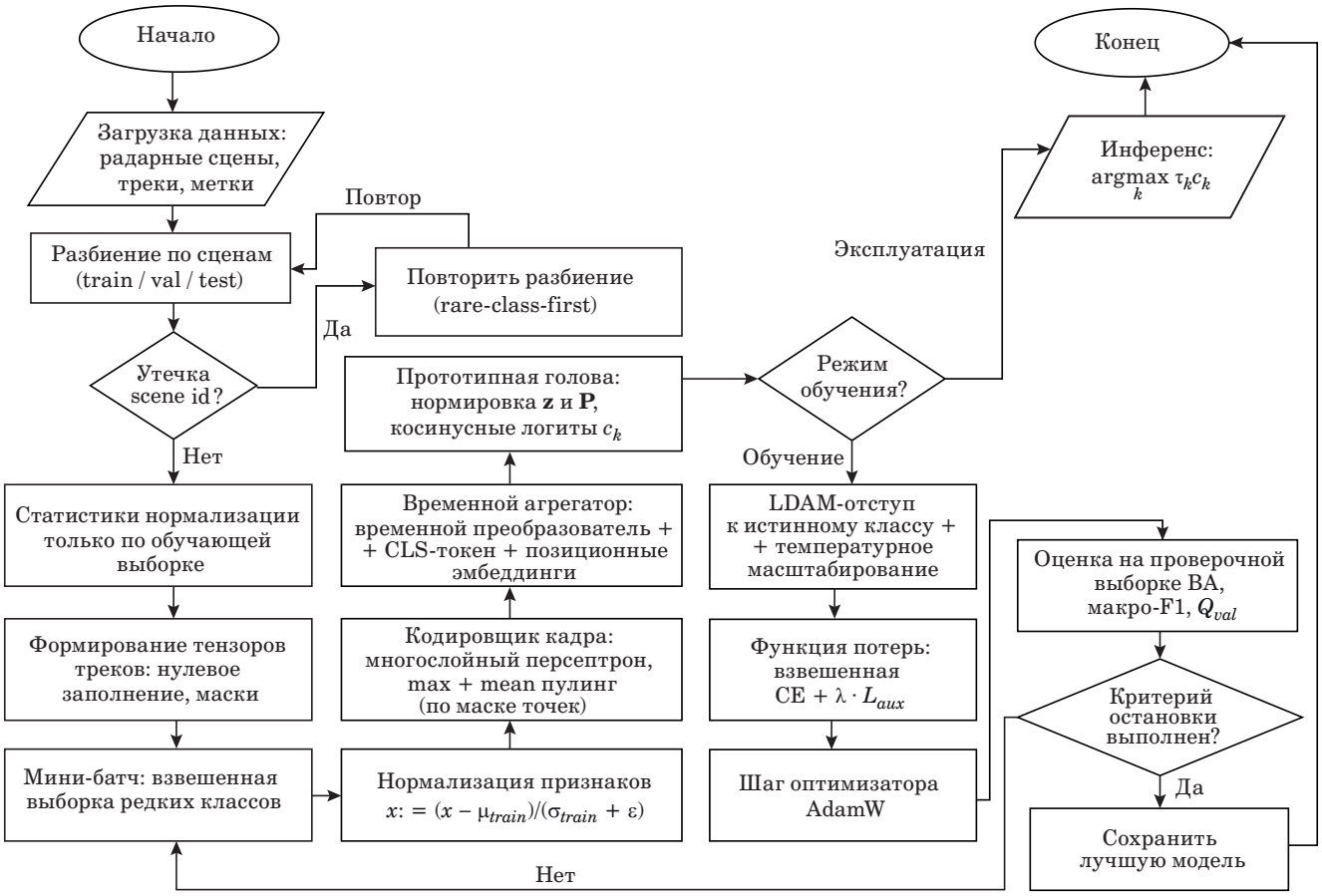
Протокол обучения

Исходные данные делятся по сценам записи в пропорции примерно 60/20/20: 20 % сцен направляются в валидационную выборку, 20 % — в тестовую, остальная часть — в обучающую. После разбиения программно проверяется отсутствие пересечений по идентификаторам сцен и треков между подмножествами: треки из одной записи не должны одновременно оказаться, например, в обучающей и тестовой выборках. Затем по обучающей части вычисляются статистики нормализации, формируются тензоры треков с нулевым заполнением и масками точек, а также рассчитываются классовые параметры n_k, E_k, τ_k, m_k и w_k, которые используются в классификационной голове и функции потерь.

Для уменьшения влияния дисбаланса классов мини-батчи формируются с помощью взвешенной случайной выборки. При таком способе треки редких классов получают большую вероятность попадания в батч, поэтому их вклад чаще учитывается при вычислении градиента. Это дополняет взвешивание функции потерь: выборка управляет составом батча, а веса w_k изменяют вклад классов уже внутри значения ошибки.

В ходе обучения для каждого мини-батча последовательно выполняются нормализация входных признаков, кодирование точек каждого кадра разделяемым многослойным персептроном, временная агрегация покадровых признаков, вычисление косинусных логитов, применение LDAM-отступа к логиту истинного класса, температурное масштабирование и расчет полной функции потерь. Параметры модели θ обновляются оптимизатором AdamW. На этапе инференса LDAM-отступ не используется, поскольку истинная метка объекта неизвестна; предсказание выполняется по температурно масштабированным косинусным логитам.

Общая структура предлагаемого алгоритма показана на рис. 1, где наглядно отображены этапы подготовки данных, их разбиение на сцены, формирование мини-батчей, нормализация признаков, покадровое кодирование, временная агрегация, прототипическая классификация, а также разделение режимов обучения и инференса. Схема отражает полный путь обработки данных: от подготовки сцен и формирования батчей до получения трекового представления, классификации и контроля качества на валидационной выборке. Отдельное ветвление подчеркивает различие между режимами обучения и инференса: при обучении используются отступ, функция потерь и обновление параметров, тогда как при инференсе выполняется только прямой проход модели с выбором класса по итоговым логитам. Возврат от критерия ранней остановки к формированию батча показывает итерационный характер обучения и связь процедуры оптимизации с валидационным контролем.



■ ***Рис. 1.*** Структурная схема алгоритма КВРТ
■ ***Fig. 1.*** Structural scheme of the temporal radar track classifier algorithm

Результаты экспериментов

Все эксперименты проводились на двух открытых наборах данных: RadarScenes (пять классов, 31 619 треков, доля редчайшего класса — twowheeler — менее 5 %) и Boreas-Objects-V1 (четыре класса, 4699 треков, экстремальный дисбаланс 66:1 по классу misc). На RadarScenes применялась скользящая проверка по записям на трех наборах (scene-level 3-fold CV) с тремя значениями инициализации генератора случайных чисел — итого девять независимых запусков. На Boreas-Objects-V1 использовалось единственное стратифицированное разбиение 70/30 по сценам (одно разбиение, два независимых запуска).

Во всех основных экспериментах модель обучалась 30 эпох с ранней остановкой по сбалансированной точности (BA) на валидационном разбиении. Оптимизатор — AdamW, скорость обучения — 3×10^{-4}, размер батча — 64, размерность итогового эмбеддинга трека (TRACK_DIM) = 256. Базовый метод отличался линейной головой классификатора, равномерными весами функции потерь, случайным разбиением треков, остальная же нейросетевая архитектура была идентична КВРТ.

Главным критерием качества является сбалансированная точность, дополненная макро-F1 и поклассовыми показателями полноты. Целевой порог полноты — recall ≥ 0,80 по каждому классу.

Общее число обучаемых параметров КВРТ составляет около 709 тыс., из которых 84 % сосредоточено в блоке временной агрегации — трехслойном преобразователе с четырьмя головами самовнимания. Кодировщик множества точек содержит 59 тыс. параметров, прототипная классификационная голова добавляет лишь 1280 параметров — по 256 на каждый из пяти классов. Вычислительная стоимость одного прямого прохода, оцененная аналитически для трека из 20 кадров при среднем числе отражений в кадре, равном двум, составляет около 29 млн операций с плавающей точкой. Поскольку классификация выполняется на уровне трека, а не отдельного кадра, задержка вывода не является узким местом: при частоте обновления радара 17 Гц трек из 20 кадров формируется примерно за 1,2 с, тогда как классификация занимает единицы миллисекунд, т. е. менее 1 % времени накопления трека. Это означает, что вычислительная нагрузка модели не является ограничивающим фактором при ее интеграции в систему помощи водителю.

Эксперименты на наборе данных RadarScenes

Модель обучалась и тестировалась с использованием трехкратной перекрестной проверки с тремя случайными инициализациями параметров, что дало девять независимых запусков. В каждом запуске вычислялись сбалансированная точность (BA), макро-F1 и поклассовые значения полноты (recall).

По результатам девяти запусков средние значения составили: BA = 0,9967±0,0048, макро-F1 = 0,9961±0,0058. Целевой порог полноты recall ≥ 0,80 был достигнут во всех классах во всех девяти запусках. В семи из девяти случаев значение BA равнялось 1,0; в двух оставшихся зафиксированы единичные ошибки классификации: два трека из 95 в разбиении 2 (инициализация 1) были перепутаны между классами largevehicle и twowheeler.

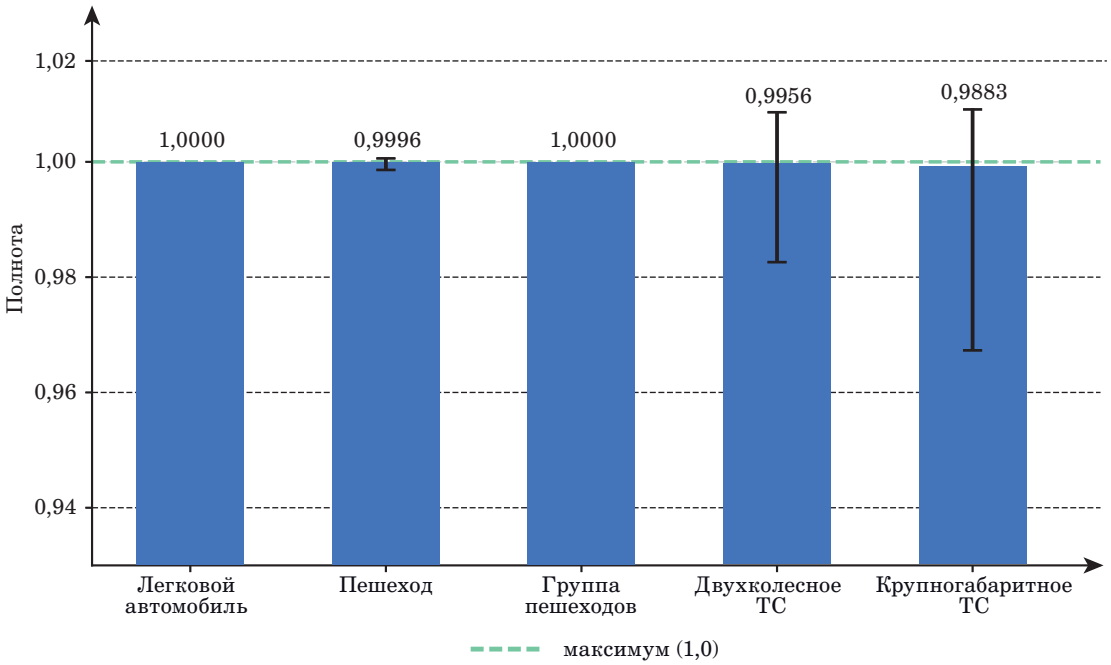
Поклассовые значения полноты представлены на рис. 2. Наиболее стабильные результаты показали классы car (легковой автомобиль) и pedgroup (группа пешеходов) (recall = 1,000±0,000 в обоих случаях). Класс pedestrian (пешеход) также демонстрирует высокую точность классификации: recall = 0,9996±0,001. Наибольший разброс наблюдается у классов largevehicle (крупногабаритное ТС) (recall = 0,9883±0,021) и twowheeler (двухколесное ТС) (recall = 0,9956±0,013), что соответствует их меньшей представленности в обучающей выборке.

Высокие показатели при низкой вариативности свидетельствуют о насыщении задачи при 30 эпохах обучения. Значение этого явления для интерпретации абляционного исследования подробно рассматривается ниже в подразделе «Абляционное исследование».

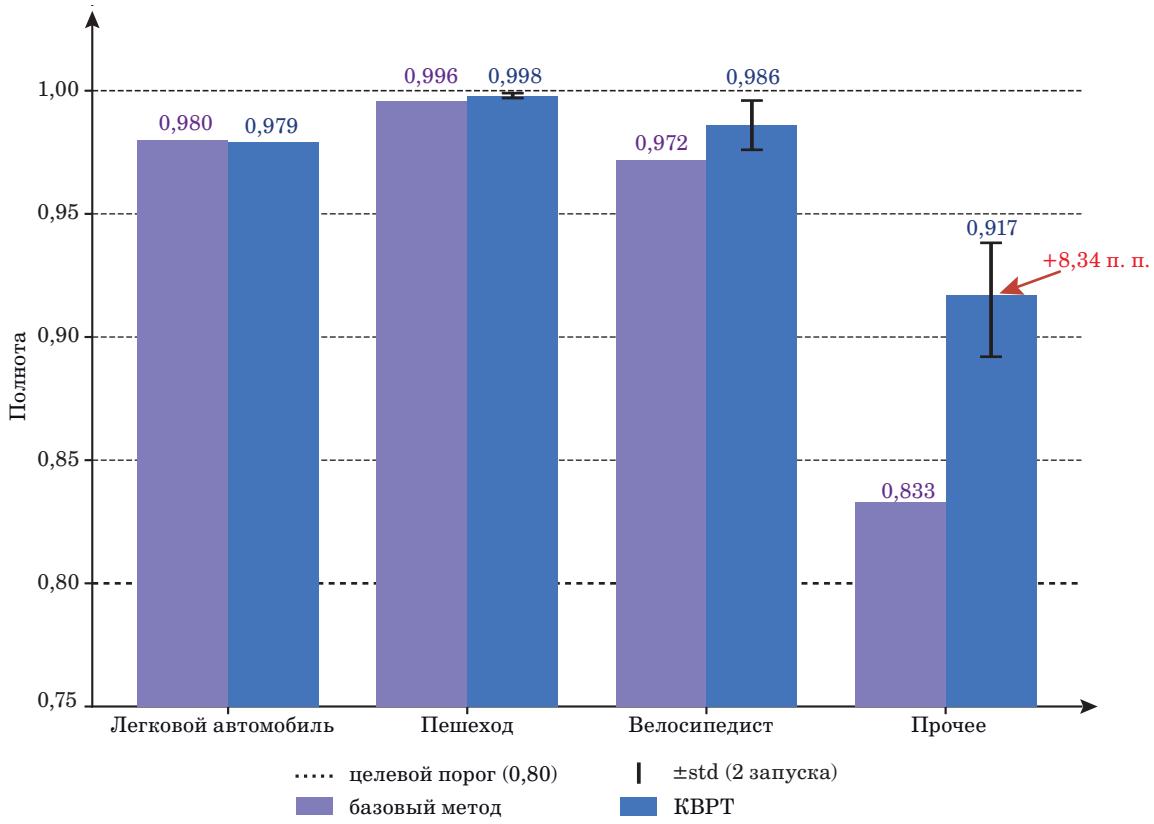
Сравнение с базовым методом на наборе данных Boreas-Objects-V1

Набор Boreas-Objects-V1 содержит 4699 треков, распределенных по четырем классам, при этом класс прочее представлен лишь 70 треками (1,5 % выборки). При стратифицированном разбиении 70/30 по сценам тестовая часть включает не более 20–22 треков этого класса, что делает данный набор подходящим для оценки вклада механизмов адаптации к дисбалансу классов. Каждый из двух методов — предложенный (КВРТ) и базовый — запускался дважды; для КВРТ приведены среднее значение и СКО по двум запускам.

Классификатор временных радарных треков превосходит базовый метод по всем ключевым метрикам. Сбалансированная точность составила 0,970±0,003 против 0,945 у базового метода (прирост 0,025), макро-F1 — 0,938±0,010 против 0,927 (прирост 0,011 п. п.). Наиболее выраженный эффект наблюдается для редчайшего класса misc: полнота возросла с 0,833 до 0,917±0,025 (прирост 0,083 п. п.). При этом адаптация к дисбалансу не привела к деградации на доминирующих классах: полнота car снизилась лишь на 0,001, тогда как pedestrian и cyclist показали прирост на 0,002 и 0,014 п. п. соответственно. Целевой порог полноты не ниже 0,80 выполнен обоими методами для всех классов.



■ ***Рис. 2.*** Поклассовые значения полноты модели КВРТ на наборе данных RadarScenes
■ ***Fig. 2.*** Per-class recall of the proposed method on RadarScenes



■ ***Рис. 3.*** Поклассовые значения полноты для КВРТ и базового метода на наборе данных Boreas-Objects-V1
■ ***Fig. 3.*** Per-class recall for the temporal radar track classifier and the baseline on Boreas-Objects-V1

Поклассовые значения полноты с планками погрешностей для КВРТ приведены на рис. 3.

Результат по классу misc (70 тестовых треков) свидетельствует о том, что базовый метод без балансировки при столь малом числе образцов стабильно недообучается на миноритарном классе, КВРТ же способно выделять паттерны данных в этих условиях.

Абляционное исследование

Для изолированной оценки вклада каждого компонента КВРТ было проведено абляционное исследование: восемь конфигураций на RadarScenes и пять конфигураций на Boreas-Objects-V1, по 30 эпох обучения, одна инициализация, одно разбиение на каждый набор данных. Конфигурации конструировались путем поочередного удаления или замены отдельных компонентов при прочих равных условиях. Исследование отвечает на три исследовательских вопроса, поставленных во введении.

RadarScenes. На наборе данных RadarScenes при 30 эпохах обучения задача классификации достигает насыщения: большинство конфигураций показывают BA = 1,000, т. е. неотличимы от полной модели. Это явление принято называть потолочным эффектом. Его причина состоит в том, что при достаточном числе эпох любая разумная конфигурация с корректной агрегацией по временной оси успешно решает данную задачу. В силу этого вклад отдельных компонентов: прототипной классификационной головы КВРТ, функции LDAM, классово-условного масштабирования температуры — не поддается изолированному измерению при 30 эпохах. Все результаты по RadarScenes сведены в табл. 1.

Приведенные данные позволяют сформулировать два вывода с различным уровнем надежности. Первый вывод, подкрепленный наиболее убедительными данными: агрегация по временной оси является структурно необходимой. При замене временного трансформера покадровой классификацией сбалансированная точность снижается с 0,998 до 0,327 ($\Delta = -0{,}671$). Класс pedgroup полностью исчезает из предсказаний: все 223 трека этого класса отнесены к car или largevehicle, что объясняется недостаточностью одной-трех точек одиночного радарного кадра для разграничения близких классов без учета временного контекста. При этом замена трансформерной агрегации простым усреднением по кадрам не приводит к деградации (BA = 1,000): принципиальную роль играет сам факт агрегации по временной оси, а не конкретная ее форма.

Измеримый вклад при 30 эпохах установлен для двух компонентов: вспомогательной головы микродоплеровских моментов и взвешивания функции потерь по классовым весам. Оба при исключении снижают сбалансированную точность на 0,007 (до 0,993) сугубо за счет ухудшения полноты класса twowheeler (с 1,000 до 0,965, т. е. –3,5 п. п.). Совпадение характера деградации в обоих случаях указывает на то, что данный редкий класс (≈ 5 % обучающей выборки) чувствителен к ослаблению градиентного сигнала, а не к конкретному механизму его усиления.

■ ***Таблица 1.*** Абляционное исследование (RadarScenes)
■ ***Table 1.*** Component exclusion study (RadarScenes)

Конфигурация	Исключаемый компонент	BA	Полнота		Примечание
			twowheeler	pedgroup	
Полная модель (КВРТ)	–	0,998	1,000	1,000	Контрольная конфигурация
Усреднение вместо трансформера	CLS-трансформер → простое усреднение	1,000	1,000	1,000	Потолочный эффект
Без прототипной головы	Голова КВРТ → линейная голова	1,000	1,000	1,000	То же
Без отступа LDAM	Аддитивный отступ m_k	1,000	1,000	1,000	– “ –
Без температурного масштабирования	Классово-условная температура τ_k	1,000	1,000	1,000	– “ –
Без вспомогательной головы	Вспомогательная голова микро-Доплера	0,993	0,965	1,000	–0,5 п. п. по BA
Без взвешенной функции потерь	Классово-условные веса функции потерь	0,993	0,965	1,000	–0,5 п. п. по BA
Покадровая классификация	Вся агрегация по временной оси	0,327	0,070	0,000	Критическое снижение

■ ***Таблица 2.*** Абляционное исследование (Boreas-Objects-V1)
■ ***Table 2.*** Component exclusion study (Boreas-Objects-V1)

Конфигурация	Исключаемый компонент	BA	Макро-F1	Полнота: cyclist
Полная модель (КВРТ)	–	0,973	0,922	0,976
Без выравнивающей функции потерь	Функция потерь на выравнивание вложений	0,973	0,922	0,976
Без проекционного адаптера входа	Адаптер размерности признакового пространства	0,973	0,922	0,976
Без временной агрегации	Трансформерная агрегация по кадрам	0,973	0,918	0,976
Без прототипной головы (межнаборная)	Голова КВРТ (межнаборная конфигурация)	0,978	0,918	1,000

Boreas-Objects-V1. На наборе данных Boreas-Objects-V1 абляционное исследование охватывало шесть конфигураций (одно случайное состояние). Ни одна из проверенных конфигураций не показала статистически значимого отклонения от полной модели: все значения сбалансированной точности находятся в диапазоне 0,972–0,978, что не превышает одного стандартного отклонения полной модели по двум запускам. Результаты приведены в табл. 2.

Отдельного пояснения требует конфигурация «без выравнивающей функции потерь»: ее исключение не изменяет ни одну из метрик. Это объясняется тем, что функция выравнивания вложений разработана для согласования признаковых пространств двух различных наборов данных при переносе модели; при обучении на одном наборе она не выполняет никакой значимой регуляризирующей функции. Конфигурация «без прототипной головы» соответствует замене прототипной головы линейным классификатором при использовании проекционного адаптера входа; показатель BA при этом не снижается, а полнота класса cyclist достигает 1,000 при незначительном снижении макро-F1.

Обсуждение

Эксперименты показали, что архитектура действительно работает там, где мы ожидали — на уровне трека, а не отдельного кадра. Покадровый блок устроен по принципу PointNet: детекции обрабатываются общим преобразованием и объединяются симметричной операцией, что дает инвариантность к порядку строк. Но, в отличие от классической задачи с плотными облаками точек, в одном радарном кадре их буквально 1–3 штуки, и именно поэтому основная различающая сила перемещается на уровень временной последовательности кадров.

Наборы данных RadarScenes и Boreas-Objects-V1 охватывают основные категории дорожных объектов: car, pedestrian, twowheeler и largevehicle. Однако они не отражают в полной мере реальный состав участников дорожного движения; ряд практически значимых категорий, таких как мотоциклисты (объединенные с велосипедистами в один класс) и дорожные рабочие, в обоих датасетах отдельно не представлены. Расширение классового состава при появлении соответствующих размеченных данных не требует переобучения модели целиком: прототипная классификационная голова допускает введение новых классов путем добавления прототипных векторов, что рассматривается как направление дальнейших исследований.

На наборе RadarScenes по трем разбиениям и трем случайным состояниям (девять запусков) была получена сбалансированная точность 0,997±0,005 и макро-F1 0,996±0,006, при этом целевой уровень полноты не ниже 0,80 был достигнут для всех пяти классов в каждом запуске.

Непосредственное сопоставление результатов с опубликованными результатами по другим методам затруднено различием входных представлений и использованных наборов данных. Метод RadarGNN [13], также примененный на RadarScenes, работает с отдельными кадрами облака точек в задачах сегментации и обнаружения объектов и достигает макроусредненной F1-меры 77,1 % при сегментации; предложенный же метод решает иную по постановке задачу — классификацию треков с явным учетом дисбаланса

классов, что не позволяет сравнивать эти числа напрямую. Метод RadarTCN [17] использует двумерные спектральные карты дальности и скорости в качестве входного представления и оценивается на наборе данных CARRADA, содержащем принципиально иной тип радарного сигнала по сравнению с разреженным облаком точек, составляющим основу RadarScenes и Boreas-Objects-V1; количественное сопоставление с RadarTCN в данных условиях методически некорректно. Таким образом, предложенный подход занимает самостоятельную нишу: в отличие от RadarGNN он ориентирован на временную последовательность кадров и адаптацию к дисбалансу, а в отличие от RadarTCN — оперирует разреженными облаками точек без предварительного формирования двумерных проекций.

Наиболее четкая разница проявилась на Boreas-Objects-V1, где класс прочее — всего 70 треков из 4699, т. е. 1,5 %. Именно здесь адаптация к дисбалансу показала себя: сбалансированная точность выросла с 0,945 до 0,970, а полнота на редчайшем классе — с 0,833 до 0,917 (+8,34 п. п.) при том, что на доминирующих классах деградации нет. Улучшение оказалось адресным — только там, где оно было нужно.

Абляционное исследование подтвердило структурную роль временной агрегации: при ее отключении и переходе к классификации на уровне отдельных кадров сбалансированная точность на RadarScenes снижается с 0,998 до 0,327, а полнота класса pedgroup обращается в ноль. Это согласуется с известной ролью CLS-агрегации в трансформерных архитектурах, но впервые количественно показано для классификации радарных треков. На RadarScenes часть упрощенных конфигураций также достигает BA = 1,000 за 30 эпох, что отражает эффект насыщения метрики и ограничивает возможность разделить вклад отдельных компонентов метрической головы на этом наборе данных.

Ограничения работы связаны с объемом валидации и областью применимости. На Boreas-Objects-V1 использовано одно разбиение и два случайных состояния, что не позволяет строго оценить статистическую значимость различий. Не проверена работа метода на треках, формируемых реальным детектором и трекером, а не на заранее размеченных треках наборов данных. Дальнейшие исследования включают воспроизведение экспериментов на Boreas-Objects-V1 в режиме «3 разбиения × 3 случайных состояния», сравнение с базовой моделью при одинаковой размерности трекового вектора, анализ чувствительности по параметрам τ_0, m_0 и β и унификацию постановки задачи для прямого сопоставления с методами семантической сегментации радарных данных, такими как RadarGNN.

Заключение

Разработан классификатор временных радарных треков с многомасштабной адаптацией к дисбалансу классов — КВРТ. Архитектура включает три блока: инвариантный к перестановке кодировщик детекций, временной трансформер с агрегирующим служебным токеном и классификационную голову на основе сравнения с обучаемыми эталонами классов. Параметры классово-условной температуры и аддитивного отступа выводятся аналитически из статистики обучающей выборки — подбирать гиперпараметры отдельно под каждый класс не нужно. Данные разбиваются на уровне сцен записи, а программная проверка исключает пересечения между подвыборками.

Проверка на двух открытых наборах данных дала устойчивые результаты в обоих режимах. На RadarScenes — сбалансированная точность 0,997±0,005, полнота не ниже 0,80 для каждого из пяти классов в каждом из девяти запусков. На Boreas-Objects-V1 при выраженном дисбалансе — прирост полноты редчайшего класса +8,340 п. п. без потерь на доминирующих категориях. Абляционное исследование зафиксировало главное: без временной агрегации задача классификации треков не решается — точность на RadarScenes падает с 0,998 до 0,327.

К направлениям дальнейших исследований относятся: статистическая валидация на Boreas-Objects-V1 в схеме «3 разбиения × 3 случайных состояния»; проверка метода на треках, формируемых реальным детектором; анализ чувствительности по ключевым гиперпараметрам классификационной головы — базовой температуре косинусных логитов τ_0, базовой величине аддитивного отступа LDAM m_0 и параметру эффективного числа образцов β, управляющему силой классово-условного перевзвешивания; унификация постановки задачи для прямого сопоставления с методами поточечной семантической сегментации радарных данных.

Финансовая поддержка

Исследование выполнено при поддержке Российского научного фонда (соглашение № 25-29-00868).

Литература

1. **Bijelic M., Gruber T., Mannan F., Kraus F., Ritter W., Dietmayer K., Heide F.** Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather. *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 11679–11689. doi:10.1109/CVPR42600.2020.01170, EDN: PADCFG
2. **Srivastav A., Mandal S.** Radars for autonomous driving: A review of deep learning methods and challenges. *IEEE Access*, 2023, vol. 11, pp. 97147–97168. doi:10.1109/ACCESS.2023.3312382, EDN: RQXILU
3. **Соколик Н. В.** Определение скорости движения и дальности быстродвижущихся объектов в РЛС с непрерывным линейно-частотно-модулированным излучением с использованием автокорреляционной схемы. *Известия высших учебных заведений России. Радиоэлектроника*, 2020, т. 23, № 2, с. 63–72. doi:10.32603/1993-8985-2020-23-2-63-72, EDN: JDERCM
4. **Сунь Х., Чжуан С., Костров А. А.** Обзор современных систем технического зрения, применяемых в транспортной отрасли. *Современные наукоемкие технологии*, 2024, № 9, с. 69–73. doi:10.17513/snt. 40150, EDN: BPHTGT
5. **Zhou Y., Liu L., Zhao H., López-Benítez M., Yu L., Yue Y.** Towards deep radar perception for autonomous driving: Datasets, methods, and challenges. *Sensors*, 2022, vol. 22, no. 11, Art. 4208. doi:10.3390/s22114208, EDN: IUZPZK
6. **Артюхин И. В., Аверин И. М., Флаксман А. Г., Рубцов А. Е.** Алгоритм оценки углов прихода сигналов в системе распределенных некогерентных автомобильных радаров. *Журнал радиоэлектроники* [электронный журнал], 2023, № 4. doi:10.30898/1684-1719.2023.4.2, EDN: XSEZVF. http://jre.cplire.ru/jre/about_e.html (дата обращения: 02.05.2025).
7. **Coppola R., Ahmed S., Alouini M. S.** Road users classification based on bi-frame micro-Doppler with 24-GHz FMCW radar. *Frontiers in Signal Processing*, 2022, vol. 2, Art. 864538. doi:10.3389/frsip.2022.864538, EDN: AUYPYP
8. **Басов О. О., Толстой И. М., Ле Ань Ту.** Классификация движущихся наземных объектов по их доплеровским портретам с помощью методов машинного обучения. *Современные наукоемкие технологии*, 2021, № 10, с. 17–22. doi:10.17513/snt.38848, EDN: WPJJSE
9. **Лук Д. В., Коновалов А. А., Хоанг Л. М.** Алгоритм распознавания малоразмерных воздушных целей по траекторным признакам в полуактивной РЛС. *Известия высших учебных заведений России. Радиоэлектроника*, 2023, т. 26, № 5, с. 76–88. doi:10.32603/1993-8985-2023-26-5-76-88, EDN: CZDSXO
10. **Qi C. R., Su H., Mo K., Guibas L. J.** PointNet: Deep learning on point sets for 3D classification and segmentation. *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 652–660. doi:10.1109/CVPR.2017.16
11. **Qi C. R., Yi L., Su H., Guibas L. J.** PointNet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in Neural Information Processing Systems (NeurIPS)*, 2017, vol. 30, pp. 5099–5108. arXiv:1706.02413
12. **Шпрехер Д. М., Минаков Е. И., Грачев А. Н., Башеров М. С.** Классификация воздушных целей в РЛС на основе нейросетевой модели с расширенным признаковым пространством. *Программные продукты и системы*, 2025, т. 38, № 4, с. 682–693. doi:10.15827/0236-235X.152.682-693, EDN: ZNAJAZ
13. **Fent F., Bauerschmidt P., Lienkamp M.** RadarGNN: Transformation invariant graph neural network for radar-based perception. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2023, pp. 182–191. doi:10.1109/CVPRW59228.2023.00023
14. **Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., Kaiser L., Polosukhin I.** Attention is all you need. *Advances in Neural Information Processing Systems (NeurIPS)*, 2017, vol. 30, pp. 5998–6008. arXiv:1706.03762
15. **Devlin J., Chang M.-W., Lee K., Toutanova K.** BERT: Pre-training of deep bidirectional transformers for language understanding. *Proc. 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, 2019, pp. 4171–4186. doi:10.18653/v1/N19-1423
16. **Люткин Д. А., Поздняков Д. В., Соловьев А. А., Жуков Д. В., Малик М. Ш. И., Игнатов Д. И.** Применение трансформеров для определения профильного врача на основе запросов пользователей. *Автоматика и телемеханика*, 2024, № 3, с. 86–100. doi:10.31857/S0005231024030076, EDN: TQAELK
17. **Li Y., Zhang M., Jing H., Liu Z.** RadarTCN: Lightweight online classification network for automotive radar targets based on TCN. *Sensors*, 2024, vol. 24, no. 9, Art. 2813. doi:10.3390/s24092813, EDN: UXADLX
18. **Schumann O., Hahn M., Scheiner N., Weishaupt F., Tilly J. F., Dickmann J., Wöhler C.** RadarScenes: A real-world radar point cloud data set for automotive applications. Proc. *24th International Conference on Information Fusion (FUSION)*, 2021, pp. 1–8. doi:10.23919/FUSION49465.2021.9627037
19. **Константинов А. Ф., Дьяконова Л. П.** Сравнительный анализ методов снижения дисбаланса классов при построении моделей машинного обучения в финансовом секторе. *Известия Кабардино-Балкарского научного центра РАН*, 2025, т. 27, № 1, с. 143–151. doi:10.35330/1991-6639-2025-27-1-143-151, EDN: XRYMDH
20. **Cui Y., Jia M., Lin T.-Y., Song Y., Belongie S.** Class-balanced loss based on effective number of samples. *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 9260–9269. doi:10.1109/CVPR.2019.00949

21. **Cao K., Wei C., Gaidon A., Arechiga N., Ma T.** Learning imbalanced datasets with label-distribution-aware margin loss. *Advances in Neural Information Processing Systems (NeurIPS)*, 2019, vol. 32, pp. 1567–1578. arXiv:1906.07413
22. **Zhang Y., Kang B., Hooi B., Yan S., Feng J.** Deep long-tailed learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, vol. 45, no. 9, pp. 10795–10816. doi:10.1109/TPAMI.2023.3268118, EDN: LSGVDA
23. **Deng J., Guo J., Yang J., Xue N., Liu Z., Gong Y., Zafeiriou S.** ArcFace: Additive angular margin loss for deep face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, vol. 44, no. 10, pp. 5962–5979. doi:10.1109/TPAMI.2021.3087709, EDN: XKIJAV
24. **Snell J., Swersky K., Zemel R.** Prototypical networks for few-shot learning. *Advances in Neural Information Processing Systems (NeurIPS)*, 2017, vol. 30, pp. 4077–4087. arXiv:1703.05175
25. **Burnett K., Yoon D. J., Wu Y., Li A. Z., Zhang H., Lu S., Qian J., Tseng W.-K., Lambert A., Leung K. Y. K., Schoellig A. P., Barfoot T. D.** Boreas: A multi-season autonomous driving dataset. *The International Journal of Robotics Research*, 2023, vol. 42, no. 1–2, pp. 33–42. doi:10.1177/02783649231160195, EDN: HIRFYI

UDC 004.93'1:621.396.96
doi:10.31799/1684-8853-2026-4-60-74
EDN: GGKFJQ

Temporal track classifier for automotive radar with multi-scale adaptation to class imbalance

E. A. Lopukhova[a], Junior Researcher, orcid.org/0000-0001-6802-6010
G. S. Voronkov[a], PhD, Tech., Associate Professor, orcid.org/0000-0002-8788-2696, voronkov.gs@ugatu.su
E. P. Topolskaya[a], PhD, Tech., Associate Professor, orcid.org/0000-0003-2207-2702
[a]Ufa University of Science and Technology, 32, Z. Validi St., 450076, Ufa, Russian Federation

Introduction: The recognition of road objects from automotive radar data is complicated by the small number of reflections in individual frames and by class imbalance, which leaves rare objects underrepresented during training. **Purpose:** To develop a radar-track classifier which is robust to class imbalance and applicable to open road-traffic radar datasets. **Methods:** We have developed and applied a radar-track classification method that combines frame-wise processing of unordered sets of radar reflections by a shared multilayer perceptron with invariant feature aggregation, temporal transformer-based aggregation with a learnable classification token, and a prototype-based classification head in a normalized cosine space. Class imbalance is considered through the use of a margin dependent on the label distribution, class-dependent temperature scaling, loss-function weights calculated from the effective number of samples, and weighted sampling during mini-batch formation; auxiliary prediction of statistical moments of radial velocity has been additionally applied. **Results:** The experiments have used RadarScenes and Boreas-Objects-V1. RadarScenes contains 158 sequences with point-wise labels for five classes, whereas Boreas-Objects-V1 contains 4,699 tracks with three-dimensional labels for four classes, including a rarest class with a 1.5% share. On Boreas-Objects-V1, balanced accuracy has increased from 0.945 to 0.970, and the recall of the rarest class has increased from 0.833 to 0.917. The ablation study has shown that removing temporal aggregation reduces balanced accuracy on RadarScenes from 0.998 to 0.327, which confirms the structural necessity of sequential information. **Practical relevance:** The classifier can be applied for training radar-based road-object recognition systems under limited rare-class examples.

Keywords — automotive radar, radar track, road-object recognition, class imbalance, temporal aggregation, metric learning, neural networks.

For citation: Lopukhova E. A., Voronkov G. S., Topolskaya E. P. Temporal track classifier for automotive radar with multi-scale adaptation to class imbalance. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2026, no. 4, pp. 60–74 (In Russian). doi:10.31799/1684-8853-2026-4-60-74, EDN: GGKFJQ

Financial support
The research is supported by Russian Science Foundation (agreement No. 25-29-00868).

References

1. Bijelic M., Gruber T., Mannan F., Kraus F., Ritter W., Dietmayer K., Heide F. Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather. *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 11679–11689. doi:10.1109/CVPR42600.2020.01170, EDN: PADCFG
2. Srivastav A., Mandal S. Radars for autonomous driving: A review of deep learning methods and challenges. *IEEE Access*, 2023, vol. 11, pp. 97147–97168. doi:10.1109/ACCESS.2023.3312382, EDN: RQXILU
3. Sokolik N. V. Velocity and range measurement of fast-moving objects in a continuous linear frequency-modulated radar using an autocorrelation scheme. *Journal of the Russian Universities. Radioelectronics*, 2020, vol. 23, no. 2, pp. 63–72 (In Russian). doi:10.32603/1993-8985-2020-23-2-63-72, EDN: JDERCM
4. Sun H., Zhuang S., Kostrov A. A. An overview of modern vision systems used in the transport industry. *Modern High Technologies*, 2024, no. 9, pp. 69–73 (In Russian). doi:10.17513/snt.40150, EDN: BPHTGT
5. Zhou Y., Liu L., Zhao H., López-Benítez M., Yu L., Yue Y. Towards deep radar perception for autonomous driving: Datasets, methods, and challenges. *Sensors*, 2022, vol. 22, no. 11, Art. 4208. doi:10.3390/s22114208, EDN: IUZPZK
6. Artyukhin I. V., Averin I. M., Flaksman A. G., Rubtsov A. E. Direction-of-Arrival Estimation Algorithm for a Distributed Non-Coherent Automotive Radar System. *Journal of Radio Electronics* [online journal], 2023, no. 4. doi:10.30898/1684-1719.2023.4.2, EDN: XSEZVF. Available at: http://jre.cplire.ru/jre/about_e.html (accessed 2 May 2025) (In Russian).
7. Coppola R., Ahmed S., Alouini M. S. Road users classification based on bi-frame micro-Doppler with 24-GHz FMCW radar. *Frontiers in Signal Processing*, 2022, vol. 2, Art. 864538. doi:10.3389/frsip.2022.864538, EDN: AUYPYP
8. Basov O. O., Tolstoy I. M., Le Anh Tu. Classification of moving ground objects by their doppler signatures using machine learning methods. *Modern High Technologies*, 2021, no. 10, pp. 17–22 (In Russian). doi:10.17513/snt.38848, EDN: WPJJSE
9. Luk D. V., Konovalov A. A., Hoang L. M. Algorithm for recognition of small-sized aerial targets by trajectory features

in a semi-active radar. *Journal of the Russian Universities. Radioelectronics*, 2023, vol. 26, no. 5, pp. 76–88 (In Russian). doi:10.32603/1993-8985-2023-26-5-76-88, EDN: CZDSXO
10. Qi C. R., Su H., Mo K., Guibas L. J. PointNet: Deep learning on point sets for 3D classification and segmentation. *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 652–660. doi:10.1109/CVPR.2017.16
11. Qi C. R., Yi L., Su H., Guibas L. J. PointNet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in Neural Information Processing Systems (NeurIPS)*, 2017, vol. 30, pp. 5099–5108. arXiv:1706.02413
12. Shprekher D. M., Minakov E. I., Grachev A. N., Basherov M. S. Radar target classification based on a neural network model with an extended feature space. *Software & Systems*, 2025, vol. 38, no. 4, pp. 682–693 (In Russian). doi:10.15827/0236-235X.152.682-693, EDN: ZNAJAZ
13. Fent F., Bauerschmidt P., Lienkamp M. RadarGNN: Transformation invariant graph neural network for radar-based perception. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2023, pp. 182–191. doi:10.1109/CVPRW59228.2023.00023
14. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., Kaiser L., Polosukhin I. Attention is all you need. *Advances in Neural Information Processing Systems (NeurIPS)*, 2017, vol. 30, pp. 5998–6008. arXiv:1706.03762
15. Devlin J., Chang M.-W., Lee K., Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. *Proc. 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, 2019, pp. 4171–4186. doi:10.18653/v1/N19-1423
16. Lyutkin D. A., Pozdnyakov D. V., Solovev A. A., Zhukov D. V., Malik M. Sh. I., Ignatov D. I. Application of transformers to specialist physician identification based on user queries. *Automation and Remote Control*, 2024, no. 3, pp. 86–100 (In Russian). doi:10.31857/S0005231024030076, EDN: TQAELK
17. Li Y., Zhang M., Jing H., Liu Z. RadarTCN: Lightweight online classification network for automotive radar targets based on TCN. *Sensors*, 2024, vol. 24, no. 9, Art. 2813. doi:10.3390/s24092813, EDN: UXADLX
18. Schumann O., Hahn M., Scheiner N., Weishaupt F., Tilly J. F., Dickmann J., Wöhler C. RadarScenes: A real-world radar point cloud data set for automotive applications. Proc. *24th International Conference on Information Fusion (FUSION)*, 2021, pp. 1–8. doi:10.23919/FUSION49465.2021.9627037
19. Konstantinov A. F., Dyakonova L. P. Comparative analysis of class imbalance reduction methods for machine learning models in the financial sector. *News of the Kabardino-Balkarian Scientific Center of RAS*, 2025, vol. 27, no. 1, pp. 143–151 (In Russian). doi:10.35330/1991-6639-2025-27-1-143-151, EDN: XRYMDH
20. Cui Y., Jia M., Lin T.-Y., Song Y., Belongie S. Class-balanced loss based on effective number of samples. *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 9260–9269. doi:10.1109/CVPR.2019.00949
21. Cao K., Wei C., Gaidon A., Arechiga N., Ma T. Learning imbalanced datasets with label-distribution-aware margin loss. *Advances in Neural Information Processing Systems (NeurIPS)*, 2019, vol. 32, pp. 1567–1578. arXiv:1906.07413
22. Zhang Y., Kang B., Hooi B., Yan S., Feng J. Deep long-tailed learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, vol. 45, no. 9, pp. 10795–10816. doi:10.1109/TPAMI.2023.3268118, EDN: LSGVDA
23. Deng J., Guo J., Yang J., Xue N., Liu Z., Gong Y., Zafeiriou S. ArcFace: Additive angular margin loss for deep face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, vol. 44, no. 10, pp. 5962–5979. doi:10.1109/TPAMI.2021.3087709, EDN: XKIJAV
24. Snell J., Swersky K., Zemel R. Prototypical networks for few-shot learning. *Advances in Neural Information Processing Systems (NeurIPS)*, 2017, vol. 30, pp. 4077–4087. arXiv:1703.05175
25. Burnett K., Yoon D. J., Wu Y., Li A. Z., Zhang H., Lu S., Qian J., Tseng W.-K., Lambert A., Leung K. Y. K., Schoellig A. P., Barfoot T. D. Boreas: A multi-season autonomous driving dataset. *The International Journal of Robotics Research*, 2023, vol. 42, no. 1–2, pp. 33–42. doi:10.1177/ 02783649231160195, EDN: HIRFYI

ПАМЯТКА ДЛЯ АВТОРОВ

Поступающие в редакцию статьи проходят обязательное рецензирование.

При наличии положительной рецензии статья рассматривается редакционной коллегией. Принятая в печать статья направляется автору для согласования редакторских правок. После согласования автор представляет в редакцию окончательный вариант текста статьи.

Процедуры согласования текста статьи могут осуществляться как непосредственно в редакции, так и по e-mail (ius.spb@gmail.com).

При отклонении статьи редакция представляет автору мотивированное заключение и рецензию, при необходимости доработать статью — рецензию.

Редакция журнала напоминает, что ответственность за достоверность и точность рекламных материалов несут рекламодатели.

АНДРЕЕВ Илья Андреевич

Оператор роты (научной) Главного управления связи, Военная академия связи им. Маршала Советского Союза С. М. Буденного, Санкт-Петербург.
В 2025 году окончил магистратуру Евразийского национального университета им. Л. Н. Гумилева и Северо-Кавказского Федерального университета по направлению «Информационная безопасность».
Является автором 10 научных публикаций.
Область научных интересов – кибербезопасность систем интернета вещей, доверенное взаимодействие устройств, методы аутентификации периферийных устройств, адаптивное обнаружение киберугроз и др.
Эл. адрес:
andreev.ilia.1984@mail.ru

ДАВЫДОВ Вячеслав Анатольевич

Доцент Московского института электроники и математики, ВШЭ, старший научный сотрудник лаборатории киберфизических систем Института проблем управления РАН, Москва.
В 1992 году окончил Ленинградский институт авиационного приборостроения по специальности «Автоматизированные системы обработки информации и управления».
В 1995 году защитил диссертацию на соискание ученой степени кандидата технических наук, 2017 году – кандидата экономических наук.
Является автором 40 научных публикаций.
Область научных интересов – алгебраическая и комбинаторная теория кодирования, теория активных систем, теория блокчейн.
Эл. адрес: v.davydov@hse.ru

КРУК Евгений Аврамович

Профессор, научный руководитель Научно-исследовательского института телекоммуникаций ВШЭ, Москва. Почетный работник науки и техники РФ, заслуженный деятель науки РФ.
В 1973 году окончил Ленинградский институт авиационного приборостроения по специальности «Автоматизированные системы управления».
В 1999 году защитил диссертацию на соискание ученой степени доктора технических наук.
Является автором более 250 научных работ, восьми монографий и 48 патентов на изобретения.
Область научных интересов – криптография, компьютерная безопасность, помехоустойчивое кодирование.
Эл. адрес: ekrouk@hse.ru

ВОРОНКОВ Григорий Сергеевич

Старший научный сотрудник НИЛ «Сенсорные системы на основе устройств интегральной фотоники» Уфимского университета науки и технологий.
В 2007 году окончил с отличием Уфимский государственный авиационный технический университет по специальности «Средства связи с подвижными объектами».
В 2017 году защитил диссертацию на соискание ученой степени кандидата технических наук.
Является автором 62 научных публикаций.
Область научных интересов – обработка сигналов, интегральная фотоника, СВЧ-сигналы.
Эл. адрес: voronkov.gs@ugatu.su

ЗАДБОЕВ Вадим Александрович

Младший научный сотрудник Военной академии связи им. Маршала Советского Союза С. М. Буденного, Санкт-Петербург.
В 2017 году окончил Северо-Кавказский федеральный университет по специальности «Информационная безопасность автоматизированных систем».
Является автором 42 научных публикаций и трех патентов на изобретения.
Область научных интересов – информационная безопасность, системы контроля и управления безопасностью информационно-телекоммуникационных и информационно-вычислительных систем.
Эл. адрес: zadboev89@mail.ru

ЛЕВАШОВА Татьяна Викторовна

Старший научный сотрудник лаборатории интегрированных систем автоматизации Санкт-Петербургского Федерального исследовательского центра РАН.
В 1986 году окончила Ленинградский электротехнический институт им. В. И. Ульянова (Ленина) по специальности «Электронные вычислительные машины».
В 2009 году защитила диссертацию на соискание ученой степени кандидата технических наук.
Является автором более 200 научных публикаций.
Область научных интересов – управление знаниями, инженерия онтологий, управление онтологиями, управление контекстом, социокиберфизические системы.
Эл. адрес:
tatiana.levashova@iias.spb.su

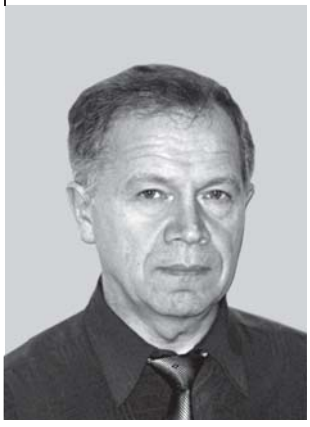
ЛИПАТНИКОВ Валерий Алексеевич

Профессор, старший научный сотрудник Военной академии связи им. Маршала Советского Союза С. М. Буденного, Санкт-Петербург, заслуженный изобретатель РФ, член-корреспондент РАЕН.
В 1974 году окончил Военную академию связи им. Маршала Советского Союза С. М. Буденного по специальности «Специальная радиотехника».
В 2000 году защитил диссертацию на соискание ученой степени доктора технических наук.
Является автором более 300 научных публикаций и 80 патентов на изобретения.
Область научных интересов – теория многоуровневой иерархической радиоэлектронной защиты, безопасности связи и информации инфотелекоммуникационных сетей.
Эл. адрес: lipatnikovanl@mail.ru

МАКАРЕНКО Денис Владимирович

Адъюнкт Военной академии связи им. Маршала Советского Союза С. М. Буденного, Санкт-Петербург.
В 2020 году окончил Военную академию связи им. Маршала Советского Союза С. М. Буденного по специальности «Применение и эксплуатация автоматизированных систем специального назначения».
Является автором 10 научных публикаций и одного патента на изобретение.
Область научных интересов – теория многоуровневой иерархической радиоэлектронной защиты, безопасности связи и информации инфотелекоммуникационных сетей.
Эл. адрес: Nsdivdeni@yandex.ru

МОЛДОВЯН Николай Андреевич



Профессор, главный научный сотрудник Научного центра информационных технологий и искусственного интеллекта АНОО ВО «Университет «Сириус».
В 1975 году окончил Кишиневский политехнический институт по специальности «Полупроводниковые приборы».
В 2001 году защитил диссертацию на соискание ученой степени доктора технических наук.
Является автором более 300 научных публикаций и 70 патентов на изобретения.
Область научных интересов – компьютерная безопасность, криптография, симметричные и асимметричные криптосистемы, электронная цифровая подпись, аутентификация, блочные шифры, псевдовероятностные шифры, постквантовая криптография.
Эл. адрес:
moldovyan.na@talantiuspeh.ru

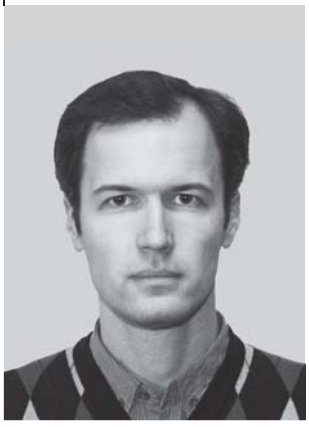
ЛОПУХОВА Екатерина Александровна

Младший научный сотрудник, аспирант кафедры телекоммуникационных систем Уфимского университета науки и технологий.
В 2021 году окончила Уфимский государственный авиационный технический университет по направлению «Инфокоммуникационные технологии и системы специальной связи».
Является автором 50 научных публикаций и восьми результатов интеллектуальной деятельности.
Область научных интересов – интеллектуальная обработка сигналов, машинное обучение и анализ данных, классификация радарных целей, нейронные сети в задачах телекоммуникаций и медицинской диагностики.
Эл. адрес: lopukhova.ea@ugatu.su

МОЛДОВЯН Александр Андреевич

Профессор, главный научный сотрудник лаборатории кибербезопасности и постквантовых криптосистем Санкт-Петербургского института информатики и автоматизации РАН.
В 1974 году окончил Ленинградский электротехнический институт им. В. И. Ульянова (Ленина) по специальности «Автоматизированные системы управления».
В 2005 году защитил диссертацию на соискание ученой степени доктора технических наук.
Является автором более 200 научных публикаций и 60 патентов на изобретения.
Область научных интересов – компьютерная безопасность, защита информации, криптография, протоколы электронной цифровой подписи.
Эл. адрес: maa1305@yandex.ru

ПОНОМАРЕВ Андрей Васильевич



Старший научный сотрудник лаборатории интегрированных систем автоматизации Санкт-Петербургского института информатики и автоматизации РАН.
В 2003 году окончил Тюменский государственный нефтегазовый университет по специальности «Автоматизированные системы обработки информации и управления».
В 2012 году защитил диссертацию на соискание ученой степени кандидата технических наук.
Является автором 38 научных публикаций.
Область научных интересов – крауд-вычисления и краудсорсинг, рекомендующие системы, машинное обучение, системы поддержки принятия решений.
Эл. адрес:
ponomarev@iias.spb.su

ПОПЕРЕЧНЫЙ Павел Сергеевич

Ведущий инженер ООО «КНС Групп», Москва.
В 2006 году окончил Томский государственный университет систем управления и радиоэлектроники по специальности «Радиоэлектронные системы».
В 2017 году защитил диссертацию на соискание ученой степени кандидата технических наук.
Является автором 17 научных публикаций, десяти патентов на изобретения и девяти патентов на полезную модель.
Область научных интересов – теория помехоустойчивого кодирования.
Эл. адрес: yncherepor@mail.ru

ТОПОЛЬСКАЯ Елизавета Павловна

Доцент кафедры телекоммуникационных систем, заведующая научно-исследовательской лабораторией «Сенсорные системы на основе устройств интегральной фотоники» Уфимского университета науки и технологий.
В 2012 году окончила Уфимский государственный авиационный технический университет по специальности «Средства связи с подвижными объектами».
В 2016 году защитила диссертацию на соискание ученой степени кандидата технических наук.
Является автором 200 научных публикаций и 13 результатов интеллектуальной деятельности.
Область научных интересов — цифровая обработка сигналов, интегральная фотоника, кремниевая радиофотоника, сенсорные системы и антенные устройства.
Эл. адрес: eorlingsbest@mail.ru

СМИРНОВ Александр Викторович

Профессор, главный научный сотрудник, заведующий лабораторией интегрированных систем автоматизации Санкт-Петербургского Федерального исследовательского центра РАН.
В 1979 году окончил Ленинградский электротехнический институт им. В. И. Ульянова (Ленина) по специальности «Системы автоматического управления».
В 1994 году защитил диссертацию на соискание ученой степени доктора технических наук.
Является автором более 350 научных публикаций.
Область научных интересов — концептуальное моделирование, управление знаниями, управление контекстом, конфигурирование систем, системы поддержки принятия решений, социокиберфизические системы.
Эл. адрес: smir@iias.spb.su

ХРУСТАЛЕВ Владимир Викторович

Аспирант кафедры аэрокосмических компьютерных и программных систем Санкт-Петербургского государственного университета аэрокосмического приборостроения.
В 2021 году окончил магистратуру Санкт-Петербургского государственного университета аэрокосмического приборостроения по специальности «Информатика и вычислительная техника».
Является автором семи научных публикаций, одного патента на изобретение и одного патента на полезную модель.
Область научных интересов – теория помехоустойчивого кодирования.
Эл. адрес: v_crys@mail.ru

Уважаемые авторы!

При подготовке рукописей статей необходимо руководствоваться следующими рекомендациями.

Статьи должны содержать изложение новых научных результатов. Название статьи должно быть кратким, но информативным. В названии недопустимо использование сокращений, кроме самых общепринятых (РАН, РФ, САПР и т. п.).

Текст рукописи должен быть оригинальным, а цитирование и самоцитирование корректно оформлено.

Объем статьи (текст, таблицы, иллюстрации и библиография) не должен превышать эквивалента в 20 страниц, напечатанных на бумаге формата А4 на одной стороне через 1,5 интервала Word шрифтом Times New Roman размером 13, поля не менее двух сантиметров.

Обязательными элементами оформления статьи являются: индекс УДК, заглавие, инициалы и фамилия автора (авторов), ученая степень, звание (при отсутствии — должность), полное название организации, аннотация и ключевые слова на русском и английском языках, ORCID и электронный адрес одного из авторов. При написании аннотации не используйте аббревиатур и не делайте ссылок на источники в списке литературы. Предоставляйте подрисуночные подписи и названия таблиц на русском и английском языках.

Статьи авторов, не имеющих ученой степени, рекомендуется публиковать в соавторстве с научным руководителем, наличие подписи научного руководителя на рукописи обязательно; в случае самостоятельной публикации обязательно предоставляйте заверенную по месту работы рекомендацию научного руководителя с указанием его фамилии, имени, отчества, места работы, должности, ученого звания, ученой степени.

Простые **формулы** набирайте в Word, сложные с помощью редактора Mathtype или Equation. Для набора одной формулы не используйте два редактора; при наборе формул в формульном редакторе знаки препинания, ограничивающие формулу, набирайте вместе с формулой; для установки размера шрифта в Mathtype никогда не пользуйтесь вкладкой Other, Smaller, Larger, используйте заводские установки редактора, не подгоняйте размер символов в формулах под размер шрифта в тексте статьи, не растягивайте и не сжимайте мышью формулы, вставленные в текст; пробелы в формуле ставьте только после запятых при перечислении с помощью Ctrl+Shift+Space (пробел); не отделяйте пробелами знаки: + = – ×, а также пространство внутри скобок; для выделения греческих символов в Mathtype полужирным начертанием используйте Style → Other → bold.

Для набора формул в Word никогда не используйте вкладки: «Уравнение», «Конструктор», «Формула» (на верхней панели: «Вставка» — «Уравнение»), так как этот ресурс предназначен только для внутреннего использования в Word и не поддерживается программами, предназначенными для изготовления оригинал-макета журнала.

При наборе символов в тексте помните, что символы, обозначаемые латинскими буквами, набираются светлым курсивом, русскими и греческими — светлым прямым, векторы и матрицы — прямым полужирным шрифтом.

Подробнее см. http://i-us.ru/index.php/ius/author-guide

Иллюстрации:

— рисунки, графики, диаграммы, блок-схемы предоставляйте в виде отдельных исходных файлов, поддающихся редактированию, используя векторные программы: Visio (*.vsd, *.vsdx); Adobe Illustrator (*.ai); Coreldraw (*.cdr, версия не выше 15); Excel (*.xls); Word (*.docx); AutoCad, Matlab (экспорт в PDF, EPS, SVG, WMF, EMF); Компас (экспорт в PDF); веб-портал DRAW.IO (экспорт в PDF); Inkscape (экспорт в PDF);

— фото и растровые — в формате *.tif, *.png с максимальным разрешением (не менее 300 pixels/inch).

Наличие подрисуночных подписей и названий таблиц на русском и английском языках обязательно (желательно не повторяющих дословно комментарии к рисункам в тексте статьи).

В редакцию предоставляются:

— сведения об авторе (фамилия, имя, отчество, место работы, должность, ученое звание, учебное заведение и год его окончания, ученая степень и год защиты диссертации, область научных интересов, количество научных публикаций, домашний и служебный адреса и телефоны, e-mail), фото авторов: анфас, в темной одежде на белом фоне, должны быть видны плечи и грудь, высокая степень четкости изображения без теней и отблесков на лице, фото можно представить в электронном виде в формате *.tif, *.png, *.jpg с максимальным разрешением — не менее 300 pixels/inch при минимальном размере фото 40×55 мм;

— экспертное заключение;

— экспортное заключение.

Список литературы составляется по порядку ссылок в тексте и оформляется следующим образом:

— для книг и сборников — фамилия и инициалы авторов, полное название книги (сборника), город, издательство, год, общее количество страниц, doi;

— для журнальных статей — фамилия и инициалы авторов, полное название статьи, название журнала, год издания, номер журнала, номера страниц, doi;

— ссылки на иностранную литературу следует давать на языке оригинала без сокращений;

— при использовании web-материалов указывайте адрес сайта и дату обращения.

Список литературы оформляйте двумя отдельными блоками по образцам lit.dot на сайте журнала (http://i-us.ru/paperrules): Литература и References.

Более подробно правила подготовки текста с образцами изложены на нашем сайте в разделе «Руководство для авторов» – http://i-us.ru/index.php/ius/author-guide.

Контакты

Куда: 190000, г. Санкт-Петербург, ул. Большая Морская, д. 67, лит. А, ГУАП, РИЦ
Кому: Редакция журнала «Информационно-управляющие системы»
Эл. почта: ius.spb@gmail.com
Сайт: www.i-us.ru