

ТРУДЫ УЧЕБНЫХ ЗАВЕДЕНИЙ СВЯЗИ

Темы номера:

- ✓ Энергетическая эффективность в многосотовых massive MIMO
- ✓ Переход к восьмому поколению технологии Wi-Fi
- ✓ Методика балансировки трафика в сетях SDN

Федеральное государственное бюджетное образовательное
учреждение высшего образования
«Санкт-Петербургский государственный университет телекоммуникаций
им. проф. М.А. Бонч-Бруевича» (СПбГУТ)

Научный журнал

ТРУДЫ
УЧЕБНЫХ ЗАВЕДЕНИЙ СВЯЗИ

Том 12. № 2

Proceedings of Telecommunication Universities

Vol. 12. Iss. 2

Санкт-Петербург

2026

Описание журнала

Научный журнал. Включен 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 (распоряжение Минобрнауки России № 21-р от 12.02.2019), по специальностям (распоряжение № 33-р от 01.02.2022):

1.2.2. Математическое моделирование, численные методы и комплексы программ

2.2.6. Оптические и оптико-электронные приборы и комплексы

2.2.13. Радиотехника, в том числе системы и устройства телевидения

2.2.14. Антенны, СВЧ-устройства и их технологии

2.2.15. Системы, сети и устройства телекоммуникаций

2.2.16. Радиолокация и радионавигация

2.3.1. Системный анализ, управление и обработка информации, статистика

2.3.6. Методы и системы защиты информации, информационная безопасность

Журнал позиционирует себя как научный, в связи с этим его целями являются ознакомление научной общественности (научного сообщества) с результатами оригинальных исследований, выполненных ведущими учеными и специалистами и их коллективами, а также апробация научных результатов, полученных при подготовке кандидатских и докторских диссертаций для повышения качества (уровня) проводимых исследований. Издание ставит перед собой задачу расширения инфокоммуникативного пространства взаимодействия российских и зарубежных ученых. Целевой аудиторией журнала являются ученые и специалисты-практики в области связи и телекоммуникаций и смежных направлениях науки и техники, а также профессорско-преподавательский состав и студенты, обучающиеся по программам аспирантуры, магистратуры, специалитета и бакалавриата профильных вузов и кафедр.

Выпускается с 1960 года. Выходит 6 раз в год. Издается на русском и английском языках.

Редакционный совет

Киричек Р.В. <i>Главный редактор</i>	д.т.н., проф., Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича (СПбГУТ), г. Санкт-Петербург, Россия
Владыко А.Г. <i>Зам. Главного редактора</i>	к.т.н., доцент, Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича (СПбГУТ), г. Санкт-Петербург, Россия
Буйневич М.В.	д.т.н., проф., Санкт-Петербургский университет государственной противопожарной службы МЧС России, г. Санкт-Петербург, Россия
Зеневич А.О.	д.т.н., проф., Белорусская государственная академия связи, г. Минск, Республика Беларусь
Розанов Н.Н.	д.ф.-м.н., проф., чл.-корр. РАН, АО «Государственный оптический институт им. С.И. Вавилова» (ГОИ), г. Санкт-Петербург, Россия
Дукельский К.В.	д.т.н., доцент, АО «Государственный оптический институт им. С.И. Вавилова» (ГОИ), г. Санкт-Петербург, Россия
Кучерявый Е.	PhD, Технологический университет Тампере, г. Тампере, Финляндия
Каримов Б.Т.	к.т.н., доцент, Институт электроники и телекоммуникаций, Кыргызский государственный технический университет И. Раззакова (КГТУ), г. Бишкек, Кыргызстан
Тиамийу О.А.	PhD, Университет Илорина, г. Илорин, Нигерия
Козин И.Д.	д.ф.-м.н., проф., Алматинский университет энергетики и связи, г. Алма-Аты, Казахстан
Самуйлов К.Е.	д.т.н., проф., Российский университет дружбы народов (РУДН), г. Москва, Россия
Степанов С.Н.	д.т.н., проф., Московский технический университет связи и информатики (МТУСИ), г. Москва, Россия
Росляков А.В.	д.т.н., проф., Поволжский государственный университет телекоммуникаций и информатики (ПГУТИ), г. Самара, Россия
Кучерявый А.Е.	д.т.н., проф., Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича (СПбГУТ), г. Санкт-Петербург, Россия
Канаев А.К.	д.т.н., проф., Петербургский университет путей сообщения имени Александра I (ПГУПС), г. Санкт-Петербург, Россия
Новиков С.Н.	д.т.н., проф., Сибирский государственный университет телекоммуникаций и информатики (СибГУТИ), г. Новосибирск, Россия
Дворников С.В.	д.т.н., проф., Военная академия связи им. Маршала Советского Союза С.М. Буденного (ВАС), г. Санкт-Петербург, Россия
Коржик В.И.	д.т.н., проф., Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича (СПбГУТ), г. Санкт-Петербург, Россия
Ковалгин Ю.А.	д.т.н., проф., Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича (СПбГУТ), г. Санкт-Петербург, Россия

Description

Scientific journal. The journal is included in the List of reviewed scientific publications, in which the main scientific results of dissertations for the degree of candidate of science and for the degree of doctor of science should be published (order of the Ministry of Education and Science of Russia No 21-r of 12 February 2019) in the field of (order of the Ministry of Education and Science of Russia No 33-r of 01 February 2022):

1.2.2. Mathematical modeling, numerical methods and complexes of programs

2.2.6. Optical and optoelectronic devices and complexes

2.2.13. Radio engineering, including television systems and devices

2.2.14. Antennas, microwave devices and its technologies

2.2.15. Systems, networks and telecommunication devices

2.2.16. Radiolocation and radio navigation

2.3.1. System analysis, management and information processing, statistics

2.3.6. Methods and systems of information security, cybersecurity

The journal positions itself as a scientific one, in this regard, its goals are to familiarize the scientific community (scientific community) with the results of original research carried out by leading scientists and specialists and their teams, as well as approbation of scientific results obtained in the preparation of candidate and doctoral dissertations to improve the quality (level) of ongoing research. The publication sets itself the task of expanding the infocommunicative space of interaction between Russian and foreign scientists. The target audience of the journal are scientists and practitioners in the field of communications & telecommunications and related fields of science & technology, as well as faculty and students enrolled in postgraduate, master's, specialisation and bachelor's programs of profiled universities and departments.

Since 1960. Published 6 times per year. Published in Russian and English.

Editorial Board

R.V. Kirichek DSc, prof., The Bonch-Bruевич Saint Petersburg State University
Editor-in-chief of Telecommunications (SPbSUT), Saint Petersburg, Russia

A.G. Vladiko PhD, associate prof., The Bonch-Bruевич Saint Petersburg State University
Deputy editor-in-chief of Telecommunications (SPbSUT), Saint Petersburg, Russia

M.V. Buinevich DSc, prof., Saint-Petersburg University of State Fire Service of EMERCOM of Russia,
Saint Petersburg, Russia

A.O. Zenevich DSc, prof., Belarusian State Academy of Communications, Minsk, Republic of Belarus

N.N. Rozanov DSc, prof., member-corr. RAS, Open Joint Stock Company «S.I. Vavilov State Optical Institute» (SOI),
Saint Petersburg, Russia

K.V. Dukel'skii DSc, associate prof., Open Joint Stock Company «S.I. Vavilov State Optical Institute» (SOI),
Saint Petersburg, Russia

Y. Koucheryavy PhD, Tampere University of Technology, Tampere, Finland

B.T. Karimov PhD, Institute of Electronics and Telecommunications, Kyrgyz State Technical University
named after I. Razzakov, Bishkek, Kyrgyzstan

O.A. Tiarniyu PhD, University of Ilorin, Ilorin, Nigeria

I.D. Kozin DSc, prof., Almaty University of Power Engineering and Telecommunications, Almaty, Kazakhstan

K.E. Samuilov DSc, prof., Peoples' Friendship University (RUDN), Moscow, Russia

S.N. Stepanov DSc, prof., Moscow Technical University of Communication and Informatics (MTUCI),
Moscow, Russia

A.V. Roslyakov DSc, prof., Povolzhskiy State University of Telecommunications and Informatics (PSUTI), Samara,
Russia

A.E. Koucheryavy DSc, prof., The Bonch-Bruевич Saint Petersburg State University of Telecommunication (SPbSUT),
Saint Petersburg, Russia

A.K. Kanaev DSc, prof., Emperor Alexander I-st Petersburg State Transport University (PSTU),
Saint Petersburg, Russia

S.N. Novikov DSc, prof., Siberian State University of Telecommunications and Information Sciences (SibSUTIS),
Novosibirsk, Russia

S.V. Dvornikov DSc, prof., Military Academy of Telecommunications named after Marshal Union S.M. Budyonny,
Saint Petersburg, Russia

V.I. Korzhik DSc, prof., The Bonch-Bruевич Saint Petersburg State University of Telecommunication (SPbSUT),
Saint Petersburg, Russia

Yu.A. Kovalgin DSc, prof., The Bonch-Bruевич Saint Petersburg State University of Telecommunication (SPbSUT),
Saint Petersburg, Russia

РЕГИСТРАЦИОННАЯ ИНФОРМАЦИЯ / REGISTRATION INFORMATION

Журнал зарегистрирован Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций: ПИ № 77-77501 от 17.01.2020 г. (пред. рег. № 77-17986 от 07.04.2004 г.)

Размещение в РИНЦ (elibrary.ru) по договору: № 59-02/2013R от 20.02.2013

Registered by Federal Service for Supervision of Communications, Information Technology and Mass Media on 17.01.2020: PI No. 77-77501 (prev. reg. on 04.07.2004: No. 77-17986)

Accommodation in RINC (elibrary.ru) by agreement on 20.02.2013: No. 59-02/2013R



Товарный знак № 929373.

Правообладатель:

Федеральное государственное бюджетное образовательное учреждение высшего образования «Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича» (СПбГУТ)

191186, Санкт-Петербург, наб. реки Мойки, 61, литера А

Trademark No. 929373.

Copyright holder:

Federal State Budget-Financed Educational Institution of Higher Education «The Bonch-Bruevich Saint Petersburg State University of Telecommunications» (SPbSUT)

191186, St. Petersburg, emb. Moika River, 61, letter A

КОНТАКТНАЯ ИНФОРМАЦИЯ / CONTACT INFORMATION

Учредитель и издатель: Федеральное государственное бюджетное образовательное учреждение высшего образования «Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича» (СПбГУТ)

Адрес учредителя: 191186, Санкт-Петербург, набережная реки Мойки, д. 61, литера А

Адрес редакции: 193232, Санкт-Петербург,

пр. Большевиков, 22/1, к. 606/1

Тел.: +7 (812) 326-31-63, м. т. 2022

E-mail: tuzs@sut.ru

Web: <http://tuzs.sut.ru>

ВК: <http://vk.com/spbtuzs>

Ответственный редактор **Татарникова И.М.**

Выпускающий редактор **Яшугин Д.Н.**

Дизайн: **Коровин В.М.**, изображение на обложке сгенерировано ИИ Freepik <https://ru.freepik.com>

Publisher: Federal State Budget-Financed Educational Institution of Higher Education «The Bonch-Bruevich Saint Petersburg State University of Telecommunications» (SPbSUT)

Publisher address: 191186, Saint Petersburg, Moika river embankment, 61-A

Post address: 193232, Saint Petersburg, Prospekt Bolshevikov, 22/1

Phone: +7 (812) 326-31-63, local 2022

E-mail: tuzs@sut.ru

Web: <http://tuzs.sut.ru>

Executive Editor **Tatarnikova I.M.**

Commissioning Editor **Yashugin D.N.**

Design: **Korovin V.M.**, cover image generated by AI Freepik <https://ru.freepik.com>

ВЫХОДНЫЕ ДАННЫЕ / IMPRINT

Дата выхода в свет: 05.01.2026
Тираж: 1000 экз. Цена свободная.

Отпечатано в типографии
Федерального государственного бюджетного образовательного учреждения высшего образования «Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича»

Release date: 05.01.2026
Circulation: 1000 copies. Free price.

Printed in the printing office
Federal State Budget-Financed Educational Institution of Higher Education «The Bonch-Bruevich Saint Petersburg State University of Telecommunications»



СОДЕРЖАНИЕ

CONTENTS

КОМПЬЮТЕРНЫЕ НАУКИ И ИНФОРМАТИКА

- | | | |
|--|---|--|
| <p>Кротов К.В.
Целочисленная оптимизация составов и расписаний выполнения пакетов в системах Flow Shop при изменяющейся периодичности обслуживания приборов</p> | 7 | <p>Krotov K.V.
Integer optimization of package compositions and schedules in Flow Shop systems with varying frequency of instrument maintenance</p> |
|--|---|--|

ЭЛЕКТРОНИКА, ФОТОНИКА, ПРИБОРОСТРОЕНИЕ И СВЯЗЬ

- | | | |
|--|----|---|
| <p>Глушак Е.В.
Анализ отказоустойчивости в IoT-системах на основе туманных вычислений</p> | 27 | <p>Glushak E.V.
Fault tolerance analysis in IoT systems based on fog computing</p> |
| <p>Дмитриева Ю.С., Елагин В.С.
Разработка и апробация методики балансировки трафика в сетях SDN на базе комбинированного параметра QoS</p> | 36 | <p>Dmitrieva Yu.S., Elagin V.S.
Development and testing of traffic balancing techniques in SDN networks based on the combined QoS parameter</p> |
| <p>Нгием В.Д., Дворкович А.В.
Энергетическая эффективность в многосотовых massive MIMO: влияние схем обработки и параметров аппаратной реализации</p> | 45 | <p>Nghiem V.D., Dvorkovich A.V.
Energy efficiency in multi-cell massive MIMO: impact of signal processing schemes and hardware implementation parameters</p> |
| <p>Никитин Ю.А., Филин В.А.
Моделирование параллельного нониусного цифроаналогового преобразователя второго типа</p> | 53 | <p>Nikitin Yu.A., Filin V.A.
Simulating parallel vernier second type digital-to-analog converter</p> |
| <p>Скоробогатова С.А., Викулов А.С., Парамонов А.И.
Анализ проблем проектирования сетей Wi-Fi в свете перехода к восьмому поколению технологии</p> | 64 | <p>Skorobogatova S.A., Vikulov A.S., Paramonov A.I.
Analysis of Wi-Fi network design problems considering transition to Wi-Fi 8</p> |

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И ТЕЛЕКОММУНИКАЦИИ

- | | | |
|---|-----|---|
| <p>Буй Н.Ч., Пащенко Ф.Ф., То Ф.Т.
Представление промышленного трафика и кодирование протокольной семантики для трансформерных моделей в малых сетях Industrial Ethernet</p> | 74 | <p>Bui N.T., Pashchenko F.F., To P.T.
Representation of industrial traffic and encoding of protocol semantics for transformer models in small industrial ethernet networks</p> |
| <p>Опарин Е.В., Прошин Ф.А.
Анализ последствий атак на элементы системы синхронизации, функционирующие по алгоритму BTCA на основе протокола точного времени</p> | 92 | <p>Oparin E.V., Proshin F.A.
An analysis of a sequence to attacks on synchronization system elements operating according to BTCA based on precision time protocol</p> |
| <p>Шевченко А.А., Задбоев В.А., Липатников В.А., Мелехов К.В., Кузин П.И.
Модель процесса мониторинга сетевой безопасности сети передачи данных в условиях многоэтапных атак</p> | 102 | <p>Shevchenko A.A., Zadboev V.A., Lipatnikov V.A., Melekhov K.V., Kuzin P.I.
A Model for the Network Security Monitoring Process in Data Transmission Networks Under Multi-Stage Attack Conditions</p> |
| <p>Шелухин О.И., Маторин Ф.А.
Синтез реалистичных синтетических данных с помощью диффузионной модели TabDDPM в задачах обнаружения сетевых атак</p> | 113 | <p>Sheluhin O.I., Matorin F.A.
Realistic synthetic data generation using the TabDDPM diffusion model for network attack detection</p> |

КОМПЬЮТЕРНЫЕ НАУКИ И ИНФОРМАТИКА

1.2.2 – Математическое моделирование, численные методы и комплексы программ

Научная статья

УДК 004:519.854

<https://doi.org/10.31854/1813-324X-2026-12-2-7-25>

EDN:VOGGBN



Целочисленная оптимизация составов и расписаний выполнения пакетов в системах Flow Shop при изменяющейся периодичности обслуживания приборов

Кирилл Викторович Кротов, krotov_k1@mail.ru

Севастопольский государственный университет,
Севастополь, 299053, Российская Федерация

Аннотация

Реализация процессов выполнения пакетов заданий в системах Flow Shop подвержена влиянию отказов приборов и их простоев при восстановлении. Поддержание заданного уровня надежности систем обеспечивается периодическим техническим обслуживанием их приборов, связанным с устранением неполадок, являющихся причинами отказов. В течение интервалов времени обслуживания приборы являются недоступными для реализации назначенных им функций. Планирование обслуживания приборов позволяет определить интервалы времени между их реализациями. Однако положение реализаций обслуживания приборов может не являться фиксированным, а только – входить во временные интервалы заданной длительности (временные «окна»). В этом случае положение реализаций обслуживания приборов оптимизируется с учетом характера процессов выполнения пакетов заданий в системах. В то же время выполняемые в системах задания, включаемые в пакеты, могут входить в состав заказов, для которых заданы директивные сроки окончания исполнения. Вследствие этого **актуальна** задача оптимизации составов пакетов заданий, включения пакетов в интервалы доступности приборов не фиксированной длительности, порядков выполнения пакетов в этих интервалах при вхождении заданий в заказы, с определенными для них директивными сроками. При незначительных размерах задач их решения могут быть получены путем использования смешанного целочисленного линейного программирования.

Цель работы состоит в формировании новой математической модели целочисленного программирования оптимизации решений указанного вида. Для ее достижения использованы **методы** построения моделей математического программирования. На первом этапе реализовано формирование нелинейной математической модели. Получены выражения, используемые для построения ограничений, соответствующих распределению всех заданий по пакетам. С целью снижения временных затрат на получение решений проведена линеаризация модели. Для ее верификации разработано приложение в среде IBM ILOG CPLEX. В ходе исследований получены **результаты**, которые показали эффективность модели для решения задач планирования выполнения пакетов заданий в системах Flow Shop при изменяющейся периодичности обслуживания и условия вхождения заданий в заказы с определенными для них директивными сроками исполнения. Результаты имеют **практическую значимость** при решении задач малой размерности построения расписаний выполнения пакетов заданий в производственных системах.

Ключевые слова: математические модели смешанного целочисленного линейного программирования, периодическое техническое обслуживание приборов, пакеты заданий, конвейерные системы, расписания

Ссылка для цитирования: Кротов К.В. Целочисленная оптимизация составов и расписаний выполнения пакетов в системах Flow Shop при изменяющейся периодичности обслуживания приборов // Труды учебных заведений связи. 2026. Т. 12. № 2. С. 7–25. DOI:10.31854/1813-324X-2026-12-2-7-25. EDN:VOGGBN

Original research

<https://doi.org/10.31854/1813-324X-2026-12-2-7-25>

EDN:VOGGBN

Integer Optimization of Package Compositions and Schedules in Flow Shop Systems with Varying Frequency of Instrument Maintenance

 Kirill V. Krotov, krotov_k1@mail.ru

Sevastopol State University,
Sevastopol, 299053, Russian Federation

Annotation

The implementation of job package execution processes in Flow Shop systems is affected by instrument failures and downtime during recovery. Maintaining a given level of reliability of the systems is ensured by periodic maintenance of their devices related to troubleshooting, which are the causes of failures. During the maintenance time intervals, the devices are unavailable for the implementation of their assigned functions. Device maintenance planning allows you to determine the time intervals between their implementations. However, the position of the instrument maintenance implementations may not be fixed, but may fall within time intervals of a given duration (time "windows"). In this case, the position of instrument maintenance implementations is optimized taking into account the nature of the processes of completing task packages in the systems. At the same time, tasks performed in systems that are included in packages may be part of orders for which specific deadlines have been set. As a result, the task of optimizing the composition of task packages, including packages in the availability intervals of devices of non-fixed duration, and the order of execution of packages in these intervals when tasks enter orders, with specific deadlines for them, is urgent. With small problem sizes, their solutions can be obtained by using mixed integer linear programming.

The purpose of the work is to form a new mathematical model of integer programming for optimizing solutions of this type. **Methods** of constructing mathematical programming models were used to achieve this goal. At the first stage, the formation of a nonlinear mathematical model was implemented. Expressions are obtained that are used to construct constraints corresponding to the distribution of all tasks into packages. In order to reduce the time spent on obtaining solutions, the model was linearized. To verify it, an application has been developed in the IBM ILOG CPLEX environment. In the course of the research, **results** were obtained that showed the effectiveness of the model for solving the tasks of planning the execution of task packages in Flow Shop systems with varying service frequency and the condition that tasks are included in orders with specific deadlines for their execution. The results are of **practical importance** in solving small-dimensional problems of constructing scenarios for completing task packages in production systems.

Keywords: mathematical models of mixed integer linear programming, periodic maintenance of devices, task packages, conveyor systems, schedules

For citation: Krotov K.V. Integer Optimization of Package Compositions and Schedules in Flow Shop Systems with Varying Frequency of Instrument Maintenance. *Proceedings of Telecommunication Universities*. 2026;12(2):7–25. (in Russ.) DOI:10.31854/1813-324X-2026-12-2-7-25. EDN:VOGGBN

Введение

Дисциплина обслуживания заданий в конвейерных системах (КС) или системах Flow Shop предполагает последовательное прохождение ими всех приборов, входящих в эти системы. Типизация заданий, выполняющихся в этих системах, осуществляется с точки зрения равных значений времени

осуществления операций с ними на приборах систем и равных значений времени переналадок приборов на реализацию действий с ними. Для увеличения производительности процессов выполнения заданий разных типов в указанных системах осуществляется формирование пакетов однотипных заданий. Задания одного типа, включенные в па-

кеты, выполняются на приборах конвейерных систем последовательно. То есть, время окончания осуществления операций с пакетом равно моменту времени окончания выполнения последнего задания в нем. Переналадка приборов реализуется при переходе от действий с заданиями в одном пакете к проведению операций с заданиями в другом пакете. Тогда пакет – это совокупность однотипных заданий, выполняемых на приборах последовательно без переналадки.

В КС (системах Flow Shop) операции с заданиями, включенными в пакеты, могут быть выполнены в течение интервалов времени функционирования приборов, после окончания каждого из которых реализуется их (приборов) периодическое техническое обслуживание (ПТО). Реализация ПТО приборов позволяет идентифицировать неисправности в них, устранять эти неисправности, что обеспечивает исключение значительных отказов приборов, восстановление после которых занимает длительное время. К политикам реализации ПТО приборов относятся: 1) проведение ПТО через заданные интервалы времени функционирования приборов; 2) оптимизация периодов между ПТО приборов для обеспечения максимальной их надежности при ограничениях на используемые ресурсы (в частности, при ограничении простоев приборов, обеспечении заданной эффективности процессов выполнения заданий в КС); 3) поддержание заданного уровня надежности системы при оптимизации эффективности процессов выполнения заданий в КС. При осуществлении политики первого вида действия, позволяющие поддерживать приборы в работоспособном состоянии, реализуются через определенные интервалы времени. Тогда в течение интервалов времени заданной длительности между ПТО приборы осуществляют операции с заданиями, при реализации ПТО являются недоступными. Однако длительности интервалов времени доступности приборов могут быть не фиксированными, моменты времени окончания этих интервалов могут входить в установленные временные интервалы, которые названы временными «окнами». В этом случае положение ПТО приборов и, соответственно, положения интервалов времени их недоступности, оптимизируются внутри этих временных «окон». Возникает задача оптимизации решений по назначению однотипных заданий в составы пакетов, включению пакетов в интервалы доступности приборов, определения порядков пакетов для проведения операций с ними в течение указанных периодов времени. Использование рассматриваемого подхода позволяет получать более эффективные результаты планирования указанных процессов выполнения пакетов заданий (ПЗ) в КС с учетом ПТО приборов, увеличить их производительность.

Могут быть определены два способа решения задач оптимизации составов ПЗ и порядков их выполнения на приборах КС. Первый предусматривает оптимизацию количества заданий в пакетах. Вычисление характеристик процессов выполнения ПЗ в КС реализуется при условии, что пакеты передаются между приборами целиком. Этот подход применен в [1] при разработке модели Mixed Integer Linear Programming (MILP) оптимизации количества заданий разных типов в пакетах и расписаний выполнения ПЗ в КС. Второй способ предусматривает вычисление характеристик процессов выполнения ПЗ в КС с учетом того, что задания передаются с предшествующего прибора на последующий сразу при окончании действий с ними на первом из них без ожидания окончания выполнения других заданий из этого пакета. Этот способ положен в основу разрабатываемой математической модели MILP оптимизации распределения заданий разных типов по пакетам и расписаний выполнения ПЗ в КС с учетом ПТО приборов. Осуществляемые исследования базируются на результатах работы [2], продолжение которых предусматривает решение задачи оптимизации составов ПЗ и расписаний их выполнения в КС при учете интервалов доступности приборов нефиксированной (изменяющейся) длительности. В то же время постановка задачи исследований предусматривает, что результаты выполнения разнотипных заданий, распределяемых по пакетам, входят в различные заказы, для которых заданы директивные сроки окончания их исполнения.

Таким образом, развитие работы [2] предусматривает построение новой математической модели MILP оптимизации составов ПЗ, их (пакетов) включения в интервалы доступности приборов нефиксированной длительности, расписаний выполнения ПЗ в рамках указанных интервалов доступности приборов, а также оптимизации положения реализаций ПТО приборов в пределах временных «окон», начало и окончание которых определены в условиях задачи. То есть оптимизация решений по составам ПЗ и расписаниям их выполнения в КС осуществляется при условии, что длительности интервалов доступности приборов могут быть изменены, положение реализаций ПТО приборов оптимизируется в пределах заданных временных «окон». При этом построение указанной модели MILP осуществляется с учетом условия, предусматривающего, что результаты выполнения в КС заданий, распределенных по пакетам, включаются в заказы, для которых определены директивные сроки окончания их исполнения. Представленные особенности отличают задачу, решаемую в рассматриваемой работе, от задачи, решенной в работе [2], и обеспечивая ее оригинальность.

Существующие результаты исследований, связанных с оптимизацией расписаний выполнения заданий и размещению реализаций ПТО приборов, позволяют решать задачи планирования: 1) выпуска продукции на производственных предприятиях при условии проведения ПТО систем в целом [3–6]; 2) выполнения единичных заданий (ЕЗ) совместно с определением моментов времени начала ПТО приборов [7–15]; 3) операций с ЕЗ в системах разного вида при известных моментах времени начала ПТО (заданной периодичности ПТО и ограничениях на доступность приборов [16–21]).

Работы [3–21] отражают результаты исследований зарубежных авторов. Исследования российских ученых в основном направлены на разработку методов оптимизации производственных расписаний выполнения ЕЗ в мелко- и среднесерийном производстве [22–24], оптимизации рейсов воздушных судов авиакомпаний [25], расписаний учебных занятий в вузах [26–29] и отдельно графиков проведения ПТО оборудования без оптимизации расписаний [30–33]. Особого интереса заслуживает статья [30], в которой оптимизируется расписание работы ремонтных бригад в соответствии с графиком ПТО оборудования. Однако отечественными исследователями не решаются задачи совместной оптимизации расписаний выполнения заданий (в том числе пакетов заданий) в обрабатывающих системах разного вида совместно с оптимизацией графика ПТО.

Анализ существующих моделей и методов оптимизации решения задач рассматриваемого вида показал, что ни одна из рассмотренных работ не содержит результаты, использование которых позволяет осуществлять оптимизацию составов ПЗ и их порядков для проведения операций с ними в КС при учете ПТО приборов [3–21]. Эти модели и методы не учитывают возможность изменения положения реализаций ПТО в пределах временных «окон» заданной длительности. Задача оптимизации составов ПЗ и расписаний их выполнения при нефиксированных длительностях интервалов времени доступности приборов, которые могут быть изменены с целью повышения эффективности процессов осуществления операций с ними в КС, не является решенной. Также при решении задачи должно быть учтено условие включения в заказы результатов выполнения заданий из пакетов, для которых определены директивные сроки окончания их исполнения.

По этой причине решение рассматриваемой задачи оптимизации составов пакетов, включения пакетов в интервалы доступности приборов нефиксированной длительности, расписаний выполнения пакетов внутри этих интервалов, определения длительностей самих интервалов времени яв-

ляется актуальным. В постановке задачи исследований определены: 1) тип обрабатываемых систем – конвейерные системы; 2) способ типизации заданий в соответствии с равными значениями времени их выполнения на приборах и переналадок приборов на их выполнение; 3) необходимость формирования пакетов и оптимизации их составов; 4) вид передачи заданий из пакетов между приборами, который предусматривает их перемещение с предшествующего прибора на последующий после окончания их выполнения на первом из приборов; 5) ограничения на интервалы времени доступности и недоступности приборов (каждый интервал времени недоступности приборов входит в соответствующее временное «окно», его положение оптимизируется в рамках этого окна с целью повышения эффективности процессов выполнения ПЗ в КС и реализаций ПТО); 6) необходимость включения результатов выполнения заданий из пакетов в заказы, для которых определены директивные сроки окончания исполнения. Требуется разработать новую математическую модель MILP оптимизации решений рассматриваемого вида и программу в среде IBM ILOG CPLEX для этой модели. На основе программы требуется провести верификацию модели и определить особенности решений по размещению реализаций ПТО внутри временных «окон» в зависимости от параметров процессов выполнения ПЗ в КС.

Математическая модель оптимизации составов и расписаний выполнения пакетов заданий в системах Flow Shop при изменяющейся периодичности обслуживания приборов

При построении математической модели, применяемой для оптимизации решений рассматриваемого вида, использованы обозначения, аналогичные введенным в работе [2], а также введены новые обозначения. Обозначения, используемые при построении новой математической модели MILP оптимизации рассматриваемых решений, представлены в таблице 1.

Особенностью задачи является включение результатов выполнения заданий каждого w -го типа в количестве n_{wk} в состав только одного заказа, то есть $\sum_{k=1}^K r_{wk} = 1$ ($w = \overline{1, W}$). Через M обозначено большое число, заданное равным 1000, чтобы гарантировать, во-первых, положительные значения правых частей ограничений при условии, что значения переменных решения и индикаторных переменных равны 1, во-вторых, отрицательные значения правых частей ограничений при условии, что значения переменных решения и индикаторных переменных равны 0.

Реализации ПТО приборов позволяют обнаружить и устранить в них неисправности и непо-

ладки, являющиеся причинами их отказов. В результате чего предполагается, что значительные отказы этих приборов, вызывающие их длительное восстановление, невозможны. В процессе функционирования приборов после ПТО могут возникнуть незначительные отказы, длительности восстановления приборов после которых много меньше времени выполнения заданий.

ТАБЛИЦА 1. Обозначения параметров и решений задачи, используемых при построении модели MILP

TABLE 1. Designations of the Parameters and Solutions of the Problem Used in the Construction of the MILP Model

Обозначение	Интерпретация
Индексы приборов, заданий, пакетов, заказов, интервалов доступности	
(1)	(2)
l	Индекс прибора
b	Индекс интервала доступности приборов, в течение которого реализуется выполнение ПЗ, предшествующий b -й реализации их ПТО
j	Номер, под которым пакет включен в последовательность выполнения ПЗ в b -м интервале доступности приборов
i	Номер для задания, включенного в пакет под номером j в b -м интервале доступности приборов
w	Тип заданий, выполняемых в составе пакетов в КС
k	Индекс заказа
Параметры, характеризующие количество приборов, количество заданий, пакетов, заказов и интервалов доступности приборов	
L	Количество приборов
J	Количество пакетов, которые могут быть назначены в b -е интервалы доступности приборов
I	Количество заданий, которые будут включены в пакеты
B	Количество реализаций ПТО приборов
W	Количество типов заданий
K	Количество заказов
Параметры, характеризующие процессы выполнения заданий на приборах КС и формирование заказов	
$\mathbf{T} = \ t_{lw}\ _{L \times W}$	Матрица длительностей выполнения заданий w -х типов на l -х приборах КС
$\mathbf{P} = \ p_{lw}\ _{L \times W}$	Матрица длительностей наладок l -х приборов на выполнение заданий w -х типов
$\mathbf{R} = \ r_{wk}\ _{W \times K}$	Матрица типов w заданий, результаты выполнения которых распределяются по заказам k -х типов ($r_{wk} = 1$, если задания w -го типа включены в заказ k -го типа, $r_{wk} = 0$ в противном случае)
$\mathbf{N} = \ n_{wk}\ _{W \times K}$	Матрица количества результатов выполнения заданий w -х типов, включаемых в заказы k -х типов
$\mathbf{D} = \ d_k\ $	Вектор директивных сроков окончания исполнения заказов

Обозначение	Интерпретация
Индексы приборов, заданий, пакетов, заказов, интервалов доступности	
(1)	(2)
Параметры, характеризующие границы временных «окон» и длительности ПТО приборов	
s_{lb}	Значение левой границы временного интервала («окна»), в котором оптимизируется положение b -й реализации ПТО l -го прибора
e_{lb}	Значение правой границы временного интервала («окна»), в котором оптимизируется положение b -й реализации ПТО l -го прибора
t_l^{pm}	Длительность ПТО l -го прибора
τ	Длительность интервала времени, применяемого при оптимизации сдвига реализаций ПТО приборов относительно левых границ временных «окон»
Переменные решения задачи	
x_{ijb}	Переменная, характеризующая распределение заданий по пакетам в интервалах времени доступности приборов ($x_{ijb} = 1$, если задание распределено под номером i в пакет, который включен под номером j в b -й интервал доступности l -х приборов, $x_{ijb} = 0$ в противном случае)
y_{wjb}	Переменная, характеризующая типы заданий, включенных в пакеты внутри интервалов доступности приборов ($y_{wjb} = 1$, если пакет, включенный под номером j в b -й интервал доступности, содержит задания w -го типа, $y_{wjb} = 0$ в противном случае)
u_{jb}	Переменная, характеризующая включение в b -е интервалы доступности l -х приборов ПЗ под номерами j ($u_{jb} = 1$, если в b -й интервал доступности l -х приборов ПЗ включен под номером j – занимает j -ю позицию в последовательностях их выполнения в b -м интервале доступности; $u_{jb} = 0$ в противном случае)
k_{lb}^{pm}	Количество интервалов времени τ , соответствующих смещению начала b -й реализации ПТО l -го прибора относительно левой границы s_{lb} интервала $[s_{lb}; e_{lb}]$ (смещение равно $k_{lb}^{pm} \tau$)
q_{ijb}^l	Момент времени начала реализации операций с заданием, включенным под номером i в пакет, размещенный под номером j в b -м интервале доступности l -го прибора
g_{jb}^l	Момент времени завершения операций на l -м приборе с ПЗ, включенным под j -м номером в его b -й интервал доступности
g_k	Момент времени окончания исполнения k -го заказа; количество типов заданий
p_k	Запаздывание с исполнением k -го заказа относительно директивного срока d_k

Поэтому незначительными отказами приборов, возникающими после ПТО, длительности восстановления после которых минимальны, можно пренебречь. Начало и окончание реализаций ПТО при-

боров не являются фиксированными, а входят в заданные интервалы времени $[s_{lb}; e_{lb}]$ (временные «окна»), в рамках которых их положение оптимизируется с целью увеличения эффективности процессов выполнения ПЗ в КС. Таким образом, $[s_{lb}; e_{lb}]$ – это интервалы времени, в пределах которых оптимизируется положение b -х реализаций ПТО l -х приборов с учетом требования повышения общей производительности процессов выполнения ПЗ в КС. Реализация ПТО может начинаться в момент времени, не меньший левой границы интервала, окончание ПТО не может превышать его правую границу. Разработка модели MILP оптимизации наряду с составами ПЗ и расписаниями их выполнения в КС также и положения ПТО приборов в рамках временных «окон» (при учете условия исполнения заказов в заданные директивные сроки) позволяет обеспечить развитие математической модели, предложенной в [2], и как следствие, новизну и оригинальность получаемых результатов.

При формировании ограничений модели, позволяющих вычислять значения q_{ijb}^l моментов времени начала выполнения i -х заданий в j -х пакетах, включенных в b -е интервалы доступности l -х приборов ($i = \overline{1, I}; j = \overline{1, J}; b = \overline{1, B+1}; l = \overline{1, L}$), использованы предположения о том, что количество заданий в пакетах должно быть не менее двух, и в b -е интервалы доступности приборов будет включен хотя бы один пакет. Кроме того, предполагается, что составы ПЗ, включенных в одни и те же b -е интервалы доступности ($b = \overline{1, B+1}$) l -х приборов ($l = \overline{1, L}$), не имеют отличий; одинаковыми являются и порядки выполнения на приборах КС пакетов заданий, включенных в одни и те же интервалы доступности.

При построении системы ограничений, позволяющих вычислить значения q_{ijb}^l , в основу положена базовая модель оптимизации решений, предложенная в работе [2, выражения (2–26), стр. 14]. Эта модель позволяет оптимизировать решения при фиксированных интервалах времени доступности приборов. С учетом введенных в рассмотрение обозначений система ограничений, используемая для вычисления значений q_{ijb}^l и соответствующая базовой модели оптимизации решений из [2, ограничения (14–22), стр. 14] (при фиксированных интервалах времени доступности приборов), имеет следующий вид:

$$q_{111}^1 \geq \sum_{w=1}^W p_{1w} y_{w11}, \quad (1)$$

$$q_{11b}^1 \geq e_{1,b-1} + \sum_{w=1}^W p_{1w} y_{w1b}, \quad b = \overline{2, B+1}, \quad (2)$$

$$q_{1jb}^1 \geq g_{j-1,b}^1 + \sum_{w=1}^W p_{1w} y_{w1b} - M \cdot (1 - u_{jb}), \quad j = \overline{2, J}, \quad b = \overline{1, B+1}, \quad (3)$$

$$q_{ijb}^1 \geq q_{i-1,j,b}^1 + \sum_{w=1}^W t_{1w} y_{wjb} - M \cdot (1 - x_{ijb}), \quad i = \overline{2, I}, \quad j = \overline{1, J}, \quad b = \overline{1, B+1}, \quad (4)$$

$$g_{jb}^l \geq q_{ijb}^l + \sum_{w=1}^W t_{lw} y_{wjb} - M \cdot (2 - u_{jb} - x_{ijb}), \quad i = \overline{1, I}, \quad j = \overline{1, J}, \quad b = \overline{1, B+1}, \quad l = \overline{1, L}, \quad (5)$$

$$q_{111}^l \geq \max(q_{111}^{l-1} + \sum_{w=1}^W t_{l-1w} y_{w11}; \sum_{w=1}^W p_{lw} y_{w11}), \quad l = \overline{2, L}, \quad (6)$$

$$q_{11b}^l \geq \max(q_{11b}^{l-1} + \sum_{w=1}^W t_{l-1w} y_{w1b};$$

$$e_{1,b-1} + \sum_{w=1}^W p_{lw} y_{w1b}), \quad b = \overline{2, B+1}, \quad l = \overline{2, L}, \quad (7)$$

$$q_{1jb}^l \geq \max(q_{1jb}^{l-1} + \sum_{w=1}^W t_{l-1w} y_{wjb};$$

$$g_{j-1,b}^l + \sum_{w=1}^W p_{lw} y_{wjb}) - M \cdot (1 - u_{jb}), \quad j = \overline{2, J}, \quad b = \overline{1, B+1}, \quad l = \overline{2, L}, \quad (8)$$

$$q_{ijb}^l \geq \max(q_{ijb}^{l-1} + \sum_{w=1}^W t_{l-1w} y_{wjb};$$

$$q_{i-1,j,b}^l + \sum_{w=1}^W t_{lw} y_{wjb}) - M \cdot (1 - x_{ijb}), \quad i = \overline{2, I}, \quad j = \overline{1, J}, \quad b = \overline{1, B+1}, \quad l = \overline{2, L}, \quad (9)$$

$$g_{jb}^l \leq s_{lb}, \quad l = \overline{1, L}, \quad b = \overline{1, B}, \quad j = \overline{1, J}. \quad (10)$$

В выражениях (1–10), представляющих собой ограничения модели оптимизации решений и позволяющих определять значения q_{ijb}^l и g_{jb}^l , учтено условие, что s_{lb} и e_{lb} являются левыми и правыми границами интервалов $[s_{lb}; e_{lb}]$ недоступности приборов, в течение которых реализуется их ПТО. То есть моменты времени начала и окончания ПТО приборов являются фиксированными, также, как являются фиксированными и интервалы доступности приборов.

В постановке задачи, решаемой в данной работе, значения s_{lb} и e_{lb} являются левыми и правыми границами временных интервалов (временных «окон»), в рамках которых оптимизируется положение ре-

лизаций ПТО приборов (в том числе этим рассматриваемая работа отличается от работы [2]). При этом начало рассматриваемых реализаций ПТО приборов смещено на k_{lb}^{pm} интервалов времени τ относительно левой границы s_{lb} рассматриваемого интервала $[s_{lb}; e_{lb}]$. Значения k_{lb}^{pm} для каждой b -й реализации ПТО каждого l -го прибора определяются в процессе оптимизации решений. Тогда значения левых границ интервалов недоступности l -х приборов, соответствующих моментам времени начала b -х реализаций их ПТО и моментам времени окончания b -х интервалов доступности приборов, определяются выражением $s_{lb} + k_{lb}^{pm}\tau$. Значения правых границ интервалов недоступности l -х приборов, соответствующих моментам времени окончания b -х реализаций их ПТО и моментам времени начала $(b+1)$ -х интервалов доступности приборов, определяются выражением $s_{lb} + k_{lb}^{pm}\tau + t_l^{pm}$. Определение в процессе оптимизации решений значений k_{lb}^{pm} ($b = \overline{1, B}; l = \overline{1, L}$) осуществляется с учетом ограничения, предусматривающего, что моменты времени окончания b -х реализаций ПТО l -х приборов не превышают значений e_{lb} правых границ соответствующих временных интервалов $[s_{lb}; e_{lb}]$, в которых их (этих реализаций) положение оптимизируется. Указанное ограничение имеет следующий вид: $s_{lb} + k_{lb}^{pm}\tau + t_l^{pm} \leq e_{lb}$ ($b = \overline{1, B}; l = \overline{1, L}$).

В соответствии с выполненными рассуждениями модифицированы выражения (2, 7, 10), входящие в модель (1-10) и используемые для вычисления значений q_{11b}^1 ($b = \overline{2, B+1}$), q_{11b}^l ($l = \overline{2, L}; b = \overline{2, B+1}$) и g_{jb}^l ($j = \overline{1, J}; b = \overline{1, B+1}$).

Модификация выражений (2, 7, 10) осуществлена следующим образом. Окончание b -х реализаций ПТО ($l = 1$)-го прибора связано с границами e_{1b} интервалов $[s_{1b}; e_{1b}]$: $s_{1b} + k_{1b}^{pm}\tau + t_1^{pm} \leq e_{1b}$. Начало выполнения ($i = 1$)-х заданий в ($j = 1$)-х пакетах в b -х интервалах ($b = \overline{2, B+1}$) доступности ($l = 1$)-го прибора возможно после его наладки на осуществление операций с заданиями w -го типа в этих пакетах, которая следует за окончанием $(b-1)$ -х реализаций ПТО.

Для этих заданий значение q_{11b}^1 ($b = \overline{2, B+1}$) определяется выражением вида:

$$q_{11b}^1 \geq s_{1,b-1} + k_{1,b-1}^{pm}\tau + t_1^{pm} + \sum_{w=1}^W p_{1w}y_{w1b}.$$

Начало b -х реализаций ПТО l -х приборов ($l = \overline{2, L}$) определяется выражением: $s_{lb} + k_{lb}^{pm}\tau$, окончание этих реализаций ПТО – выражением: $s_{lb} + k_{lb}^{pm}\tau + t_l^{pm}$, где k_{lb}^{pm} – переменная решения. Окончание b -х реализаций ПТО l -х приборов связано с границами e_{lb} интервалов $[s_{lb}; e_{lb}]$: $s_{lb} + k_{lb}^{pm}\tau + t_l^{pm} \leq e_{lb}$. Начало выполнения ($i = 1$)-х

заданий в ($j = 1$)-х пакетах в b -х интервалах доступности l -х приборов ($l = \overline{2, L}$) возможно после их наладки на осуществления операций с заданиями w -х типов, которая следует за окончанием $(b-1)$ -х реализаций ПТО приборов ($b = \overline{2, B+1}$).

Тогда значения q_{11b}^l для ($i = 1$)-х заданий в ($j = 1$)-х пакетах в b -х интервалах доступности l -х приборов определяются выражением:

$$q_{11b}^l \geq \max (q_{11b}^{l-1} + \sum_{w=1}^W t_{l-1w}y_{w1b}; (s_{l,b-1} + k_{l,b-1}^{pm}\tau + t_l^{pm}) + \sum_{w=1}^W p_{lw}y_{w1b}).$$

где $b = \overline{2, B+1}$, $l = \overline{2, L}$.

Первая компонента операции $\max$ аналогична такой же компоненте в выражении (7), вторая компонента определяет момент времени окончания наладки l -го прибора после завершения $(b-1)$ -й реализации его ПТО.

Для значений g_{jb}^l ограничения (5) модифицированы следующим образом: $g_{jb}^l \leq s_{lb} + k_{lb}^{pm}\tau$ ($b = \overline{1, B}$). Модификация ограничений (5) предусматривает, что ПЗ должны быть выполнены строго в b -е интервалы доступности l -х приборов, окончание которых определяется выражением $s_{lb} + k_{lb}^{pm}\tau$. Вместе с тем моменты времени окончания выполнения ПЗ в $(B+1)$ -х интервалах доступности l -х приборов ($l = \overline{1, L}$) превышают момент времени окончания B -й реализации их ПТО: $g_{j,B+1}^l \geq s_{lB} + k_{lB}^{pm}\tau + t_l^{pm}$ ($j = \overline{1, J}; l = \overline{1, L}$).

С целью введения ограничений, определяющих требование распределения по пакетам всех заданий w -х типов ($w = \overline{1, W}$) в количестве n_{wk} , включенных в k -е заказы ($k = \overline{1, K}$), введены индикаторные переменные. То есть в пакеты, включающиеся под j -ми номерами в b -е интервалы доступности l -х приборов, требуется распределять все n_{wk} заданий w -х типов ($w = \overline{1, W}$), входящих в k -е заказы ($k = \overline{1, K}$). Для построения ограничений, соответствующих выполнению этого условия, введены индикаторные переменные v_{ijbkw} ($i = \overline{1, I}; j = \overline{1, J}; b = \overline{1, B+1}; w = \overline{1, W}; k = \overline{1, K}$). В том случае, если задание назначено под i -м номером в пакет, включенный под j -м номером в b -й интервал доступности приборов ($x_{ijb} = 1$), в этот j -й пакет в b -м интервале доступности назначены задания w -го типа ($y_{wjb} = 1$), при этом задания w -го типа включены в k -й заказ ($r_{wk} = 1$), то $v_{ijbkw} = 1$.

Таким образом, если $x_{ijb}y_{wjb}r_{wk} = 1$ (i -е задание из пакета, включенного под j -м номером в b -й интервал доступности приборов, относится к w -у

типу заданий из k -го заказа), то $v_{ijbkw} = 1$. Если $x_{ijb}y_{wjb}r_{wk} = 0$ (i -е задание из пакета, включенного под j -м номером в b -й интервал доступности l -го прибора, не относится к w -у типу либо не соответствует типу заданий, включенных в k -й заказ), то $v_{ijbkw} = 0$.

Сформирован способ идентификации значений индикаторной переменной y , соответствующих выполнению / не выполнению ограничений вида $f(x_1, x_2, \dots, x_n) \geq b$. Полученные выражения, позволяющие определять значения y , имеют вид:

$$\begin{aligned} f(x_1, x_2, \dots, x_n) &\geq b - M \cdot (1 - y), \\ f(x_1, x_2, \dots, x_n) &\leq b - 1 + My. \end{aligned} \quad (11)$$

Для условий: 1) $v_{ijbkw} = 1$, когда $x_{ijb}y_{wjb}r_{wk} = 1$, 2) $v_{ijbkw} = 0$, когда $x_{ijb}y_{wjb}r_{wk} = 0$, ограничения, используемые для определения значений индикаторных переменных v_{ijbkw} , получены на основе (11) в виде:

$$\begin{aligned} x_{ijb}y_{wjb}r_{wk} &\geq 1 - M \cdot (1 - v_{ijbkw}), \\ x_{ijb}y_{wjb}r_{wk} &\leq Mv_{ijbkw}, \end{aligned}$$

где $i = \overline{1, I}$, $j = \overline{1, J}$, $b = \overline{1, B + 1}$, $w = \overline{1, W}$, $k = \overline{1, K}$.

Тогда ограничения, определяющие требование включения в состав пакетов всех заданий w -х типов из k -х заказов в количестве n_{wk} , имеют вид:

$$\sum_{b=1}^{B+1} \sum_{j=1}^J \sum_{i=1}^I v_{ijbkw} = n_{wk},$$

где $w = \overline{1, W}$, $k = \overline{1, K}$.

Момент времени окончания исполнения k -х заказов определяется как максимум среди моментов времени окончания операций со всеми зданиями w -х типов, входящими в эти заказы.

Для вычисления g_k получено выражение:

$$\begin{aligned} g_k &\geq \left(q_{ijb}^L + \sum_{w=1}^W t_{Lw}y_{wjb} \right) - M \cdot \left(1 - \sum_{w=1}^W v_{ijbkw} \right), \\ i &= \overline{1, I}, \quad j = \overline{1, J}, \quad b = \overline{1, B + 1}, \quad k = \overline{1, K}, \end{aligned}$$

где выражение $(q_{ijb}^L + \sum_{w=1}^W t_{Lw}y_{wjb})$ позволяет вычислить значения моментов времени окончания выполнения i -х заданий из j -х пакетов, включенных в b -е интервал доступности L -го прибора, а выражение $M \cdot (1 - \sum_{w=1}^W v_{ijbkw})$ – осуществлять вычисления только для заданий w -х типов, включенных в k -й заказ. Запоздывание p_k с окончанием исполнения k -х заказов относительно заданных для них директивных сроков определяется выражением: $p_k \geq \max(0, g_k - d_k)$ ($k = \overline{1, K}$), где d_k – значения директивных сроков окончания исполнения k -х заказов. Критерий оптимизации решений представляет собой сумму запаздываний с исполнением k -х

заказов, которая минимизируется. В этом случае критерий оптимизации решений рассматриваемого вида имеет вид $f = \sum_{k=1}^K p_k$. Учет в разрабатываемой модели MILP условия включение результатов выполнения заданий, распределенных по пакетам, в заказы, для которых заданы директивные сроки окончания исполнения, обуславливает ее отличие от модели, рассмотренной в [2], и тем самым ее новизну.

Включение пакетов в b -е интервалы доступности l -х приборов предполагает следующие ограничения:

1) задания с i -и номерами могут быть распределены в пакеты в том случае, если пакеты под j -ми номерами включены в b -е интервалы доступности:

$$\sum_{i=1}^I x_{ijb} \leq Mu_{jb},$$

где $j = \overline{1, J}$, $b = \overline{1, B + 1}$;

2) в j -е пакеты, включенные в b -е интервалы доступности l -х приборов, должно быть распределено не менее двух заданий:

$$\sum_{i=1}^I x_{ijb} \geq 2 - M \cdot (1 - u_{jb}),$$

где $j = \overline{1, J}$, $b = \overline{1, B + 1}$;

3) в следующие друг за другом пакеты, включенные в b -е интервалы доступности l -х приборов, распределены задания разных типов:

$$y_{wjb} + y_{w,j+1,b} \leq 1,$$

где $j = \overline{1, J - 1}$, $w = \overline{1, W}$, $b = \overline{1, B + 1}$;

4) в каждом b -м интервале доступности размещается не менее одного ПЗ:

$$\sum_{j=1}^J u_{jb} \geq 1,$$

где $b = \overline{1, B + 1}$;

5) пакет, включенный под j -м номером в b -й интервал доступности l -х приборов, может содержать задания одного w -го типа и задания могут быть добавлены в пакет только в случае, если он включен в этот интервал:

$$\sum_{w=1}^W y_{wjb} = u_{jb},$$

где $j = \overline{1, J}$, $b = \overline{1, B + 1}$.

Также формирование ограничений на включение заданий под i -ми номерами в пакеты, размещаемые с j -ми идентификаторами в b -х интервалах доступности l -х приборов, осуществляется с использованием подхода, основывающегося на логических условиях [34].

Первое логическое условие: если задание содержится под $(i + 1)$ -м номером в пакете, включенном

под j -м номером в b -й интервал доступности l -х приборов ($x_{i+1,jb} = 1$), то в этом же пакете должно содержаться задание с i -м номером ($x_{ijb} = 1$).

Вид логического условия:

$$\text{ЕСЛИ } x_{i+1,jb} = 1, \text{ ТО } x_{ijb} = 1,$$

где $i = \overline{1, I-1}$, $j = \overline{1, J}$, $b = \overline{1, B+1}$.

Форма ограничений для этого условия:

$$x_{ijb} - x_{i+1,jb} \geq 0.$$

Аналогично сформированы ограничения по включению в b -е интервалы доступности двух идущих подряд пакетов:

$$u_{jb} - u_{j+1,b} \geq 0,$$

где $j = \overline{1, J-1}$, $b = \overline{1, B+1}$.

В соответствии с выполненными рассуждениями, сформирована новая математическая модель оптимизации назначения заданий в составы пакетов, включения пакетов в b -е интервалы доступности приборов не фиксированной длительности, размещения реализаций ПТО приборов в заданных временных интервалах (временных «окнах») при исполнении заказов в заданные сроки.

Вид полученной модели следующий:

1) критерий оптимизации решений:

$$f = \sum_{k=1}^K p_k; \quad (12)$$

2) ограничения:

$$\sum_{i=1}^I x_{ijb} \geq 2 - M \cdot (1 - u_{jb}), \quad b = \overline{1, B+1}, j = \overline{1, J}; \quad (13)$$

$$\sum_{i=1}^I x_{ijb} \leq M u_{jb}, \quad b = \overline{1, B+1}, j = \overline{1, J}; \quad (14)$$

$$\sum_{j=1}^J u_{jb} \geq 1, \quad b = \overline{1, B+1}; \quad (15)$$

$$\sum_{w=1}^W y_{wjb} = u_{jb}, \quad b = \overline{1, B+1}, j = \overline{1, J}; \quad (16)$$

$$y_{wjb} + y_{w,j+1,b} \leq 1, \quad b = \overline{1, B+1}, j = \overline{1, J-1}, w = \overline{1, W}; \quad (17)$$

$$x_{ijb} - x_{i+1,jb} \geq 0, \quad i = \overline{1, I-1}, j = \overline{1, J}, b = \overline{1, B+1} \quad (18)$$

$$u_{jb} - u_{j+1,b} \geq 0, \quad j = \overline{1, J-1}, b = \overline{1, B+1}; \quad (19)$$

$$x_{ijb} y_{wjb} r_{wk} \geq 1 - M \cdot (1 - v_{ijbwk}), \quad (20)$$

$$x_{ijb} y_{wjb} r_{wk} \leq M v_{ijbwk}, \quad i = \overline{1, I},$$

$$j = \overline{1, J}, b = \overline{1, B+1}, w = \overline{1, W}, k = \overline{1, K}$$

$$\sum_{b=1}^{B+1} \sum_{j=1}^J \sum_{i=1}^I v_{ijbwk} = n_{wk}, \quad w = \overline{1, W}, k = \overline{1, K}; \quad (21)$$

$$q_{111}^1 \geq \sum_{w=1}^W p_{1w} y_{w11}; \quad (22)$$

$$q_{11b}^1 \geq s_{1,b-1} + k_{1,b-1}^{pm} \tau + t_1^{pm} + \sum_{w=1}^W p_{1w} y_{w1b}, \quad b = \overline{2, B+1}; \quad (23)$$

$$q_{1jb}^1 \geq g_{j-1,b}^1 + \sum_{w=1}^W p_{1w} y_{w1b} - M \cdot (1 - u_{jb}), \quad j = \overline{2, J}, b = \overline{1, B+1}; \quad (24)$$

$$q_{ijb}^1 \geq q_{i-1,j,b}^1 + \sum_{w=1}^W t_{1w} y_{wjb} - M \cdot (1 - x_{ijb}), \quad i = \overline{2, I}, j = \overline{1, J}, b = \overline{1, B+1}; \quad (25)$$

$$g_{jb}^l \geq q_{ijb}^l + \sum_{w=1}^W t_{lw} y_{wjb} - M \cdot (2 - u_{jb} - x_{ijb}), \quad i = \overline{1, I}, j = \overline{1, J}, b = \overline{1, B+1}, l = \overline{1, L}; \quad (26)$$

$$q_{111}^l \geq \max(q_{111}^{l-1} + \sum_{w=1}^W t_{l-1w} y_{w11}; \sum_{w=1}^W p_{lw} y_{w11}), \quad l = \overline{2, L}; \quad (27)$$

$$q_{11b}^l \geq \max(q_{11b}^{l-1} + \sum_{w=1}^W t_{l-1w} y_{w1b}, (s_{l,b-1} + k_{l,b-1}^{pm} \tau + t_l^{pm}) + \sum_{w=1}^W p_{lw} y_{w1b}), \quad b = \overline{2, B+1}, l = \overline{2, L} \quad (28)$$

$$q_{1jb}^l \geq \max(q_{1jb}^{l-1} + \sum_{w=1}^W t_{l-1w} y_{wjb}; g_{j-1,b}^l + \sum_{w=1}^W p_{lw} y_{wjb}) - M \cdot (1 - u_{jb}), \quad j = \overline{2, J}, l = \overline{2, L}; \quad (29)$$

$$q_{ijb}^l \geq \max(q_{ijb}^{l-1} + \sum_{w=1}^W t_{l-1w} y_{wjb}, q_{i-1,j,b}^l + \sum_{w=1}^W t_{lw} y_{wjb}) - M \cdot (1 - x_{ijb}), \quad i = \overline{2, I}, j = \overline{1, J}, b = \overline{1, B+1}, l = \overline{2, L}; \quad (30)$$

$$g_k \geq \left(q_{ijb}^l + \sum_{w=1}^W t_{Lw} y_{wjb} \right) - M \times \left(1 - \sum_{w=1}^W v_{ijbwk} \right), \quad (31)$$

$$i = \overline{1, I}, j = \overline{1, J}, b = \overline{1, B+1}, k = \overline{1, K};$$

$$p_k \geq \max(0, g_k - d_k), \quad k = \overline{1, K}; \quad (32)$$

$$s_{lb} + k_{lb}^{pm} \tau + t_{lb}^{pm} \leq e_{lb}, \quad b = \overline{1, B}; \quad (33)$$

$$g_{jb}^l \leq s_{lb} + k_{lb}^{pm} \tau, \quad b = \overline{1, B}; \quad (34)$$

$$g_{j, B+1}^l \geq s_{l, B} + k_{l, B}^{pm} \tau + t_l^{pm}, \quad j = \overline{1, J}, l = \overline{1, L}; \quad (35)$$

$$g_k \geq 0, \quad k = \overline{1, K}, k_{lb}^{pm} \geq 0, l = \overline{1, L}, b = \overline{1, B}; \quad (36)$$

$$q_{ijb}^l \geq 0, \quad g_{jb}^l \geq 0, \quad (37)$$

$$i = \overline{1, I}, j = \overline{1, J}, b = \overline{1, B+1}, l = \overline{1, L};$$

$$x_{ijb} \in \{0, 1\}, y_{wjb} \in \{0, 1\}, u_{jb} \in \{0, 1\},$$

$$v_{ijbwk}^l \in \{0, 1\}, \quad i = \overline{1, I}, j = \overline{1, J}, \quad (38)$$

$$b = \overline{1, B+1}, w = \overline{1, W}, k = \overline{1, K}$$

Линеаризация модели оптимизации составов и расписаний выполнения пакетов заданий в системе Flow Shop при изменяющейся периодичности обслуживания приборов

Модель (12–38) является нелинейной. Нелинейными являются выражения (20, 27–30, 32). Для линеаризации ограничений (20) в рассмотрение введены переменные z_{ijbw} : $z_{ijbw} = x_{ijb} y_{wjb}$ ($i = \overline{1, I}, j = \overline{1, J}, b = \overline{1, B+1}, w = \overline{1, W}$) и соответствующие ограничения, обеспечивающие корректность замены [35]:

$$z_{ijbw} \leq y_{wjb}, \quad z_{ijbw} \leq x_{ijb}, \quad z_{ijbw} \geq y_{wjb} + x_{ijb} - 1, \\ z_{ijbw} \in \{0, 1\}.$$

С целью линеаризации (27) введены индикаторные переменные m_l ($l = \overline{2, L}$), интерпретация которых аналогична [2], и дополнительные ограничения, которые аналогичны базовой модели [2, ограничения (27–30), стр. 17], обеспечивающие корректность введения в модель этих переменных.

Для линеаризации (28) введены переменные m_b^l ($l = \overline{2, L}, b = \overline{2, B+1}$) и следующие дополнительные ограничения, обеспечивающие корректность введения в модель этих переменных:

$$[(s_{l, b-1} + k_{l, b-1}^{pm} \tau + t_l^{pm}) + \sum_{w=1}^W p_{lw} y_{w1b}) - (q_{11b}^{l-1} + \sum_{w=1}^W t_{l-1w} y_{w1b})] + M m_b^l \geq 0, \\ [(q_{11b}^{l-1} + \sum_{w=1}^W t_{l-1w} y_{w1b})] - ((s_{l, b-1} + k_{l, b-1}^{pm} \tau + t_l^{pm}) +$$

$$+ \sum_{w=1}^W p_{lw} y_{w1b})] + M \cdot (1 - m_b^l) \geq 0,$$

$$q_{11b}^l \geq (q_{11b}^{l-1} + \sum_{w=1}^W t_{l-1w} y_{w1b}) - M \cdot (1 - m_b^l),$$

$$q_{11b}^l \geq ((s_{l, b-1} + k_{l, b-1}^{pm} \tau + t_l^{pm}) + \sum_{w=1}^W p_{lw} y_{w1b}) - M m_b^l.$$

Для линеаризации (29) введены индикаторные переменные m_{jb}^l ($j = \overline{2, J}, b = \overline{1, B+1}, l = \overline{2, L}$), интерпретация которых аналогична [2], и дополнительные ограничения, которые аналогичны базовой модели [2, ограничения (35–38), стр. 17].

Для линеаризации (30) в рассмотрение введены индикаторные переменные m_{ijb}^l ($i = \overline{2, I}, j = \overline{1, J}, b = \overline{1, B+1}, l = \overline{2, L}$), интерпретация которых аналогична [2], и дополнительные ограничения, которые аналогичны базовой модели [2, ограничения (39–42), стр. 17], обеспечивающие корректность введения в модель этих переменных.

Для линеаризации выражений (32) введены переменные $z1_k$ ($k = \overline{1, K}$) и дополнительные ограничения:

$$(d_k - g_k) + M z1_k \geq 0, \\ (g_k - d_k) + M(1 - z1_k) \geq 0, \\ p_k \geq (g_k - d_k) - M(1 - z1_k), \\ p_k \geq -M z1_k.$$

Итоговый вид модели MILP оптимизации решений по назначению заданий в пакеты, включению пакетов в интервалы доступности приборов, порядков выполнения ПЗ в интервалах доступности приборов, размещению реализаций ПТО приборов в рамках заданных временных интервалов при учете исполнения заказов в заданные директивные сроки:

1) критерий оптимизации решений:

$$f = \sum_{k=1}^K p_k; \quad (39)$$

2) ограничения:

$$\sum_{i=1}^I x_{ijb} \geq 2 - M \cdot (1 - u_{jb}), \quad (40)$$

$$b = \overline{1, B+1}, j = \overline{1, J};$$

$$\sum_{i=1}^I x_{ijb} \leq M u_{jb}, \quad b = \overline{1, B+1}, j = \overline{1, J}; \quad (41)$$

$$\sum_{j=1}^J u_{jb} \geq 1, \quad b = \overline{1, B+1}; \quad (42)$$

$$\sum_{w=1}^W y_{wjb} = u_{jb}, \quad b = \overline{1, B+1}, j = \overline{1, J}; \quad (43)$$

$$y_{wjb} + y_{w,j+1,b} \leq 1, \quad b = \overline{1, B+1}, j = \overline{1, J-1}, w = \overline{1, W}; \quad (44)$$

$$x_{ijb} - x_{i+1,jb} \geq 0, \quad i = \overline{1, I-1}, j = \overline{1, J}, b = \overline{1, B+1} \quad (45)$$

$$u_{jb} - u_{j+1,b} \geq 0, \quad j = \overline{1, J-1}, b = \overline{1, B+1}; \quad (46)$$

$$z_{ijbw} r_{wk} \geq 1 - M \cdot (1 - v_{ijbwk})$$

$$z_{ijbw} r_{wk} \leq M v_{ijbwk}, \quad i = \overline{1, I}, \quad (47)$$

$$j = \overline{1, J}, b = \overline{1, B+1}, w = \overline{1, W}, k = \overline{1, K}$$

$$\sum_{b=1}^{B+1} \sum_{j=1}^J \sum_{i=1}^I v_{ijbwk} = n_{wk}, \quad w = \overline{1, W}, k = \overline{1, K}; \quad (48)$$

$$z_{ijbw} \leq y_{wjb}, \quad z_{ijbw} \leq x_{ijb}, \quad (49)$$

$$z_{ijbw} \geq y_{wjb} + x_{ijb} - 1;$$

$$q_{111}^1 \geq \sum_{w=1}^W p_{1w} y_{w11}; \quad (50)$$

$$q_{11b}^1 \geq s_{1,b-1} + k_{1,b-1}^{pm} \tau + t_1^{pm} + \sum_{w=1}^W p_{1w} y_{w1b}, \quad (51)$$

$$b = \overline{2, B+1};$$

$$q_{1jb}^1 \geq g_{j-1,b}^1 + \sum_{w=1}^W p_{1w} y_{w1b} - M \cdot (1 - u_{jb}), \quad (52)$$

$$j = \overline{2, J}, b = \overline{1, B+1};$$

$$q_{ijb}^1 \geq q_{i-1,j,b}^1 + \sum_{w=1}^W t_{1w} y_{wjb} - M \cdot (1 - x_{ijb}), \quad (53)$$

$$i = \overline{2, I}, j = \overline{1, J}, b = \overline{1, B+1};$$

$$g_{jb}^l \geq q_{ijb}^l + \sum_{w=1}^W t_{lw} y_{wjb} - M \cdot (2 - u_{jb} - x_{ijb}), \quad (54)$$

$$i = \overline{1, I}, j = \overline{1, J}, b = \overline{1, B+1}, l = \overline{1, L};$$

$$[\sum_{w=1}^W p_{lw} y_{w11} - (q_{111}^{l-1} + \sum_{w=1}^W t_{l-1w} y_{w11})] + M m_l \geq 0, \quad (55)$$

$$l = \overline{2, L};$$

$$[(q_{111}^{l-1} + \sum_{w=1}^W t_{l-1w} y_{w11})] - \sum_{w=1}^W p_{lw} y_{w11} +$$

$$+ M(1 - m_l) \geq 0, \quad l = \overline{2, L}; \quad (56)$$

$$q_{111}^l \geq (q_{111}^{l-1} + \sum_{w=1}^W t_{l-1w} y_{w11}) - M \cdot (1 - m_l), \quad (57)$$

$$l = \overline{2, L};$$

$$q_{111}^l \geq \sum_{w=1}^W p_{lw} y_{w11} - M m_l, \quad l = \overline{2, L}; \quad (58)$$

$$[((s_{l,b-1} + k_{l,b-1}^{pm} \tau + t_l^{pm}) + \sum_{w=1}^W p_{lw} y_{w1b}) -$$

$$- (q_{11b}^{l-1} + \sum_{w=1}^W t_{l-1w} y_{w1b})] + M m_b^l \geq 0, \quad (59)$$

$$b = \overline{2, B}, l = \overline{2, L};$$

$$[(q_{11b}^{l-1} + \sum_{w=1}^W t_{l-1w} y_{w1b})] -$$

$$- ((s_{l,b-1} + k_{l,b-1}^{pm} \tau + t_l^{pm}) + \sum_{w=1}^W p_{lw} y_{w1b})] +$$

$$+ M \cdot (1 - m_b^l) \geq 0, \quad b = \overline{2, B+1}, l = \overline{2, L}; \quad (60)$$

$$q_{11b}^l \geq (q_{11b}^{l-1} + \sum_{w=1}^W t_{l-1w} y_{w1b}) - M \cdot (1 - m_b^l), \quad (61)$$

$$b = \overline{2, B+1}, l = \overline{2, L};$$

$$q_{11b}^l \geq ((s_{l,b-1} + k_{l,b-1}^{pm} \tau + t_l^{pm}) +$$

$$+ \sum_{w=1}^W p_{lw} y_{w1b}) - M m_b^l, \quad b = \overline{2, B+1}, l = \overline{2, L}; \quad (62)$$

$$[(g_{j-1,b}^l + \sum_{w=1}^W p_{lw} y_{wjb}) -$$

$$- (q_{1jb}^{l-1} + \sum_{w=1}^W t_{l-1w} y_{wjb})] + M m_{jb}^l \geq 0, \quad (63)$$

$$j = \overline{2, J}, b = \overline{1, B+1}, l = \overline{2, L}.$$

$$[(q_{1jb}^{l-1} + \sum_{w=1}^W t_{l-1w} y_{wjb}) -$$

$$- (g_{j-1,b}^l + \sum_{w=1}^W p_{lw} y_{wjb})] + M(1 - m_{jb}^l) \geq 0, \quad (64)$$

$$j = \overline{2, J}, b = \overline{1, B+1}, l = \overline{2, L}.$$

$$q_{1jb}^l \geq (q_{1jb}^{l-1} + \sum_{w=1}^W t_{l-1w} y_{wjb}) -$$

$$- M \cdot (2 - u_{jb} - m_{jb}^l), \quad (65)$$

$$j = \overline{2, J}, b = \overline{1, B+1}, l = \overline{2, L};$$

$$q_{1jb}^l \geq (g_{j-1,b}^l + \sum_{w=1}^W p_{lw} y_{wjb}) -$$

$$- M \cdot (1 - u_{jb} + m_{jb}^l), \quad (66)$$

$$j = \overline{2, J}, b = \overline{1, B+1}, l = \overline{2, L};$$

$$[(q_{i-1,j,b}^l + \sum_{w=1}^W t_{lw} y_{wjb}) - (q_{ijb}^{l-1} + \sum_{w=1}^W t_{l-1w} y_{wjb})] + M m_{ijb}^l \geq 0, \quad (67)$$

$$i = \overline{2, I}, j = \overline{1, J}, b = \overline{1, B+1}, l = \overline{2, L};$$

$$[(q_{ijb}^{l-1} + \sum_{w=1}^W t_{l-1w} y_{wjb}) - (q_{i-1,j,b}^l + \sum_{w=1}^W t_{lw} y_{wjb})] + M(1 - m_{ijb}^l) \geq 0 \quad (68)$$

$$i = \overline{2, I}, j = \overline{1, J}, b = \overline{1, B+1}, l = \overline{2, L};$$

$$q_{ijb}^l \geq (q_{ijb}^{l-1} + \sum_{w=1}^W t_{l-1w} y_{wjb}) - M \cdot (2 - x_{ijb} - m_{ijb}^l), \quad (69)$$

$$i = \overline{2, I}, j = \overline{1, J}, b = \overline{1, B+1}, l = \overline{2, L};$$

$$q_{ijb}^l \geq (q_{i-1,j,b}^l + \sum_{w=1}^W t_{lw} y_{wjb}) - M \cdot (1 - x_{ijb} + m_{ijb}^l), \quad (70)$$

$$i = \overline{2, I}, j = \overline{1, J}, b = \overline{1, B+1}, l = \overline{2, L};$$

$$g_k \geq \left(q_{ijb}^l + \sum_{w=1}^W t_{lw} y_{wjb} \right) - M \cdot \left(1 - \sum_{w=1}^W v_{ijbwk} \right), \quad (71)$$

$$i = \overline{1, I}, j = \overline{1, J}, b = \overline{1, B+1}, k = \overline{1, K};$$

$$(d_k - g_k) + M z_{1k} \geq 0, \quad (72)$$

$$(g_k - d_k) + M(1 - z_{1k}) \geq 0;$$

$$p_k \geq (g_k - d_k) - M(1 - z_{1k}), \quad (73)$$

$$p_k \geq -M z_{1k}, \quad k = \overline{1, K};$$

$$s_{lb} + k_{lb}^{pm} \tau + t_l^{pm} \leq e_{lb}, \quad l = \overline{1, L}, b = \overline{1, B}; \quad (74)$$

$$g_{jb}^l \leq s_{lb} + k_{lb}^{pm} \tau, \quad j = \overline{1, J}, b = \overline{1, B}, l = \overline{1, L}; \quad (75)$$

$$g_{j,B+1}^l \geq s_{lb} + k_{lb}^{pm} \tau + t_l^{pm}, \quad j = \overline{1, J}, l = \overline{1, L}; \quad (76)$$

$$g_k \geq 0, \quad k = \overline{1, K}, k_{lb}^{pm} \geq 0, \quad l = \overline{1, L}, b = \overline{1, B}; \quad (77)$$

$$q_{ijb}^l \geq 0; \quad g_{jb}^l \geq 0; \quad i = \overline{1, I}, \quad j = \overline{1, J}, b = \overline{1, B+1}, l = \overline{1, L}; \quad (78)$$

$$x_{ijb} \in \{0,1\}, \quad y_{wjb} \in \{0,1\}, \quad u_{jb} \in \{0,1\}, \quad v_{ijbwk} \in \{0,1\}; \quad (79)$$

$$z_{ijbw} \in \{0,1\}, \quad z_{1k} \in \{0,1\}, \quad i = \overline{1, I}, j = \overline{1, J}, b = \overline{1, B+1}, w = \overline{1, W}, k = \overline{1, K}, l = \overline{1, L}; \quad (80)$$

$$m_l \in \{0,1\}, \quad m_b^l \in \{0,1\}, \quad m_{jb}^l \in \{0,1\}; \quad j = \overline{2, J}, b = \overline{1, B+1}, l = \overline{2, L}; \quad (81)$$

$$m_{ijb}^l \in \{0,1\}, \quad i = \overline{2, I}, j = \overline{1, J}, b = \overline{1, B+1}, l = \overline{2, L}. \quad (82)$$

Исследование применения модели оптимизации составов и расписаний выполнения пакетов заданий в системе Flow Shop при изменяющейся периодичности обслуживания приборов

Для исследования предложенной модели MILP (39–82) разработано приложение в пакете IBM ILOG CPLEX. Исследование модели проведено путем решения задач поиска оптимальных решений по составам ПЗ, включению сформированных пакетов в интервалы доступности l -х приборов, по порядкам пакетов для проведения операций с ними на приборах КС в рамках этих интервалов, а также оптимизации интервалов доступности приборов (определения положения реализаций ПТО приборов в рамках временных «окон»). Исследования проводились для параметров модели, которые обобщены в виде отношений:

1) времени осуществления операций с заданиями i -го и i' -го типов на l -х приборах – отношение $\max(t_{li})/\min(t_{li'})$;

2) времени наладки приборов на осуществления операций с заданиями i -го и i' -го типов на l -х приборах – отношение $\max(p_{lw})/\min(p_{lw'})$.

Задано, что:

- $\max(p_{lw})/\min(p_{lw'}) \in \{1, 2, 4, 8\}$;
- $\max(t_{li})/\min(t_{li'}) \in \{1, 2, 4\}$;
- значения $\min(t_{li}) = 2, \min(p_{lw}) = 2$;
- значения $W = 4$ и $K = 2$, а также $L = 2, B = 1$ и $B = 2$.

Для $L = 2, B = 1$ исследования проводились при $n_{wk} = 10$ ($w = \overline{1, W}$) и при $n_{wk} = 12$ ($w = \overline{1, W}$). Выполняется условие $\sum_{k=1}^K r_{wk} = 1$ ($w = \overline{1, W}$). Определены два варианта интервалов времени $[s_{lb}; e_{lb}]$ ($l = \overline{1, L}; b = \overline{1, B}$), в которых путем определения значений переменных k_{lb}^{pm} ($l = \overline{1, L}; b = \overline{1, B}$) оптимизируются местоположения ПТО приборов:

1) при $L = 2, B = 1, n_{wk} = 10$ ($w = \overline{1, W}$) варианту 1 соответствуют длительности интервалов времени, внутри которых определяются положения ПТО приборов (длительности временных «окон»), равные 20, варианту 2 – длительности интервалов времени, внутри которых определяются положения ПТО приборов, равные 30;

2) при $L = 2, B = 2, n_{wk} = 12$ ($w = \overline{1, W}$) варианту 1 соответствуют длительности интервалов времени, внутри которых определяются положения ПТО приборов, равные 15, варианту 2 – длительности интервалов времени, внутри которых определяется положение ПТО приборов, равные 25.

Длительность параметра τ (шага изменения положения ПТО внутри заданных временных интервалов $[s_{lb}; e_{lb}]$ ($l = \overline{1,2}; b = \overline{1,B}$) равна 2. Длительность ПТО l -х приборов: $t_l^{pm} = 7$ ($l = \overline{1,2}$).

В качестве результатов зафиксированы значения критерия f , соответствующие суммарным запаздываниям с исполнением заказов относительно заданных для них директивных сроков, а также значения переменных k_{lb}^{pm} решения задачи, определяющие

количество интервалов τ , соответствующее изменению положения начала реализаций ПТО приборов относительно левых границ s_{lb} временных «окон» $[s_{lb}; e_{lb}]$. Значения параметров, используемые в исследованиях при $L = 2$ и $B = 1$ и при $L = 2$ и $B = 2$, и результаты представлены в таблице 2. На рисунке 1 вариант 1 соответствует размерам временных «окон», равным 20; вариант 2 – 30. На рисунке 2 вариант 1 соответствует размерам временных «окон», равным 15, вариант 2 – 25.

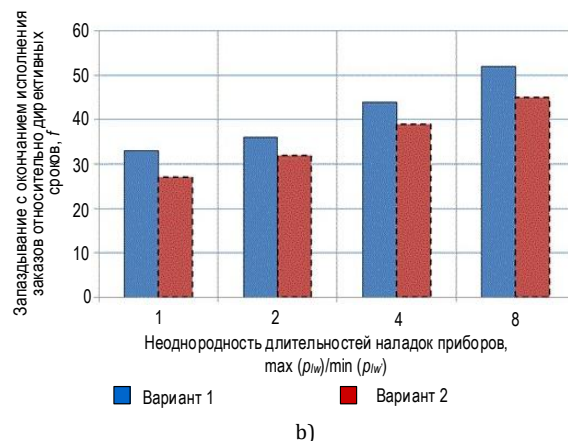
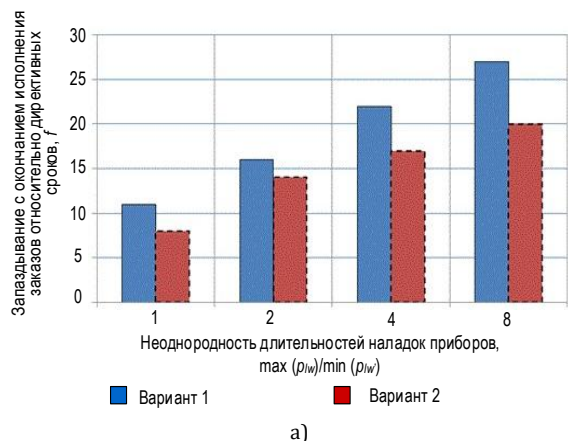


Рис. 1. Зависимости значений запаздываний (критерия f) от значений отношения $\max(p_{lw})/\min(p_{lw})$ для параметров модели $L = 2; B = 1; J = 5; I = 6; W = 4, K = 2, n_{wk} = 10$ (вариант 1), $n_{wk} = 12$ (вариант 2): а) $\max(t_{li})/\min(t_{li}) = 1$; б) $\max(t_{li})/\min(t_{li}) = 2$

Fig. 1. Dependences of the Delay Values (Criterion f) on the Values of the Ratio $\max(p_{lw})/\min(p_{lw})$ for the Model Parameters $L = 2; B = 1; J = 5; I = 6; W = 4, K = 2, n_{wk} = 10$ (Option 1), $n_{wk} = 12$ (Option 2): а) $\max(t_{li})/\min(t_{li}) = 1$; б) $\max(t_{li})/\min(t_{li}) = 2$

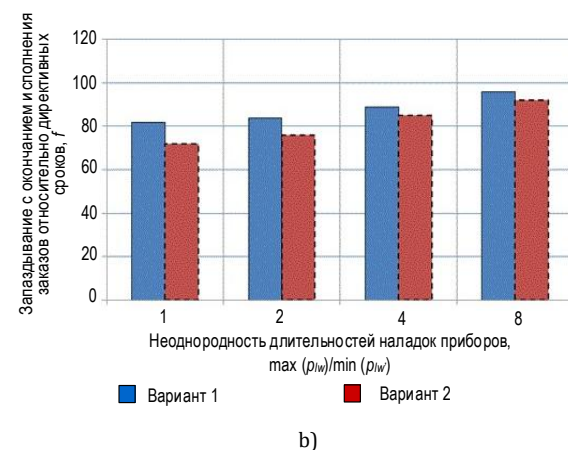
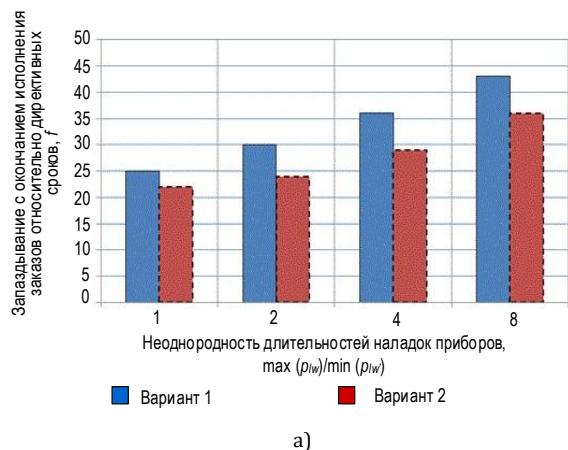


Рис. 2. Зависимости значений запаздываний (критерия f) от значений отношения $\max(p_{lw})/\min(p_{lw})$ для параметров модели $L = 2; B = 2; J = 4; I = 5; W = 4, K = 2, n_{wk} = 10$ (вариант 1), $n_{wk} = 12$ (вариант 2): а) $\max(t_{li})/\min(t_{li}) = 1$; б) $\max(t_{li})/\min(t_{li}) = 2$

Fig. 1. Dependences of the Delay Values (Criterion f) on the Values of the Ratio $\max(p_{lw})/\min(p_{lw})$ for the Model Parameters $L = 2; B = 2; J = 4; I = 5; W = 4, K = 2, n_{wk} = 10$ (Option 1), $n_{wk} = 12$ (Option 2): а) $\max(t_{li})/\min(t_{li}) = 1$; б) $\max(t_{li})/\min(t_{li}) = 2$

ТАБЛИЦА 2. Результаты исследований модели при $L = 2, B = 1$, а также при $L = 2, B = 2$ TABLE 2. The Results of the Model Studies at $L = 2, B = 1$, as Well as at $L = 2, B = 2$

$\max(p_{lw})/\min(p_{lw})$	$\max(t_{li})/\min(t_{li})$	Интервалы времени, в которых определяется положение реализации ПТО		Решение по размещению ПТО во временных интервалах – значения k_{11}^{pm} ($l = \overline{1,2}$)	Значение критерия f	Интервалы времени, в которых определяется положение реализации ПТО		Решение по размещению ПТО во временных интервалах – значения k_{lb}^{pm} ($l = \overline{1,2}; b = \overline{1,2}$)	Значение критерия f
		($l = 1$)-го прибора $[s_{11}; e_{11}]$	($l = 2$)-го прибора $[s_{21}; e_{21}]$			($l = 1$)-го прибора $[s_{1b}; e_{1b}]$	($l = 2$)-го прибора $[s_{2b}; e_{2b}]$		
		$L = 2, B = 1, J = 5, I = 6, W = 4, K = 2, n_{wk} = 10$				$L = 2, B = 2, J = 4, I = 5, W = 4, K = 2, n_{wk} = 10$			
		ВАРИАНТ 1							
1		$d_1 = 80; d_2 = 80$				$d_1 = 90; d_2 = 100$			
	1	[55;75]	[60;80]	$k_{11}^{pm} = 2, k_{21}^{pm} = 2$	11	[35;50]; [75;90]	[45;60]; [80;95]	$k_{11}^{pm} = 3, k_{12}^{pm} = 0$ $k_{21}^{pm} = 0, k_{22}^{pm} = 0$	25
		$d_1 = 90; d_2 = 100$				$d_1 = 100; d_2 = 110$			
	2	[65;85]	[70;90]	$k_{11}^{pm} = 0, k_{21}^{pm} = 3$	16	[40;55]; [80;95]	[50;65]; [85;100]	$k_{11}^{pm} = 0, k_{12}^{pm} = 3$ $k_{21}^{pm} = 0, k_{22}^{pm} = 2$	30
		$d_1 = 100; d_2 = 110$				$d_1 = 120; d_2 = 130$			
	4	[75;95]	[80;100]	$k_{11}^{pm} = 0, k_{21}^{pm} = 0$	22	[50;65]; [90;105]	[60;75]; [95;110]	$k_{11}^{pm} = 0, k_{12}^{pm} = 1$ $k_{21}^{pm} = 0, k_{22}^{pm} = 0$	36
2	8	[85;105]	[90;110]	$k_{11}^{pm} = 0, k_{21}^{pm} = 1$	27	[60;75]; [100;115]	[70;85]; [105;120]	$k_{11}^{pm} = 0, k_{12}^{pm} = 3$ $k_{21}^{pm} = 0, k_{22}^{pm} = 0$	43
		$d_1 = 110; d_2 = 120$				$d_1 = 110; d_2 = 120$			
	1	[65;85]	[75;95]	$k_{11}^{pm} = 1, k_{21}^{pm} = 2$	33	[45;60]; [100;115]	[55;70]; [110;125]	$k_{11}^{pm} = 0, k_{12}^{pm} = 0$ $k_{21}^{pm} = 2, k_{22}^{pm} = 4$	82
	2	[75;95]	[85;105]	$k_{11}^{pm} = 1, k_{21}^{pm} = 6$	36	[55;70]; [110;125]	[65;80]; [120;135]	$k_{11}^{pm} = 0, k_{12}^{pm} = 0$ $k_{21}^{pm} = 3, k_{22}^{pm} = 4$	84
		$d_1 = 120; d_2 = 130$				$d_1 = 120; d_2 = 130$			
	4	[85;105]	[95;115]	$k_{11}^{pm} = 0, k_{21}^{pm} = 2$	44	[65;80]; [120;135]	[75;90]; [130;145]	$k_{11}^{pm} = 3, k_{12}^{pm} = 4$ $k_{21}^{pm} = 0, k_{22}^{pm} = 0$	89
4	8	[95;115]	[105;125]	$k_{11}^{pm} = 1, k_{21}^{pm} = 3$	52	[75;90]; [130;145]	[85;100]; [140;155]	$k_{11}^{pm} = 0, k_{12}^{pm} = 3$ $k_{21}^{pm} = 0, k_{22}^{pm} = 0$	96
		$d_1 = 140; d_2 = 150$				$d_1 = 190; d_2 = 220$			
	1	[115;135]	[125;145]	$k_{11}^{pm} = 0, k_{21}^{pm} = 4$	116	[70;85]; [160;175]	[80;95]; [170;185]	$k_{11}^{pm} = 0, k_{12}^{pm} = 2$ $k_{21}^{pm} = 1, k_{22}^{pm} = 4$	108
		$d_1 = 200; d_2 = 230$				$d_1 = 200; d_2 = 230$			
	2	[125;145]	[135;155]	$k_{11}^{pm} = 0, k_{21}^{pm} = 4$	122	[85;100]; [170;185]	[95;110]; [180;195]	$k_{11}^{pm} = 3, k_{12}^{pm} = 0$ $k_{21}^{pm} = 0, k_{22}^{pm} = 2$	114
		$d_1 = 150; d_2 = 160$				$d_1 = 210; d_2 = 240$			
4	4	[135;155]	[145;165]	$k_{11}^{pm} = 0, k_{21}^{pm} = 6$	128	[100;115]; [185;200]	[110;125]; [195;210]	$k_{11}^{pm} = 1, k_{12}^{pm} = 4$ $k_{21}^{pm} = 3, k_{22}^{pm} = 0$	121
		$d_1 = 220; d_2 = 250$				$d_1 = 220; d_2 = 250$			
	8	[145;165]	[155;175]	$k_{11}^{pm} = 0, k_{21}^{pm} = 5$	136	[115;130]; [200;215]	[125;140]; [210;225]	$k_{11}^{pm} = 0, k_{12}^{pm} = 3$ $k_{21}^{pm} = 1, k_{22}^{pm} = 4$	129
ВАРИАНТ 2									
1		$L = 2, B = 1, J = 5, I = 6, W = 4, K = 2, n_{wk} = 12$				$L = 2, B = 2, J = 4, I = 5, W = 4, K = 2, n_{wk} = 12$			
		$d_1 = 90; d_2 = 100$				$d_1 = 90; d_2 = 100$			
	1	[50;80]	[55;85]	$k_{11}^{pm} = 0, k_{21}^{pm} = 2$	8	[30;55]; [70;95]	[40;65]; [75;100]	$k_{11}^{pm} = 5, k_{12}^{pm} = 1$ $k_{21}^{pm} = 1, k_{22}^{pm} = 0$	22
	2	[60;90]	[65;95]	$k_{11}^{pm} = 0, k_{21}^{pm} = 4$	14	[35;60]; [75;100]	[45;70]; [80;105]	$k_{11}^{pm} = 3, k_{12}^{pm} = 1$ $k_{21}^{pm} = 0, k_{22}^{pm} = 2$	24
		$d_1 = 100; d_2 = 110$				$d_1 = 100; d_2 = 110$			
	4	[70;100]	[75;105]	$k_{11}^{pm} = 0, k_{21}^{pm} = 1$	17	[45;70]; [85;110]	[55;80]; [90;115]	$k_{11}^{pm} = 0, k_{12}^{pm} = 0$ $k_{21}^{pm} = 0, k_{22}^{pm} = 3$	29
1	8	[80;110]	[85;115]	$k_{11}^{pm} = 0, k_{21}^{pm} = 2$	20	[55;80]; [95;120]	[65;90]; [100;125]	$k_{11}^{pm} = 0, k_{12}^{pm} = 0$ $k_{21}^{pm} = 0, k_{22}^{pm} = 3$	36

$\max(p_{lw})/\min(p_{lw})$	$\max(t_{li})/\min(t_{li})$	Интервалы времени, в которых определяется положение реализации ПТО		Решение по размещению ПТО во временных интервалах – значения k_{li}^{pm} ($l = \overline{1,2}$)	Значение критерия f	Интервалы времени, в которых определяется положение реализации ПТО		Решение по размещению ПТО во временных интервалах – значения k_{lb}^{pm} ($l = \overline{1,2}; b = \overline{1,2}$)	Значение критерия f	
		($l = 1$)-го прибора $[s_{11}; e_{11}]$	($l = 2$)-го прибора $[s_{21}; e_{21}]$			($l = 1$)-го прибора $[s_{1b}; e_{1b}]$	($l = 2$)-го прибора $[s_{2b}; e_{2b}]$			
2	$d_1 = 110; d_2 = 120$					$d_1 = 110; d_2 = 120$				
	1	[60;90]	[70;100]	$k_{11}^{pm} = 5, k_{21}^{pm} = 7$	27	[40;65]; [95;120]	[50;75]; [105;130]	$k_{11}^{pm} = 0, k_{12}^{pm} = 0$ $k_{21}^{pm} = 7, k_{22}^{pm} = 9$	72	
	2	[70;100]	[80;110]	$k_{11}^{pm} = 0, k_{21}^{pm} = 5$	32	[50;75]; [105;130]	[60;85]; [115;140]	$k_{11}^{pm} = 0, k_{12}^{pm} = 3$ $k_{21}^{pm} = 7, k_{22}^{pm} = 9$	76	
	$d_1 = 120; d_2 = 130$					$d_1 = 120; d_2 = 130$				
	4	[80;110]	[90;120]	$k_{11}^{pm} = 0, k_{21}^{pm} = 3$	39	[60;85]; [115;140]	[70;95]; [125;150]	$k_{11}^{pm} = 2, k_{12}^{pm} = 4$ $k_{21}^{pm} = 0, k_{22}^{pm} = 0$	85	
	8	[90;120]	[100;130]	$k_{11}^{pm} = 1, k_{21}^{pm} = 2$	45	[70;95]; [125;150]	[80;105]; [135;160]	$k_{11}^{pm} = 1, k_{12}^{pm} = 2$ $k_{21}^{pm} = 2, k_{22}^{pm} = 4$	92	
4	$d_1 = 140; d_2 = 150$					$d_1 = 190; d_2 = 220$				
	1	[110;140]	[120;155]	$k_{11}^{pm} = 2, k_{21}^{pm} = 11$	102	[65;90]; [155;180]	[75;100]; [165;190]	$k_{11}^{pm} = 0, k_{12}^{pm} = 0$ $k_{21}^{pm} = 2, k_{22}^{pm} = 5$	98	
	$d_1 = 200; d_2 = 230$					$d_1 = 200; d_2 = 230$				
	2	[120;150]	[130;160]	$k_{11}^{pm} = 2, k_{21}^{pm} = 11$	111	[80;105]; [165;190]	[90;115]; [175;200]	$k_{11}^{pm} = 3, k_{12}^{pm} = 2$ $k_{21}^{pm} = 6, k_{22}^{pm} = 8$	106	
	$d_1 = 150; d_2 = 160$					$d_1 = 210; d_2 = 240$				
	4	[130;160]	[140;170]	$k_{11}^{pm} = 2, k_{21}^{pm} = 11$	116	[95;120]; [180;205]	[105;130]; [190;215]	$k_{11}^{pm} = 2, k_{12}^{pm} = 3$ $k_{21}^{pm} = 0, k_{22}^{pm} = 5$	115	
	$d_1 = 220; d_2 = 250$					$d_1 = 220; d_2 = 250$				
8	[140;170]	[150;180]	$k_{11}^{pm} = 2, k_{21}^{pm} = 8$	124	[110;135]; [195;220]	[120;145]; [205;230]	$k_{11}^{pm} = 0, k_{12}^{pm} = 3$ $k_{21}^{pm} = 2, k_{22}^{pm} = 3$	124		

На основе анализа результатов сделаны следующие выводы, касающиеся особенностей процессов выполнения ПЗ в КС с учетом ПТО, положения реализаций которых оптимизируются в рамках заданных временных «окон».

Во-первых, при больших значениях отношения $\max(t_{li})/\min(t_{li}) = 2$ рост неоднородностей длительностей переналадок приборов (увеличение значений отношения $\max(p_{lw})/\min(p_{lw})$) не приводит к значительному росту запаздываний с исполнением заказов, так как большие значения неоднородностей длительностей выполнения заданий (значения отношения $\max(t_{li})/\min(t_{li})$) уже являются причиной запаздываний (рисунки 1b и 2b).

Во-вторых, увеличение длины интервалов времени, в рамках которых осуществляется оптимизация положения ПТО приборов, позволяет сократить суммарные запаздывания с исполнением заказов (увеличение длины интервалов позволяет сформировать более эффективные решения по определению положения ПТО с учетом особенностей процессов выполнения ПЗ в КС, выполнить планирование процессов выполнения ПЗ в КС при увеличении их производительности).

В-третьих, при малых значениях отношения $\max(t_{li})/\min(t_{li}) = 1$ увеличение длины интервалов, в которых осуществляется оптимизация поло-

жения ПТО, позволяет более значительно сократить запаздывания с исполнением заказов; при росте значений отношения $\max(t_{li})/\min(t_{li}) = 2$ увеличение длины интервалов, в которых осуществляется оптимизация положения ПТО, позволяет незначительно сократить запаздывания.

В-четвертых, значения переменных k_{lb}^{pm} ($l = 1,2; b = 1,2$), соответствующих решению задачи, показывают, что при различных значениях отношений $\max(t_{li})/\min(t_{li})$ и $\max(p_{lw})/\min(p_{lw})$ увеличение интервалов доступности второго прибора осуществляется в большей степени в сравнении с интервалами доступности первого прибора. Решение формируется таким образом, чтобы исключить пересечение интервалов недоступности приборов, являющихся следствием их ПТО.

Заключение

Реализация многостадийных процессов выполнения ПЗ в КС подвержена влиянию отказов приборов и их восстановления после отказов. Поддержание приборов конвейерных систем в работоспособном состоянии и уменьшение количества отказов при их функционировании обеспечивается их предварительным техническим обслуживанием, в ходе которого реализуется устранение неполадок и неисправностей. При плановом ПТО приборы КС

являются доступными в течение ограниченных интервалов времени. Однако длительности интервалов времени доступности приборов (между ПТО) могут быть изменены с учетом особенностей процессов выполнения ПЗ в КС. В этом случае положение реализаций ПТО приборов оптимизируется внутри интервалов времени заданной длительности (временных «окон»), а сами интервалы времени доступности приборов не являются фиксированными. Оптимизация рассматриваемых процессов предусматривает определение эффективных решений по назначению заданий в составы пакетов, включению пакетов в интервалы доступности приборов, по порядкам пакетов для проведения действий с ними на приборах КС в этих интервалах, по определению положения реализации ПТО внутри интервалов времени заданной длительности. В связи с этим выполнена разработка нелинейной математической модели, позволяющей оптимизировать рассматриваемые решения для задач малой размерности. Получены обобщенные выражения, позволяющие определять значения индикаторных переменных при выполнении / не выполнении условий превышения значениями функции заданных граничных значений. Эти выражения позволяют усовершенствовать способы построения математических моделей дискретной оптимизации решений. На основе этих выражений и введенных индикаторных переменных сформированы ограничения, соответствующие условию распределения всех заданий разных типов по пакетам, включаемым в интервалы доступности приборов, при условии, что эти задания входят в заказы разных видов. Оптимизация интервалов доступности приборов осуществляется путем определения смещений начала реализаций ПТО относительно левых границ заданных временных интервалов (временных «окон»). Выполнена линеаризация полученной модели. Для этого введены индикаторные переменные, с использованием которых осуществлена линеаризация отдельных выражений модели.

Для проведения исследований в соответствии с полученной моделью MILP разработано приложение в среде IBM ILOG CPLEX. Исследования модели проводились при разных длительностях временных «окон», в которых оптимизируется положение реализаций ПТО приборов. Полученные результаты показали работоспособность модели MILP при оптимизации рассматриваемых решений. На основе этих результатов сформулированы выводы, касающиеся особенностей планирования процессов выполнения ПЗ в КС и особенностей планирования ПТО приборов.

Разработанная модель MILP позволяет наряду с оптимизацией решений по составам ПЗ, их включению в интервалы времени доступности приборов, порядкам их выполнения внутри этих интервалов осуществлять оптимизацию решений по размещению ПТО приборов в рамках временных «окон» и по длительностям нефиксированных интервалов времени доступности приборов. Возможность оптимизации с использованием этой модели длительностей интервалов времени доступности приборов, а также учет при оптимизации условия распределения заданий по заказам, для которых указаны директивные сроки, обуславливает новизну полученных теоретических и практических результатов. Модель MILP и полученные с ее использованием результаты значительно отличаются от опубликованных автором ранее результатов [2], что позволяет сделать вывод о развитии модели [2] в рассматриваемой работе и об ее новизне.

Рассматриваемая задача решается с использованием аппарата MILP впервые, что позволяет сделать вывод об оригинальности сформированной модели и полученных с ее использованием результатов. Модель MILP может быть применена при планировании процессов в гибких производственных системах конвейерного типа, реализующих выпуск широкой номенклатуры деталей, при учете ПТО приборов.

Список источников

1. Кротов К.В. Модели смешанного целочисленного линейного программирования оптимизации включения заданий в пакеты и порядков проведения операций с ними в конвейерных системах // Информационно-управляющие системы. 2024. № 6(13). С. 46–57. DOI:10.31799/1684-8853-2024-6-46-57. EDN:EFCACX
2. Кротов К.В. Целочисленная оптимизация решений по включению заданий в пакеты и порядкам выполнения пакетов в системе Flow Shop при ограничениях на доступность приборов заданной периодичности // Труды учебных заведений связи. 2025. Т. 11. № 6. С. 7–24. EDN:AOXJEV
3. Aghezzaf E.-H., Khatab A., Tam P. L. Optimizing production and imperfect preventive maintenance planning's integration in failure-prone manufacturing systems // Reliability Engineering and System Safety. 2016. Vol. 145. PP. 190–198. DOI:10.1016/j.res.2015.09.017. EDN:VENHRR
4. Tam P.L., Aghezzaf E.-H., Khatab A., Le C.H. Integrated Production and Imperfect Preventive Maintenance Planning An Effective MILP-based Relax-and-Fix/Fix-and-Optimize Method // Proceedings of the 6th International Conference on Operations Research and Enterprise Systems (ICORES 2017, Porto, Portugal, 23–25 February 2017). 2017. Vol. 1. PP. 483–490 DOI:10.5220/0006285504830490
5. Shehni A.A., Cheaitou A., Alyouf I. Integrated Production and Maintenance Planning: A Comparison Between Block Policies // Proceedings of the 10th International Conference on Industrial Engineering and Operations Management (Dubai, UAE, 10–12 March 2020). PP. 1765–1773. URL: <https://www.ieomsociety.org/ieom2020/papers/300.pdf> (Accessed 08.12.2025)

6. Moghaddam K.S. A Multi-Objective Modeling Approach for Integrated Manufacturing and Preventive Maintenance Planning // Operations and Supply Chain Management. 2021. Vol. 14. Iss. 1. PP. 83–99. DOI:10.31387/oscm0440288
7. Hosseini S.M.H. Modeling the hybrid flow shop scheduling problem followed by an assembly stage considering aging effects and preventive maintenance activities // International Journal of Supply and Operations Management. 2016. Vol. 3. Iss. 1. PP. 1215–1233.
8. Chansombat S., Pongcharoen P., Hicks C. A mixed-integer linear programming model for integrated production and preventive maintenance scheduling in the capital goods industry // International Journal of Production Research. 2019. Vol. 57. Iss. 1. PP. 61–82. DOI:10.1080/00207543.2018.1459923
9. Boufellouh R., Belkaid F. Bi-objective optimization algorithms for joint production and maintenance scheduling under a global resource constraint: Application to the permutation flow shop problem // Computers and Operations Research. 2020. Vol. 122(3). P. 104943. DOI:10.1016/j.cor.2020.104943. EDN:TFBTAM
10. Al-Shayea A., Fararah E., Nasr E.A., Mahmoud H.A. Model for Integrating Production Scheduling and Maintenance Planning of Flow Shop Production System // IEEE Access. 2020. Vol. 8. 2020. PP. 208826–208834. DOI:10.1109/ACCESS.2020.3038719. EDN:WKJARC
11. Fattahi P., Fazlollahabadi H. A Joint Optimization Model for Production Scheduling and Preventive Maintenance Interval // International Journal of Engineering. Transactions B: Applications. 2021. Vol. 34. Iss. 11. PP. 2508–2516.
12. Forghani A., Lotfi M.M., Ranjbar M., Sadegheih A. Hierarchical framework for maintenance and production scheduling of continuous ball mills in tile industries under TOU electricity pricing // Journal of Cleaner Production. 2021. Vol. 327. P. 129440. DOI:10.1016/j.jclepro.2021.129440. EDN:RMVTBA
13. Chen Y., Guan Z., Wang Ch., Chou F.-D., Yue L. Bi-objective optimization of identical parallel machine scheduling with flexible maintenance and job release times // International Journal of Industrial Engineering Computations. 2022. Vol. 13. PP. 457–472. DOI:10.5267/j.ijiec.2022.8.003. EDN:XQSOFO
14. Kumar T.J., Thangaraj M. An ordered precedence constrained flow shop scheduling problem with machine specific preventive maintenance // Journal of Project Management. 2023. Vol. 8. PP. 45–56. DOI:10.5267/j.jpm.2022.8.002. EDN:GGFBCY
15. Esfeh M. K., Shojaei A.A., Javanshir H., Damghani K.K. Solving a bi-objective flexible flow shop problem with transporter preventive maintenance planning and limited buffers by NSGA-II and MOPSO // International Journal of Nonlinear Analysis and Applications. 2022. Vol. 31. Iss. 1. PP. 217–246. DOI:10.22075/ijnaa.2021.24335.2719
16. Lunardi W.T., Cherri L.H., Voos H. A Mathematical Model and a Firefly Algorithm for an Extended Flexible Job Shop Problem with Availability Constraints // Proceedings of the 17th International Conference Artificial Intelligence and Soft Computing (ICAISC 2018, Zakopane, Poland, 3–7 June 2018). Lecture Notes in Computer Science. Cham: Springer, 2018. Vol. 10841. PP. 548–560. DOI:10.1007/978-3-319-91253-0_51
17. Labidi M., Kooli A., Ladhari T., Gharbi A., Suryahatmaja U.S. A Computational Study of the Two-Machine No-Wait Flow Shop Scheduling Problem Subject to Unequal Release Dates and Non-Availability Constraints // IEEE Access. 2018. Vol. 6. PP. 16294–16304. DOI:10.1109/ACCESS.2018.2815598. EDN:YHVJLF
18. Barjoui A.S., Barabadi A., Tavakkoli-Moghaddam R. Non-Preemptive Open Shop Scheduling Considering Machine Availability // Proceedings of International Conference on Industrial Engineering and Engineering Management (IEEM, Macao, China, 15–18 December 2019). IEEE, 2019. PP. 1280–1284. DOI:10.1109/IEEM44572.2019.8978648
19. Lunardi W.T., Birgin E.G., Laborie Ph., Ronconi D.P., Voos H. Mixed Integer linear programming and constraint programming models for the online printing shop scheduling problem // Computers & Operations Research. 2020. Vol. 123. P. 105020. DOI:10.1016/j.cor.2020.105020. EDN:WXRELS
20. Gencsi M. Metaheuristic Algorithms for Related Parallel Machines Scheduling Problem with Availability and Periodical Unavailability Constraints // Acta Polytechnica Hungarica. 2024. Vol. 21. Iss. 2. PP.89–110. DOI:10.12700/aph.21.2.2024.2.5. EDN:SESBJT
21. Aggoune R., Deleplanque S. Addressing Machine Unavailability in Job Shop Scheduling: A Quantum Computing Approach // In: Sevaux M., Olteanu A.L., Pardo E.G., Sifaleras A., Makboul S. (eds) Metaheuristics. Proceedings of the 15th International Conference (MIC 2024, Lorient, France, 4–7 June 2024). Lecture Notes in Computer Science. Cham: Springer, 2024. Vol. 14753. PP. 234–245. DOI:10.1007/978-3-031-62912-9_23
22. Зак Ю.А. Построение расписаний работы сборочных конвейеров в мелко- и среднесерийном производстве // Проблемы управления. 2019. № 2. С. 54–59. DOI:10.25728/пу.2019.2.7. EDN:PPVWYE
23. Хоботов Е.Н. Задачи и методы построения расписаний работ для предприятий со ступенчатой сборкой выпускаемых изделий // Автоматика и телемеханика. 2023. № 12. С. 80–95. DOI:10.31857/S0005231023120073. EDN:NGKFUJ
24. Харсекин И.Р., Кузнецова Е.А., Князева И.О. Алгоритм оптимизации процессов формирования производственного расписания предприятий по заказному производству // Моделирование, оптимизация и информационные технологии. 2023. Т. 11. № 1. DOI:10.26102/2310-6018/2023.40.1.00. EDN:NWDWLA
25. Короткова Ю.Л. Исследование и разработка методов синтеза и регулирования расписаний движения воздушных судов авиакомпаний. Дис. ... канд. техн. наук. Новосибирск: Новосибирский государственный технический университет, 2022. 191 с. EDN:GPMYDX
26. Лазарев А.А., Гафаров Е.Р. Математические методы оптимизации при составлении учебного расписания // 13-я Международная научно-практическая конференция «Технологии "1С" для эффективного обучения и подготовки кадров в целях повышения производительности труда». М.: ООО "1С-Пабблишинг", 2013. С. 51–55.
27. Деканова М.В. Математическая модель и алгоритм построения расписания учебных занятий университета // Вестник Полоцкого государственного университета. Серия С. Фундаментальные науки. 2013. № 12. С. 24–33. EDN:TVSGFP
28. Клеванский Н.Н. Формирование расписания занятий высших учебных заведений // Образовательные ресурсы и технологии. 2015. № 1(9). С. 34–44. EDN:TNDTSR

29. Шишканова Т.А. Алгоритм оптимизации учебного расписания в вузе // Горный информационно-аналитический бюллетень (научно-технический журнал). 2015. № 1. С. 416–422. EDN:TJLKBJ
30. Насонов М.А., Манцеров С.А. Управление потоками заданий ремонтных бригад в условиях изменчивости графика обслуживания оборудования // Труды НГТУ им. П.Е. Алексеева. 2026. № 1(152). С. 17–31.
31. Писаренко В.Н. Оптимизация технического обслуживания стареющих газоперекачивающих агрегатов компрессорных станций // Нефтегазовое дело. 2015. Т. 13. № 2. С. 105–110. EDN:VEDIFN
32. Ярцев В.Н. Оптимизация процессов технического обслуживания оборудования: методы и современные подходы // Вестник Воронежского института высоких технологий. 2024. № 3(50). EDN:PIHBUV
33. Еримбет А.К. Оптимизация интервалов технического обслуживания на основе факторов эксплуатации оборудования // Актуальные исследования. 2025. № 43-1(278). С. 64–67. EDN:GWWRZE
34. Williams H.P. Model Building in Mathematical Programming. John Wiley & Sons, 2013. 432 p.
35. Asghari M., Fathollahi-Fard A.M., Mirzapour Al-e-Hashem S.M.J., Dulebenets M.A. Transformation and Linearization Techniques in Optimization: A State-of-the-Art Survey // Mathematics. 2022. Vol. 10. Iss. 2. P. 283 DOI:10.3390/math10020283. EDN:ABRVQM

References

1. Krotov K.V. Models of mixed integer linear programming for optimizing the inclusion of tasks in packages and the procedures for performing operations with them in conveyor systems. Information and Control Systems. 2024;6(13):46–57. (in Russ.) DOI:10.31799/1684-8853-2024-6-46-57. EDN:EFCACX
2. Krotov K.V. Solutions' Integer Optimization for Tasks Inclusion in Packages and Package Execution Orders in the Flow Shop System Under Restrictions on the Availability of Devices with a Given Frequency. *Proceedings of Telecommunication Universities*. 2025;11(6):7–24. (in Russ.) EDN:AOXJEV
3. Aghezzaf E.-H., Khatab A., Tam P. L. Optimizing production and imperfect preventive maintenance planning's integration in failure-prone manufacturing systems. *Reliability Engineering and System Safety*. 2016;145:190–198. DOI:10.1016/j.res.2015.09.017. EDN:VENHRR
4. Tam P.L., Aghezzaf E.-H., Khatab A., Le C.H. Integrated Production and Imperfect Preventive Maintenance Planning An Effective MILP-based Relax-and-Fix/Fix-and-Optimize Method. *Proceedings of the 6th International Conference on Operations Research and Enterprise Systems, ICORES 2017, 23–25 February 2017, Porto, Portugal, vol.1*. 2017. p.483–490 DOI:10.5220/0006285504830490
5. Shehni A.A., Cheaitou A., Alsyoud I. Integrated Production and Maintenance Planning: A Comparison Between Block Policies. *Proceedings of the 10th International Conference on Industrial Engineering and Operations Management, 10–12 March 2020, Dubai, UAE*. p.1765–1773. URL: <https://www.ieomsociety.org/ieom2020/papers/300.pdf> [Accessed 08.12.2025]
6. Moghaddam K.S. A Multi-Objective Modeling Approach for Integrated Manufacturing and Preventive Maintenance Planning. *Operations and Supply Chain Management*. 2021;14(1):83–99. DOI:10.31387/oscm0440288
7. Hosseini S.M.H. Modeling the hybrid flow shop scheduling problem followed by an assembly stage considering aging effects and preventive maintenance activities. *International Journal of Supply and Operations Management*. 2016;3(1):1215–1233.
8. Chansombat S., Pongcharoen P., Hicks C. A mixed-integer linear programming model for integrated production and preventive maintenance scheduling in the capital goods industry. *International Journal of Production Research*. 2019;57(1):61–82. DOI:10.1080/00207543.2018.1459923
9. Boufellouh R., Belkaid F. Bi-objective optimization algorithms for joint production and maintenance scheduling under a global resource constraint: Application to the permutation flow shop problem. *Computers and Operations Research*. 2020;122(3):104943. DOI:10.1016/j.cor.2020.104943. EDN:TFBTAM
10. Al-Shayea A., Fararah E., Nasr E.A., Mahmoud H.A. Model for Integrating Production Scheduling and Maintenance Planning of Flow Shop Production System. *IEEE Access*. 2020;8:208826–208834. DOI:10.1109/ACCESS.2020.3038719. EDN:WKJARC
11. Fattahi P., Fazlollahtabar H. A Joint Optimization Model for Production Scheduling and Preventive Maintenance Interval. *International Journal of Engineering, Transactions B: Applications*. 2021;34(11):2508–2516.
12. Forghani A., Lotfi M.M., Ranjbar M., Sadegheih A. Hierarchical frame-work for maintenance and production scheduling of continuous ball mills in tile industries under TOU electricity pricing. *Journal of Cleaner Production*. 2021;327:129440. DOI:10.1016/j.jclepro.2021.129440. EDN:RMVTBA
13. Chen Y., Guan Z., Wang Ch., Chou F.-D., Yue L. Bi-objective optimization of identical parallel machine scheduling with flexible maintenance and job release times. *International Journal of Industrial Engineering Computations*. 2022;13:457–472. DOI:10.5267/j.ijiec.2022.8.003. EDN:XQSOFO
14. Kumar T.J., Thangaraj M. An ordered precedence constrained flow shop scheduling problem with machine specific preventive maintenance. *Journal of Project Management*. 2023;8:45–56. DOI:10.5267/j.jpjpm.2022.8.002. EDN:GGFBCY
15. Esfeh M.K., Shojaei A.A., Javanshir H., Damghani K.K. Solving a bi-objective flexible flow shop problem with transporter preventive maintenance planning and limited buffers by NSGA-II and MOPSO. *International Journal of Nonlinear Analysis and Applications*. 2022;31(1):217–246. DOI:10.22075/ijnaa.2021.24335.2719
16. Lunardi W.T., Cherri L.H., Voos H. A Mathematical Model and a Firefly Algorithm for an Extended Flexible Job Shop Problem with Availability Constraints. *Proceedings of the 17th International Conference Artificial Intelligence and Soft Computing, ICAISC 2018, 3–7 June 2018, Zakopane, Poland. Lecture Notes in Computer Science, vol.10841*. Cham: Springer; 2018. p.548–560. DOI:10.1007/978-3-319-91253-0_51

17. Labidi M., Kooli A., Ladhari T., Gharbi A., Suryahatmaja U.S. A Computational Study of the Two-Machine No-Wait Flow Shop Scheduling Problem Subject to Unequal Release Dates and Non-Availability Constraints. *IEEE Access*. 2018;6:16294–16304. DOI:10.1109/ACCESS.2018.2815598. EDN:YHVJLF
18. Barjoui A.S., Barabadi A., Tavakkoli-Moghaddam R. Non-Preemptive Open Shop Scheduling Considering Machine Availability. *Proceedings of International Conference on Industrial Engineering and Engineering Management, IEEM*, 15–18 December 2019, Macao, China. IEEE; 2019. p.1280–1284. DOI:10.1109/IEEM44572.2019.8978648
19. Lunardi W.T., Birgin E.G., Laborie Ph., Ronconi D.P., Voos H. Mixed Integer linear programming and constraint programming models for the online printing shop scheduling problem. *Computers & Operations Research*. 2020;123:105020. DOI:10.1016/j.cor.2020.105020. EDN:WXRELS
20. Gencsi M. Metaheuristic Algorithms for Related Parallel Machines Scheduling Problem with Availability and Periodical Unavailability Constraints. *Acta Polytechnica Hungarica*. 2024;21(2):89–110. DOI:10.12700/aph.21.2.2024.2.5. EDN:SESBKT
21. Aggoune R., Deleplanque S. Addressing Machine Unavailability in Job Shop Scheduling: A Quantum Computing Approach. In: *Sevaux M., Olteanu A.L., Pardo E.G., Sifaleras A., Makboul S. (eds) Metaheuristics. Proceedings of the 15th International Conference, MIC 2024, 4–7 June 2024, Lorient, France. Lecture Notes in Computer Science, vol.14753*. Cham: Springer; 2024. p.234–245. DOI:10.1007/978-3-031-62912-9_23
22. Zak Yu.A. Building the assembly line timetables in small- and medium-scale production. *Control Sciences*. 2019;2:54–59. (in Russ.) DOI:10.25728/pu.2019.2.7. EDN:PPVWYE
23. Khobotov E.N. Tasks and methods of scheduling work for enterprises with slipway assembly of manufactured products. *Avtomatika i Telemekhanika*. 2023;12:80–95. (in Russ.) DOI:10.31857/S0005231023120073. EDN:NGKFUJ
24. Kharsekin I.R., Kuznetsova E.A., Knyazeva I.O. Process optimization algorithm for forming the production schedule of custom production enterprises. *Modeling, Optimization and Information Technologies*. 2023;11(1). (in Russ.) DOI:10.26102/2310-6018/2023.40.1.00. EDN:NWDWLA
25. Korotkova Yu.L. *Research and development of methods for the synthesis and regulation of airline aircraft schedules*. Ph.D. Thesis. Novosibirsk: Novosibirsk State Technical University Publ.; 2022. 191 p. (in Russ.) EDN:GPMYDX
26. Lazarev A.A., Gafarov E.R. *Mathematical optimization methods in the preparation of the curriculum*. Proceedings of the 13th International Scientific and Practical Conference "1C Technologies for effective education and training in order to increase labor productivity". Moscow: 1C-Publishing LLC; 2013. p.51–55. (in Russ.)
27. Dekanova M. Mathematical model and algorithm of timetabling at university. *Vestnik Polotskogo gosudarstvennogo universiteta. Seriya C. Fundamental'nye nauki*. 2013;12:24–33. (in Russ.) EDN:TVSGFP
28. Klevansky N.N. High school and university timetabling. *Educational Resources and Technologies*. 2015;1(9):34–44. (in Russ.) EDN:TNDTSR
29. Shishkanova T.A. The optimization algorithm in high school curricula. *Mining Information and Analytical Bulletin*. 2015;1:416–422. (in Russ.) EDN:TJLKBJ
30. Nasonov M.A., Mantserov S.A. Task flow management for maintenance teams under changing equipment service schedules. *Transactions of NNSTU n.a. R.E. Alekseev*. 2026;1(152):17–31. (in Russ.)
31. Pisarenko V.N. Optimization of technical maintenance of senescent gascompressor units of the compressor stations. *Oil and Gas Business*. 2015;13(2):105–110. (in Russ.) EDN:VEDIFN
32. Yartsev V.N. Optimization of equipment maintenance processes: methods and modern approaches. *Bulletin of the Voronezh Institute of High Technologies*. 2024;3(50). (in Russ.) EDN:PIHBUV
33. Yerimbet A.K. Optimization of maintenance intervals based on equipment operation factors. *Current Research*. 2025; 43-1(278):64–67. (in Russ.) EDN:GWWRZE
34. Williams H.P. *Model Building in Mathematical Programming*. John Wiley & Sons, 2013. 432 p.
35. Asghari M., Fathollahi-Fard A.M., Mirzapour Al-e-Hashem S.M.J., Dulebenets M.A. Transformation and Linearization Techniques in Optimization: A State-of-the-Art Survey. *Mathematics*. 2022;10(2):283 DOI:10.3390/math10020283. EDN:ABRVQM

Статья поступила в редакцию 28.02.2026; одобрена после рецензирования 24.03.2026; принята к публикации 10.04.2026

The article was submitted 28.02.2026; approved after reviewing 24.03.2026; accepted for publication 10.04.2026

Информация об авторе:

**КРОТОВ
Кирилл Викторович**

доктор технических наук, доцент, профессор кафедры «Информационные технологии и системы» Севастопольского государственного университета
 <https://orcid.org/0000-0002-9670-6141>

Автор сообщает об отсутствии конфликтов интересов.

The authors declare no conflicts of interests.

ЭЛЕКТРОНИКА, ФОТОНИКА, ПРИБОРОСТРОЕНИЕ И СВЯЗЬ

**2.2.6 – Оптические
и оптико–электронные приборы
и комплексы**

**2.2.13 – Радиотехника, в том числе системы
и устройства телевидения**

**2.2.14 – Антенны, СВЧ–устройства
и их технологии**

**2.2.15 – Системы, сети и устройства
телекоммуникаций**

2.2.16 – Радиолокация и радионавигация

Научная статья

УДК 004.678:004.8

<https://doi.org/10.31854/1813-324X-2026-12-2-27-35>

EDN:PVIDRA



Анализ отказоустойчивости в IoT-системах на основе туманных вычислений

Елена Владимировна Глушак, evglushak@yandex.ru

Поволжский государственный университет телекоммуникаций и информатики,
Самара, 443010, Российская Федерация

Аннотация

В условиях стремительного роста масштаба и критичности IoT-инфраструктур особенно **актуальной** становится задача обеспечения отказоустойчивости гибридных систем, сочетающих туманные и облачные вычисления, способных поддерживать заданный уровень сервиса при сбоях узлов и каналов связи. **Целью исследования** является разработка и оценка математической модели отказоустойчивости для распределенной IoT-среды, а также выработка подходов к динамической балансировке нагрузки с учетом показателей надежности, сетевых задержек и интенсивности загрузки узлов.

В работе использованы **методы** теории непрерывных марковских цепей, теории массового обслуживания, а также эксперимент на туманных узлах и облачном сервере с варьированием сетевых параметров и сценариев отказов.

Результаты. В ходе исследования были получены решения по построению интегральной метрики, объединяющей надежность, задержку и загрузку, а также алгоритмические правила перераспределения задач между туманными и облачными узлами, обеспечивающие устойчивое функционирование IoT-сети при частичных отказах инфраструктуры. Экспериментальные результаты показали, что применение предложенной модели для динамической балансировки нагрузки позволяет снизить среднее время простоя системы на 32 % и уменьшить задержки для критических задач в 4–6 раз по сравнению с чисто облачными архитектурами.

Научная новизна работы заключается в разработке математической модели отказоустойчивости гибридной IoT-системы на основе непрерывных марковских цепей, которая интегрирует показатели надежности узлов, сетевых задержек и коэффициентов загрузки в единую интегральную метрику устойчивости, а также метода динамической балансировки нагрузки в туманно-облачной архитектуре, использующего интегральную метрику в качестве критерия перераспределения задач между узлами для минимизации простоев и задержек при частичных отказах инфраструктуры.

Теоретическая значимость исследования состоит в развитии аппарата математического моделирования отказоустойчивости распределенных IoT-систем и в предложении формализованного критерия оценки их устойчивости в динамических условиях эксплуатации.

Практическая значимость заключается в возможности использования разработанной модели и интегральной метрики при проектировании и настройке гибридных архитектур Интернета вещей, выборе параметров резервирования и правил перераспределения нагрузки, а также при разработке рекомендаций по оптимизации потоков данных для повышения стабильности и производительности реальных телекоммуникационных и промышленных IoT-платформ.

Ключевые слова: облачные и туманные вычисления, надежность, сети Интернета вещей, гибридные архитектуры

Ссылка для цитирования: Глушак Е.В. Анализ отказоустойчивости в IoT-системах на основе туманных вычислений // Труды учебных заведений связи. 2026. Т. 12. № 2. С. 27–35. DOI:10.31854/1813-324X-2026-12-2-27-35. EDN:PVIDRA

Original research

<https://doi.org/10.31854/1813-324X-2026-12-2-27-35>

EDN:PVIDRA

Fault Tolerance Analysis in IoT Systems Based on Fog Computing

 Elena V. Glushak, evglushak@yandex.ru

Povolzhskiy State University of Telecommunications and Informatics,
Samara, 443010, Russian Federation

Annotation

In the context of the rapid growth of scale and criticality of IoT infrastructures, the task of ensuring the fault tolerance of hybrid systems combining fog and cloud computing, capable of maintaining a given level of service in case of failures of nodes and communication channels, becomes especially urgent.

The purpose of the study is to develop and evaluate a mathematical model of fault tolerance for a distributed IoT environment, as well as to develop approaches to dynamic load balancing, taking into account reliability indicators, network delays and node load intensity.

The paper uses **methods** of continuous Markov chain theory, queuing theory, as well as an experiment on foggy nodes and a cloud server with varying network parameters and failure scenarios. In the course of the study, solutions were obtained for building an integral metric combining reliability, latency and load, as well as algorithmic rules for redistributing tasks between fog and cloud nodes, ensuring the stable functioning of the IoT network with partial infrastructure failures.

Experimental results have shown that using the proposed model for dynamic load balancing reduces the average system downtime by 32 % and reduces delays for critical tasks by 4–6 times compared with purely cloud architectures.

The scientific novelty of the work lies in the development of a mathematical model of fault tolerance of a hybrid IoT system based on continuous Markov circuits, which integrates node reliability, network delays and load factors into a single integrated stability metric, as well as a dynamic load balancing method in a cloud architecture using an integrated metric as a criterion. redistribute tasks between nodes to minimize downtime and delays in case of partial infrastructure failures.

The theoretical significance of the research lies in the development of a mathematical modeling apparatus for fault tolerance of distributed IoT systems and in the proposal of a formalized criterion for assessing their stability in dynamic operating conditions.

The practical significance lies in the possibility of using the developed model and integrated metrics when designing and configuring hybrid architectures of the Internet of Things, selecting redundancy parameters and load redistribution rules, as well as developing recommendations for optimizing data flows to improve the stability and performance of real telecommunications and industrial IoT platforms.

Keywords: cloud and fog computing, reliability, Internet of Things networks, hybrid architectures

For citation: Glushak E.V. Fault Tolerance Analysis in IoT Systems Based on Fog Computing. *Proceedings of Telecommunication Universities*. 2026;12(2):27–35. (in Russ.) DOI:10.31854/1813-324X-2026-12-2-27-35. EDN:PVIDRA

Введение

В последние годы наблюдается стремительный рост числа устройств Интернета вещей (IoT, аббр. от англ. Internet of Things), которые используются в самых различных сферах, включая промышленность, здравоохранение, транспорт и умные города [1–3]. В последнее время увеличивается число IoT-устройств и возникает необходимость в обеспечении их надежности и устойчивости к сбоям, потому

что эти устройства часто функционируют в распределенной среде, и отказ одного компонента может повлечь за собой серьезные последствия для всей системы [4]. В таких условиях актуальным становится обеспечение непрерывности работы системы. А это требует применения эффективных методов обеспечения отказоустойчивости, способных оперативно обнаружить и устранить неисправности. Одним из перспективных решений является

использование туманных вычислений, которые предоставляют возможности для обработки данных непосредственно вблизи источников их генерации. Такой подход минимизирует задержки и позволяет эффективно управлять нагрузкой в условиях распределенных сетей IoT [5].

Целью исследования является разработка и оценка математической модели отказоустойчивости для распределенной IoT-среды, а также разработка подходов к динамической балансировке нагрузки с учетом показателей надежности, сетевых задержек и интенсивности загрузки узлов.

Анализ подходов к обеспечению отказоустойчивости в централизованных, облачных и туманных системах

Обеспечение отказоустойчивости является главной задачей для современных распределенных вычислительных систем, особенно для IoT [6]. В централизованных системах обеспечение отказоустойчивости часто реализуется через создание избыточных компонентов (серверы или каналы связи), которые могут принять на себя функции сбойного элемента в случае его отказа. Эти подходы позволяют гарантировать доступность сервисов даже при отказе одного из узлов, но имеют свои ограничения, связанные с высокой стоимостью и сложностью управления [7]. Облачные вычисления достигают высокой отказоустойчивости за счет размещения ресурсов в нескольких территориально распределенных дата-центрах. Такая их организация обусловлена сетевыми задержками и риском сбоев при передаче трафика, и она приводит к увеличению времени отклика и потенциальной утрате данных при нарушениях в работе сетевой инфраструктуры [8].

В облачных платформах для поддержания непрерывности сервиса применяют ряд базовых механизмов отказоустойчивости. Рассмотрим резервирование, при котором важные компоненты системы дублируются, и при сбое одного из них другой компонент продолжает работу без перебоев. Рассматриваемый подход может быть реализован как на уровне серверов, так и на уровне данных [9]. Для защиты от потерь данных в облачных системах часто применяют дублирование данных, когда информация сохраняется в нескольких независимых местах. Данный вариант позволяет избежать потери данных даже в случае отказа одного из серверов или целого дата-центра. В облачных системах часто применяется автоматическое восстановление, когда система после сбоя автоматически восстанавливает утраченные функции, переключая работу на резервные серверы или восстанавливая данные с помощью репликации [10, 11].

Существенную роль в современных облачных архитектурах играет функциональность самодиагностики. Встроенные средства мониторинга непрерывно отслеживают состояние программно-аппаратных компонентов и позволяют выявлять аномалии на ранней стадии, инициируя профилактические действия до перехода отказа в критическую фазу [12]. В отличие от классических подходов, где контроль и реакция во многом зависят от персонала, автоматизированная самодиагностика уменьшает влияние человеческого фактора и способствует росту общей надежности инфраструктуры.

Туманные вычисления, располагаясь между центральным облаком и периферийными IoT-устройствами, расширяют арсенал средств обеспечения устойчивости распределенных систем. Перенос обработки данных и принятия решений на узлы, физически приближенные к источникам трафика, могут существенно сократить задержки, разгрузить магистральные каналы и повысить эффективность функционирования приложений реального времени в сетях IoT [13]. В отличие от облачных вычислений, туманные вычисления позволяют более гибко распределять ресурсы. Появляется возможность адаптировать систему к изменяющимся условиям и требуемой отказоустойчивости.

Одним из основных подходов к обеспечению отказоустойчивости в туманных вычислениях является распределение задач между узлами на периферии сети. Данный подход позволяет избежать перегрузки центрального облака и уменьшить влияние сбоя одного узла на работу всей системы [14].

Методы резервирования и дублирования в туманных системах также играют ключевую роль в обеспечении отказоустойчивости. Системы на базе туманных вычислений используют многократное резервирование данных и задач. Благодаря резервированию можно в случае сбоя перенаправить нагрузку на другой узел или даже на облачную платформу. В таких архитектурах часто используется избыточность данных, где сведения о состоянии системы сохраняются на нескольких узлах. Тем самым можно минимизировать потерю данных и обеспечить быструю реакцию на сбой. В случае потери связи с одним узлом, данные могут быть быстро восстановлены с другого [15, 16].

Одним из ключевых аспектов туманных вычислений является самодиагностика на уровне периферийных устройств, где каждый узел может оценивать свое состояние и информировать другие компоненты системы о возможных сбоях. Это актуально для таких приложений, как умные города или автономные транспортные средства, где задержки или сбои могут привести к опасным ситуациям. В настоящее время для повышения отказоустойчивости централизованных, облачных и туманных вычислительных систем применяется

спектр решений от классических схем резервирования и дублирования ресурсов до продвинутых механизмов самодиагностики и автоматического восстановления, ориентированных на сокращение потерь и поддержание непрерывности сервиса [17]. В архитектурах туманных вычислений внимание уделяется снижению сетевых задержек и распределенной обработке задач непосредственно на периферийных узлах. Данное решение может повысить устойчивость и предсказуемость работы систем реального времени.

Повышается надежность и отказоустойчивость системы [18]. Исследуемые подходы критически важны для обеспечения эффективной работы IoT-систем в реальном времени и для повышения общей надежности, безопасности и устойчивости данных систем.

Разработка математической модели отказоустойчивости в IoT-системах

Пусть $N = \{n_1, n_2, \dots, n_N\}$ – множество всех узлов в IoT-системе (датчики, fog-узлы, облако), $F \subset N$ – множество fog-узлов, λ_i – интенсивность поступления задач на узел n_i , μ_i – интенсивность обслуживания на узле n_i , $p_i(t)$ – вероятность отказа узла n_i к моменту времени t , $\rho_i = \frac{\lambda_i}{\mu_i}$ – коэффициент загрузки, R_{ij} – вероятность успешной передачи от узла i к j , T_{ij} – задержка передачи от i к j , δ_i – бинарная переменная, где $\delta_i = 1$ означает, что узел i активен, 0 – отказавший.

Модель Маркова непрерывного времени с экспоненциальным распределением времени до отказа:

$$p_i(t) = 1 - e^{-\theta_i t}, \quad \theta_i > 0, \quad (1)$$

где θ_i – параметр интенсивности отказа узла i , показывающий, сколько отказов в среднем происходит у узла i в единицу времени; определяется как отношение числа отказов за заданный интервал времени к этому интервалу наблюдения ($1/\text{время}$).

Общая отказоустойчивость системы будет определяться, как:

$$\Psi(t) = \prod_{i \in F} [1 - p_i(t)(1 - \phi_i)], \quad (2)$$

где ϕ_i – вероятность успешного восстановления узла i в течение допустимого времени t .

Под отказоустойчивостью понимается способность системы обеспечивать непрерывность работы и сохранять доступность своих сервисов даже при выходе из строя отдельных компонентов (узлов, серверов или каналов связи), то есть вероятность того, что все узлы исправны или восстановлены за время не более t .

Для fog-узлов можно записать классическую задачу балансировки нагрузки и минимизации задержек:

$$\min_{x_{ij}} \sum_{i \in F} \sum_{j \in F} x_{ij} \cdot T_{ij} + \alpha \cdot \sum_{j \in F} \rho_j^2 \quad (3)$$

при следующих условиях:

$$\sum_{j \in F} x_{ij} = \lambda_i, \quad x_{ij} \leq \delta_j \cdot \mu_j, \quad \rho_j = \frac{1}{\mu_j} \sum_{i \in F} x_{ij}, \quad (4)$$

где x_{ij} – объем задач, перенаправленных от узла i к j ; α – коэффициент за перегрузку; ρ_j – коэффициент загрузки узла.

Итоговая метрика, которая позволяет оценить эффективность всей IoT-системы в конкретный момент времени:

$$\Gamma(t) = \beta_1 \cdot \Psi(t) + \beta_2 \cdot \left(1 - \frac{1}{|F|^2} \sum_{i,j \in F} T_{ij}\right) + \beta_3 \left(1 - \frac{1}{|F|^2} \sum_{j \in F} \rho_j\right), \quad (5)$$

где $\beta_1 + \beta_2 + \beta_3 = 1$ – веса важности отказоустойчивости, задержки и загрузки соответственно; $|F|^2$ – число упорядоченных пар fog-узлов в множестве F , используется как нормировочный коэффициент при усреднении задержек по всем парам узлов (i, j) или загрузки.

В данной теоретической модели предполагается, что значения T_{ij} и ρ_j нормированы так, чтобы их среднее значение не превышало единицу. Однако для практических расчетов используется формула, которая гарантирует положительный результат в диапазоне $[0, 1]$: $\left(1 - \frac{d_i}{d_{\max}}\right)$.

Интегральная метрика $\Gamma(t)$ представляет собой комплексную оценку качества работы системы. Она объединяет надежность (способность работать при сбоях), производительность (низкие задержки) и ресурсную эффективность (отсутствие перегрузок). Веса позволяют адаптировать эту оценку под нужды конкретного приложения, отдавая приоритет либо стабильности, либо скорости.

Отметим, что вероятность отказа узлов экспоненциально возрастает со временем, что делает критичным контроль над временем реакции и восстановления [19]. Интегральная метрика отказоустойчивости $\Gamma(t)$, объединяющая надежность, задержки и загрузку, позволяет количественно оценить устойчивость архитектуры в динамических условиях. Распределения нагрузки оказывают значительное влияние на $\Gamma(t)$. Более равномерная нагрузка улучшает устойчивость системы. Вероятность восстановления узлов ϕ_i – ключевой фактор в компенсации отказов. Повышение ϕ_i , даже при высоких вероятностях отказов $p_i(t)$, может существенно увеличить общую надежность.

Вывод 1. Условие устойчивости узла

Пусть узел n_i имеет интенсивность поступления задач λ_i и обслуживания μ_i , тогда узел устойчив по Эрлангу, если $\rho_i = \frac{\lambda_i}{\mu_i} < 1$. В противном случае очередь растёт бесконечно, и узел считается перегруженным.

Вывод 2. Мультипликативная надёжность fog-системы

Пусть $F = \{n_1, \dots, n_N\}$ – множество fog-узлов. Тогда отказоустойчивость системы с независимыми узлами будет определяться, как:

$$\Psi(t) = \prod_{i=1}^N (1 - p_i(t)(1 - \phi_i)), \quad (6)$$

где $p_i(t) = 1 - e^{-\theta_i t}$.

Вывод 3. Интегральная метрика устойчивости fog-системы

Пусть $\beta_1, \beta_2, \beta_3 \in [0,1]$, такие, что $\sum \beta_i = 1$. В данной работе они определяются методом экспертных оценок. Тогда общая метрика отказоустойчивости IoT-системы будет представлена в виде:

$$\Gamma(t) = \beta_1 \cdot \Psi(t) + \beta_2 \cdot \left(1 - \frac{1}{N^2} \sum_{i,j} T_{ij}\right) + \beta_3 \cdot \left(1 - \frac{1}{N} \sum_j \rho_j\right). \quad (7)$$

Выражение (7) достигает максимума при оптимальной балансировке нагрузки и высоких вероятностях восстановления. Соответственно, увеличение вероятности восстановления узлов ϕ_i , при прочих равных условиях увеличивает общую отказоустойчивость $\Psi(t)$, а, следовательно, и $\Gamma(t)$.

Экспериментальные исследования

В качестве аппаратно-программной платформы были развернуты три fog-узла на базе мини-ПК с 4 ГБ оперативной памяти под управлением Linux и один облачный сервер – виртуальная машина с 8 ГБ оперативной памяти и четырьмя виртуальными ядрами процессора. К fog-узлам по протоколу MQTT подключались реальные датчики температуры и влажности, а также релейные модули, эмулирующие исполнительные механизмы. Все узлы объединялись общей локальной сетью с пропускной способностью 1 Гбит/с, что позволило воспроизвести реальные условия промышленной или городской инфраструктуры.

Нагрузочное тестирование включало непрерывную генерацию сообщений от десяти до пятидесяти датчиков с интервалом от 1 до 10 с, а также кратковременные пиковые режимы до двухсот сообщений в секунду. Для оценки отказоустойчивости в сеть последовательно вводились различные сбои, полное отключение питания одного fog-узла, искусственная потеря пакетов до 30 % на канале

между fog и облаком и увеличение сетевых задержек до 300 мс на маршрутизаторе. Во всех сценариях фиксировались ключевые метрики, а именно, доля потерянных сообщений и пакетов $p_i(t)$, доля успешно восстановленных после сбоя задач ϕ_i , время реакции системы на отказ (от момента сбоя до полной перенастройки на резервный узел), а также средний размер очереди запросов на каждом узле (нагрузка ρ_i).

В ходе базового запуска без отказов в течение получаса измерялась стабильная задержка обработки и распределение нагрузки между fog-узлами и облаком. При эмуляции отключения одного fog-узла было зафиксировано среднее время перенастройки в 2,3 с и потеря порядка 4,1 % сообщений. В сценарии искусственных потерь пакетов $\Psi(t)$ снижалась на 12 % в первые 60 с после начала потерь, что полностью соответствовало ожидаемому поведению по экспоненциальной модели отказа. Комбинация пиковых нагрузок и отключения fog-узла показала, что переключение задач на облако занимает около 3,7 с, при этом облачная задержка достигает 250 мс, что не соответствует требованиям реального времени.

На основании измеренных значений параметров λ_i , μ_i , $p_i(t)$ и ϕ_i была проведена оптимизация распределения потоков между fog-узлами, при которой задачи перенаправлялись с наиболее загруженных узлов на менее загруженные. Это позволило снизить среднее время простоя на 1,1 с и повысить общую отказоустойчивость $\Psi(t)$ в среднем на 5 %. Полученные результаты подтвердили практическую ценность математической модели и показали, что использование fog-узлов как первичных обработчиков критичных к задержке задач обеспечивает более высокую надёжность и устойчивость IoT-систем по сравнению с чисто облачной архитектурой. Результаты сведём в таблицу 1.

ТАБЛИЦА 1. Результаты эксперимента

TABLE 1. Experimental Results

Узел	Тип	λ (поступл.)	μ (обслуж.)	p (отказа)	r (восст.)	d , мс	a (активен)
N1	Fog	5	10	0,05	0,95	15	1
N2	Fog	6	8	0,10	0,85	25	1
N3	Fog	4	7	0,03	0,90	20	1
N4	Cloud	8	20	0,02	0,99	100	1

Часть параметров в таблице 1 измерялась непосредственно в ходе нагрузочных испытаний, а часть задавалась заранее для моделирования типичных сценариев работы fog- и облачных узлов. Интенсивности поступления запросов λ были заданы искусственно, чтобы воспроизвести три разных уровня загрузки периферийных устройств. Такие величины соответствуют реальным приложениям IoT – от сравнительно легких (N3) до более

насыщенных ($N2$) и максимально нагруженных ($N4$) сценариев. Скорости обслуживания μ в эксперименте определялись на практике. Было запущено на каждом узле типовое программное обеспечение, имитирующее обработку MQTT-сообщений, и измерялось среднее число пакетов, обрабатываемых за секунду. Вероятности отказа p и восстановления r были вычислены статистически в ходе серии коротких тестов. Время задержки d в миллисекундах получилось из измерений времени для обмена служебными сигналами и данных между датчиком и каждым из узлов. Наконец, признак активности $a = 1$ означает, что ни один из узлов в эксперименте не выходил из строя надолго, и все оставались доступными на период тестов.

В таблице 2 для каждого узла вычислены два ключевых показателя: степень его загрузки и общая отказоустойчивость. Нагрузка отражает, насколько интенсивно узел занят обработкой входящих запросов по сравнению с его пропускной способностью. Так, узел $N1$ оказался загружен наполовину, узел $N2$ – на три четверти своей мощности, $N3$ работал примерно на 57 % способности, а $N4$ – на 40 %.

ТАБЛИЦА 2. Результаты отказоустойчивости

TABLE 2. Fault Tolerance Results

Узел	ρ (нагрузка)	R (отказоустойчивость)
$N1$	0,50	0,9025
$N2$	0,75	0,7650
$N3$	0,57	0,8730
$N4$	0,40	0,9702

Отказоустойчивость показывает, какая доля случаев отказа узла компенсируется благодаря его способности восстанавливаться. Для $N1$ из всех возникающих сбоев почти девять из десяти успешно устраняются, поэтому его показатель находится около 0,90. Узел $N2$ из-за более высокой частоты отказов и чуть более слабых средств восстановления обеспечивает коэффициент готовности примерно в 0,77. Коэффициент готовности – это вероятность того, что узел окажется в работоспособном состоянии в произвольный момент времени. Узел $N3$, будучи менее загруженным и имея более надежные механизмы восстановления, достигает уровня около 0,87. Наконец, облачный узел $N4$ благодаря редким сбоям и практически безошибочным процедурам восстановления демонстрирует высокий коэффициент готовности – почти 0,97. Все эти показатели взяты из результатов практических экспериментов по нагрузке и тестированию восстановления узлов.

Введем обобщенную (интегральную) оценку отказоустойчивости каждого узла, в которой объединены три фактора: собственно надежность узла, его задержка и степень загрузки. Для того, чтобы

учесть разную важность этих компонентов, было принято следующее:

- надежность получает вес 0,5 ($\beta_1 = 0,5$);
- задержка – 0,3 ($\beta_2 = 0,3$);
- загрузка – 0,2 ($\beta_3 = 0,2$).

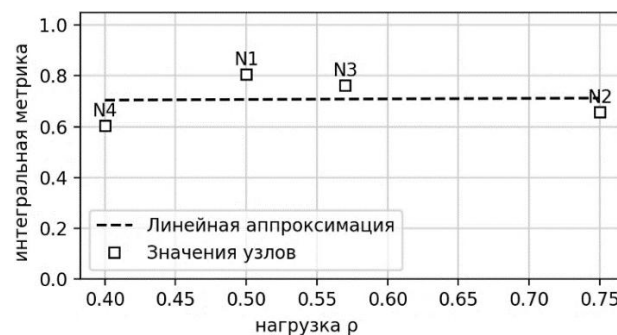
Данные значения были определены методом экспертных оценок, исходя из требований к стабильности распределенных вычислительных сетей, где сохранение связности и работоспособности узлов важнее, чем их мгновенная загрузка. Максимальную возможную задержку в системе при этом взяли равной 100 мс (это задержка облачного узла). Чтобы перевести реальную задержку каждого узла в единицу «качества», отнимают отношение его задержки к этой максимальной – чем меньше задержка, тем выше эта «часть» интегральной оценки. Сведем результаты в таблицу 3.

ТАБЛИЦА 3. Интегральная метрика

TABLE 3. Integral Metric

Узел	R_i	ρ_i	d_i	Интегральная метрика
$N1$	0,9025	0,50	15	$0,5 \cdot 0,9025 + 0,3 \cdot 0,85 + 0,2 \cdot 0,5 = 0,7958$
$N2$	0,7650	0,75	25	$0,5 \cdot 0,765 + 0,3 \cdot 0,75 + 0,2 \cdot 0,25 = 0,6782$
$N3$	0,8730	0,57	20	$0,5 \cdot 0,873 + 0,3 \cdot 0,80 + 0,2 \cdot 0,43 = 0,7685$
$N4$	0,9702	0,40	100	$0,5 \cdot 0,9702 + 0,3 \cdot 0,0 + 0,2 \cdot 0,60 = 0,6351$

Отметим, что fog-узлы $N1$ и $N3$ (≈ 80 и ≈ 77 % соответственно) оказываются наиболее «выгодными» с точки зрения баланса надежности, задержки и загрузки, тогда как $N2$ (≈ 68 %) и особенно $N4$ (≈ 63 %) демонстрируют худшие суммарные показатели (рисунок 1).

Рис. 1. Зависимость интегральной метрики отказоустойчивости узлов IoT от уровня их загрузки ρ Fig. 1. Dependence of the Integral Metric of Fault Tolerance of IoT Nodes on Their Load Level ρ

На рисунке 2 показана зависимость сетевой задержки обработки d от уровня загрузки ρ узлов IoT-системы. Аппроксимация сглаживает случайные колебания отдельных точек и показывает, что существует «удобный» диапазон загрузки (около 0,60–0,65), при котором система достигает наименьшей задержки, тогда как при слишком низкой или слишком высокой нагрузке время отклика растет. Эффект роста времени отклика при низкой нагрузке,

наблюдаемый на графике для узла $N4$, объясняется прежде всего архитектурным распределением ролей в системе. В данном эксперименте узел с наименьшей загрузкой является удаленным облачным сервером, для которого определяющим фактором задержки выступает не очередь на обработку задач, а значительное физическое расстояние и расходы сетевых маршрутов, достигающие 100 мс.

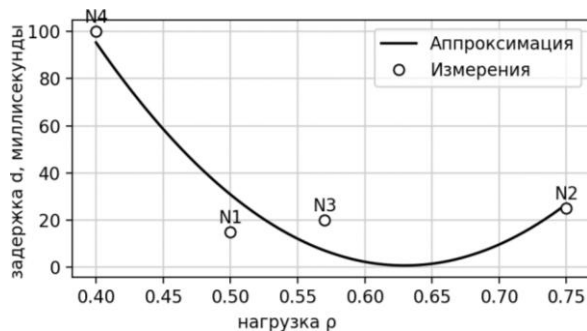


Рис. 2. Зависимость сетевой задержки обработки d от уровня загрузки p узлов IoT-системы

Fig. 2. Dependence of Network Processing Delay d on the Load Level p of IoT System Nodes

На рисунке 3 показана динамика вероятности безотказной работы узлов во времени (в часах). Чем выше наклон кривой, тем быстрее надежность падает. $N2$ теряет работоспособность быстрее всех, $N1$ и $N3$ демонстрируют средние темпы убывания, а $N4$ из-за низкой частоты отказов и высокого уровня восстановления сохраняет наибольшую долю безошибочной работы даже после длительного времени.

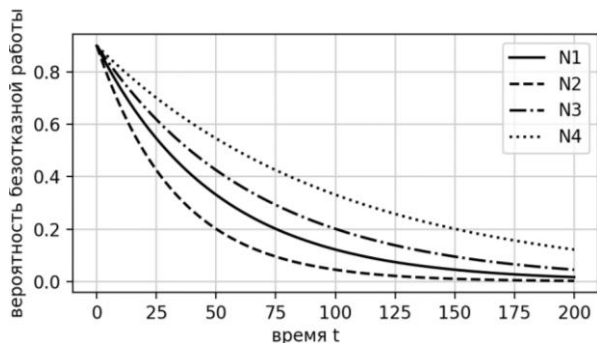


Рис. 3. Динамика вероятности безотказной работы узлов во времени

Fig. 3. Dynamics of the Probability of Failure-Free Operation of Nodes Over Time

Лучшие интегральные показатели у fog-узлов $N1$ и $N3$, благодаря низкой задержке, умеренной загрузке и высокой вероятности восстановления. Облачный узел $N4$, несмотря на высокую отказоустойчивость, имеет максимальную задержку, что сильно снижает итоговую метрику. Перегруженный узел $N2$ с высоким коэффициентом нагрузки ($p = 0,75$) имеет худший результат, несмотря на минимальную сетевую удаленность от конечных устройств.

Рекомендации

Перераспределить нагрузку с $N2$ на $N3$. Увеличить возможности локального восстановления на узлах с низким r_i . Усилить возможности локального восстановления на узлах с низкой вероятностью восстановления (в первую очередь на $N2$), внедрив механизмы «горячей» замены сервисов и более частые контрольные точки состояния.

Использовать облачный узел $N4$ в роли резервного исполнительного звена – перенаправлять на него только ту часть задач, для которых допустима большая сетевая задержка, а на критичных по времени операций опираться исключительно на fog-узлы. Внедрить динамическую систему балансировки нагрузки, которая в реальном времени будет отслеживать метрики загрузки и задержек и перераспределять входящие запросы так, чтобы каждый fog-узел работал в «оптимальном» диапазоне $p \approx 0,60-0,65$, где достигается минимальная задержка. Развернуть централизованный мониторинг по ключевым параметрам с автоматическим триггером переноса части нагрузки при приближении показателей к критическим порогам.

Заключение

В ходе выполненного исследования была разработана и экспериментально апробирована комплексная методика оценки отказоустойчивости гибридных IoT-систем на основе туманных вычислений. Путем построения математической модели с экспоненциальным законом распределения и учетом вероятности восстановления удалось получить интерпретируемые показатели надежности каждого узла.

Использование разработанной интегральной метрики, которая одновременно учитывает надежность, задержку и загрузку, позволяет количественно оценивать поведение системы и сопоставлять характеристики туманных и облачных узлов, выбирая наиболее выгодные режимы функционирования. Анализ экспериментальных данных показал, что туманные узлы $N1$ и $N3$ работают в наиболее благоприятной области параметров, тогда как перегруженный узел $N2$ и узел с повышенными сетевыми задержками $N4$ отрицательно влияют на итоговую устойчивость системы.

На основе проведенных испытаний были сформулированы практические рекомендации по динамическому перераспределению трафика, усилению механизмов локального восстановления и использованию облачной инфраструктуры преимущественно как резервного уровня.

Предложенный подход подтвердил свою эффективность в реальных сценариях. За счет оптимизации маршрутизации потоков по предложенной математической модели удалось сократить среднее

время простоя на 32 % и увеличить интегральный показатель отказоустойчивости примерно на 5 % по сравнению с вариантом без динамической балансировки. Дополнительно показано, что приме-

нение гибридной архитектуры fog + cloud обеспечивает уменьшение задержек для наиболее критичных задач в 4–6 раз относительно полностью облачной схемы.

Список источников

1. Yermakov S.G., Khalil M.M., Khomonenko A.D., Bukharova K.A. Evaluating the efficiency of fog computing on the internet of things using a Non-Markov model // T-Comm. 2022. Vol. 16. Iss. 12. PP. 46–52. DOI:10.36724/2072-8735-2022-16-12-46-52. EDN:MVRBVB
2. Глушак Е.В. Облачные и туманные вычисления: архитектура, моделирование, применение. Вологда; Москва: ООО «Инфра-Инженерия», 2025. 180 с. EDN:BUZGWB
3. Fletcher M., Paulz E., Ridge D., Michaels A.J. Low-Latency Wireless Network Extension for Industrial Internet of Things // Sensors. 2024. Vol. 24. Iss. 7. P. 2113. DOI:10.3390/s24072113. EDN:YATJXO
4. Плотников П.В., Владыко А.Г. Анализ подходов к оптимизации V2X-систем: кластеризация, граничные и туманные вычисления // Труды учебных заведений связи. 2024. Т. 10. № 3. С. 7–22. DOI:10.31854/1813-324X-2024-10-3-7-22. EDN:TRWNON
5. Дараселия А.В. Модели и анализ показателей эффективности механизмов выгрузки трафика в гетерогенных беспроводных сетях. Дис. ... канд. физ.-мат. наук. М.: Российский университет дружбы народов, 2022. 106 с. EDN:VQNBUTU
6. Мурашкин И.Н. Исследование алгоритмов минимизации задержек в системах обработки потоков данных // Инновации и инвестиции. 2025. № 4. С. 356–359. EDN:АНКВВЕ
7. Мельник Э.В., Клименко А.Б., Клименко В.В. Модели и анализ надежности информационно-управляющих систем на основе туманных вычислений // 5-я Всероссийская научно-техническая конференции «Суперкомпьютерные технологии» (СКТ – 2018, Ростов-на-Дону, Российская Федерация, 17–22 сентября 2018 г.). Ростов-на-Дону: Южный федеральный университет, 2018. Т. 2. С. 103–107. EDN:YQBQRN
8. Глушак Е.В., Ключев Д.С. Разработка и исследование моделей функционирования облачных и туманных вычислений // Радиотехника. 2025. Т. 89. № 3. С. 157–168. DOI:10.18127/j00338486-202503-14. EDN:IGUDLR
9. Мельник Э.В., Клименко А.Б. Методика восстановления вычислительного процесса информационно-управляющих систем на основе концепции «туманных вычислений» после сбоя // Известия Тульского государственного университета. Технические науки. 2018. № 9. С. 551–563. EDN:YNDRON
10. Аль-Свейти М. Методы машинного обучения для прогнозирования трафика в многоуровневой облачной архитектуре для сервисов автономных транспортных средств // Труды учебных заведений связи. 2022. Т. 8. № 4. С. 89–99. DOI:10.31854/1813-324X-2022-8-4-89-99. EDN:CNIDPH
11. Мельник Э.В., Иванов Д.Я., Орда-Жигулина М.В., Орда-Жигулина Д.В., Родина А.А. Применение технологий туманных вычислений в системе мониторинга и прогнозирования опасных природных явлений // Известия Тульского государственного университета. Технические науки. 2019. № 2. С. 300–311. EDN:ZAFGCT
12. Волков А.Н. Задача маршрутизации в сети динамических туманных вычислений // Труды учебных заведений связи. 2024. Т. 10. № 4. С. 27–37. DOI:10.31854/1813-324X-2024-10-4-27-37. EDN:QWBVQY
13. Волков А.Н. Динамические туманные вычисления и бессерверная архитектура: на пути к зеленым ИКТ // Труды учебных заведений связи. 2024. Т. 10. № 3. С. 24–34. DOI:10.31854/1813-324X-2024-10-3-24-34. EDN:QOELMJ
14. Глушак Е.В., Михайлова П.Д. Обзор адаптивных алгоритмов распределения потоков данных Интернета вещей в облачных и туманных средах // Инфокоммуникационные технологии. 2024. Т. 22. № 4(88). С. 15–22. DOI:10.18469/ikt.2024.22.4.02. EDN:ALXFGY
15. Волков А.Н. Разработка и исследование комплекса моделей и методов построения сетей связи на основе туманных вычислений и предоставления услуг телеприсутствия. Автореферат дисс. ... докт. техн. наук. СПб.: Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича, 2024. 54 с.
16. Куприянов Д.О. Математическое моделирование потока заявок к облачному вычислительному кластеру // T-Comm: Телекоммуникации и транспорт. 2020. Т. 14. № 10. С. 39–44. DOI:10.36724/2072-8735-2020-14-10-39-44. EDN:MWCXTC
17. Мельник Э.В., Клименко А.Б., Иванов Д.Я. Модель задачи распределения вычислительной нагрузки для информационно-управляющих систем на базе концепции туманных вычислений // Известия Тульского государственного университета. Технические науки. 2018. № 2. С. 174–187. EDN:YXJCRP
18. Глушак Е.В., Ключев Д.С., Воловач В.И. Анализ распределения задач в системе облачных и туманных вычислений // T-Comm: Телекоммуникации и транспорт. 2025. Т. 19. № 7. С. 25–33. DOI:10.36724/2072-8735-2025-19-7-25-33. EDN:WHDPWF
19. Воробьев С.П. Математическая модель оптимизации сетевой инфраструктуры распределенной корпоративной системы на базе облачных, туманных и граничных технологий // Моделирование, оптимизация и информационные технологии. 2019. Т. 7. № 3(26). С. 4. DOI:10.26102/2310-6018/2019.26.3.003. EDN:LIPTF

References

1. Yermakov S.G., Khalil M.M., Khomonenko A.D., Bukharova K.A. Evaluating the efficiency of fog computing on the internet of things using a Non-Markov model. *T-Comm*. 2022;16(12):46–52. DOI:10.36724/2072-8735-2022-16-12-46-52. EDN:MVRBVB
2. Glushak E.V. *Cloud and fog computing: architecture, modeling, application*. Vologda; Moscow: Infra-Engineering Publ.; 2025. 180 p. (in Russ.) EDN:BUZGWB

3. Fletcher M., Paulz E., Ridge D., Michaels A.J. Low-Latency Wireless Network Extension for Industrial Internet of Things. *Sensors*. 2024;24(7):2113. DOI:10.3390/s24072113. EDN:YATJXO
4. Plotnikov P.V., Vladyko A.G. Analysis of Approaches to Optimization of V2X Systems: Clustering, Edge and Fog Computing. *Proceedings of Telecommunication Universities*. 2024;10(3):7–22. (in Russ.) DOI:10.31854/1813-324X-2024-10-3-7-22. EDN:TRWNON
5. Daraselia A.V. *Models and analysis of performance indicators of traffic offloading mechanisms in heterogeneous wireless networks*. Ph.D. Thesis. Moscow: Peoples' Friendship University of Russia Publ.; 2022. 106 p. (in Russ.) EDN:VQNBTU
6. Murashkin I.N. Research of latency minimization algorithms in data stream processing systems. *Innovations and Investments*. 2025;4:356–359. (in Russ.) EDN:AHKBWE
7. Melnik E.V., Klimenko A.B., Klimenko V.V. Models and reliability analysis of information and control systems based on fog computing. *Proceedings of the 5th All-Russian Scientific and Technical Conference at Supercomputer Technologies, SKT – 2018, 17–22 September 2018, Rostov-on-Don, Russian Federation, vol.2*. Rostov-on-Don: Southern Federal University Publ.; 2018. p.103–107. (in Russ.) EDN:YQBQRN
8. Glushak E.V., Klyuev D.S. Development and research of models for the functioning of cloud and fog computing. *Radioengineering*. 2025;89(3):157–168. (in Russ.) DOI:10.18127/j00338486-202503-14. EDN:IGUDLR
9. Melnik E.V., Klimenko A.B. A recovery technique of the fog-computing-based information and control system computational process. *Izvestiya Tula State University*. 2018;9:551–563. (in Russ.) EDN:YNDROH
10. Alsweiti M. Deep Learning Approaches for Traffic Prediction Forecasting in Multi-Level Cloud Architecture for Autonomous Vehicle Services. *Proceedings of Telecommunication Universities*. 2022;8(4):89–99. (in Russ.) DOI:10.31854/1813-324X-2022-8-4-89-99. EDN:CNIDPH
11. Melnik E.V., Ivanov D.Ya., Orda-Zhigulina M.V., Orda-Zhigulina D.V., Rodina A.A. Application of fog computing for monitoring and forecasting system of hazardous natural phenomena. *Izvestiya Tula State University*. 2019;2:300–311. (in Russ.) EDN:ZAFGCT
12. Volkov A.N. Routing Task in Dynamic Fog Computing Network. *Proceedings of Telecommunication Universities*. 2024;10(4):27–37. (in Russ.) DOI:10.31854/1813-324X-2024-10-4-27-37. EDN:QWBVQY
13. Volkov A.N. Dynamic Fog Computing Towards Green ICT. *Proceedings of Telecommunication Universities* 2024;10(3):24–34. (in Russ.) DOI:10.31854/1813-324X-2024-10-3-24-34. EDN:QOELMJ
14. Glushak E.V., Mikhailova P.D. Overview of the adaptive algorithms for distributing internet of things data flows in cloud and foggy media. *Infokommunikacionnye Tehnologii*. 2024;22(4-88):15–22. (in Russ.) DOI:10.18469/ikt.2024.22.4.02. EDN:ALXFGY
15. Volkov A.N. *Development and study of a set of models and methods for constructing communication networks based on fog computing and providing telepresence services*. DSc Thesis. St. Petersburg: The Bonch-Bruевич Saint Petersburg State University of Telecommunications Publ.; 2024. 54 p. (in Russ.)
16. Kupriyanov D.O. Mathematical modeling of requests flow to cloud compute cluster. *T-Comm*. 2020;14(10):39–44. (in Russ.) DOI:10.36724/2072-8735-2020-14-10-39-44. EDN:MWCXTC
17. Melnik E.V., Klimenko A.B., Ivanov D.Ya. A model of the computational load distribution problem for information and control systems based on the fog computing concept. *Izvestiya Tula State University*. 2018;2:174–187. (in Russ.) EDN:YXJCRP
18. Glushak E.V., Klyuev D.S., Volovach V.I. Analysis of distribution of tasks in the cloud and fog computing system. *T-Comm*. 2025;19(7):25–33. (in Russ.) DOI:10.36724/2072-8735-2025-19-7-25-33. EDN:WHDPWF
19. Vorobyov S.P. Mathematical model of optimization of the network infrastructure of a distributed enterprise system on a cloud, misty and edge technologies. *Modeling, Optimization, and Information Technology*. 2019(7);3-26:4. (in Russ.) DOI:10.26102/2310-6018/2019.26.3.003. EDN:LIPTF

Статья поступила в редакцию 13.03.2025; одобрена после рецензирования 03.04.2025; принята к публикации 07.04.2026

The article was submitted 13.03.2025; approved after reviewing 03.04.2025; accepted for publication 07.04.2026

Информация об авторе:

**ГЛУШАК
Елена Владимировна**

кандидат технических наук, доцент кафедры сетей и систем связи Поволжского государственного университета телекоммуникаций и информатики
 <https://orcid.org/0009-0000-5494-9746>

Автор сообщает об отсутствии конфликтов интересов.

The author declares no conflicts of interests.

Научная статья

УДК 004.7

<https://doi.org/10.31854/1813-324X-2026-12-2-36-44>

EDN:FOVZUL



Разработка и апробация методики балансировки трафика в сетях SDN на базе комбинированного параметра QoS

Юлия Сергеевна Дмитриева✉, dmitrieva@sut.ru

Василий Сергеевич Елагин, v.elagin@sut.ru

Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича,
Санкт-Петербург, 193232, Российская Федерация

Аннотация

Актуальность. Рост трафика реального времени (видеосвязь, стриминг, облачные сервисы) предъявляет повышенные требования к качеству обслуживания в сетях. Традиционные методы балансировки нагрузки, ориентированные только на загрузку узлов сети, не учитывают потерю пакетов, задержку и джиттер. Существует потребность в усовершенствованных алгоритмах балансировки для программно-конфигурируемых сетей, способных динамически учитывать, как нагрузку трафика на узлы сети, так и его качественные характеристики.

Цель: разработка, теоретическое обоснование и экспериментальная апробация методики балансировки сетевого трафика на базе комбинированного параметра качества для программно-конфигурируемых сетей, обеспечивающей соблюдение требований к качеству обслуживания клиентского трафика.

Методы: теоретический анализ SLA операторов связи для выявления ключевых параметров качества обслуживания в сетях, математическое моделирование и формализация процедуры нормализации разнородных параметров качества, метод балансировки трафика на основе модифицированного алгоритма Enhanced Weighted Round-Robin, с использованием разработанного программного обеспечения имитационного моделирования балансировки трафика в сети SDN.

Решение. Разработана методика динамической балансировки, с интегрированным расчетом комбинированного параметра качества обслуживания в сетях. Предложен механизм нормализации параметров качества в единый весовой коэффициент на основе анализа отраслевых стандартов. Экспериментально доказана эффективность: методика обеспечила равномерное распределение нагрузки и улучшение комбинированного показателя качества обслуживания в сетях на 46 % по сравнению с методикой балансировки только по нагрузке.

Новизна. Впервые разработана оригинальная адаптивная модель математической нормализации разнородных параметров Quality of Service. Создано специализированное программное обеспечение для имитационного моделирования и визуализации метода балансировки с учетом Quality of Service.

Практическая значимость. Разработка готового инструмента для моделирования и тестирования устойчивости инфраструктуры. Демонстрация практической реализации интеллектуальной балансировки для внедрения в контроллеры SDN.

Ключевые слова: SDN, балансировка нагрузки, QoS, алгоритм Enhanced Weighted Round-Robin, нормализация параметров QoS

Ссылка для цитирования: Дмитриева Ю.С., Елагин В.С. Разработка и апробация методики балансировки трафика в сетях SDN на базе комбинированного параметра QoS // Труды учебных заведений связи. 2026. Т. 12. № 2. С. 36–44. DOI:10.31854/1813-324X-2026-12-2-36-44. EDN:FOVZUL

Original research

<https://doi.org/10.31854/1813-324X-2026-12-2-36-44>

EDN:FOVZUL

Development and Testing of Traffic Balancing Techniques in SDN Networks Based on the Combined QoS Parameter

 Yulia S. Dmitrieva , dmitrieva@sut.ru

 Vasily S. Elagin, v.elagin@sut.ru

The Bonch-Bruевич Saint Petersburg State University of Telecommunications,
St. Petersburg, 193232, Russian Federation

Annotation

Actuality. The growth of real-time traffic places increased demands on the quality of service in networks. Traditional load balancing methods, focused only on network node loading, do not take packet loss, delay, and jitter into account. There is a need for advanced balancing algorithms for software-configurable networks capable of dynamically taking into account both the traffic load on network nodes and its quality characteristics.

Purpose: development, theoretical justification, and experimental testing of network traffic balancing techniques based on a combined quality parameter for software-configurable networks, ensuring compliance with client traffic service quality requirements.

Methods: theoretical analysis of SLA communication operators to identify key parameters of the quality of service to all networks, mathematical modeling and formalization of the normalization procedure of various quality parameters, traffic balancing method based on the modified Enhanced Weighted Round-Robin algorithm, with the use of developed software for simulating traffic balancing in SDN networks. The solution. Development analysis of dynamic balancing techniques, with integrated calculation of the combined service quality parameter in networks. A mechanism for normalization of quality parameters is proposed.

Experimentally proven effectiveness: the algorithm ensured uniform distribution of the load and improved the combined indicator of the quality of service in networks by 46 % compared to the method of load balancing only.

Novelty. For the first time, an original adaptive model of mathematical normalization of various Quality of Service parameters was developed. Specialized software was created for simulation modeling and visualization of the balancing method taking into account Quality of Service.

Practical significance. Development of a ready-made tool for modeling and testing infrastructure stability. Demonstration of practical implementation of intellectual balancing for implementation in SDN controllers.

Keywords: SDN, load balancing, QoS, Enhanced Weighted Round-Robin algorithm, normalization of QoS parameters

For citation: Dmitrieva Yu.S., Elagin V.S. Development and Testing of Traffic Balancing Techniques in SDN Networks Based on the Combined QoS Parameter. *Proceedings of Telecommunication Universities*. 2026;12(2):36–44. (in Russ.) DOI:10.31854/1813-324X-2026-12-2-36-44. EDN:FOVZUL

Введение

Развитие программно-конфигурируемых сетей (SDN, аббр. от англ. Software-Defined Networking) открыло новые возможности для гибкого управления сетевыми ресурсами, однако задача эффективного распределения трафика остается актуальной [1–4]. Традиционные методы балансировки нагрузки часто ориентированы исключительно на загрузку узлов сети, не учитывая качества обслуживания (QoS, аббр. от англ. Quality of Service),

которое напрямую влияет на качество предоставляемых сервисов клиенту. Такие параметры, как потери пакетов (loss of packet), задержка (ping) и джиттер (jitter) являются критически важными для современных сервисов, особенно для трафика реального времени (от англ. Real Time).

В настоящей работе представлена методика балансировки сетевого трафика на базе комбинированного параметра QoS, разработанная для SDN. Целью исследования является разработка и апро-

бация методики динамической балансировки нагрузки, который наряду с загрузкой коммутаторов учитывает ключевые параметры QoS. Методика включает в себя механизмы непрерывного мониторинга, нормализации разнородных параметров QoS и принятия решений на основе алгоритма Enhanced Weighted Round-Robin.

Для проверки эффективности предложенного подхода была разработана и зарегистрирована программа для ЭВМ «Клиентская панель управления, настройки и визуализации результатов экспериментов имитационного моделирования», позволяющая проводить эксперименты в контролируемых условиях и визуализировать их результаты в реальном времени. Работа направлена на повышение качества клиентских сервисов за счет интеллектуального распределения трафика, учитывающего как нагрузку на инфраструктуру, так и параметры QoS клиентского трафика.

Принципы работы методики на базе комбинированного параметра QoS

Методика динамического перераспределения нагрузки функционирует по образцу непрерывного циклического процесса, который можно представить, как последовательность этапов: мониторинг, анализ, принятие решения, выполнение действия, мониторинг. Цикличность методики позволяет системе постоянно адаптироваться к изменяющимся условиям сети, реагируя на возникающие или прогнозируемые перегрузки.

Основные принципы работы методики

1) Оптимизация ресурсов: главная цель – равномерно распределить нагрузку по всем доступным коммутаторам, избегая концентрации трафика на отдельных коммутаторах, что может привести к их перегрузке и деградации сервиса.

2) Автономность: после запуска алгоритм функционирует без вмешательства, автоматически принимая решения на основе собранных данных.

3) Простота и вычислительная эффективность: базовая реализация алгоритма Enhanced Weighted Round-Robin [5–6] ориентирована на непрерывную обработку данных и принятие решений, что важно для высоконагруженных систем.

Цикличность методики обеспечивает непрерывную обработку данных, где каждое принятое решение приводит к изменению состояния сети, которое затем снова анализируется для следующего цикла принятия решений.

Методика балансировки трафика коммутаторов по комбинированному параметру QoS

Рассмотрим этапы методики балансировки трафика коммутаторов, используя алгоритм EWRR (рисунок 1, где X – максимальный порог нагрузки на коммутатор).

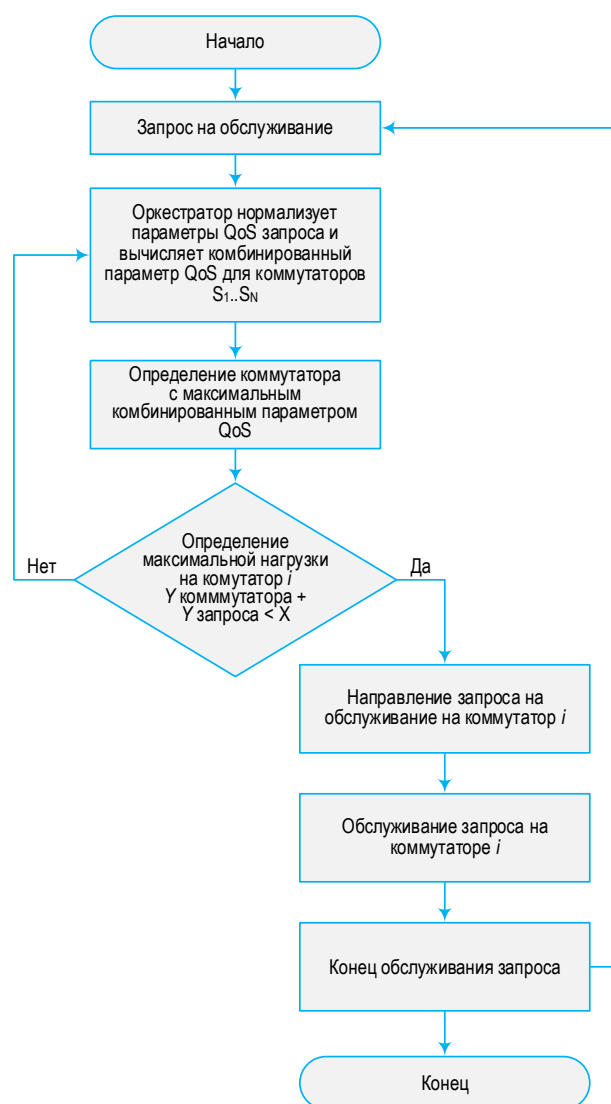


Рис. 1. Методика балансировки трафика коммутаторов по комбинированному параметру QoS

Fig. 1. The Method of Balancing Switch Traffic on the Combined QoS Parameter

Основой для принятия решения является актуальная и точная информация о состоянии сети. В разработанной модели реализован непрерывный механизм сбора и первичного анализа параметров коммутаторов (нагрузка, параметры QoS). Параметры коммутатора анализируются с заданной частотой (например, несколько раз в секунду или через фиксированные интервалы указанного модельного времени). Это обеспечивает актуальность данных, что критически важно для оперативного реагирования на быстрые изменения нагрузки и параметров QoS.

Сначала поступает поток запросов на обслуживание. Далее осуществляется нормализация параметров QoS запроса и вычисление комбинированного параметра для коммутаторов $S1..Sn$.

Для балансировки нагрузки необходимо учитывать параметры QoS коммутаторов по значимым в данной сети характеристикам. В интересах повышения качества клиентских сервисов требуется принимать во внимание не только нагрузку на коммутаторе, но и параметры QoS для запроса [7–8]. Как образец требований к качеству клиентских сервисов возможно взять общедоступные договора российских операторов связи, в них указываются конкретные параметры QoS для разного типа трафика и уровня SLA [8]. Анализ договоров показал, что все операторы связи указывают требования к трем параметрам QoS: jitter, ping, loss of packet.

При балансировке сетевого трафика на базе параметров QoS учитываются величины с разными единицами измерения и возникает вопрос о нормализации параметров QoS, так как они различны по

своим единицам измерения. Параметры QoS оказывают определенное влияние на сервис, поэтому необходимо оценить «вес» каждого из них при балансировке трафика, для чего определяем нормализующие коэффициенты C_{ij} для всех параметров.

Методика балансировки нагрузки коммутаторов на базе параметров QoS будет выглядеть следующим образом: чтобы решить, на какой коммутатор перенаправить соединение, оркестратору необходимо сравнить значения соответствующих параметров для каждого коммутатора. Для этого необходимо сформировать из собранных параметров матрицу B (1) с b_{ij} , где i – порядковый номер параметра ($i = 1, \dots, k$), $k = 3$; j – номер коммутатора ($j = 1, \dots, n$); $b_{\ell p}$, b_{ping} , b_{jitter} – значения параметров loss of packet, ping и jitter, соответственно; n – количество коммутаторов.

$$B = \|b_{ij}\|_{(3 \times n)} = \begin{pmatrix} C_{11}b_{\ell p_1} & C_{21}b_{\ell p_2} & \dots & C_{n1}b_{\ell p_n} \\ C_{12}b_{\text{ping}_1} & C_{21}b_{\text{ping}_2} & \dots & C_{n2}b_{\text{ping}_n} \\ C_{13}b_{\text{jitter}_1} & C_{23}b_{\text{jitter}_2} & \dots & C_{n3}b_{\text{jitter}_n} \end{pmatrix}. \quad (1)$$

Элемент b_{ij} в матрице B (1) умножается на нормализующий коэффициент C_{ij} , который определяется с помощью следующего выражения:

$$\sum_{j=1}^n \sum_{i=1}^3 C_{ij} = 1.$$

Таким образом, получаем информацию о коммутаторе с наилучшим комбинированным параметром QoS.

Алгоритм нормализации параметров QoS коммутатора

Произведем поэтапно расчет нормализованной суммы параметров QoS для коммутатора:

$$w_{\text{QoS}} = w_{\ell p} + w_{\text{ping}} + w_{\text{jitter}}, \quad (2)$$

где $w_{\ell p}$, w_{ping} , w_{jitter} – нормализованный «вес» значений параметров loss of packet, ping и jitter, соответственно.

Для нормализации параметров QoS необходимо привести значения в разных единицах измерения к одному виду. Введем пределы параметров loss of packet, ping и jitter (для примера возьмем параметры QoS из договора крупного оператора связи).

Этап 1. Введем ограничения для рассматриваемых параметров:

ℓp от 0 до 0,1 %;
ping от 1 до 100 мс;
jitter от 0 до 20 мс.

Этап 2. Произведем нормализацию, т. е. приведем к единому виду параметры, имеющие разные единицы измерения:

ℓp от 0 до 0,1 % умножаем на 10000 и делим на 10;
ping от 1 до 100 мс делим на 100;
jitter от 0 до 20 мс делим на 20.

Этап 3. Применим нормализующий коэффициент потери пакетов (из договоров операторов связи для трафика Real Time значения параметров QoS самые строгие и равнозначны между собой):

$$C_{1j} = 0,33; C_{2j} = 0,34; C_{3j} = 0,33,$$

где нормализующие коэффициенты: C_{1j} – потери пакетов, C_{2j} – задержка; C_{3j} – джиттер.

Выполнив первые три действия, получим весовой коэффициент параметров QoS коммутатора для (2):

$$w_{\text{QoS}} = \frac{\ell p \times 10000}{10} \times 0,33 + \frac{\text{ping}}{100} \times 0,33 + \frac{\text{jitter}}{20} \times 0,34.$$

Этап 4. Определим перегрузки – не превышает ли значение нагрузки с данным запросом на коммутаторе максимально возможную:

$$W_{\text{Ком}} + W_{\text{Ожидаемая запроса}} < X.$$

Этап 5. Направим запрос на выбранный коммутатор с максимальным QoS или в оркестратор для повторной обработки (при перегрузке коммутаторов).

Методика обеспечивает в режиме реального времени мониторинг состояния всех моделируемых коммутаторов сетевой инфраструктуры, который служит основой для принятия решений алгоритмом.

Программная инструментализация методики

Для проверки методики балансировки нагрузки на базе комбинированного параметра QoS создана программа для ЭВМ под названием «Клиентская панель управления, настройки и визуализации результатов экспериментов имитационного моделирования» [9]. Она позволяет пользователю взаимодействовать с серверной системой через интуитивно понятный интерфейс, обеспечивая визуализацию результатов имитационного моделирования в виде графиков. Настройка параметров эксперимента обеспечивается с помощью удобных ползунков и панели управления, позволяя гибко изменять условия моделирования. Кроме того, программа для ЭВМ способствует ускоренному анализу данных, предоставляя пользователю возможность наблюдать динамическое поведение моделей в реальном времени, и осуществляет установку соединения с сервером и передачу параметров эксперимента, обеспечивая точную настройку и мониторинг работы имитационной модели.

Разработанная имитационная модель представляет собой программную реализацию, предназначенную для воспроизведения поведения сетевой инфраструктуры в контролируемой среде. Ее архитектура построена на принципе дискретно-событийного моделирования, что позволяет эффективно моделировать прохождение трафика через типовые сегменты сети. Сетевая инфраструктура моделируется как совокупность независимых узлов сети, характеристики которых настраиваются индивидуально. Взаимодействие между узлами осуществляется через механизмы динамического перенаправления трафика, имитирующие изменение маршрутизации в реальных сетях. Ключевое внимание уделяется моделированию источников нагрузки, которые характеризуются интенсивностью генерации трафика и возможностью динамического переключения целевого узла сети. Это позволяет методике балансировки перенаправлять потоки трафика между узлами в реальном времени.

Для обеспечения управления системой непрерывно собираются основные метрики состояния каждого узла (загрузка, параметры QoS). На основе этих данных работает алгоритм динамического перераспределения нагрузки, функционирующий по циклическому принципу: мониторинг, анализ, принятие решения, выполнение действия, мониторинг. Этот процесс позволяет системе адаптироваться к изменяющимся условиям сети и минимизировать перегрузки.

Программная реализация системы построена по клиент-серверной архитектуре. Клиентская часть отвечает за взаимодействие с пользователем, настройку экспериментов и визуализацию данных. Серверная часть реализует логику имитационного

моделирования, расчет метрик и работу методики балансировки. Взаимодействие между компонентами обеспечивает управление моделированием в реальном времени, визуализацию состояния сети и сохранение результатов для последующего анализа. В левой части экрана находится скрываемая панель управления экспериментом (рисунок 2), а в верхней части меню «Setup» расположены кнопки команд:

«create» – создание эксперимента с текущими характеристиками;
«start» – запуск эксперимента;
«stop» – завершение эксперимента;
«download» – загрузка результатов эксперимента в формате Excel.

В части меню «Board settings» позволяет пользователю управлять характеристиками эксперимента. При изменении любого из параметров необходимо пересоздать эксперимент, чтобы настройки стали актуальными для сервера.

Для настройки моделей эксперимента используются слайдеры настройки:

- количества моделей (минимальное значение – 1, максимальное – 10);
- минимального значения диапазона количества приходящих в модель агентов (минимальное значение – 1, максимальное – 20);
- максимального значения диапазона количества приходящих в модель агентов (минимальное значение – 5, максимальное – 100);
- интервала времени прихода агентов в модели (минимальное значение – 100 мс, максимальное – 10000 мс);
- интервала времени взятия статистики с моделей (минимальное значение – 100 мс, максимальное – 10000 мс);
- количества Source элементов в каждой модели (минимальное значение – 1, максимальное – 10);
- минимального значения диапазона количества мест в Queue элементе у моделей (минимальное значение – 1, максимальное – 25);
- максимального значения диапазона количества мест в Queue элементе у моделей (минимальное значение – 5, максимальное – 100);
- минимального значения диапазона количества мест в Delay элементе у моделей (минимальное значение – 1, максимальное – 25);
- максимального значения диапазона количества мест в Delay элементе у моделей (минимальное значение – 5, максимальное – 100);
- значения времени обслуживания агента в Delay элементе у моделей (минимальное значение – 100 мс, максимальное – 10000 мс);
- значения коэффициента загруженности, при котором происходит принятие решения о переводе потока агентов (минимальное значение – 0,1, максимальное – 1).

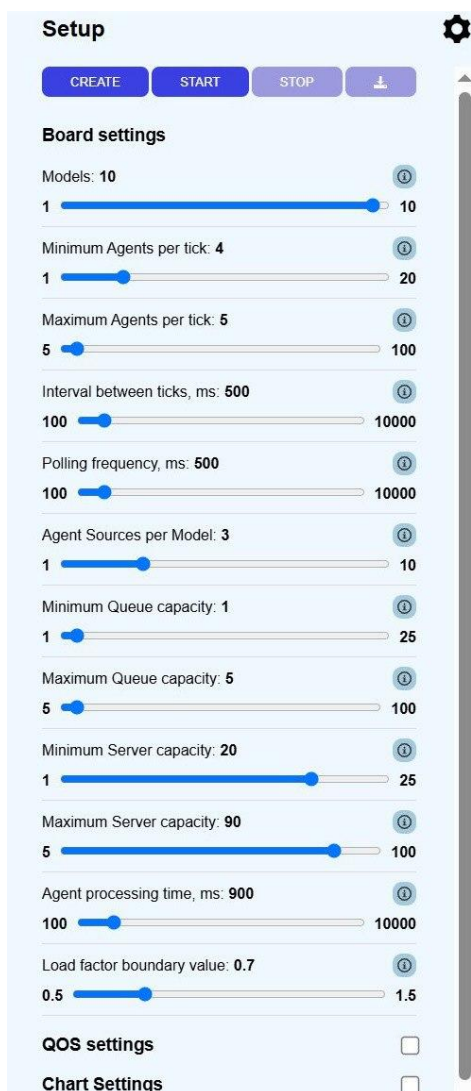


Рис. 2. Панель управления и конфигурации параметров эксперимента

Fig. 2. Control Panel and Configuration of Experiment Parameters

Балансировка нагрузки на коммутаторах по параметрам QoS включается при помощи флага. При активном флаге предоставляется возможность управлять параметрами контроля по QoS в части меню «QoS settings» (рисунок 3).

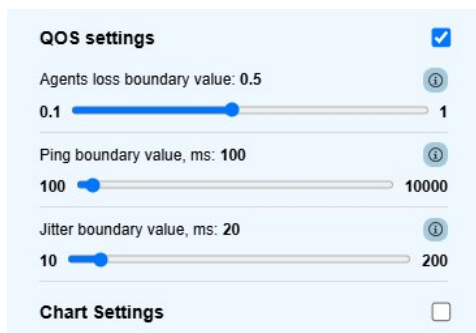


Рис. 3. Параметры QoS эксперимента

Fig. 3. QoS Parameters of the Experiment

Для настройки QoS в эксперименте используются слайдеры настройки для агентов:

- величина loss of packet, параметр задает значение QoS loss of packet (минимальное значение находится в диапазоне от 0 до 0,1 %, максимальное – от 0 до 1 %);
- значения ping, параметр задает значение QoS ping (минимальное значение находится в диапазоне от 0 до 100 мс, максимальное – от 0 до 10000 мс);
- значения jitter, параметр задает значение QoS jitter (минимальное значение находится в диапазоне от 0 до 10 мс, максимальное значение – от 0 до 200 мс);
- количества хранимых и отображаемых на графиках данных эксперимента.

Если флаг настроек системы визуализации статистики не активен, то все данные будут отображаться на графиках и не будут сохранены в файл ModelsParametersStatistic.exsx. При активном флаге появляется слайдер для контроля за сохраняемыми и отображаемыми данными на графиках эксперимента. Область эксперимента – блоки информации о каждой модели – представлена на рисунке 4, где отображаются два графика: нагрузка на моделируемый коммутатор и очередь на обслуживание. Область обновляется в реальном времени, интервал обновления задается в панели управления.

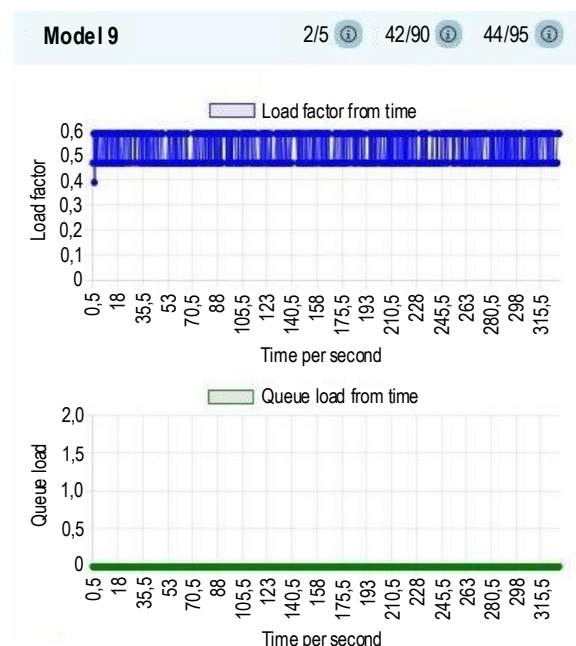


Рис. 4. Область эксперимента на примере модели 9

Fig. 4. The Field of Experiment on the Example of Model 9

Экспериментальная апробация

Проведем эксперимент с помощью программы «Клиентская панель управления, настройки и визуализации результатов экспериментов имитационного моделирования» для проверки методики.

Зададим параметры имитационного моделирования в программе:

- 1) эксперимент проводится на 10 коммутаторах;
- 2) значения комбинированного параметра QoS:
 - $lp_1, lp_2, lp_3, \dots, lp_{10}$ (параметр loss of packet в 0,1 %, $\Delta = 0..10$);
 - $p_1, p_2, p_3, \dots, p_{10}$ (параметр ping в мс $\Delta = 1..100$);
 - $j_1, j_2, j_3, \dots, j_{10}$ (параметр jitter в мс $\Delta = 0..20$);
- 3) максимальная загрузка коммутаторов 70 %.

С указанными параметрами на снимке экрана (рисунок 5) провели эксперимент.

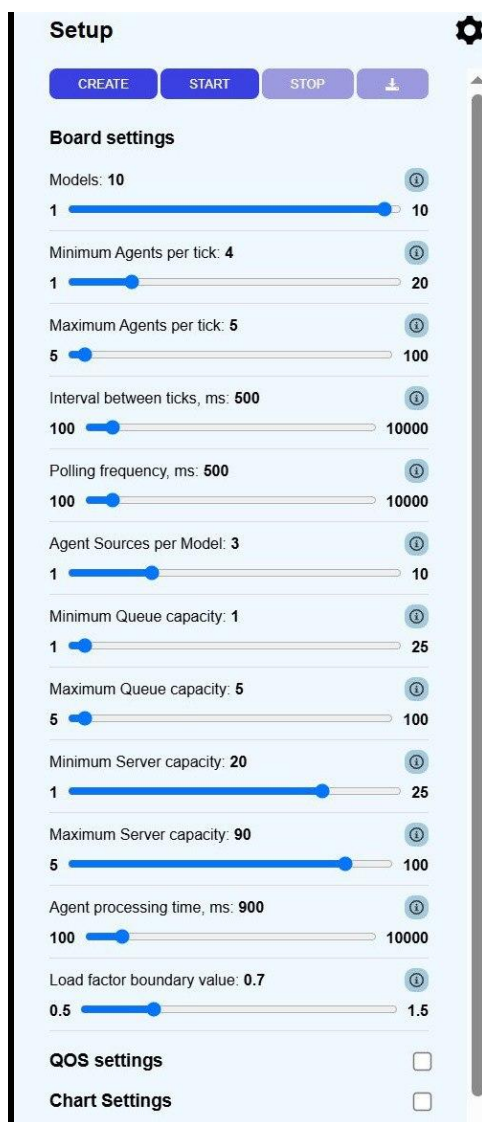


Рис. 5. Панель с конфигурацией параметров эксперимента

Fig. 5. Panel with Configuration of Experiment Parameters

Графическое отображение эксперимента представлено на рисунке 6а, где продемонстрирована балансировка трафика для модели 5. Видно, что на коммутатор номер 5 в начале эксперимента поступает большое количество заявок на обслуживание и образуется очередь из агентов на обслуживание.

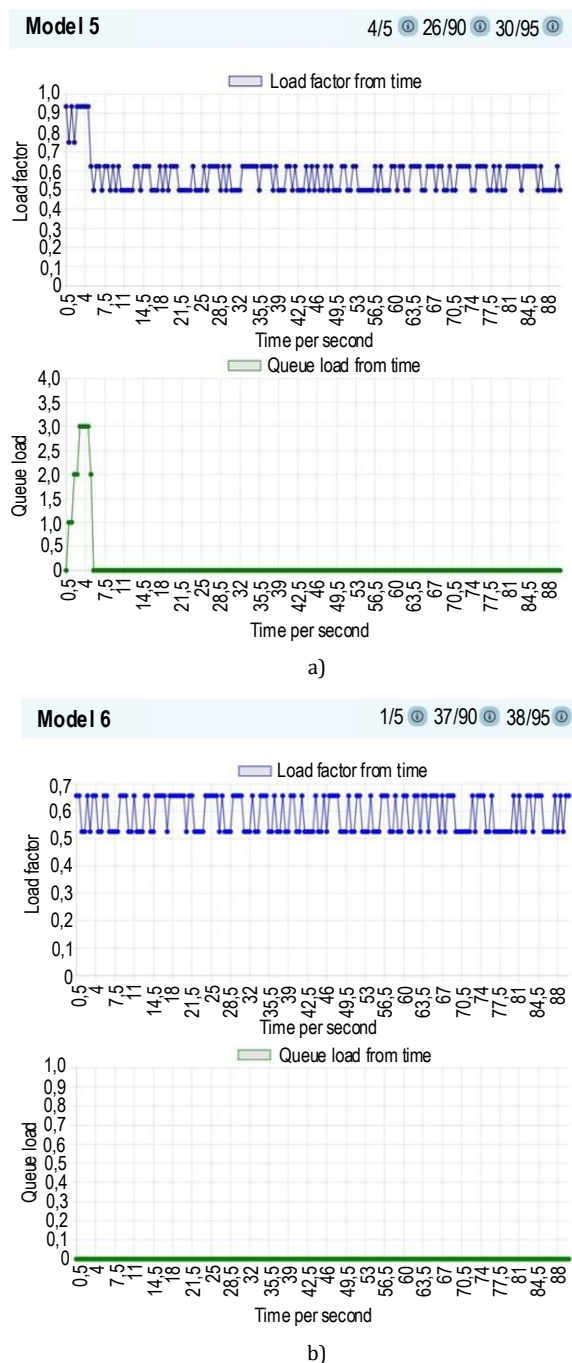


Рис. 6. Графическое изображение эксперимента на примере модели 5 (а) и модели 6 (б)

Fig. 6. Graphic Representation of the Experiment Using the Example of Model 5 (a) and Model 6 (b)

Нагрузка на модель 5 равна 0,95–4 с и далее становится равномерной на всем этапе эксперимента. С 5-й секунды алгоритм равномерно распределяет нагрузку (агентов) по моделям, и график нагрузки становится равномерным и однородным для всего остального времени эксперимента. В то же время с 5-й секунды равномерное распределение трафика по другим агентам демонстрирует, что очередь на данную модель становится равной 0, т. е. каждый

поступивший запрос обрабатывается сразу – следовательно, нагрузка на все коммутаторы распределяется корректно и равномерно. Таким образом, график (см. рисунок 6а) отображает корректную работу алгоритма в части балансировки трафика и распределения поступающих запросов на все модели эксперимента с равномерной нагрузкой на каждую модель из 10.

На модель 6, представленную на снимке экрана (рисунок 6б) с начала эксперимента, равномерно распределяется трафик, очередь из агентов (запросов) отсутствует. Параметры нагрузки на модель 6 соответствуют заданным значениям эксперимента и не выходят за пределы обозначенной максимальной нагрузки 0,7.

По результатам работы программы «Клиентская панель управления, настройки и визуализации результатов экспериментов имитационного моделирования» проведено два эксперимента по балансировке нагрузки между коммутаторами. Один эксперимент с учетом нагрузки, а второй по параметрам QoS. Создается файл с результатами статистических данных эксперимента с названием ModelsParametersStatistic.exsx. На основании этих данных построены графики (рисунок 7).

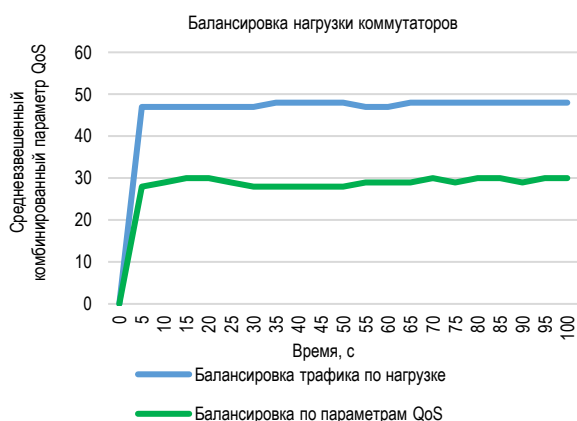


Рис. 7. Сравнительный анализ методов балансировки по нагрузке и на базе параметров QoS

Fig. 7. Comparative Analysis of Balancing Methods by Load and Based on QoS Parameters

При использовании методики балансировки только по нагрузке комбинированный показатель QoS остается на относительно низком уровне, с заметными колебаниями, что свидетельствует о нестабильности качества обслуживания. При применении предложенной методики балансировки на базе комбинированного параметра QoS показатель значительно повышается, демонстрируя устойчивое улучшение качества обслуживания. Предложенная методика обеспечила уменьшение комбинированного показателя QoS на 46 % по сравнению

с методом балансировки (чем меньше этот параметр, тем выше качество предоставляемых услуг), учитывающим только нагрузку на узлы сети. Это подтверждает эффективность интеграции параметров QoS (loss of packet, ping и jitter) в алгоритм принятия решений при балансировке трафика в сетях SDN.

Заключение

В ходе проведенного исследования была успешно разработана и апробирована методика балансировки трафика на базе комбинированного параметра QoS для сетей SDN, основанная на принципах цикличности (мониторинг – анализ – решение – действие – мониторинг), автономности и вычислительной эффективности. В ее основу положен модифицированный алгоритм Enhanced Weighted Round-Robin, дополненный механизмом учета трех ключевых параметров QoS: loss of packet, ping и jitter.

Кроме того, предложен механизм нормализации разнородных параметров QoS, позволяющий приводить значения с разными единицами измерения к единой безразмерной шкале. На основе анализа SLA крупных российских операторов связи были определены весовые коэффициенты для каждого параметра, что обеспечивает объективность и адаптивность принятия решений.

Проведена экспериментальная апробация методики в контролируемых условиях (с использованием программы для ЭВМ «Клиентская панель управления, настройки и визуализации результатов экспериментов имитационного моделирования» [2]). Результаты экспериментов подтвердили эффективность предложенной методики: 1) обеспечено равномерное распределение нагрузки между коммутаторами; 2) продемонстрирована способность системы оперативно реагировать на пиковые нагрузки и перераспределять трафик для предотвращения перегрузок отдельных узлов сети; 3) комбинированный показатель QoS уменьшился на 46 % по сравнению с балансировкой, учитывающей только нагрузку узлов.

Таким образом, предложенная методика балансировки трафика на базе комбинированного параметра QoS доказала свою практическую значимость и эффективность. Она позволяет не только оптимизировать использование сетевых ресурсов, но и гарантировать соблюдение согласованных уровней качества обслуживания (SLA), что является критически важным для операторов связи и поставщиков облачных сервисов. Дальнейшие исследования могут быть направлены на адаптацию методики для более сложных сетевых топологий и интеграцию с системами прогнозирования нагрузки.

Список источников

1. McKeown N., Anderson T., Balakrishnan H., Parulkar G., Peterson L., Rexford J., et al. OpenFlow: enabling innovation in campus networks. *ACM SIGCOMM Computer Communication Review*. 2008;38(2):69–74. DOI:10.1145/1355734.1355746.
2. ONF TR-502. SDN architecture. 2014. URL: https://www.opennetworking.org/wp-content/uploads/2013/02/TR_SDN_ARCH_1.0_06062014.pdf (Accessed 20.01.2026)
3. RFC 7426 Software-Defined Networking (SDN): Layers and Architecture Terminology // IETF.org. 2015. URL: <https://tools.ietf.org/html/rfc7426> (Accessed 20.01.2026)
4. SDN Architecture Overview. Version 1.0. Open Networking Foundation. 2013. URL: https://opennetworking.org/wp-content/uploads/2014/10/TR-521_SDN_Architecture_overview.pdf (Accessed 20.01.2026)
5. Sinha G., Sinha D.K. Enhanced Weighted Round Robin Algorithm to Balance the Load for Effective Utilization of Resource in Cloud Environment // *EAI Endorsed Transactions on Cloud Systems*. 2020. Vol. 6. Iss. 18. DOI:10.4108/eai.7-9-2020.166284
6. Priya S.S., Rajendran T. Enhanced Weighted Round Robin: A New Paradigm in Cloud Load Balancing // *Indian Journal of Science and Technology*. 2025. Vol. 18. Iss. 15. PP. 1220–1228. DOI:10.17485/IJST/v18i15.3976
7. Rec. ITU-Y.1541. Network performance objectives for IP-based services. 2011.
8. Рекомендация МСЭ-Т E.800 Определения терминов, связанных с качеством обслуживания. 2008.
9. Дмитриева Ю.С., Елагин В.С., Рыбин И.П., Сербин А.А. Клиентская панель управления, настройки и визуализации результатов экспериментов имитационного моделирования. Свидетельство о государственной регистрации программы для ЭВМ № RU 2025662025 от 29.04.2025. Оpubл. 16.05.2025. EDN:NNHGGO

References

1. McKeown N., Anderson T., Balakrishnan H., Parulkar G., Peterson L., Rexford J., et al. OpenFlow: enabling innovation in campus networks. *ACM SIGCOMM Computer Communication Review*. 2008;38(2):69–74. DOI:10.1145/1355734.1355746.
2. ONF TR-502. SDN architecture. 2014. URL: https://www.opennetworking.org/wp-content/uploads/2013/02/TR_SDN_ARCH_1.0_06062014.pdf [Accessed 20.01.2026]
3. IETF.org. RFC 7426 Software-Defined Networking (SDN): Layers and Architecture Terminology. 2015. URL: <https://tools.ietf.org/html/rfc7426> (Accessed 20.01.2026)
4. SDN Architecture Overview. Version 1.0. Open Networking Foundation. 2013. URL: https://opennetworking.org/wp-content/uploads/2014/10/TR-521_SDN_Architecture_overview.pdf [Accessed 20.01.2026]
5. Sinha G., Sinha D.K. Enhanced Weighted Round Robin Algorithm to Balance the Load for Effective Utilization of Resource in Cloud Environment. *EAI Endorsed Transactions on Cloud Systems*. 2020;6(18). DOI:10.4108/eai.7-9-2020.166284
6. Priya S.S., Rajendran T. Enhanced Weighted Round Robin: A New Paradigm in Cloud Load Balancing. *Indian Journal of Science and Technology*. 2025;18(15):1220–1228. DOI:10.17485/IJST/v18i15.3976
7. Rec. ITU-Y.1541. Network performance objectives for IP-based services. 2011
8. Rec. ITU-T E.800 Definitions of terms related to quality of service. 2008.
9. Dmitrieva Yu.S., Elagin V.S., Rybin I.P., Serbin A.A. Client Panel for Control, Settings and Visualization of the Results of Simulation Modeling Experiments. Patent RF, no 2025662025, 29.04.2025. (in Russ.) EDN:NNHGGO

Статья поступила в редакцию 16.01.2026; одобрена после рецензирования 26.01.2026; принята к публикации 09.02.2026

The article was submitted 16.01.2026; approved after reviewing 26.01.2026; accepted for publication 09.02.2026

Информация об авторах:

**ДМИТРИЕВА
Юлия Сергеевна**

старший преподаватель кафедры инфокоммуникационных систем Санкт-Петербургского государственного университета телекоммуникаций им. проф. М.А. Бонч-Бруевича
 <https://orcid.org/0000-0002-7736-7121>

**ЕЛАГИН
Василий Сергеевич**

кандидат технических наук, доцент, доцент кафедры инфокоммуникационных систем Санкт-Петербургского государственного университета телекоммуникаций им. проф. М.А. Бонч-Бруевича
 <https://orcid.org/0000-0003-4077-6869>

Авторы сообщают об отсутствии конфликтов интересов.

The authors declare no conflicts of interests.

Научная статья
УДК 004.732.056
<https://doi.org/10.31854/1813-324X-2026-12-2-45-52>
EDN:DCXPUT

Энергетическая эффективность в многосотовых massive MIMO: влияние схем обработки и параметров аппаратной реализации

Ван Донг Нгием✉, ngiem.v@phystech.edu
Александр Викторович Дворкович, dvork.alex@gmail.com

Московский физико-технический институт (национальный исследовательский университет),
Москва, 117303, Российская Федерация

Аннотация

В работе исследуется компромисс между энергетической и спектральной эффективностью в многосотовых системах massive MIMO. Актуальность исследования обусловлена необходимостью одновременного повышения пропускной способности и снижения энергопотребления базовых станций в беспроводных сетях следующего поколения при росте числа антенн и обслуживаемых пользователей.

Цель исследования. Выявить влияние числа антенн базовой станции, количества обслуживаемых пользователей, схем линейной обработки сигналов и параметров аппаратной реализации на энергетическую эффективность многосотовых систем massive MIMO и определить конфигурации, обеспечивающие наилучший компромисс между энергетической и спектральной эффективностью.

Используемые методы. В статье разработана математическая модель системы massive MIMO, учитывающая число антенн базовой станции M , количество пользовательских устройств K , а также различные схемы линейной обработки сигналов (MR, ZF, RZF, S-MMSE, M-MMSE). В модель включены параметры потребляемой мощности, отражающие особенности аппаратной реализации, рассмотренные в виде двух наборов характеристик элементной базы. Оценка энергетической и спектральной эффективности выполнена на основе имитационного моделирования для различных конфигураций системы.

Результаты. Показано, что оптимальное соотношение числа антенн к количеству пользователей $M/K \approx 3-4$ обеспечивает максимальную энергетическую эффективность без существенного снижения спектральной производительности. Установлено, что алгоритмы M-MMSE и S-MMSE демонстрируют наилучшие показатели энергетической эффективности при умеренной вычислительной сложности, особенно при использовании улучшенной аппаратной базы. Полученные зависимости подтверждают наличие выраженного оптимума энергетической эффективности при увеличении числа антенн базовой станции.

Новизна работы заключается в комплексном анализе компромисса между энергетической и спектральной эффективностью в многосотовых системах massive MIMO с одновременным учетом схем линейной обработки сигналов и параметров аппаратной реализации, что позволяет обосновать практические рекомендации по выбору конфигурации базовых станций в условиях технологических ограничений.

Практическая значимость. Результаты исследования могут быть использованы при проектировании и оптимизации энергоэффективных многосотовых систем massive MIMO для беспроводных сетей связи следующего поколения с учетом технологических ограничений аппаратной реализации и требований к качеству обслуживания.

Ключевые слова: massive MIMO, спектральная эффективность, энергетическая эффективность, обработка сигналов, беспроводные сети 5G

Ссылка для цитирования: Нгием В.Д., Дворкович А.В. Энергетическая эффективность в многосотовых massive MIMO: влияние схем обработки и параметров аппаратной реализации// Труды учебных заведений связи. 2026. Т. 12. № 2. С. 45–52. DOI:10.31854/1813-324X-2026-12-2-45-52. EDN:DCXPUT

Original research

<https://doi.org/10.31854/1813-324X-2026-12-2-45-52>

EDN:DCXPUT

Energy Efficiency in Multi-Cell Massive MIMO: Impact of Signal Processing Schemes and Hardware Implementation Parameters

✉ Van Dong Nghiem, ngiem.v@phystech.edu

✉ Aleksandr V. Dvorkovich, dvork.alex@gmail.com

Moscow Institute of Physics and Technology,
Moscow, 117303, Russian Federation

Annotation

This paper investigates the trade-off between energy efficiency and spectral efficiency in multi-cell massive MIMO systems. The **relevance** of the study is driven by the need to simultaneously increase throughput and reduce the energy consumption of base stations in next-generation wireless networks, given the growing number of antennas and served users.

The purpose of the study is to determine how the number of base-station antennas, the number of served users, linear signal processing schemes, and hardware implementation parameters affect the energy efficiency of multi-cell massive MIMO systems, and to identify configurations that provide the best trade-off between energy and spectral efficiency.

Methods. A mathematical model of a massive MIMO system is developed, taking into account the number of base station antennas M , the number of user equipment K , and various linear signal processing schemes, including MR, ZF, RZF, S-MMSE, and M-MMSE. The model incorporates power consumption parameters reflecting hardware implementation characteristics, represented by two different sets of component specifications. The evaluation of energy and spectral efficiency is carried out using simulation-based analysis for various system configurations.

Results. The results show that an optimal antenna-to-user ratio of $M/K \approx 3-4$ achieves maximum energy efficiency without a significant reduction in spectral efficiency. It is demonstrated that the M-MMSE and S-MMSE algorithms provide the highest energy efficiency performance with moderate computational complexity, particularly when improved hardware components are employed. The obtained results confirm the existence of a pronounced energy efficiency optimum as the number of base station antennas increases.

The novelty of this work lies in the comprehensive analysis of the energy-spectral efficiency trade-off in multi-cell massive MIMO systems while jointly accounting for linear signal processing schemes and hardware implementation parameters, which enables the formulation of practical recommendations for base station configuration under technological constraints.

Practical significance. The findings of this study can be applied to the design and optimization of energy-efficient multi-cell massive MIMO systems for next-generation wireless communication networks, taking into account hardware implementation constraints and quality-of-service requirements.

Keywords: massive MIMO, spectral efficiency, energy efficiency, signal processing, 5G wireless networks

For citation: Nghiem V.D., Dvorkovich A.V. Energy Efficiency in Multi-Cell Massive MIMO: Impact of Signal Processing Schemes and Hardware Implementation Parameters. *Proceedings of Telecommunication Universities*. 2026;12(2):45–52. (in Russ.) DOI:10.31854/1813-324X-2026-12-2-45-52. EDN:DCXPUT

Введение

Современные беспроводные сети сталкиваются с вызовами, связанными с необходимостью обеспечения высокой пропускной способности при одновременном снижении энергопотребления.

Технология massive MIMO, предполагающая использование большого числа антенн на базовых станциях, предлагает решение этой проблемы за счет пространственного мультиплексирования сигналов и рассматривается многими российскими

ми учеными [1–4]. Однако увеличение числа антенн и пользователей приводит к росту энергопотребления, что требует тщательного анализа и оптимизации компромисса между энергетической (ЕЕ) и спектральной (SE) эффективностью [5]. Исследования в области компромисса между ЕЕ и SE в системах massive MIMO активно развиваются. В работе [6] предложен подход к оптимизации выбора этого компромисса с использованием многокритериального адаптивного генетического алгоритма. Авторы [7] рассматривают методы повышения ЕЕ в сетях 5G с использованием massive MIMO, выделяя ключевые проблемы и возможные решения. В [8] представлен анализ компромисса ЕЕ–SE в нисходящем канале massive MIMO с учетом распределения мощности и числа антенн. Работа [9] фокусируется на оптимизации компромисса ЕЕ–SE в системах massive MIMO с использованием статистической информации о канале. Наконец, в [10] рассматривается влияние различных конфигураций массивов антенн на ЕЕ систем massive MIMO.

Постановка задачи

Целью данной работы является разработка методов оптимизации ЕЕ и SE в системах massive MIMO. Необходимо определить оптимальное соотношение между числом антенн на базовых станциях и количеством обслуживаемых пользователей, при котором достигается максимальная ЕЕ без значительного снижения SE. Также требуется анализ влияния различных схем обработки сигналов на указанные метрики.

Методы исследования

Для достижения поставленной цели использованы следующие методы:

- 1) математическое моделирование систем massive MIMO с различными параметрами (число антенн, количество пользователей, схемы обработки сигналов);
- 2) анализ ЕЕ и SE при различных конфигурациях системы;
- 3) применение многокритериальных методов оптимизации для поиска рационального компромисса между ЕЕ и SE;
- 4) имитационное моделирование для валидации полученных результатов.

Порядок проведения исследования включает несколько последовательных этапов, направленных на всесторонний анализ характеристик систем massive MIMO. На начальном этапе была разработана математическая модель системы, учитывающая ключевые параметры: количество антенн на базовой станции (M), число пользователей в системе (K), а также различные схемы обработки сигналов, включая минимизацию среднеквадратичной ошибки с учетом мультипользовательского искажения

(М-MMSE, аббр. от англ. Multi-cell Minimum Mean Square Error), упрощенную минимизацию (S-MMSE, аббр. от англ. Single-cell Minimum Mean Square Error), регуляризованное «форсирование нуля» (RZF, аббр. от англ. Regularized Zero Forcing), классическое «форсирование нуля» (ZF, аббр. от англ. Zero Forcing) и метод максимального отношения (MR, аббр. от англ. Maximal Ratio).

Далее исследование было сосредоточено на анализе ЕЕ и SE в зависимости от значений M и K . Особое внимание уделялось выявлению оптимального соотношения между числом антенн и количеством пользователей, обеспечивающего максимальную ЕЕ без значительной потери SE.

Следующим шагом стало сравнительное исследование указанных схем обработки сигналов с точки зрения показателей ЕЕ и SE. Это позволило установить, какие алгоритмы обеспечивают наилучший компромисс между расходом энергии и пропускной способностью системы при различных условиях эксплуатации.

Для проверки достоверности теоретических выкладок было проведено имитационное моделирование, результаты которого подтвердили корректность предложенной модели и допущений. Анализ полученных данных показал, что в условиях преднамеренных внешних помех при правильной настройке параметров системы возможно достичь повышения помехоустойчивости каналов связи в massive MIMO, что является важным результатом для повышения надежности и устойчивости современных беспроводных сетей.

Изложение основного материала

В данном исследовании рассматривается многосотовая система massive MIMO. В каждой ячейке базовая станция оснащена M антеннами и одновременно обслуживает K пользовательских устройств (UE). Канал стационарен внутри когерентного блока длиной τ_c символов. Для оценки канала используются ортогональные пилотные последовательности длиной τ_p символов.

Коэффициент повторного использования пилотов принят равным $f = 1$. При таком выборе во всех ячейках применяется одинаковый набор пилотов. Внутри одной ячейки для исключения внутрисотовой интерференции пилоты назначаются ортогонально, что минимально требует $\tau_p = K$. Данное условие гарантирует корректную оценку внутрисотовых каналов и позволяет фокусировать внимание на влиянии межсотового переиспользования пилотов на помехоустойчивость и пропускную способность. Альтернативные варианты $f > 1$ увеличивают τ_p и снижают долю полезных данных в τ_c , поэтому в рамках сравнения выбран базовый вариант $f = 1$ с минимально необходимой длиной пилотных сигналов.

SE оценивается через среднюю пропускную способность на ячейку R_{cell} (Мбит/с) с учетом доли пилотных сигналов $(1 - \tau_p/\tau_c)$ и схемы обработки сигналов. EE определяется как отношение суммарной полезной пропускной способности к полной потребляемой мощности P_{tot} , измеряемой в Вт (Дж/с), поэтому единица измерения EE – Мбит/Дж:

$$EE = R_{\text{cell}}/P_{\text{tot}}. \quad (1)$$

Для анализа масштабируемости дополнительно рассматривается пропускная способность на единицу площади R_{area} при фиксированной плотности размещения ячеек.

Рассмотрим модель схмотехнического энергопотребления (CP, аббр. от англ. Circuit Power).

Полная потребляемая мощность представляется в виде:

$$P_{\text{tot}} = P_{\text{TX}} + P_{\text{PA}} + P_{\text{BB}}(M, K, S) + P_{\text{RF}}(M) + P_{\text{FH}} + P_{\text{FIX}}, \quad (2)$$

где S – выбранная схема обработки; P_{TX} – излучаемая радиочастотная мощность на антенну, суммированная по всем антеннам; $P_{\text{PA}} = P_{\text{RF,out}}/\eta_{\text{PA}}$ – мощность, расходуемая усилителями мощности; η_{PA} – коэффициент полезного действия (КПД) усилителей мощности; $P_{\text{BB}}(M, K, S)$ – мощность, потребляемая для цифровой обработки сигнала в базовой полосе, включающей оценку канала, формирование / разрешение лучей и детектирование, и зависящая от алгоритмической сложности выбранной схемы; $P_{\text{RF}}(M)$ – потребление ВЧ-тракта на антенну (фильтры, преобразователи сигналов, АЦП / ЦАП), масштабируется по M ; P_{FH} – мощность, потребляемая транспортной (транзитной) сетью, соединяющей базовую станцию с центром обработки; P_{FIX} – нагрузка, не зависящая от M и K (системные накладные расходы, охлаждение и т. п.).

В исследовании используются два набора параметров модели CP (таблица 1). Первый соответствует консервативным оценкам элементной базы. Второй отражает улучшенную аппаратную реализацию (уменьшенные удельные мощности ВЧ-трактов и вычислительных блоков), что позволяет оценить чувствительность EE к технологическим изменениям.

В рамках исследования проведено сравнение наиболее распространенных линейных алгоритмов обработки сигналов в системах massive MIMO. Рассмотрены следующие варианты [11]:

1) MR – метод согласования сигнала с каналом, при котором вектор передачи формируется пропорционально отклику канала, характеризующийся минимальной вычислительной сложностью и низкой способностью к подавлению межпользовательских помех;

2) ZF – алгоритм, обеспечивающий подавление межпользовательских помех путем обращения

матрицы канала; для его корректной работы необходимо выполнение условия $M > K$;

3) RZF – модификация ZF с регуляризацией, позволяющая балансировать между усилением полезного сигнала и подавлением помех;

4) S-MMSE – метод минимизации среднеквадратичной ошибки с использованием только внутрисотовой информации о каналах;

5) M-MMSE – усовершенствованный алгоритм, учитывающий межсотовую статистическую информацию, что повышает эффективность подавления помех, но сопровождается увеличением вычислительных затрат.

ТАБЛИЦА 1. Наборы параметров модели CP для двух вариантов аппаратной реализации

TABLE 1. CP Model Parameter Sets for Two Hardware Implementation Scenarios

Компонент мощности	Обозначение	Консервативный набор	Улучшенный набор
Излучаемая мощность	P_{TX} , Вт/антенну	0,1	0,1
КПД усилителя мощности	η_{PA}	0,25	0,4
Потребляемая мощность ВЧ-тракта на антенну	$P_{\text{RF}}(M)$, Вт/антенну	0,3	0,15
Потребляемая мощность цифровой обработки на одного абонента	$P_{\text{BB}}(M, K, S)$, Вт/UE	0,5	0,25
Потребляемая мощность транспортной сети	P_{FH} , Вт	5	3
Системные накладные расходы	P_{FIX} , Вт	10	5

С точки зрения вычислительной сложности, обработка в схеме MR требует минимальных ресурсов. Алгоритмы ZF и RZF используют матрицы размером $K \times K$, тогда как S-MMSE и M-MMSE работают с матрицами размером $M \times M$ и дополнительно используют расширенный набор статистических данных, что повышает нагрузку на блок базовой полосы P_{BB} и, следовательно, влияет на энергетическую эффективность системы.

Анализ экспериментальных условий проводится для двух сценариев: $K = 10$ и $K = 20$ при изменении числа антенн M от 10 до 200 с шагом 10. Для алгоритма ZF дополнительно соблюдается условие $M > K$. Доля времени когерентного блока, затрачиваемая на передачу пилотных последовательностей, определяется как $\tau_p/\tau_c = K/\tau_c$. Структура пилотных сигналов задается коэффициентом повторного использования $f = 1$, что предполагает идентичный набор пилотов во всех ячейках и

обеспечивает минимальную длину пилотной последовательности $\tau_p = K$.

Коэффициент полезного действия усилителей мощности η_{PA} принят выше 0,25. Такое значение объясняется распределением суммарной излучаемой мощности между большим числом антенн, что снижает требования к линейности отдельных усилителей и позволяет использовать более эффективные схемы усиления, включая архитектуры класса Doherty и усилители с отслеживанием огибающей. Остальные компоненты полной потребляемой мощности P_{tot} параметризованы через два набора значений – «консервативный» и «улучшенный», что обеспечивает воспроизводимость расчетов и возможность оценки влияния технологических факторов на энергетическую эффективность.

Для каждой тройки параметров (M, K, S) , где S – схема обработки, выполняются следующие шаги.

Шаг 1. Определяется SE с учетом времени, затраченного на передачу пилотных сигналов, и межсотового переиспользования пилотов:

$$SE = \left(1 - \frac{\tau_p}{\tau_c}\right) \cdot \sum_{k=1}^K \log_2(1 + \text{SINR}_k), \quad (3)$$

где τ_p – длина пилотной последовательности; τ_c – длина когерентного блока; SINR_k – отношение сигнал / шум + помеха для k -го пользователя.

Шаг 2. На основе модели CP рассчитывается полная мощность P_{tot} :

$$P_{tot} = M \cdot \left(\frac{P_{TX}}{\eta_{PA}} + P_{RF}\right) + K \cdot P_{BB}(M, K, S) + P_{FH} + P_{FIX}. \quad (4)$$

3) Вычисляется EE как отношение суммарной полезной пропускной способности к P_{tot} :

$$EE = \frac{R_{cell}}{P_{tot}}, \quad (5)$$

где R_{cell} – суммарная пропускная способность на ячейку.

Если ширина полосы B известна, то:

$$R_{cell} = B \cdot SE_{norm}, \quad (6)$$

где B – это скорость передачи символов (ширина полосы в Гц); SE_{norm} – нормализованная спектральная эффективность в бит/(с·Гц), т.е. уже учтена доля пилотных сигналов и схемы обработки.

Подставляя выражение для SE, получаем:

$$R_{cell} = B \cdot \left(1 - \frac{K}{\tau_c}\right) \cdot \sum_{k=1}^K \log_2(1 + \text{SINR}_k). \quad (7)$$

Если B не указана, то R_{cell} может быть выражена в бит/символ и численно совпадает со значением SE.

Шаг 4. Строится зависимость EE от SE при изменении числа антенн M при фиксированном K .

Шаг 5. Определяется диапазон значений M , при котором достигаются значения EE, близкие к максимальным, без существенного снижения SE:

$$EE(M) \approx \max_M EE(M) \quad \text{при } SE(M) \geq SE_{min}. \quad (8)$$

Шаг 6. Выполняется сравнительный анализ схем обработки по критерию компромисса EE–SE по формулам (3, 5).

Для оценки влияния числа антенн базовой станции на EE и SE при фиксированном числе пользователей проведено моделирование по описанной выше методике. На рисунке 1 представлена зависимость EE от средней пропускной способности на одну ячейку для различных схем обработки сигналов. Моделирование выполнено для сценария с $K = 10$ обслуживаемыми пользователями при изменении числа антенн базовой станции M от 10 до 200 с шагом 10. Параметры системы соответствуют ранее описанным условиям, включая коэффициент повторного использования пилотов $f = 1$, длину пилотной последовательности $\tau_p = K$ и два набора параметров CP.

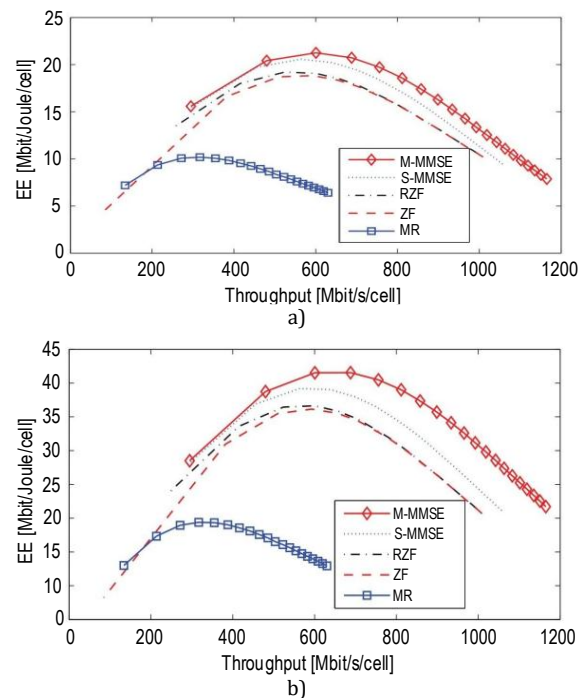


Рис. 1. EE в зависимости от средней пропускной способности на ячейку при всех схемах обработки ($K = 10$): а) с первым набором значений; б) со вторым набором

Fig. 1. EE Versus the Average Throughput Per Cell for All Processing Schemes ($K = 10$): a) First Parameter Set; b) Second Parameter Set

Сравнительный анализ схем обработки показал, что M-MMSE обеспечивает наибольшие значения EE во всем диапазоне моделирования. Максимум 21,26 Мбит/Дж достигается при $M = 30$ (соответствует средней пропускной способности 600 Мбит/с

на ячейку или 9,6 Гбит/с·км²). При $M = 40$ ЕЕ снижается незначительно – до 20,73 Мбит/Дж – при росте пропускной способности до 11 Гбит/с·км².

S-MMSE демонстрирует близкие результаты: оптимум достигается также при $M = 30$, однако ЕЕ на 3,2 % ниже, а пропускная способность на 6 % меньше по сравнению с M-MMSE.

RZF и ZF показывают сопоставимые значения, максимизируя ЕЕ на уровне около 19 Мбит/Дж при $M = 30$ и соответствующей плотности пропускной способности 8,38 Гбит/с·км².

MR демонстрирует наименьшие значения ЕЕ в исследуемом диапазоне параметров.

Отмечено, что при увеличении пропускной способности схемы RZF и ZF начинают приближаться по эффективности к M-MMSE и S-MMSE. Тем не менее, для сценария с $K = 10$ дополнительные вычислительные затраты, связанные с применением M-MMSE, оправданы за счет одновременно высокого уровня ЕЕ и SE.

Обратим внимание, что для ZF ЕЕ быстро ухудшается при пропускной способности менее 380 Мбит/с на ячейку (наблюдается при $M \approx K$, где эта схема, как известно, работает плохо). Для MR максимальная ЕЕ составляет 10,18 Мбит/Дж при $M = 40$ и пропускной способности 5,07 Гбит/с·км².

Данные рисунка 1b получены со вторым набором значений из таблицы 1. По сравнению с рисунком 1a ЕЕ всех схем увеличивается примерно вдвое, поскольку большинство параметров модели СР во втором наборе уменьшено примерно в два раза, но общие тенденции сохраняются. Для M-MMSE и S-MMSE значения ЕЕ достигают 41,52 и 39,2 Мбит/Дж соответственно при $M = 30$. В таблице 2 представлены значения ЕЕ и пропускной способности по площади для схем M-MMSE, RZF и MR при $K = 10$ и $M = 40$. В этом случае ЕЕ достигает максимума примерно при $M = 30$ или 40 независимо от значений параметров СР. Хотя указанные значения числа антенн M существенно ниже типичных прогнозов для систем massive MIMO [12–14], максимальная ЕЕ достигается при ожидаемом соотношении $M/K \approx 3$ –4. При увеличении K оптимальное значение M возрастает, особенно при использовании более энергоэффективного оборудования.

ТАБЛИЦА 2. Максимальная ЕЕ на ячейку с двумя наборами значений СР для M-MMSE, RZF и MR

TABLE 2. Maximum Cell EE for Two CP Parameter Sets for M-MMSE, RZF, and MR

Сценарий	Схема обработки	ЕЕ, Мбит/Дж		Пропускная способность по площади, Гбит/с/км ²
		набор 1	набор 2	
$K = 10$, $M = 40$	M-MMSE	23,73	41,52	11
	RZF	19,07	36,63	9,6
	MR	10,18	19,38	5,07
$K = 20$, $M = 60$	M-MMSE	20,12	38,45	12,3
	RZF	18,34	34,92	10,8
	MR	9,67	17,21	6,02

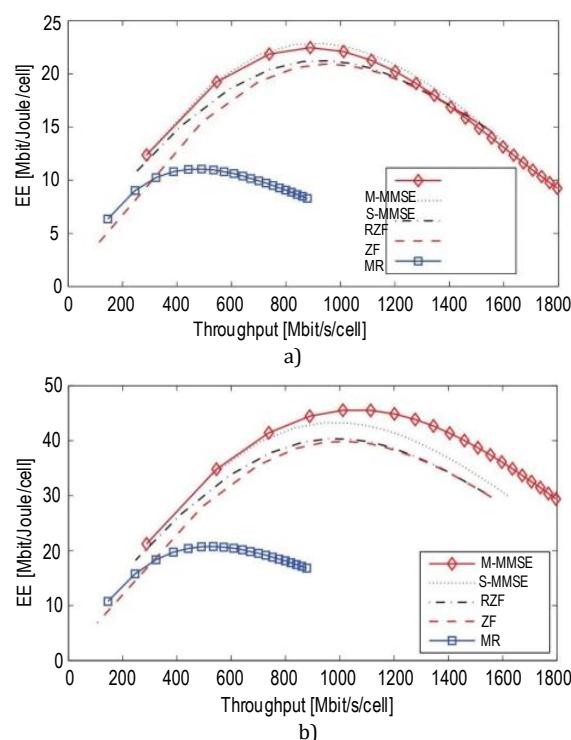


Рис. 2. ЕЕ в зависимости от средней пропускной способности на ячейку при всех схемах обработки ($K = 20$): с первым (а) и со вторым (б) набором значений в таблице 1

Fig. 2. EE Versus the Average Throughput Per Cell for All Processing Schemes ($K = 20$): (a) First Parameter Set; (b) Second Parameter Set from Table 1

На рисунке 2 показаны результаты для $K = 20$. Значения пропускной способности получены для всех схем обработки, кроме ZF, при изменении M от 10 до 200 с шагом 10; для ZF рассматривались только значения $M > K$, поэтому M изменялось от 20 до 200. Рисунок 2b показывает, что увеличение числа UE в ячейке может положительно влиять на ЕЕ, тогда как из рисунка 2a следует, что при $K = 20$ и первом наборе параметров СР наибольшая ЕЕ достигается не для M-MMSE, а для S-MMSE. С количественной точки зрения S-MMSE обеспечивает максимальную ЕЕ 22,86 Мбит/Дж при $M = 50$ и пропускной способности 15,05 Гбит/с·км², демонстрируя наилучший компромисс между ЕЕ и SE в данном сценарии. Это связано с тем, что вычислительная составляющая СР для M-MMSE значительно выше, чем для S-MMSE, RZF и ZF, и тем самым частично компенсирует выигрыш M-MMSE по SE.

Для второго набора значений СР можно сделать различные наблюдения, как видно на рисунке 2b. В этом случае общие тенденции соответствуют графикам на рисунке 1, где M-MMSE обеспечивает наивысший уровень ЕЕ и пропускной способности. Кроме того, увеличение количества UE на ячейку положительно влияет на ЕЕ всех цепей, которая больше при любом значении пропускной способности. В отличие от сценария с $K = 10$ на рисунке 1b, где максимальная ЕЕ достигается при $M = 30$ или 40,

при $K = 20$ наибольшая ЕЕ наблюдается при $M = 50$ или 60.

При обслуживании большого количества UE и одновременном увеличении числа антенн на базовой станции (для компенсации межпользовательских помех) ЕЕ сети может быть повышена, но только при условии, что преимущества и затраты на развертывание дополнительного оборудования сбалансированы.

Разработанная модель СР в этом разделе существенно зависит от аппаратной конфигурации (т. е. от количества антенн M и пользователей K), а также от выбора параметров модели, приведенных в таблице 1.

Сравнительный анализ схем обработки показал следующее:

- S-MMSE требует больше вычислительных ресурсов, чем RZF и ZF, поскольку использует матрицы размером $M \times M$, тогда как RZF и ZF – только $K \times K$;

- M-MMSE по вычислительной сложности близок к S-MMSE, однако его дополнительная нагрузка на СР частично компенсируется более эффективной оценкой межсотовых каналов;

- MR обладает наименьшей вычислительной сложностью и, соответственно, минимальным СР, поскольку не требует обращения матриц.

Тем не менее, благодаря высокой вычислительной эффективности современных систем, различия в вычислительной нагрузке между схемами

обработки оказывают лишь незначительное влияние на общий СР – он остается примерно одинаковым для всех алгоритмов.

В структуре полной потребляемой мощности наибольший вклад вносит приемо-передающий тракт (усилители мощности, ВЧ-фильтры, АЦП / ЦАП). За ним следует фиксированная мощность – системные накладные расходы, охлаждение и прочие постоянные компоненты. Далее – оценка канала, которая одинакова для всех схем обработки при фиксированном числе пользователей.

Мощность, потребляемая транспортной сетью, а также кодированием и декодированием, составляет лишь незначительную долю общего СР в massive MIMO, особенно при большом числе антенн на базовой станции.

Заключение

Проведенное исследование показало, что в системах massive MIMO существует оптимальное соотношение между числом антенн на базовой станции и количеством обслуживаемых пользователей, при котором достигается максимальная энергетическая эффективность без существенного снижения спектральной эффективности. Наиболее эффективными по показателю ЕЕ оказались схемы обработки сигналов M-MMSE и S-MMSE. Результаты моделирования могут быть использованы для проектирования и оптимизации энергосберегающих беспроводных сетей следующего поколения.

Список источников

1. Арсамерзуева А.У., Тасуева Х.З.А., Чадаев А.К. Применение технологий MIMO для повышения эффективности в инфокоммуникационных сетях // XII Всероссийская научно-практическая конференция «Молодежь, наука, инновации» (Грозный, Российская Федерация, 18 октября 2023 г.). Грозный: Грозненский государственный нефтяной технический университет им. М.Д. Миллионщикова, 2023. С. 31–35. DOI:10.26200/GSTOU.2023.95.35.004. EDN:LNURUDH
2. Алиев И.В. Спектральная эффективность в massive MIMO // Всероссийская научно-техническая конференция с международным участием «Современные проблемы телекоммуникаций» (Новосибирск, Российская Федерация, 19–20 апреля 2023 г.). Новосибирск: Сибирский государственный университет телекоммуникаций и информатики, 2023. С. 380–385. DOI:10.55648/978-5-91434-084-8-2023-274-278 EDN:YUWOKI
3. Свиридова И.В., Хорошайлов Р.Н., Лялин Д.В. Анализ методов MIMO для 5G и последующих технологий: преимущества и недостатки // Радиотехника. 2023. Т. 87. № 8. С. 100–104. DOI:10.18127/j00338486-202308-16. EDN:YIBXZQ
4. Аль-Рави М., Аль-Рави М. Анализ эффективности массовой MIMO системы восходящей связи при использовании канала с замираниями Накагами-М // Известия высших учебных заведений. Радиоэлектроника. 2017. Т. 60. № 1(655). С. 18–24. DOI:10.20535/S0021347017010022. EDN:ZQMSFN
5. Севрюков А.Е., Осипов Н.С. Влияние конфигурации систем MIMO на энергетические параметры радиолинии // III Всероссийская научно-практическая конференция «Инфокоммуникации и информационная безопасность: состояние, проблемы и пути решения», Курск, Российская Федерация, 11–13 мая 2016 г.). Курск: Юго-Западный государственный университет, 2016. С. 105–110. EDN:WXQXMX
6. Björnson E., Larsson E.G., Marzetta T.L. Massive MIMO: ten myths and one critical question // IEEE Communications Magazine. 2016. Vol. 54. Iss. 2. PP. 114–123. DOI:10.1109/MCOM.2016.7402270
7. Halbauer H., Weber A., Wiegner D., Wild T. (2019). Energy Efficient Massive MIMO Array Configurations // IEEE Globecom Workshops (GC Wkshps, Abu Dhabi, United Arab Emirates, 09–13 December 2018). IEEE, 2019. DOI:10.1109/GLOCOMW.2018.8644331
8. Hoydis J., ten Brink S., Debbah M. Massive MIMO in the UL/DL of Cellular Networks: How Many Antennas Do We Need? // IEEE Journal on Selected Areas in Communications. 2013. Vol. 31. Iss. 2. PP. 160–171. DOI:10.1109/JSAC.2013.130205
9. Liu W., Han S., Yang C. Is massive MIMO energy efficient? arXiv:1505.07187. 2015.
10. Ngo H.Q., Larsson E.G., Marzetta T.L. Energy and Spectral Efficiency of Very Large Multiuser MIMO Systems // IEEE Transactions on Communications. 2013. Vol. 61. Iss. 4. PP. 1436–1449. DOI:10.1109/TCOMM.2013.020413.110848

11. Prasad K.N.R.S.V., Hossain E., Bhargava V.K. Energy Efficiency in Massive MIMO-Based 5G Networks: Opportunities and Challenges // arXiv:1511.08689. 2015.
12. Salh A., Audah L., Shah N.S.M., Hamzah S.A. Trade-Off Energy and Spectral Efficiency in a Downlink Massive MIMO System // *Wireless Personal Communications*. 2019. Vol. 106. PP. 897–910. DOI:10.1007/s11277-019-06194-4. EDN:OHCHSG
13. Sanguinetti L., Björnson E., Hoydis J. Toward Massive MIMO 2.0: Understanding Spatial Correlation, Interference Suppression, and Pilot Contamination // *IEEE Transactions on Communications*. 2020. Vol. 68. Iss. 1. PP. 232–257. DOI:10.1109/tcomm.2019.2945792. EDN:JGHAGN
14. Hei Y., Zhang C., Song W., Kou Y. Energy and spectral efficiency tradeoff in massive MIMO systems with multi-objective adaptive genetic algorithm // *Soft Computing*. 2019. Vol. 23. PP. 7163–7179. DOI:10.1007/s00500-018-3356-x. EDN:XJLOUS

References

1. Arsamerzueva A.U., Tasueva Kh.Z.A., Chadaev A.K. Application of MIMO Technologies to Improve Efficiency in Infocommunication Networks. *Proceedings of the XII All-Russian Scientific and Practical Conference "Youth, Science, Innovation", 18 October 2023, Grozny, Russian Federation*. Grozny: GSTOU named after M.D. Millionshchikov Publ.; 2023. p.31–35. (in Russ.) DOI:10.26200/GSTOU.2023.95.35.004. EDN:LNRUDH
2. Aliyev I.V. Spectral Efficiency in Massive MIMO. *Proceedings of the All-Russian Scientific and Technical Conference with International Participation on Modern Problems of Telecommunications, 19–20 April 2023, Novosibirsk, Russian Federation*. Novosibirsk: Siberian State University of Telecommunications and Informatics Publ.; 2023. p.380–385. (in Russ.) DOI:10.55648/978-5-91434-084-8-2023-274-278 EDN:YUWOKI
3. Sviridova I.V., Khoroshailov R.N., Lyalin D.V. Analysis of MIMO Methods for 5G and Subsequent Technologies: Advantages and Disadvantages. *Radioengineering*. 2023;87(8):100–104. (in Russ.) DOI:10.18127/j00338486-202308-16. EDN:YIBXZQ
4. AL-Rawi M., AL-Rawi M. Performance of Massive MIMO Uplink System Over Nakagami-m Fading Channel. *Radioelectronics and Communications Systems* 2017;60(1):18–24. (in Russ.) DOI:10.20535/S0021347017010022. EDN:ZQMSFN
5. Sevryukov A.E., Osipov N.S. Influence Configuration MIMO Radio Systems for Energy Parameters. *Proceedings of the IIIrd All-Russian Scientific and Practical Conference "Infocommunications and Information Security: Status, Problems and Solutions", 11–13 May 2016, Kursk, Russian Federation*. Kursk: Southwest State University Publ.; 2016. p.105–110. (in Russ.) EDN:WXQXMX
6. Björnson E., Larsson E.G., Marzetta T.L. Massive MIMO: ten myths and one critical question. *IEEE Communications Magazine*. 2016;54(2):114–123. DOI:10.1109/MCOM.2016.7402270
7. Halbauer H., Weber A., Wiegner D., Wild T. (2019). Energy Efficient Massive MIMO Array Configurations. *IEEE Globecom Workshops, GC Wkshps, 09–13 December 2018, Abu Dhabi, United Arab Emirates*. IEEE; 2019. DOI:10.1109/GLOCOMW.2018.8644331
8. Hoydis J., ten Brink S., Debbah M. Massive MIMO in the UL/DL of Cellular Networks: How Many Antennas Do We Need? *IEEE Journal on Selected Areas in Communications*. 2013;31(2):160–171. DOI:10.1109/JSAC.2013.130205
9. Liu W., Han S., Yang C. Is massive MIMO energy efficient? *arXiv:1505.07187*. 2015.
10. Ngo H.Q., Larsson E.G., Marzetta T.L. Energy and Spectral Efficiency of Very Large Multiuser MIMO Systems. *IEEE Transactions on Communications*. 2013;61(4):1436–1449. DOI:10.1109/TCOMM.2013.020413.110848
11. Prasad K.N.R.S.V., Hossain E., Bhargava V.K. Energy Efficiency in Massive MIMO-Based 5G Networks: Opportunities and Challenges. *arXiv:1511.08689*. 2015.
12. Salh A., Audah L., Shah N.S.M., Hamzah S.A. Trade-Off Energy and Spectral Efficiency in a Downlink Massive MIMO System. *Wireless Personal Communications*. 2019;106:897–910. DOI:10.1007/s11277-019-06194-4. EDN:OHCHSG
13. Sanguinetti L., Björnson E., Hoydis J. Toward Massive MIMO 2.0: Understanding Spatial Correlation, Interference Suppression, and Pilot Contamination. *IEEE Transactions on Communications*. 2020;68(1):232–257. DOI:10.1109/tcomm.2019.2945792. EDN:JGHAGN
14. Hei Y., Zhang C., Song W., Kou Y. Energy and spectral efficiency tradeoff in massive MIMO systems with multi-objective adaptive genetic algorithm. *Soft Computing*. 2019;23:7163–7179. DOI:10.1007/s00500-018-3356-x. EDN:XJLOUS

Статья поступила в редакцию 14.01.2026; одобрена после рецензирования 04.03.2026; принята к публикации 16.03.2026

The article was submitted 14.01.2026; approved after reviewing 04.03.2026; accepted for publication 16.03.2026

Информация об авторах:

НГИЕМ
Ван Донг

аспирант кафедры мультимедийных технологий и телекоммуникаций Московского физико-технического института (национального исследовательского университета)
<https://orcid.org/0009-0006-6318-294X>

ДВОРКОВИЧ
Александр Викторович

доктор технических наук, профессор, профессор РАН, член-корреспондент РАН, заведующий кафедрой мультимедийных технологий и телекоммуникаций Московского физико-технического института (национального исследовательского университета)
<https://orcid.org/0000-0003-1190-3582>

Авторы сообщают об отсутствии конфликтов интересов.

The authors declare no conflicts of interests.

Научная статья

УДК 004.387:621.3.087.93

<https://doi.org/10.31854/1813-324X-2026-12-2-53-63>

EDN:DAQBCB



Моделирование параллельного нониусного цифроаналогового преобразователя второго типа

✉ Юрий Александрович Никитин, nikitin.ua@sut.ru✉ Владимир Алексеевич Филин, filin.va@sut.ru

Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича,
Санкт-Петербург, 193232, Российская Федерация

Аннотация

Цифроаналоговые преобразователи широко и эффективно применяются в радиоэлектронной аппаратуре различного назначения, когда необходимо преобразование цифрового кода управления в аналоговый параметр – ток или напряжение. Их используют, в том числе и в цифроаналоговых синтезаторах частоты для получения требуемой формы огибающей синтезируемого сигнала, широко применяют в видео- и аудиоинтерфейсах, в цифровых осциллографах и в источниках прецизионных сигналов. В настоящее время основными проблемами при построении прецизионных и (или) быстродействующих цифроаналоговых преобразователей являются технологические ограничения производства, а именно – конечная точность реализации аналоговых элементов – прежде всего, аналоговых ключей напряжения (тока) и резисторов матрицы $R-2R$. Поэтому **актуален** структурный метод преодоления существующих технологических проблем.

Цель настоящей работы заключается в теоретическом обосновании нового подхода к идеологии цифроаналогового преобразования и в построении структур цифроаналоговых преобразователей повышенной точности и (или) быстродействия.

Решение проблемы заключается в увеличении числа опорных сигналов на входах парциальных цифроаналоговых преобразователей вне зависимости от их внутренней структуры, но при безусловном обеспечении жесткого дробно-кратного (нониусного) соотношения величин опорных сигналов. При этом сопряжение неонусных опорных сигналов необходимо производить всего в двух точках – в начале и в конце шкалы, причем на постоянном токе. Точность сопряжения неонусных шкал (опорных сигналов) должна соответствовать конечной точности цифроаналогового преобразования.

Новизна и оригинальность предлагаемого метода заключается в существенном расширении диапазона преобразования цифрового кода управления в аналоговый параметр, т. е. в получении нового качества цифроаналогового преобразования.

Возможность практической реализации новой структуры цифроаналогового преобразователя подтверждена теоретическими расчетами и схемотехническим моделированием с помощью пакета Microcap12, которое показало корректность разработанного метода. Предлагаемое решение позволяет обойти технологические ограничения производства на потенциально достижимую точность преобразования при изготовлении микросхем цифроаналогового преобразователя и обеспечивает качественно новые возможности техники цифроаналогового преобразования.

Ключевые слова: цифроаналоговый преобразователь, управляющий двоичный код, шаг квантования, арифметический сумматор, аналоговый сумматор, неонус, опорный сигнал, матрица $R-2R$, обращенная матрица, переключатели тока или напряжения

Ссылка для цитирования: Никитин Ю.А., Филин В.А. Моделирование параллельного неонусного цифроаналогового преобразователя второго типа // Труды учебных заведений связи. 2026. Т. 12. № 2. С. 53–63. DOI:10.31854/1813-324X-2026-12-2-53-63. EDN:DAQBCB

Original research

<https://doi.org/10.31854/1813-324X-2026-12-2-53-63>

EDN:DAQBCB

Simulating Parallel Vernier Second Type Digital-to-Analog Converter

 Yuri A. Nikitin✉, nikitin.ua@sut.ru Vladimir A. Filin, filin.va@sut.ru

The Bonch-Bruевич Saint Petersburg State University of Telecommunications,
St. Petersburg, 193232, Russian Federation

Annotation

Digital-to-analog converters are widely and effectively used in radio electronic equipment for various purposes when it is necessary to convert a digital control code into an analog parameter – current or voltage. They are also used in digital-to-analog frequency synthesizers to obtain the desired envelope shape of the synthesized signal, and are widely used in video and audio interfaces, digital oscilloscopes, and precision signal sources. Currently, the main problems in the construction of precision and (or) high-speed digital-to-analog converters are technological limitations of production, namely, the final accuracy of the implementation of analog elements, primarily analog voltage (current) switches and R-2R matrix resistors. Therefore, the structural method of overcoming existing technological problems is relevant.

The purpose of this work is to theoretically substantiate a new approach to the ideology of digital-to-analog conversion and to build structures of digital-to-analog converters with increased accuracy and (or) high performance.

The solution to the problem is to increase the number of reference signals at the inputs of partial digital-to-analog converters, regardless of their internal structure, but while unconditionally ensuring a strict fractional-multiple (vernier) ratio of the values of the reference signals. At the same time, the vernier reference signals must be coupled at only two points – at the beginning and at the end of the scale, and at direct current. The accuracy of the vernier scale coupling (the accuracy of the reference signal coupling) must correspond to the final accuracy of the digital-to-analog conversion.

The novelty and originality of the proposed method lies in the significant expansion of the range of conversion of the digital control code into an analog parameter, i. e. in obtaining a new quality of digital-to-analog conversion.

The possibility of practical implementation of the new digital-to-analog converter structure is confirmed by theoretical calculations and circuit modeling using the Microcap12 package, which demonstrated the correctness of the proposed method. The proposed solution makes it possible to circumvent the technological limitations of production on the potentially achievable conversion accuracy in the manufacture of digital-to-analog converters chips and provides qualitatively new possibilities for digital-to-analog conversion technology.

Keywords: digital-to-analog converter, binary control code, quantization step, arithmetic adder, analog adder, vernier, reference signal, R-2R matrix, inverted matrix, current or voltage switches

For citation: Nikitin Yu.A., Filin V.A. Simulating Parallel Vernier Second Type Digital-to-Analog Converter. *Proceedings of Telecommunication Universities*. 2026;12(2):53–63. (in Russ.) DOI:10.31854/1813-324X-2026-12-2-53-63. EDN:DAQBCB

Введение

Цифроаналоговые преобразователи (ЦАП) цифрового кода управления в аналоговый параметр – ток или напряжение являются неотъемлемыми элементами цифровых структур обработки сигналов – начиная от инфранизких частот геолокации, гидроакустики, звукового и радиовещания, обра-

ботки видеосигналов и заканчивая оконечной аппаратурой радиолокационных станций. В стандартных ЦАП из единственного опорного сигнала U формируют сигналы с соответствующими значениями (весами) разрядов входного управляющего кода X , которые затем суммируют и образуют дискретные значения выходной аналоговой величины

Z – напряжения или тока [1–8]. Будем считать, что цифровой код управления – позиционный, а система счисления с основанием a – двоичная ($a = 2$).

Для двоичного ЦАП с единственным опорным сигналом U выходной сигнал определяется следующим образом:

$$Z = gUX = gU(a_{n-1}2^{n-1} + a_{n-2}2^{n-2} + \dots + a_02^0), \quad (1)$$

где U – опорный (эталонный) сигнал; g – коэффициент пропорциональности; $q_n \in (a_0, a_1, \dots, a_{n-1}) \equiv [0, 1]$ – коэффициенты двоичных разрядов, принимающие значения «0» или «1»; $n = 0, 1, \dots, \psi-1$ – номер двоичного разряда.

Величина ступеньки (шага квантования) на выходе такого ЦАП равна $h = U/(a^\psi - 1)$. В этом случае $Z = UX$ и точность преобразования является линейной функцией точности элементарных электронных компонентов ЦАП – резистивных матриц, ключей тока или напряжения и т. д., т. е. определяется технологией изготовления конкретной мик-

росхемы ЦАП и конечной точностью (разбросом) номинальных значений его внутренних элементов [1]. Основные структуры ЦАП базируются на суммировании парциальных токов или напряжений; типичные структуры таких ЦАП приведены на рисунках 1–4 [1, 6, 7].

Заметим, что в ЦАП с линейным суммированием токов (см. рисунок 1) парциальные токи ветвей изменяются в 2^{n-1} раз, что приводит к необходимости изменять топологию и площади p – n переходов в ключах SW для выравнивания значений сопротивлений открытых ключей и падения напряжения на p – n переходах открытых ключей.

На рисунках 1 и 2 ключ SW1 подключает (переключает) старший разряд ЦАП. Однако транзисторы генераторов тока I_n и ключей SW $_n$ (см. рисунок 2) имеют разную топологию, определяемую протекающими токами. Соответственно, разной будет схемная и конструктивная емкость и, следовательно, разное быстродействие двоичных разрядов.

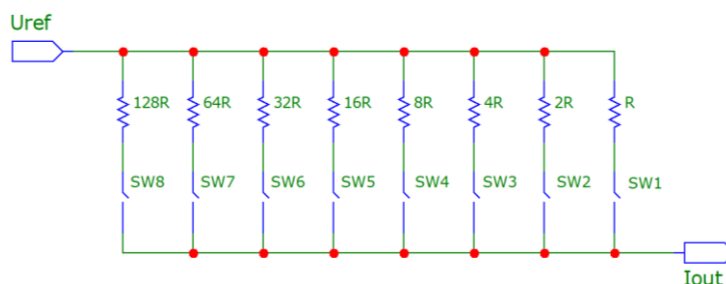


Рис. 1. 8-битный ЦАП с линейным суммированием токов

Fig. 1. 8-bit DAC with Linear Current Summation

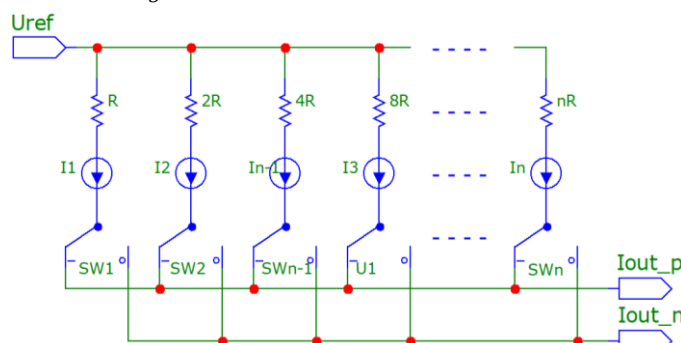


Рис. 2. ЦАП с суммированием двоично-взвешенных токов, $n = q_n 2^{-n}$, $n = 0, 1, 2, \dots, \psi-1$

Fig. 2. DAC with Binary-Weighted Current Summation, $n = q_n 2^{-n}$, $n = 0, 1, 2, \dots, \psi-1$

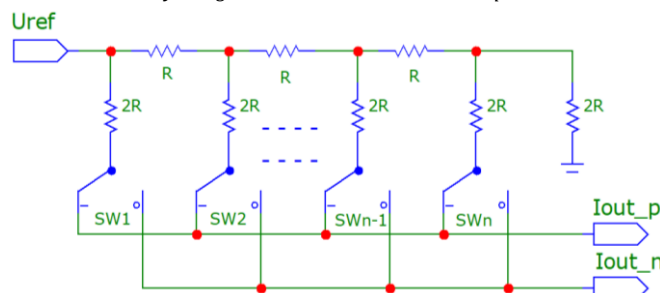
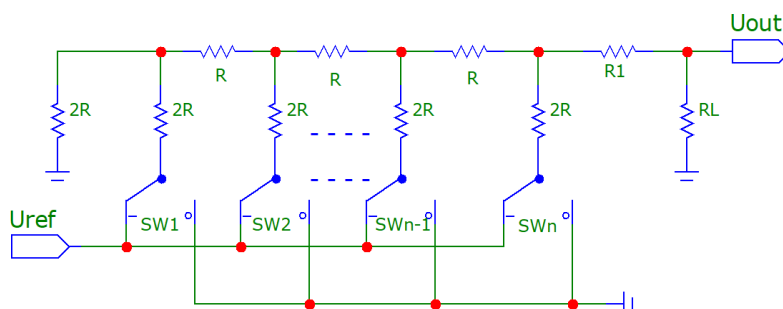


Рис. 3. ЦАП с суммированием весовых токов с помощью матрицы R2R

Fig. 3. DAC with Summation of Weight Currents Using an R2R Matrix

Рис. 4. ЦАП с суммированием напряжений на основе инверсной матрицы $R2R$ Fig. 4. Voltage Summing DAC Based on an Inverse $R2R$ Matrix

На рисунках 3 и 4 ключ SW1 переключает младший разряд ЦАП. Аналогичные проблемы у ключей SW ЦАП с суммированием весовых токов с помощью матрицы $R-2R$ (см. рисунок 3) и у ЦАП на основе суммирования напряжений с помощью обращенной матрицы $R2R$ (см. рисунок 4).

Развитие технологий производства биполярных транзисторов – основы ключей тока или напряжения, позволило уменьшить напряжение насыщения открытого ключа напряжения на биполярном транзисторе до 6–20 мВ в зависимости от протекающего тока [9], но для многоразрядных ЦАП этого недостаточно. Действительно, если принять величину опорного напряжения $U = 10,24$ В и разрядность ЦАП 8–10 бит, получим номинальную величину ступеньки такого ЦАП $h = 40$ –10 мВ, что соизмеримо с напряжением насыщения ключей SW (см. рисунок 5). Подобная коллизия приводит к необходимости параллельного включения транзисторов ключей SW старших разрядов (SW1, SW2) с целью уменьшения сопротивления в открытом состоянии для обеспечения требуемой точности всего ЦАП. В случае меньшего напряжения U проблемы возрастают. При этом величина коэффициента усиления транзисторов постоянна в широких пределах изменения тока коллектора (рисунок 6).

Погрешность преобразования ЦАП вследствие указанных причин в течение последних 30 лет почти не изменилась и не превышает 12–16 бит (www.analog.com/media/en/technical-documentation/data-sheets/ad5663.pdf; www.analog.com/media/en/technical-documentation/data-sheets/ad5664r.pdf). Поэтому характеристики преобразования реальных ЦАП отличаются от идеальных формой, величиной ступеней квантования и их расположением относительно осей координат (рисунок 7).

Для увеличения точности преобразования можно усложнять внутреннюю структуру ЦАП, увеличивать точность и стабильность параметров используемых элементов, но только в пределах технологических возможностей производства. В настоящее время указанные причины ограничивают максимально достижимую точность параллельного ЦАП 18–20 двоичными разрядами (битами) кода управления X .

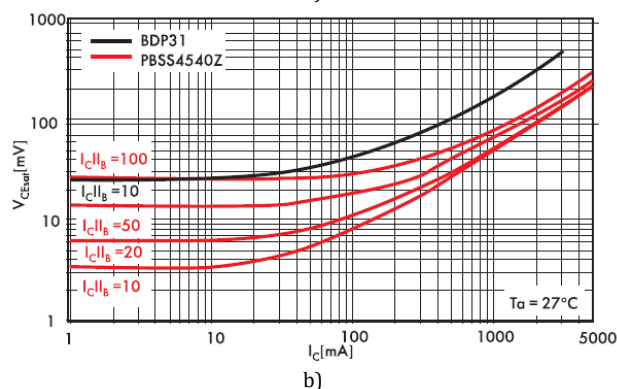
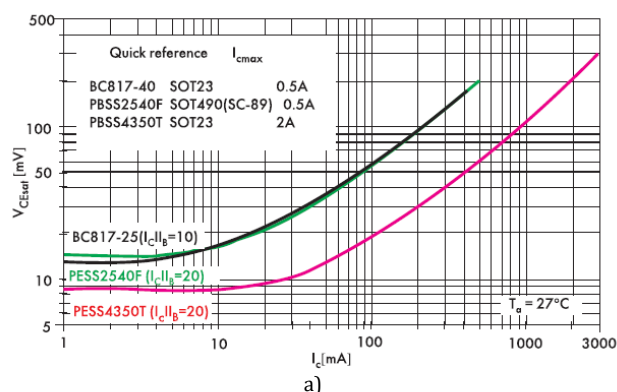
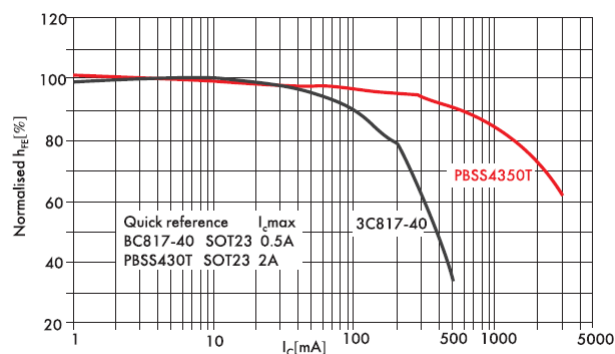


Рис. 5. Типовые значения для транзисторов серии BISS: а) напряжение насыщения для разных токов коллектора; б) зависимость напряжения насыщения от тока базы

Fig. 5. Typical Values for BISS Series Transistors: a) Saturation Voltage for Different Collector Currents; b) Dependence of Saturation Voltage on Base Current

Рис. 6. Зависимость коэффициента усиления h_{21} транзистора ключа от разрядного токаFig. 6. Dependence of the Gain h_{21} of the Key Transistor on the Discharge Current

Наиболее точные параллельные ЦАП реализуют заявленные параметры с помощью комбинации матрицы $R-2R$ стандартной разрядности 12–14 бит и дополнительного набора коммутируемых гене-

раторов «единичных» токов, число которых варьируется от 15 до 63 (рисунок 8), что дает дополнительно 4–6 двоичных разрядов до итоговой точности преобразования 18–20 бит <https://www.farnell.com/datasheets/2149287.pdf>.

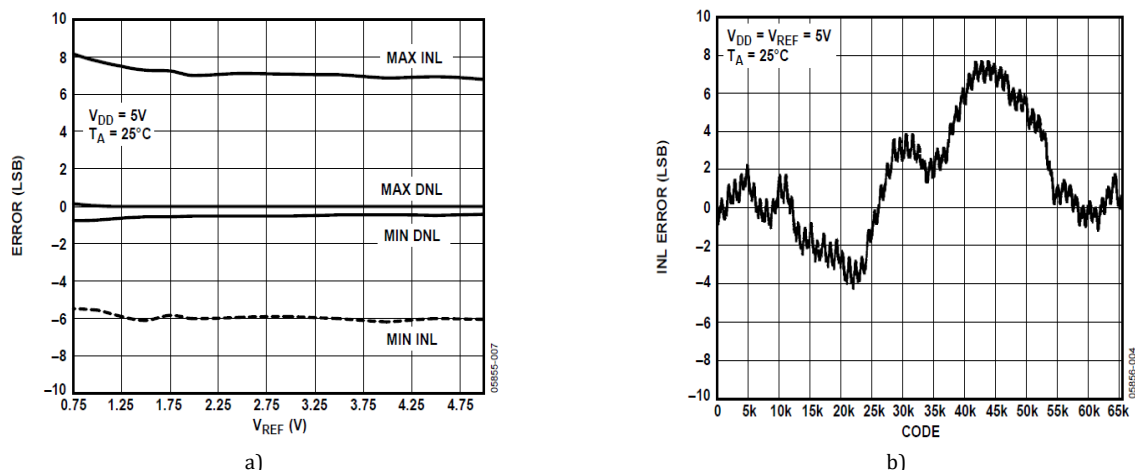


Рис. 7. Погрешности прецизионных ЦАП AD5663 (a) и AD5664R (b)

Fig. 7. Errors of the AD5663 (a) and AD5664R (b) Precision DACs

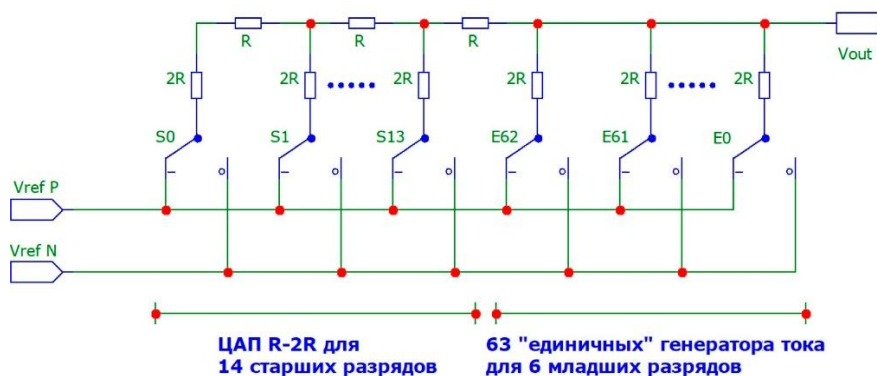


Рис. 8. Комбинированная структура матриц формирования весовых токов ЦАП AD5790

Fig. 8. Combined Structure of the AD5790 DAC Weighting Current Generation Matrices

Однако в рассмотренном способе цифроаналогового преобразования (см. рисунок 8) необходимо обеспечить высокую точность и 14-битовой матрицы $R-2R$, и 63 резисторов $2R$ матрицы токов 6-ти младших разрядов. К этим двум аналоговым узлам структуры предъявляются жесткие требования по точности и температурной стабильности во всем диапазоне эксплуатационных воздействий. В дополнение к данной структуре необходим дешифратор двоичного 6-битового кода управления младшими разрядами в линейную комбинацию токов. Кроме того, значительно увеличивается площадь кристалла.

Постановка задачи

Возможны разные способы решения данных проблем [10, 11]. В основу одного из них (рисунок 9) положен принцип нониусного преобразования [11, 12] – использование не одного, единственного, но

двух опорных сигналов, находящихся в жестком дробно-кратном соотношении, т.е. осуществление тройки преобразований:

$$X \Rightarrow Z_1; X \Rightarrow Z_2; X \Rightarrow Z_3; Z_1, Z_2, Z_3 \Rightarrow Z_0.$$

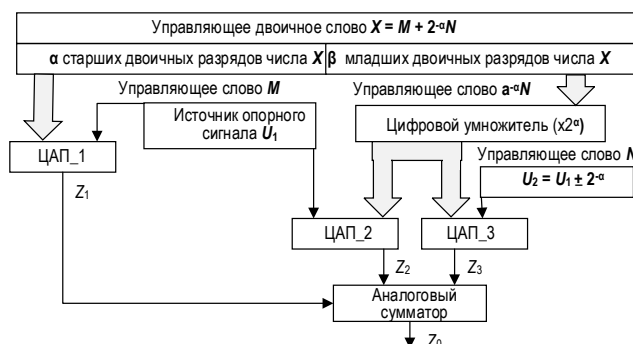


Рис. 9. Второй возможный способ нониусного цифроаналогового преобразования

Fig. 9. The Second Possible Method of Vernier Digital-to-Analog Conversion

Такое решение, как будет показано далее, позволяет резко повысить точность преобразования без существенного усложнения структур и схем используемых парциальных ЦАП, т. е. получить новое качество преобразования.

Иными словами, вместо единственного опорного сигнала присутствует пара (именно пара) опорных сигналов U_1 и U_2 , связанных жестким и однозначным соотношением:

$$U_2 = U_1(1 \pm 2^{-\alpha}), \quad (2)$$

где α – число двоичных разрядов, на которое сдвинут код управления $2^{-\alpha}N$, после чего производят суммирование преобразованных аналоговых сигналов Z_1 , Z_2 и Z_3 в выходной сигнал Z_0 .

Структуру числа X можно представить следующим образом:

$$X = X_{\alpha+\beta}X_{\alpha+\beta-1} \dots X_2X_1 \equiv 2^{-\alpha}M_{\alpha}N_{\beta},$$

где M_{α} – группа старших двоичных разрядов длины α , а N_{β} – группа младших двоичных разрядов длины β управляющего слова (числа) X . При этом желательно соблюсти условие $\alpha \geq \beta$: пусть для десятичной системы счисления $X = 98765432$; тогда разбить его можно по-разному, лучше всего на числа примерно равной длины ($M_{\alpha} = 9876$, $N_{\beta} = 5432$) при этом $\alpha = \beta = 4$.

При представлении управляющего слова в десятичной системе счисления составляющие формулы (1) изменятся:

$$Z = gUX = gU(a_{n-1}10^{n-1} + a_{n-2}10^{n-2} + \dots + a_010^0), \quad (3)$$

где $q_n \in (a_0, a_1, \dots, a_{n-1}) \equiv [0, 1, 2, \dots, 9]$ – коэффициенты десятичных разрядов, принимающие значения от «0» до «9».

Выражение (2) с учетом (3) примет вид:

$$U_2 = U_1(1 \pm 10^{-\alpha}) = U_1(1 \pm 10^{-4}), \quad (4)$$

где точность соотношения должна быть не менее $10^{-\psi}$, т. е. 10^{-8} от полной шкалы или 100 нВ при $U_1 = 10$ В, а точность цифроаналогового преобразования составит 8 десятичных разрядов при 4-х разрядных десятичных парциальных ЦАП.

Теория работы

Практически число N_{β} есть результат умножения числа $2^{-\alpha}N_{\beta}$ в 2^{α} раз, (сдвинутое на α разрядов влево), т. е. умноженное цифровым способом в 2^{α} раз. Тогда можно записать значения парциальных сигналов Z в виде следующих выражений:

$$Z_1 = U_1M = U_1 \sum_{n=1}^{\alpha} q_{\alpha} 2^{-\alpha};$$

$$Z_2 = U_1N = U_1 \sum_{n=1}^{\beta} q_{\beta} 2^{-\beta};$$

$$Z_3 = U_2N = U_1(1 \pm 2^{-\alpha}) =$$

$$= U_1 \left(\sum_{n=1}^{\beta} q_{\beta} 2^{-\beta} \pm 2^{-\alpha} \sum_{n=1}^{\beta} q_{\beta} 2^{-\beta} \right);$$

$$Z_0 = Z_1 \mp Z_2 \pm Z_3 = U_1M \mp U_1N \pm U_2N =$$

$$= U_1 \left[\sum_{n=1}^{\alpha} q_{\alpha} 2^{-\alpha} \mp \sum_{n=1}^{\beta} q_{\beta} 2^{-\beta} \right] \pm$$

$$\pm U_1(1 \pm 2^{-\alpha}) \left[\sum_{n=1}^{\beta} q_{\beta} 2^{-\beta} \right] =$$

$$= U_1 \left\{ \sum_{n=1}^{\alpha} q_{\alpha} 2^{-\alpha} \mp \sum_{n=1}^{\beta} q_{\beta} 2^{-\beta} \pm \sum_{n=1}^{\beta} q_{\beta} 2^{-\beta} \right.$$

$$\left. + 2^{-\alpha} \sum_{n=1}^{\beta} q_{\beta} 2^{-\beta} \right\}.$$

При этом необходимо и достаточно соблюдение двух условий.

Во-первых, при алгебраическом суммировании сигналов Z_1 , Z_2 и Z_3 в выходном аналоговом сумматоре сигнал Z_3 должен вычитаться из сигнала Z_2 , для случая второго опорного сигнала $U_2 = U_1(1 - 2^{-\alpha})$ в соответствии с выражением:

$$Z_0 = Z_1 + Z_2 - Z_3 =$$

$$= U_1 \left[\sum_{n=1}^{\alpha} q_{\alpha} 2^{-\alpha} + \sum_{n=1}^{\beta} q_{\beta} 2^{-\beta} \right] -$$

$$- U_1(1 - 2^{-\alpha}) \left[\sum_{n=1}^{\beta} q_{\beta} 2^{-\beta} \right] =$$

$$= U_1 \left\{ \sum_{n=1}^{\alpha} q_{\alpha} 2^{-\alpha} + 2^{-\alpha} \sum_{n=1}^{\beta} q_{\beta} 2^{-\beta} \right\}.$$

Во-вторых, при алгебраическом суммировании сигналов Z_1 , Z_2 и Z_3 в выходном аналоговом сумматоре сигнал Z_2 должен вычитаться из сигнала Z_3 в соответствии с выражением:

$$Z_0 = Z_1 - Z_2 + Z_3 = U_1 \left[\sum_{n=1}^{\alpha} q_{\alpha} 2^{-\alpha} - \sum_{n=1}^{\beta} q_{\beta} 2^{-\beta} \right] +$$

$$+ U_1(1 + 2^{-\alpha}) \left[\sum_{n=1}^{\beta} q_{\beta} 2^{-\beta} \right] =$$

$$= U_1 \left\{ \sum_{n=1}^{\alpha} q_{\alpha} 2^{-\alpha} + 2^{-\alpha} \sum_{n=1}^{\beta} q_{\beta} 2^{-\beta} \right\},$$

но при этом $U_2 = U_1(1 + 2^{-\alpha})$.

Как следует из приведенных формул, данные преобразования приводят к одинаковому результату, что обеспечивает требуемую точность преобразования цифрового кода управления X в аналоговую величину Z_0 :

$$Z_0 = U_1 \left\{ \sum_{n=1}^{\alpha} q_{\alpha} 2^{-\alpha} + 2^{-\alpha} \sum_{n=1}^{\beta} q_{\beta} 2^{-\beta} \right\}.$$

Структура одного из возможных вариантов построения нониусного ЦАП, реализующего предлагаемый способ [11, 12], приведена на рисунке 9, где цифровой умножитель, в котором младшие β $2^{-\alpha}N_{\beta}$ разрядов управляющего слова $2^{-\alpha}N_{\beta}$ подвергаются цифровому умножению в 2^{α} раз (сдвиг влево на α разрядов); а также имеются источник опорного сигнала U_1 и источник опорного сигнала U_2 , парциальный ЦАП_1, аналоговый сумматор, парциальные ЦАП_2 и ЦАП_3; при этом опорные сигналы U_1 и U_2 связаны соотношением (2), и на выходе аналогового сумматора имеем выходной сигнал Z_0 .

В качестве числового примера рассмотрим вариант нониусного преобразования двухразрядного десятичного числа $X_{(10)} \leftrightarrow (M_{\alpha} + 10^{-1} N_{\beta})_{(10)}$ в абстрактный аналоговый параметр Z . Поскольку система счисления десятичная, число $X_{(10)}$ разбиваем на два одноразрядных: M_{α} и N_{β} , а опорные сигналы U в соответствии с (3) примут вид:

$$U_1 = 1,0; U_2 = 1,1 \times U_1.$$

Характерных вариантов будет два: для случая $5 \leq M_1 < 10$ и для случая $0 \leq M_1 < 5$. Пусть $M_1 = 7$ и $M_2 = 3$, а N принимает значения от 0 до 9. Сведем этапы вычислений и преобразований в таблицу 1.

ТАБЛИЦА 1. Нониус с суммированием

TABLE 1. Vernier with Summation

M_1	N	$Z_1 = U_1 \times M$	$Z_2 = U_1 \times N$	$Z_3 = U_2 \times N$	$Z_0 = Z_1 - Z_2 + Z_3$
$M_1 = 7; 5 \leq M_1 < 10$					
7	0	7	0	0,0	7,0
7	1	7	1	1,1	7,1
7	2	7	2	2,2	7,2
7	3	7	3	3,3	7,3
7	4	7	4	4,4	7,4
7	5	7	5	5,5	7,5
7	6	7	6	6,6	7,6
7	7	7	7	7,7	7,7
7	8	7	8	8,8	7,8
7	9	7	9	9,9	7,9
$M_1 = 3; 0 \leq M_1 < 5$					
3	0	3	0	0,0	3,0
3	1	3	1	1,1	3,1
3	2	3	2	2,2	3,2

M_1	N	$Z_1 = U_1 \times M$	$Z_2 = U_1 \times N$	$Z_3 = U_2 \times N$	$Z_0 = Z_1 - Z_2 + Z_3$
3	3	3	3	3,3	3,3
3	4	3	4	4,4	3,4
3	5	3	5	5,5	3,5
3	6	3	6	6,6	3,6
3	7	3	7	7,7	3,7
3	8	3	8	8,8	3,8
3	9	3	9	9,9	3,9

Единственными узлами нониусного ЦАП (см. рисунок 9), к которым следует предъявить жесткие требования по обеспечению точности, являются выходной аналоговый сумматор и источники опорных сигналов U_1 и U_2 , причем трудности по их реализации не являются непреодолимыми, более того, стандартными для (высоко) прецизионной электроники.

Так, например, при двадцатиразрядном входном двоичном слове ($\alpha = \beta = 10$) и $U_1 = 10$ В, требуемая относительная точность аналогового сумматора и источника $U_1 \delta_a \leq 2^{-20} \approx 10^{-6}$ (абсолютная точность – 9,5 мкВ), что вполне реализуемо на современной микросхемной элементной базе. В случае, если $U_2 = 0,9 \times U_1$ и при тех же числовых параметрах, получим $Y_2 = 0,9 \times Y_1$, $U_2 = 0,9 \times U_1$ (таблица 2).

ТАБЛИЦА 2. Нониус с вычитанием

TABLE 2. Vernier with Subtraction

M	N	$Z_1 = Y_1 \times M$	$Z_2 = Y_1 \times N$	$Z_3 = Y_2 \times N$	$Z_0 = Z_1 + Z_2 - Z_3$
$M_1 = 7; 5 \leq M_1 < 10$					
7	0	7	0	0,0	7,0
7	1	7	1	0,9	7,1
7	2	7	2	1,8	7,2
7	3	7	3	2,7	7,3
7	4	7	4	3,6	7,4
7	5	7	5	4,5	7,5
7	6	7	6	5,4	7,6
7	7	7	7	6,3	7,7
7	8	7	8	7,2	7,8
7	9	7	9	8,1	7,9
$M_1 = 3; 0 \leq M_1 < 5$					
3	0	3	0	0,0	3,0
3	1	3	1	0,9	3,1
3	2	3	2	1,8	3,2
3	3	3	3	2,7	3,3
3	4	3	4	3,6	3,4
3	5	3	5	4,5	3,5
3	6	3	6	5,4	3,6
3	7	3	7	6,3	3,7
3	8	3	8	7,2	3,8
3	9	3	9	8,1	3,9

В предлагаемом способе преобразования при любом соотношении чисел N и M не происходит потери разрядов при *росте точности* цифроаналогового преобразования в $\approx a^a$ раз, т. к. происходит значительное уменьшение погрешности, т. е. качественное увеличение точности преобразования параллельного цифроаналогового преобразователя преобразования без ужесточения требований к технологии изготовления элементов парциальных ЦАП этого преобразователя.

Моделирование

На рисунке 10 приведены: первый вариант 8-битовой модели такого ЦАП и второй вариант модели, а на рисунке 11 – временные диаграммы работы при точном сопряжении опорных шкал и запуске от

8-битового счетчика импульсов. Величина ступеньки выходного напряжения 30 мВ при величине ступенек парциальных ЦАП равна 480 и 512 мВ.

Заметим, что требования к высокой точности предъявляются только к источникам опорного сигнала – напряжений U_1 и U_2 и к соблюдению соотношения (2) на *постоянном токе* и в *двух точках шкалы* – в начале и в конце. Итоговая точность преобразования в предлагаемом способе равна Ψ бит или 2^Ψ единиц полной шкалы при точности преобразования используемых ЦАП $\Psi/2$ бит. На рисунке 12 приведена упрощенная 8-битовая модель такого ЦАП с ошибкой воспроизведения шкал, а на рисунке 13 – временные диаграммы его работы при неточной настройке шкал с ошибкой и запуске от 8-битового счетчика импульсов.

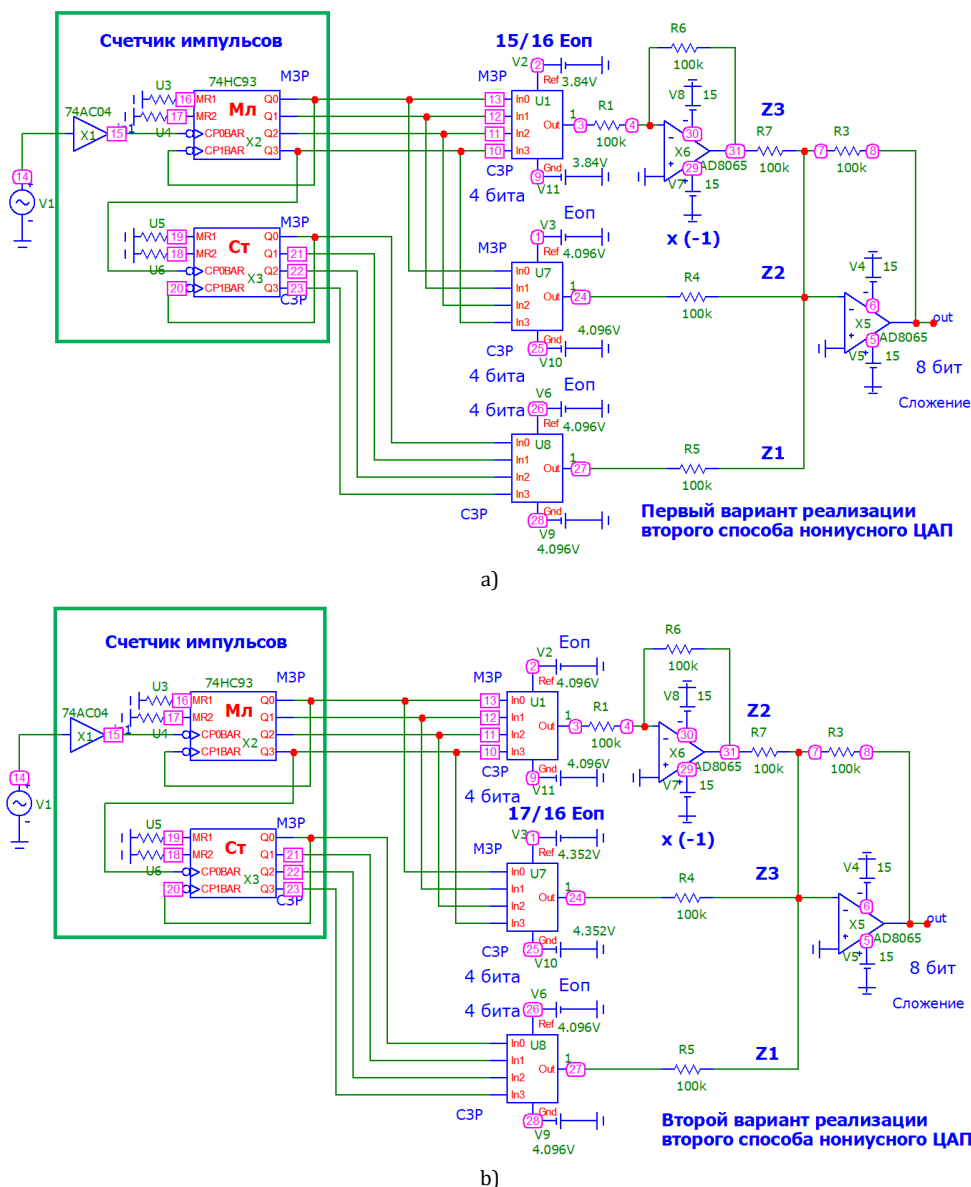


Рис. 10. Модель нониусного ЦАП второго типа с точным сопряжением опорных шкал: первый (а) и второй (б) вариант

Fig. 10. Model of a Second Type Vernier DAC with Precise Matching of Reference Scales: First (a) and Second (b) Variants

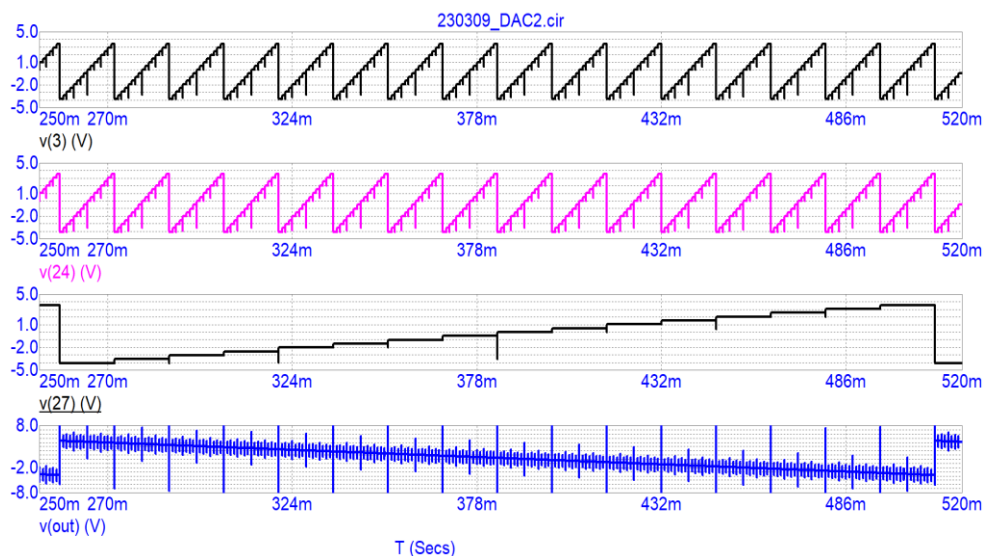


Рис. 11. Временные диаграммы работы нониусного ЦАП второго типа с точным сопряжением опорных шкал
Fig. 11. Timing Diagrams of the Operation of a Vernier DAC of the Second Type with Precise Matching of Reference Scales

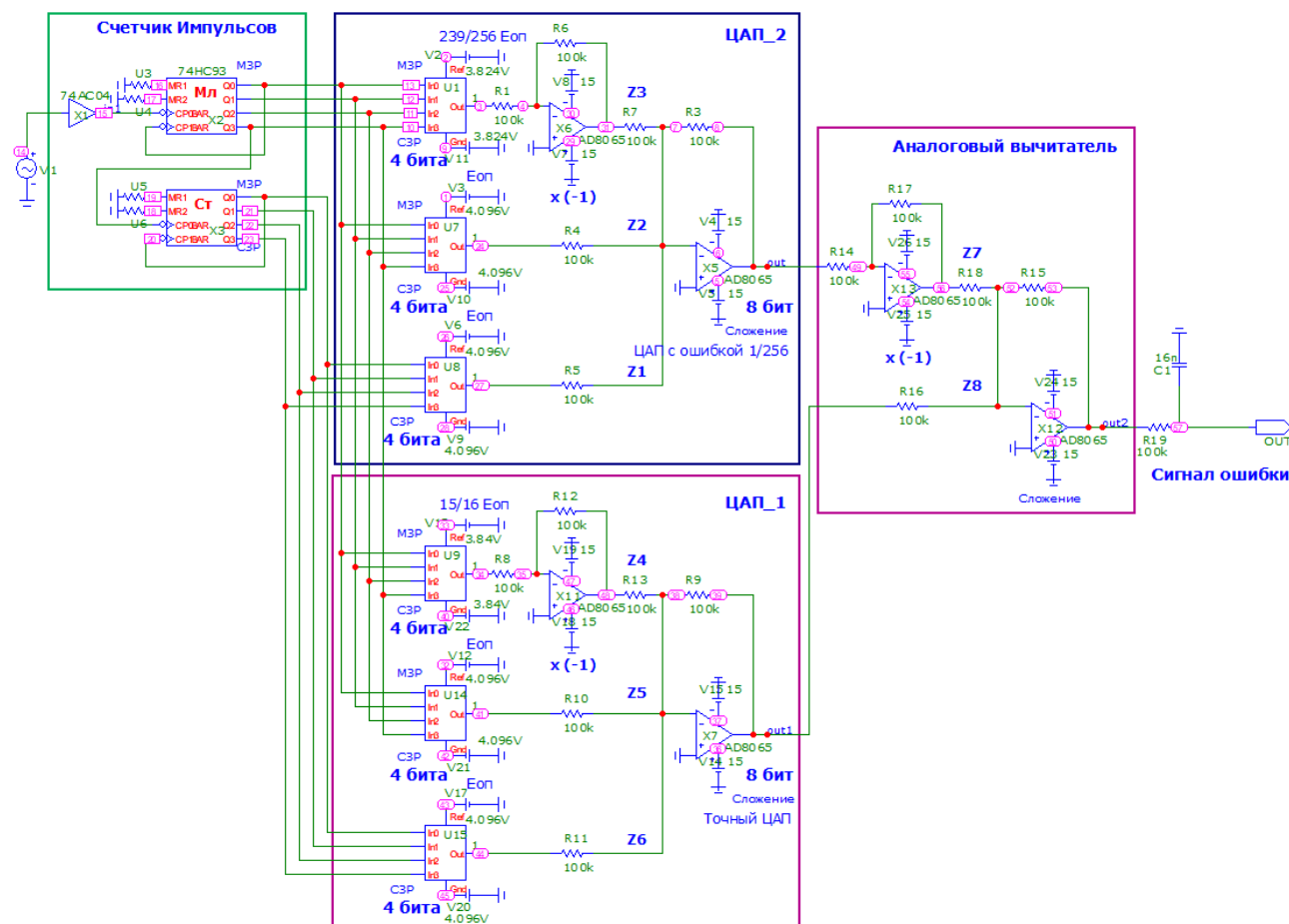


Рис. 12. Модель нониусного ЦАП второго типа с неточной настройкой
Fig. 12. Model of a Second Type Vernier DAC with Imprecise Tuning

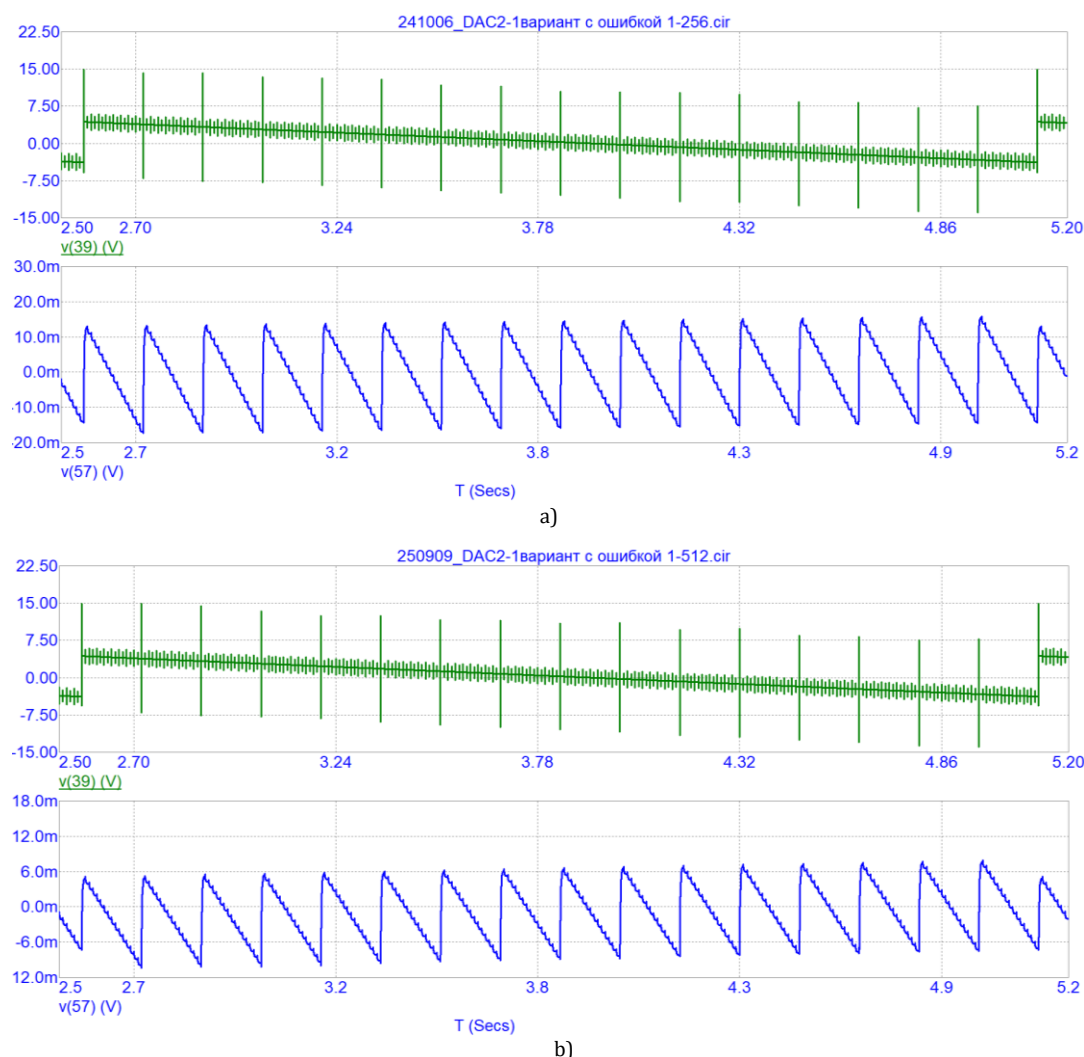


Рис. 13. Временные диаграммы его работы с ошибкой сопряжения шкал 1 (а) и 0,5 МЗР (b): 1/256 и 1/512 полной шкалы

Fig. 13. Timing Diagrams of its Operation with a Scale Conjugation Error of 1 (a) and 0.5 LSB (b): 1/256 and 1/512 of the Full Scale

Выводы

Анализ результатов показывает линейную зависимость воспроизведения выходного напряжения от ошибки сопряжения в (2). При производстве нониусных ЦАП потребуются АЦП той же точности, что и у «конечного» ЦАП. В настоящее время выпускаются микросхемы сигма-дельта АЦП с разрядностью 24 и 32 бит (www.analog.com/media/en/technical-documentation/data-sheets/ad4630-24.pdf; www.analog.com/media/en/technical-documentation/data-sheets/AD7771.pdf; www.analog.com/media/en/technical-documentation/data-sheets/250032fb.pdf).

Имея пару опорных сигналов, жестко связанных соотношением (2), можно получить новое *качество* преобразования при сохранении технологических ограничений производства на точность входящих в структуру (см. рисунок 9) аналоговых узлов.

Рассматриваемый способ нониусного ЦАП позволяет, при необходимости, увеличить быстродействие ЦАП, уменьшив вдвое разрядность исходных элементов (т. е. в $2^{\Psi/2}$ раз) [13, 14].

Список источников

1. Титце У., Шенк К. Полупроводниковая схемотехника: справочник. Пер. с нем. М.: Мир, 1982. 512 с.
2. Потоцкий А.П., Сафьянников Н.М., Смолов В.Б., Угрюмов Е.П. Преобразователь код-аналог. Патент на изобретение СССР No 1508347 A1 от 08.12.1986. Оpubл. 20.03.1999.
3. Смолов В.Б., Угрюмов Е.П., Герасимов И.В., Рачев Б.Д., Фархи О.А. Элемент с цифро-управляемой проводимостью. Патент на изобретение СССР No 1182543 A1 от 26.04.1984. Оpubл. 30.09.1985.
4. Федорков Б.Г., Телец В.П. Микросхемы ЦАП и АЦП: функционирование, параметры, применение. М.: Энергомашиздат, 1990. 320 с.

5. Смоллов В.Б. Функциональные преобразователи информации. Л.: ЭНЕРГОИЗДАТ, 1981. 247 с.
6. Марцинкявичус А.-Й. Быстродействующие интегральные микросхемы ЦАП и АЦП и измерение их параметров. М.: Радио и связь, 1988. 224 с.
7. Смоллов В.Б. Микроэлектронные цифро-аналоговые и аналого-цифровые преобразователи информации. Л.: Энергия, 1976. 336 с.
8. Никитин Ю.А. Моделирование параллельного нониусного цифроаналогового преобразователя первого типа // Труды учебных заведений связи. 2025. Т. 11. № 4. С. 7–16. DOI:10.31854/1813-324X-2025-11-4-7-16. EDN:BAUBXX
9. Шелехнев А. Прорыв в технологии малосигнальных транзисторов // Компоненты и технологии. 2003. № 6(32). С. 52–55. EDN:LMGIXK
10. Никитин Ю.А. Способ нониусного цифроаналогового преобразования. Патент на изобретение RU 2703228 C1 от 18.03.2019. Оpubл. 15.10.2019. EDN: VEPXBS
11. Никитин Ю.А. Способ нониусного цифроаналогового преобразования. Патент на изобретение RU 2726911 C1 от 02.07.2019. Оpubл. 16.07.2020. EDN:NXAXDF
12. Nikitin Y.A. Method of vernier digital-to-analog conversion. Patent USA, no. 11.736.117 B2. 10.11.2022.
13. Никитин Ю.А. Построение многозарядного параллельного цифроаналогового преобразователя // IX Международная научно-техническая и научно-методическая конференция «Актуальные проблемы инфотелекоммуникаций в науке и образовании» (Санкт-Петербург, Российская Федерация, 26–27 февраля 2020 г.). СПб.: СПбГУТ, 2020. Т. 3. С. 476–480. EDN:WOEOEQ
14. Никитин Ю.А. Теория цифроаналогового синтеза частот с помощью конечных автоматов. СПб.: Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича, 2024. 342 с.

References

1. Tiietze U., Schenk Ch. *Semiconductor circuit engineering*. Translated from German. Moscow: Mir Publ.; 1982. 512 p. (in Russ.)
2. Potocki A.P., Safyannikov N.M., Smolov V.B., Ugryumov E.P. *Code-to-analog converter*. Patent USSR, no. 1508347 A1, 20.03.1999. (in Russ.)
3. Smolov V.B., Ugryumov E.P., Gerasimov I.V., Rachev B.D., Farhi O.A. *Element with digitally controlled conductivity*. Patent USSR, no. 1182543 A1, 30.09.1985. (in Russ.)
4. Fedorkov B.G., Telets V.P. *DAC and ADC microcircuits: operation, parameters, application*. Moscow: Energomashizdat Publ.; 1990. 320 p. (in Russ.)
5. Smolov V.B. *Functional information converters*. Leningrad: Energoizdat Publ.; 1981. 247 p. (in Russ.)
6. Marcinkevicius A.-J. *High-speed integrated circuits DAC and ADC and measurement of their parameters*. Moscow: Radio i svyaz Publ.; 1988. 224 p. (in Russ.)
7. Smolov V.B. *Microelectronic digital-to-analog and analog-to-digital information converters*. Leningrad: Energiya Publ.; 1976. 336 p. (in Russ.)
8. Nikitin Yu.A. Simulation of the First Type Parallel Vernier Digital-to-Analog Converter. *Proceedings of Telecommunication Universities*. 2025;11(4):7–16. (in Russ.) DOI:10.31854/1813-324X-2025-11-4-7-16. EDN:BAUBXX
9. Shelekhnev A. Breakthrough in Small-Signal Transistor Technology. *Components and Technologies*. 2003;6(32):52–55. (in Russ.) EDN:LMGIXK
10. Nikitin Yu. A. *Vernier digital-to-analog conversion method*. Patent RF, no. 2703228, 15.10.2019. (in Russ.)
11. Nikitin Yu. A. *Method of vernier digital-to-analog conversion*. Patent RF, no. 2726911, 16.07.2020. (in Russ.)
12. Nikitin Y.A. *Method of vernier digital-to-analog conversion*. Patent USA, no. 11.736.117 B2, 10.11.2022.
13. Nikitin Yu. Construction of a multi-bit parallel digital-to-analog converter. *Proceedings of the IXth International Scientific, Technical and Scientific-Methodical Conference on Actual Problems of Infotelecommunications in Science and Education, 6–27 February 2020, St. Petersburg, Russian Federation, vol.3*. St. Petersburg: SPbSUT Publ.; 2020. p.476–480. (in Russ.) EDN:WOEOEQ
14. Nikitin Yu.A. *Theory of digital-to-analog frequency synthesis using finite automata*. St. Petersburg: The Bonch-Bruevich Saint Petersburg State University of Telecommunications Publ.; 2024. 342 p. (in Russ.).

Статья поступила в редакцию 18.11.2025; одобрена после рецензирования 20.01.2026; принята к публикации 22.01.2026

The article was submitted 18.11.2025; approved after reviewing 20.01.2026; accepted for publication 22.01.2026

Информация об авторах:

НИКИТИН Юрий Александрович	кандидат технических наук, доцент, доцент кафедры электроники Санкт-Петербургского государственного университета телекоммуникаций им. проф. М.А. Бонч-Бруевича  https://orcid.org/0000-0002-0749-9751
ФИЛИН Владимир Алексеевич	доктор технических наук, профессор кафедры электроники Санкт-Петербургского государственного университета телекоммуникаций им. проф. М.А. Бонч-Бруевича  https://orcid.org/0009-0006-5964-0194

Авторы сообщают об отсутствии конфликтов интересов.

The authors declare no conflicts of interests.

Обзорная статья

УДК 004.725.5

<https://doi.org/10.31854/1813-324X-2026-12-2-64-72>

EDN:ARIEBU



Анализ проблем проектирования сетей Wi-Fi в свете перехода к восьмому поколению технологии

✉ Светлана Алексеевна Скоробогатова, skorobogatova.sa@sut.ru

✉ Антон Сергеевич Викулов, vikulov.as@sut.ru

✉ Александр Иванович Парамонов, paramonov@sut.ru

Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича,
Санкт-Петербург, 193232, Российская Федерация

Аннотация

Актуальность работы обуславливается широким распространением сетей стандарта IEEE 802.11. В статье рассматриваются ключевые нововведения, связанные с переходом от седьмого поколения стандарта Wi-Fi IEEE 802.11be, известного как Wi-Fi 7, к восьмому поколению стандарта IEEE 802.11bn, получившему название Wi-Fi 8. В частности, акцент сделан на улучшениях пропускной способности, внедрении более совершенных методов модуляции и поддержке многоканальных передач, что позволяет значительно повысить скорость и стабильность соединений. Восьмое поколение стандарта должно привести принципиально новые подходы к организации взаимодействия между клиентским устройством и точкой доступа. В отличие от 802.11be, в 802.11bn предполагается реализация возможности одновременной работы одного клиента с несколькими точками доступа, и наоборот, что принципиально изменит большинство существующих механизмов управления радиосредой, но то же время обеспечит недостижимую ранее потенциальную непрерывность работы в рамках распределенной в пространстве беспроводной инфраструктуры. Рассматриваются также актуальные задачи, стоящие перед индустрией, включая проблему совместимости с предыдущими поколениями стандартов и сложности, возникающие при проектировании беспроводных локальных вычислительных сетей.

Целью данной работы является систематизированный обзор основных особенностей последнего дополнения стандарта IEEE 802.11be, известного как Wi-Fi 7, и разрабатываемого в настоящий момент следующего поколения IEEE 802.11bn, получившего название Wi-Fi 8.

Методы: анализ профессиональной и научной литературы, текстов стандартов и практики их использования в отрасли.

В ходе **решения** поставленной задачи в работе представлен обзор основных улучшений, ожидаемых в стандарте IEEE 802.11bn. Особое внимание уделяется изменению подходов к проектированию беспроводных локальных вычислительных сетей в контексте новых технологий, таких как Distributed Multi-Link Operation и Multi-AP Coordination.

Теоретическая значимость состоит в анализе основных аспектов нововведений IEEE 802.11be и IEEE 802.11bn. Обозначены научные проблемы, которые требуют переработки и усложнения существующих моделей и методов. К ним относятся: динамическое распределение ресурсов в многодиапазонных сетях, оптимизация использования спектрального ресурса, развитие алгоритмов управления радиоресурсами, развитие алгоритмов обеспечения хендвера.

Практическая значимость: полученный результат может быть использован при постановке задач проектирования сетей Wi-Fi седьмого и восьмого поколений.

Ключевые слова: Wi-Fi 7, Wi-Fi 8, проектирование сетей Wi-Fi

Ссылка для цитирования: Скоробогатова С.А., Викулов А.С., Парамонов А.И. Анализ проблем проектирования сетей Wi-Fi в свете перехода к восьмому поколению технологии // Труды учебных заведений связи. 2026. Т. 12. № 2. С. 64–72. DOI:10.31854/1813-324X-2026-12-2-64-72. EDN:ARIEBU

Review research

<https://doi.org/10.31854/1813-324X-2026-12-2-64-72>

EDN:ARIEBU

Analysis of the Wi-Fi Network Design Problems Considering Transition to Wi-Fi 8

✉ **Svetlana A. Skorobogatova**, skorobogatova.sa@sut.ru

✉ **Anton S. Vikulov**, vikulov.as@sut.ru

✉ **Aleksandr I. Paramonov**, paramonov@sut.ru

The Bonch-Bruевич Saint Petersburg State University of Telecommunications,
St. Petersburg, 193232, Russian Federation

Annotation

The relevance of this work is driven by the widespread adoption of IEEE 802.11 networks. This article examines the key innovations associated with the transition from the seventh-generation IEEE 802.11be Wi-Fi standard, known as Wi-Fi 7, to the eighth-generation IEEE 802.11bn standard, known as Wi-Fi 8. Specifically, attention is paid to the throughput improvements, the implementation of more advanced modulation methods, and support for multi-channel transmissions, which significantly improves the speed and stability of connections. The eighth-generation standard is expected to introduce fundamentally new approaches to organizing client-access point interactions. Unlike 802.11be, 802.11bn is expected to enable simultaneous operation of a single client with multiple access points and vice versa. This will fundamentally change most of existing radio environment management mechanisms, but at the same time, it will provide previously unreachable mobility within a distributed wireless infrastructure. Current challenges facing the industry are also considered, including compatibility with previous generations of standards and the complexities that arise during wireless local area networks design.

Purpose: A systematic review of the key features of the latest addition to the IEEE 802.11be standard, known as Wi-Fi 7, and the next generation of IEEE 802.11bn, known as Wi-Fi 8, currently under development.

Methods: Analysis of professional and scientific literature, standards texts, and industry practices.

Results. In addressing the stated objective, this paper presents an overview of the key improvements expected in the IEEE 802.11bn standard. Particular attention is paid to changing approaches to WLAN design in the context of new technologies such as Distributed Multi-Link Operation and Multi-AP Coordination.

Theoretical significance lies in the analysis of the key aspects of the innovations in IEEE 802.11be and IEEE 802.11bn. Research problems requiring solutions that require the revision and sophistication of existing models and methods are identified. These include: dynamic resource allocation in multi-band networks, optimization of spectrum resource use, development of radio resource management algorithms, and development of handover algorithms.

Practical significance: The obtained results can be used when formulating design problems for seventh- and eighth-generation Wi-Fi networks.

Keywords: Wi-Fi 7, Wi-Fi 8, Wi-Fi Networks Design

For citation: Skorobogatova S.A., Vikulov A.S., Paramonov A.I. Analysis of Wi-Fi Network Design Problems Considering Transition to Wi-Fi 8 transition. *Proceedings of Telecommunication Universities*. 2026;12(2):64–72. (in Russ.) DOI:10.31854/1813-324X-2026-12-2-64-72. EDN:ARIEBU

Введение

Современные приложения, высокоскоростные вычисления, облачное хранилище, видеозвонки и потоковая передача в формате 4K, а также услуги телеприсутствия и интерактивное взаимодействие с метавселенной [1] требуют беспроводного доступа, обеспечивающего не только высокую пропускную способность, но и стабильное соединение, низкую задержку и возможность работы с большим

числом подключенных устройств одновременно. В инженерной литературе можно встретить понятие «емкости» беспроводной сети. Под ним в широком смысле понимается число пользовательских устройств, работу которых сетевая инфраструктура может обеспечить при заданных критериях качества обслуживания. Для удовлетворения этих растущих требований происходит непрерывное развитие стандарта IEEE.

Ранее авторами был проведен анализ эволюционного развития стандарта IEEE 802.11 [2] до появления 802.11be (Wi-Fi 7).

В данной статье рассматриваются ключевые нововведения, внедренные в 802.11be, а также нерешенные проблемы, которые требуют внимания в контексте будущих поколений стандарта. Основной акцент сделан на анализе дополнения стандарта 802.11bn, которое сейчас находится на ранней стадии разработки. Необходимо отметить, что IEEE 802.11be в целом ориентирован на увеличение скорости и производительности сети, тогда как IEEE 802.11bn (режим обозначается как UHR, *аббр. от англ. Ultra High Reliability*) нацелен на повышение надежности и сокращение задержек, при этом сохраняя все улучшения, предоставленные IEEE 802.11be. Поэтому в данной работе уделим внимание текущему состоянию IEEE 802.11be, ожидаемым нововведениям в IEEE 802.11bn, а также тому, как они могут повлиять на будущее технологии беспроводных локальных сетей и подходы к их проектированию. Последнее представляется особенно важным, т. к. есть основания полагать, что нововведения IEEE 802.11bn будут несомненно революционными.

Статья посвящена ожиданиям отрасли от последующих версий стандарта Wi-Fi, а также основным научным, инженерным и технологическим задачам, с которыми предстоит столкнуться сетям данного стандарта в течение ближайших 4–5 лет.

1. Анализ основных нововведений IEEE 802.11be (EHT)

Работа над дополнением IEEE 802.11be началась в мае 2019 г. незадолго до принятия дополнения IEEE 802.11ax. Как следует из названия режима EHT, основной целью было еще большее увеличение пропускной способности и производительности. Изначально запланированные даты принятия дополнения неоднократно переносились, а окончательно стандарт был утвержден в июле 2025 г. Тем не менее, как это уже имело место с ранее выходящими дополнениями, например IEEE 802.11ac и IEEE 802.11ax, многие поставщики предлагают продукты и решения с поддержкой еще не вышедшего стандарта. Чаще всего такие реализации основаны на draft-версиях стандарта и в последующем дорабатываются программно. Далее будут представлены ключевые изменения, предложенные в стандарте IEEE 802.11be.

1.1. Увеличение ширины канала до 320 МГц в диапазоне 6 ГГц

Расширение возможной ширины канала до 320 МГц (т. е. канал типа EHT320) представляет собой дальнейшее развитие предыдущих версий стандартов IEEE 802.11ac/ax, где максимальная ширина канала составляла 160 МГц (тип канала HE160) в диапазоне 5 ГГц (UNII-1,2,3). Эта функция доступна в поддиапазонах UNII-5,6,7,8 и значительно повышает теоретически достижимую скорость передачи данных. В то же время число таких каналов в частотном плане не может превышать трех, что говорит о крайней ограниченности возможностей таких режимов для построения распределенных сетей данного стандарта. То же самое можно было ранее сказать и о каналах EHT160 в поддиапазонах UNII-1,2,3.

1.2. Режим 4096-QAM

Дополнение IEEE 802.11be вводит использование модуляции 4096-QAM, которая увеличивает скорость передачи данных на 20 % по сравнению с максимальной для IEEE 802.11ax модуляцией 1024-QAM, передавая 12 бит на символ в расчете на каждую поднесущую. Таким образом, обновленная таблица режимов модуляции и кодирования (MCS) для канала шириной 20 МГц при одном пространственно-временном потоке и защитном интервале 800 нс содержит максимальную канальную скорость передачи – 172,1 Мбит/с. Отметим, что наиболее скоростные режимы, с одной стороны, часто являются опциональными, а с другой – самыми энергозатратными, по причине больших требуемых для их работы вычислительных ресурсов. Поэтому на практике даже при выполнении всех условий, необходимых для задействования наиболее скоростных MCS, они используются не постоянно. Тем не менее, ретроспективно оглядываясь на процесс развития стандарта, хорошо видно, что используемые на практике скорости передачи на принятых в стандарте MCS устойчиво растут.

1.3. Multi-Link Operation (MLO)

В технологии MLO технологии MLO определено несколько режимов работ, каждый из которых имеет свои особенности¹.

Multi-Link Single Radio (MLSR)

В режиме MLSR используется только один радиомодуль. Устройство может прослушивать несколько каналов, но в каждый момент времени активен только один из них, то есть одновременная передача и прием не поддерживаются.

¹ Trivedi S., Doshi N. Wi-Fi 7's Multi-Link Operation (MLO) // CISCO. 2025. URL: <https://blogs.cisco.com/networking/wi-fi-7s-multi-link-operation-mlo-dissection-from-packets-to-performance> (Accessed 13 May 2025)

Enhanced Multi-Link Single Radio (EMLSR)

Улучшенная версия MLSR, которая позволяет динамически переключаться между каналами, снижая энергопотребление и задержки. EMLSR позволяет устройствам одновременно прослушивать несколько каналов, но передача или прием данных возможны только по одному каналу. Также EMLSR позволяет динамически переключаться между диапазонами для улучшения энергоэффективности и оптимизации использования ресурсов. EMLSR может использовать сканирующие радиомодули в дополнение к основным, чтобы поддерживать работу на разных каналах (например, на 6 ГГц и 5 ГГц) [3].

Simultaneous Transmit/Receive (STR)

В режиме STR устройства могут одновременно передавать и принимать данные на нескольких каналах. Это значительно повышает пропускную способность и снижает задержки, так как устройство может работать с несколькими каналами одновременно.

Non-Simultaneous Transmit/Receive (NSTR)

В данном режиме передача и прием данных происходят с синхронизацией, что предотвращает коллизии и помехи между каналами. Синхронизация передачи и приема данных помогает избежать проблем с помехами и снижает риски для стабильности связи. Такой режим полезен в условиях высоконагруженных сетей, где важно поддержание устойчивого соединения.

Enhanced Multi-Link Multi-Radio (EMLMR)

EMLMR требует наличия более двух радиомодулей и представляет собой усовершенствованную версию режима STR, позволяя динамически адаптироваться к доступным частотам. Это особенно полезно в условиях изменяющейся загрузки сети и изменяющейся картины помех в спектре.

Как показано в [4], режим STR демонстрирует более высокую пропускную способность и меньшие задержки по сравнению с EMLSR, что делает его предпочтительным для высоконагруженных сетей. В то же время EMLSR обеспечивает баланс между производительностью и энергопотреблением, что важно для мобильных устройств. При этом выбор между асинхронной и синхронной передачей в MLO представляет собой компромисс: асинхронный режим обеспечивает лучшую утилизацию многоканального потенциала, тогда как синхронный – большую стабильность соединения за счет координации доступа к каналу.

1.4. Multi-Resource Units (MRU)

В дополнении стандарта IEEE 802.11ax была представлена технология OFDMA, которая позволила делить спектральный ресурс канала на группы поднесущих и назначать их в виде ресурс-

ных блоков для одновременной передачи данных нескольким устройствам. В IEEE 802.11be эта концепция была улучшена с помощью механизма Multi-Resource Units (MRU). MRU позволяют назначать несколько ресурсных блоков одному устройству, повышая гибкость управления спектром и увеличивая спектральную эффективность. При этом ресурсные блоки могут быть разного размера и несмежными в спектре канала. Это особенно ценно в условиях высокой плотности устройств, так как в широком смысле улучшается эффективность использования спектра и конечные характеристики качества соединения.

1.5. Restricted Target Wake Time (R-TWT)

Энергосберегающая технология TWT (*аббр. от англ. Target Wake Time*), впервые представленная в 802.11ax, была дополнительно улучшена в 802.11be. TWT позволяет устройствам согласовывать периоды активности и энергосбережения (сна), сокращая потребление энергии – что особенно важно для устройств, относящихся к Интернету вещей, и мобильных устройств. R-TWT представляет собой дальнейшее усовершенствование подхода TWT, обеспечивая более точный контроль задержек и резервирование сетевых ресурсов на конкретные временные периоды. Это нацелено как на улучшение управления энергопотреблением мобильного устройства, так и на рост производительности.

2. Нерешенные проблемы в IEEE 802.11be

Несмотря на значительные улучшения, представленные дополнением IEEE 802.11be (режимы работы ЕНТ), остаются пока не полностью решенными задачи, которые требуют дальнейшего технологического развития. Некоторые из них изначально планировались для внедрения в ЕНТ, но их реализация была отложена до будущих дополнений стандарта. Так, например, в ранних публикациях на тему ЕНТ можно встретить упоминания о планах на реализацию распределенного MLO [5], а также о потенциале full-duplex [6].

Рассмотрим основные нерешенные вопросы.

1) Переключение клиентских устройств (КУ) между точками доступа (хендовер), работающими под управлением одного контроллера. Быстрое переключение КУ между точками доступа особенно важно, когда устройства работают в разных частотных каналах. Эффективная координация необходима для предотвращения прерывания обслуживания на прикладном уровне во время процесса хендовера. Различные механизмы кеширования РМК-ключа, дополнение IEEE 802.11g и ряд других нововведений облегчают хендовер, но проблема остается неполностью решенной как ввиду необходимости поддержки со стороны КУ, так и по причине

существенной зависимости от алгоритмов, реализованных их производителем.

2) Сложность управления канальным ресурсом. Добавление новых функций и протоколов существенно усложняет контроль доступа к среде передачи (MAC, *аббр. от англ. Media Access Control*), в основе которого лежит метод CSMA/CA, изначально реализованный в самой первой версии стандарта (функция DCF).

Современные КУ, поддерживающие режимы до ЕНТ включительно, могут работать в среде передачи с использованием самых разных комбинаций режимов, к числу которых могут относиться задействование следующих возможных режимов:

- DCA/EDCA/HCCA и их модификации;
- RTS/CTS; greenfield mode;
- MIMO;
- TxBF;
- MU-MIMO;
- OFDMA;
- 802.11w;
- BSS Coloring;
- MRU;
- первичных / вторичных каналов и многих других аспектов.

В случае смеси разнородных КУ в канале точки доступа (ТД), возникающее таким образом «наслоение» протокольных взаимодействий значительно усложняет управление временными и спектральными ресурсами канала. Для многих приложений высокоприоритетные кадры должны передаваться максимально быстро, что может противоречить общей логике алгоритма доступа к каналу.

3) Частичным следствием сложности механизмов доступа к среде является потенциально высокая задержка. Несмотря на то, что стандарт предлагает более высокие скорости передачи данных, задержка при этом остается значимой проблемой. Даже небольшие значения задержки могут негативно сказываться на приложениях, требующих работы в реальном времени, таких как видеоконференции, киберспорт, интернет-навыков [7] или AR / VR (дополненная и виртуальная реальность).

4) Обеспечение совместимости с предыдущими версиями. Совместимость устройств между различными стандартами остается важной проблемой. Поддержка устаревших устройств наряду с новыми технологиями усложняет интеграцию КУ в сеть и может стать барьером для развития инфраструктур.

Все вышеперечисленное подчеркивает необходимость дальнейших улучшений в таких областях, как процесс хендвера, снижение задержек, эффективное управление трафиком и обеспечение совместимости с предыдущими версиями, что, в свою очередь, окажет влияние на развитие беспроводных технологий в рамках стандарта в ближайшем будущем.

3. Краткий обзор IEEE 802.11bn

Стандарт IEEE 802.11bn (UHR) разрабатывается с целью повышения надежности и стабильности беспроводных соединений, что особенно важно для современных приложений. В данном контексте под надежностью понимается устойчивость к потерям пакетов, минимизация задержек, обеспечение резервирования каналов и избыточности. Ожидается, что стандарт обеспечит более высокий уровень надежности передачи данных, снизит задержки и повысит устойчивость к помехам – все это критически важно для приложений, где требуется работа в реальном времени без перебоев.

Рассмотрим ключевые изменения и технологии, запланированные для внедрения в IEEE 802.11bn.

3.1. Hybrid ARQ (HARQ)

Стандарт IEEE 802.11bn внедряет механизм гибридного ARQ (HARQ, *аббр. от англ. Hybrid Automatic Repeat Request*), относящегося к каналам с гибридной решающей обратной связью. Основное отличие HARQ от классического ARQ заключается в том, что при HARQ неправильно принятые блоки данных не отбрасываются, а сохраняются для последующей коррекции (рисунок 1).

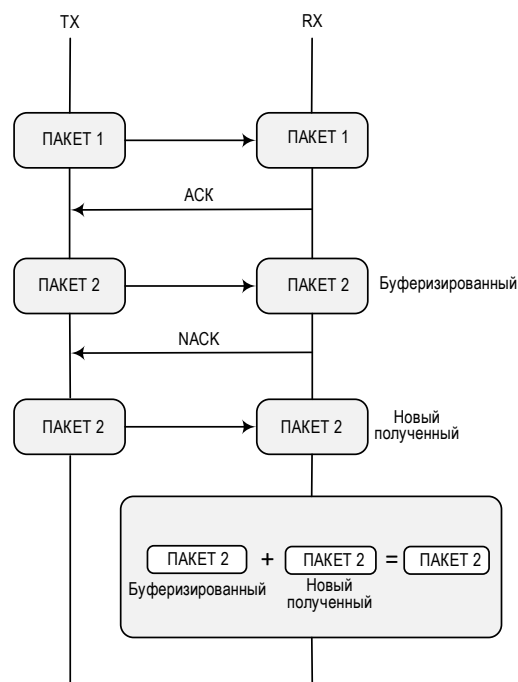


Рис. 1. Концепция HARQ

Fig. 1. HARQ Concept

Когда на принимающее устройство поступает повторно отправленный блок данных, он комбинируется с ранее полученными блоками, что позволяет восстановить данные с большей вероятностью [8]. Внедрение HARQ должно снизить PER (уровень ошибок пакетов), минимизировать риски снижения индекса используемой схемы модуляции

и кодирования (MCS) из-за повторных передач и, таким образом, обеспечить более высокую пропускную способность.

Отметим, что такой подход существенно повышает требования к аппаратным возможностям как КУ, так и точек доступа, одновременно влияя на их энергопотребление, что может иметь влияние на многие аспекты, возникающие как при проектировании беспроводных локальных вычислительных сетей (БЛВС), так и на длительность автономной работы.

3.2. Распределенный MLO (Distributed Multi-Link Operation)

Расширение архитектуры MLO с точки зрения КУ в распределенную структуру позволит устройствам MLD (*аббр. от англ. Multi-Link Device*) работать одновременно с несколькими физическими ТД. Это создает распределенные виртуальные ячейки, в которых мобильность устройств отрабатывается более эффективно. Устройство может одновременно подключаться к нескольким точкам доступа, расположенным в разных зонах, обеспечивая постоянную связь хотя бы по одному каналу [9]. Этот метод интегрирует поддержку быстрого хендовера КУ напрямую в стандарт IEEE 802.11bn, что значительно повышает надежность соединений и улучшает общее качество восприятия. В некотором смысле это можно обозначить, как «взгляд со стороны клиента» на совместную работу ТД.

Необходимо отметить, что для работы технологий MLO, координированного пространственного перераспределения и координированного формирования луча требуется тактовая синхронизация. Это важно, поскольку раньше такая функция была доступна только в сотовых сетях.

Процесс переключения КУ между точками доступа в стандарте IEEE 802.11bn потребует значительных изменений в подходах к управлению беспроводной средой. В конфигурации MLO КУ не всегда может одновременно переключаться между точками доступа на всех доступных частотных диапазонах. Для успешного переключения необходима эффективная координация между точками доступа; без этого могут возникать задержки или потеря пакетов, что негативно сказывается на качестве соединения.

Интересно отметить, что идея координации точек доступа в контексте распределения ресурсов уже реализовывалась на рынке. В первую очередь стоит упомянуть различные сетевые архитектуры mesh, которые отчасти применяли этот подход. Другим примером развития технологии является

концепция Meru Virtual Cell, обеспечивающая координированную работу точек доступа. В Meru Virtual Cell² точки доступа работают как единое целое, что позволяет обеспечить непрерывное переключение между точками доступа без ухудшения качества соединения, поскольку вся БЛВС «видна» КУ как одна ТД. Этот подход продолжает развивать компания Fortinet³.

3.3. Multi Access Point Coordination (MAPC)

Технология MAPC, с точки зрения точки доступа, объединяет MLO и координацию нескольких точек доступа, что позволяет более эффективно управлять сетевыми ресурсами и повышать общее качество обслуживания (рисунок 2). Это можно аналогично представить как «взгляд со стороны ТД».

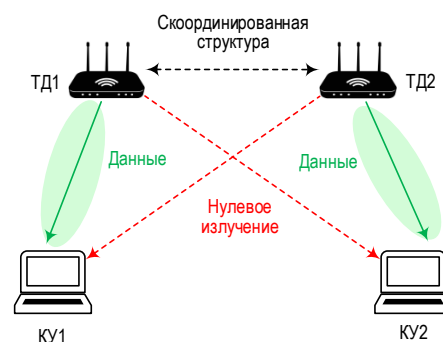


Рис. 2. Схема работы MAPC

Fig. 2. MAPC Operation

Возможные схемы MAPC включают:

- координированный OFDMA;
- координированное пространственное перераспределение;
- координированное формирование луча;
- совместную передачу.

3.4. Другие аспекты

Среди других планируемых улучшений необходимо отметить следующие.

Увеличение максимального числа пространственно-временных потоков до 16 значительно увеличивает потенциальный объем передаваемых данных благодаря использованию технологии многопоточной передачи, что крайне важно для современных приложений. Это нововведение потенциально обеспечивает значительно более высокую пропускную способность.

Стандарт IEEE 802.11bn предполагает значительный шаг вперед в части приоритизации трафика по сравнению с традиционным WMM (*аббр. от англ. Wi-Fi Multimedia*), предлагая больше классов обслуживания, чем существующие четыре

² Best Practices Guide for High-Density Design and Deployment. Meru Networks, 2012. 103 p.

³ Dash K., Gnanasambandam H., Banerjee V. The Power of Virtual Cell Wi-Fi // Fortinet. 2017. URL: <https://www.fortinet.com/blog/industry-trends/the-power-of-virtual-cell-wi-fi> (Accessed 14 April 2025)

(video, voice, best effort, background), что будет способствовать минимизации задержки при передаче.

Эволюция механизмов управления доступом к каналному ресурсу в настоящее время ведет к тому, что с текущим уровнем сложности среды ожидается использование методов искусственного интеллекта и машинного обучения, что позволит улучшить обеспечение мобильности клиентских устройств и управление соединениями. При этом возможен анализ больших объемов данных по использованию спектра, уровню сигнала, картине помех, трафику, что позволит обеспечить RRM для проактивного реагирования на изменения в среде передачи.

Первоначальные реализации стандарта использовали такие модели, как функция распределенной координации (DCF) и функция точечной координации (PCF). С введением стандарта IEEE 802.11e и WMM, были внедрены новые методы, такие как гибридная функция координации с управляемым доступом (HCCA) и расширенный распределенный доступ к каналам (EDCA). Эти методы обеспечивают улучшенную поддержку качества обслуживания в Wi-Fi сетях. WMM, являющийся частью 802.11e, позволяет управлять приоритетами трафика в зависимости от его типа. Например, трафик голосовых и видеозвонков может получать более высокий приоритет, чем фоновый трафик, что обеспечивается через EDCA, который классифицирует трафик по категориям доступа и использует разные параметры конкуренции для доступа к беспроводному каналу⁴. Качество обслуживания в Wi-Fi имеет вероятностную природу, поскольку доступ к беспроводному каналу и успешная передача данных зависят от множества факторов, таких как интенсивность трафика, перегрузка канала и его конкурентное использование.

4. Возникающие научные задачи в связи с внедрением новых технологий

Новый стандарт сталкивается с рядом задач, требующих значительных изменений в проектировании БЛВС. Одним из наиболее критичных аспектов являются изменения в MAC-уровне. Модификации на MAC-уровне значительно усложняют алгоритмы оценки пропускной способности, фундаментально изменяя расчеты сетевой емкости. Ранее расчеты пропускной способности были относительно простыми, поскольку основывались на операциях с одним каналом. В условиях DMLO и возможности одновременного использования нескольких частотных диапазонов сети потребуются динамическое

распределение ресурсов между различными каналами и диапазонами. Это потребует гораздо более сложных алгоритмов, способных управлять распределением трафика в реальном времени и адаптироваться к нагрузке сети. Старые методы статической оценки пропускной способности больше не будут давать реалистичную оценку.

Критерии оценки радиопокрытия с позиций MAPC существенно изменятся, поскольку традиционный подход с требованиями к уровню мощности принимаемого сигнала от первичной (primary) и вторичной (secondary) ТД более не будет однозначно отвечать решению задачи⁵:

1) в рамках MAPC, для скоординированного обслуживания нескольких соединений КУ, должны одновременно работать несколько ТД;

2) одновременное использование нескольких частотных диапазонов потребует усложненных механизмов взаимодействия в спектре, что, в свою очередь, может стать основой для научной задачи оценки пропускной способности сетей с несколькими частотными диапазонами.

Таким образом, от проектировщика БЛВС требуется пересмотреть подходы к расчетам с позиции организации многодиапазонного покрытия.

Для решения задачи частотного планирования потребуются принимать во внимание существенно более широкий спектральный ресурс. Кроме того, внедрение MAPC и DMLO повлияет на развитие алгоритмов управления радиоресурсами (RRM). Эти технологии принесут новый уровень сложности управления, поскольку от RRM будет теперь требоваться не только управление основным каналом для клиента, но также и, как минимум, вторичным, необходимым для резервирования или параллельной работы. Одной из научных задач станет оптимизация сети с использованием MAPC, что может включать разработку алгоритмов для эффективного распределения ресурсов между основными и вторичными каналами, минимизации потерь и улучшения координации между диапазонами.

Важность корректной оценки возможностей КУ (так называемого «профилирования») теперь резко возрастает уже на ранних этапах работы над проектом. В прошлом диапазон возможностей КУ был сравнительно небольшим и в целом предсказуемым, что делало процесс проектирования БЛВС в этом смысле более простым. Однако сейчас обеспечение совместимости проектируемой БЛВС с широким спектром КУ является само по себе отдельной задачей, не всегда очевидной с точки зрения ее постановки. В связи с этим возникает научная задача

⁴ Palayur S., Val Beitia I., Schelstraete S. Quality of Service (QoS) Mechanisms in Wi-Fi // MaxLinear. 24 May 2024. URL: [https://www.maxlinear.com/news/quality-of-service-\(qos\)-mechanisms-in-wi-fi](https://www.maxlinear.com/news/quality-of-service-(qos)-mechanisms-in-wi-fi) (Accessed 1 May 2025)

⁵ Викулов А.С. Скоробогатова С.А. Краткое руководство по планированию Wi-Fi-покрытия. 2024. URL: https://files.qtech.ru/upload/wireless/wi-fi-cover_install_guide.pdf (дата обращения 27 апреля 2026)

разработки подходов к частотному планированию и профилированию устройств, которые учитывают разнообразие характеристик КУ, включая устройства, поддерживающие три радиоканала и работающие с MLO.

Обеспечение обратной совместимости с клиентами предыдущих версий стандарта становится существенно более сложной задачей. Новые режимы должны сосуществовать с более старыми устройствами, обеспечивая соответствующую поддержку предыдущих версий. Планирование работы БЛВС в условиях необходимости обеспечения такой совместимости теперь должно стать неотъемлемой частью процесса проектирования на всех этапах работы над решением проектной задачи.

5. Выводы

Подводя итоги вышесказанному, необходимо отметить, что:

1) дополнения стандартов IEEE 802.11be и 802.11bn – это важные шаги в развитии технологии БЛВС, которые отвечают и будут отвечать растущим требованиям современных приложений;

2) несмотря на заметные достижения, к настоящему времени дополнение IEEE 802.11be имеет

нерешенные проблемы, такие как: сложность управления, высокие задержки и сложности обеспечения совместимости между устройствами разных поколений;

3) основные ожидаемые функции дополнения IEEE 802.11bn, такие как Hybrid ARQ, 16 пространственно-временных потоков, DMLO и MAPS будут востребованы в свете роста требований к производительности и задержке для многих приложений клиентских устройств;

4) внедрение новых стандартов IEEE 802.11be и 802.11bn также ставит перед сообществом необходимость в новых подходах к решению таких задач, как динамическое распределение ресурсов в многодиапазонных сетях, оптимизация использования спектрального ресурса, развитие алгоритмов управления радиоресурсами и обеспечения хендвера, что потребует переработку и усложнение существующих моделей и методов.

Решение этих задач потребует создания новых подходов к частотному планированию и профилированию КУ с учетом их разнообразия и способности поддерживать современные технологии, такие как MLO.

Список источников

1. Volkov A., Muthanna A., Paramonov A., Koucheryavy A., Elgendy I.A. Optimized Data Transmission and Signal Processing for Telepresence Suits in Multiverse Interactions // *Journal of Sensor and Actuator Networks*. 2024. Vol. 13. Iss. 6. P. 82. DOI:10.3390/jsan13060082. EDN:VBEMLB
2. Скоробогатова С.А., Викулов А.С., Парамонов А.И. Анализ эволюционного развития сетей Wi-Fi за первую четверть 21 века // *Труды учебных заведений связи*. 2025. Т. 11. № 6. С. 68–77. DOI:10.31854/1813-324X-2025-11-6-68-77. EDN:YUSEEH
3. Lan X., Zu X., Yang J. Enhanced Multilink Single-Radio Operation for the Next-Generation IEEE 802.11 BE Wi-Fi Systems // *Security and Communication Networks*. 2022. DOI:10.1155/2022/7018360. EDN:JVAZVS
4. Choerakuzhiyil A., Ho K., Reyes S. Comparison of STR and EMLSR Performance in Wi-Fi 7 MLO // *arXiv:2501.04149*. 2025. DOI:10.48550/arXiv.2501.04149
5. Lopez-Perez D., Garcia-Rodriguez A., Galati Giordano L., Kasslin M., Doppler K. IEEE 802.11be Extremely High Throughput: The Next Generation of Wi-Fi Technology Beyond 802.11ax // *arXiv:1902.04320*. 2019. DOI:10.48550/arXiv.1902.04320
6. Deng C., Fang X., Han X., Wang X., Yan L., He R., et al. IEEE 802.11be-Wi-Fi 7: New Challenges and Opportunities // *arXiv:2007.13401*. 2020. DOI:10.48550/arXiv.2007.13401
7. Кучерявый А.Е., Кучерявый Е.А., Киричек Р.В., Бородин А.С., Маколкина М.А., Выборнова А.И. и др. Интернет навыков // *Электросвязь*. 2018. № 1. С. 29–32. EDN:YSUVWI
8. Shishir G., Sonigra R., Seshadri N., Koilpillai R.D. Hybrid-ARQ Protocol for Next Generation Wi-Fi Systems // *Proceedings of the International Conference on COMMunication Systems & NETWORKS (COMSNETS, Bangalore, India, 05–09 January 2021)*. IEEE, 2021. PP. 342–350. DOI:10.1109/COMSNETS51098.2021.9352833
9. Galati Giordano L., Geraci G., Carrascosa M., Bellalta B. What Will Wi-Fi 8 Be? A Primer on IEEE 802.11bn Ultra High Reliability // *IEEE Communications Magazine*. 2023. Vol. 62. Iss. 8. PP. 126–132. DOI:10.1109/MCOM.001.2300728

References

1. Volkov A., Muthanna A., Paramonov A., Koucheryavy A., Elgendy I.A. Optimized Data Transmission and Signal Processing for Telepresence Suits in Multiverse Interactions. *Journal of Sensor and Actuator Networks*. 2024;13(6):82. DOI:10.3390/jsan13060082. EDN:VBEMLB
2. Skorobogatova S.A., Vikulov A.S., Paramonov A.I. Analysis of Wi-Fi Networks Evolution During the First Quarter of the 21st Century. *Proceedings of Telecommunication Universities*. 2025;11(6):68–77. (in Russ.) DOI:10.31854/1813-324X-2025-11-6-68-77. EDN:YUSEEH
3. Lan X., Zu X., Yang J. Enhanced Multilink Single-Radio Operation for the Next-Generation IEEE 802.11be Wi-Fi Systems. *Security and Communication Networks*. 2022. DOI:10.1155/2022/7018360. EDN:JVAZVS

4. Choorakuzhiyil A., Ho K., Reyes S. Comparison of STR and EMLSR Performance in Wi-Fi 7 MLO. *arXiv:2501.04149*. 2025. DOI:10.48550/arXiv.2501.04149
5. Lopez-Perez D., Garcia-Rodriguez A., Galati Giordano L., Kasslin M., Doppler K. IEEE 802.11be Extremely High Throughput: The Next Generation of Wi-Fi Technology Beyond 802.11ax. *arXiv:1902.04320*. 2019. DOI:10.48550/arXiv.1902.04320
6. Deng C., Fang X., Han X., Wang X., Yan L., He R., et al. IEEE 802.11be – Wi-Fi 7: New Challenges and Opportunities. *arXiv:2007.13401*. 2020. DOI:10.48550/arXiv.2007.13401
7. Koucheryavy A.Eu., Koucheryavy Y.A., Kirichek R.V., Borodin A.S., Makolkina M.A., Vybornova A.I., et al. *Electrosvyaz*. 2018;1:29–32. (in Russ.) EDN:YSUVWI
8. Shishir G., Sonigra R., Seshadri N., Koilpillai R.D. Hybrid-ARQ Protocol for Next Generation Wi-Fi Systems. *Proceedings of the International Conference on COMmunication Systems & NETworks, COMSNETS, 05–09 January 2021, Bangalore, India*. IEEE; 2021. p.342–350. DOI:10.1109/COMSNETS51098.2021.9352833
9. Galati Giordano L., Geraci G., Carrascosa M., Bellalta B. What Will Wi-Fi 8 Be? A Primer on IEEE 802.11bn Ultra High Reliability. *IEEE Communications Magazine*. 2023;62(8):126–132. DOI:10.1109/MCOM.001.2300728

Статья поступила в редакцию 22.05.2025; одобрена после рецензирования 28.08.2025; принята к публикации 19.03.2026

The article was submitted 22.05.2025; approved after reviewing 28.08.2025; accepted for publication 19.03.2026

Информация об авторах:

СКОРОБОГАТОВА
Светлана Алексеевна

аспирант кафедры сетей связи и передачи данных Санкт-Петербургского государственного университета телекоммуникаций им. проф. М.А. Бонч-Бруевича
 <https://orcid.org/0009-0000-1013-617X>

ВИКУЛОВ
Антон Сергеевич

кандидат технических наук, доцент кафедры сетей связи и передачи данных Санкт-Петербургского государственного университета телекоммуникаций им. проф. М.А. Бонч-Бруевича
 <https://orcid.org/0000-0002-6671-9267>

ПАРАМОНОВ
Александр Иванович

доктор технических наук, доцент, профессор кафедры сетей связи и передачи данных Санкт-Петербургского государственного университета телекоммуникаций им. проф. М.А. Бонч-Бруевича
 <https://orcid.org/0000-0002-4104-3504>

Авторы сообщают об отсутствии конфликтов интересов.

The authors declare no conflicts of interests.

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И ТЕЛЕКОММУНИКАЦИИ

**2.3.1 – Системный анализ,
управление и обработка
информации, статистика**

**2.3.6 – Методы и системы защиты
информации, информационная
безопасность**

Научная статья

УДК 004.056

<https://doi.org/10.31854/1813-324X-2026-12-2-74-91>

EDN:GEUZWL



Представление промышленного трафика и кодирование протокольной семантики для трансформерных моделей в малых сетях Industrial Ethernet

Нгуен Чунг Буи¹ ✉, bui.nch@phystech.eduФедор Федорович Пашенко^{1,2}, pif-70@yandex.ruФи Тхань То¹, to.ft@phystech.edu

¹Московский физико-технический институт (национальный исследовательский университет),
Москва, 117303, Российская Федерация

²Институт проблем управления РАН им. академика В.А. Трапезникова,
Москва, 117342, Российская Федерация

Аннотация

В современных индустриальных сетях Industrial Ethernet обнаружение аномалий и кибератак требует учета протокольной семантики трафика. **Актуальность** исследования обусловлена ростом числа атак на промышленные системы управления, усилением интеграции производственных и корпоративных сетей, а также широким использованием промышленных протоколов, часть которых изначально не ориентировалась на современные требования к кибербезопасности. Традиционные сигнатурные средства обнаружения ограниченно выявляют ранее неизвестные и скрытые атаки, тогда как подходы, основанные только на статистических характеристиках потоков, нередко теряют важную информацию о логике обмена, ролях сообщений и особенностях прикладного уровня промышленных протоколов. Дополнительную сложность создает дефицит размеченных данных, характерный для реальных промышленных объектов, что затрудняет обучение устойчивых моделей обнаружения атак.

Цель: разработать способ представления промышленного трафика в виде токенизированных последовательностей, пригодных для применения трансформерных моделей в малых сетях Industrial Ethernet.

Методы. Использованы потоково-сессионное агрегирование пакетов, семантическая токенизация полей протоколов Modbus/TCP и OPC UA, эмбединги с позиционным кодированием, предварительное обучение в self-supervised режиме и последующее дообучение модели на задаче классификации сессий. Оценка подхода выполнена на наборе данных Malware in Smart Factory.

Результаты. Разработано протоколно-ориентированное представление сетевых сессий, сохраняющее контекст обмена и особенности промышленных протоколов. На наборе данных Malware in Smart Factory трансформерная модель достигла F1-меры 99,3 % при обнаружении атак и превзошла модели LSTM и CNN; предварительное обучение дополнительно повысило качество классификации.

Теоретическая и практическая значимость. Предложенный подход формирует единый входной формат для анализа трафика Modbus/TCP и OPC UA и может использоваться при построении систем обнаружения аномалий и вторжений в малых промышленных сетях.

Ключевые слова: Industrial Ethernet, Modbus/TCP, OPC UA, трансформер, сетевой трафик, токенизация, позиционное кодирование, обнаружение вторжений, self-supervised обучение

Ссылка для цитирования: Буи Н.Ч., Пашенко Ф.Ф., То Ф.Т. Представление промышленного трафика и кодирование протокольной семантики для трансформерных моделей в малых сетях Industrial Ethernet // Труды учебных заведений связи. 2026. Т. 12. № 2. С. 74–91. DOI:10.31854/1813-324X-2026-12-2-74-91. EDN:GEUZWL

Original research

<https://doi.org/10.31854/1813-324X-2026-12-2-74-91>

EDN:GEUZWL

Representation of Industrial Traffic and Encoding of Protocol Semantics for Transformer Models in Small Industrial Ethernet Networks

 **Nguyen Trung Bui**¹ , bui.nch@phystech.edu

 **Fedor Fedorovich Pashchenko**^{1, 2}, pif-70@yandex.ru

 **Phi Thanh To**¹, to.ft@phystech.edu

¹Moscow Institute of Physics and Technology,
Moscow, 117303, Russian Federation

²Institute of Control Sciences RAS,
Moscow, 190031, Russian Federation

Annotation

Relevance. In modern Industrial Ethernet networks, anomaly and cyberattack detection requires taking the protocol semantics of network traffic into account. The relevance of this study is determined by the growing number of attacks on industrial control systems, the increasing integration of production and corporate networks, and the widespread use of industrial protocols, some of which were not originally designed to meet modern cybersecurity requirements. Traditional signature-based detection tools are limited in their ability to identify previously unknown and stealthy attacks, while approaches based solely on statistical flow characteristics often lose important information about communication logic, message roles, and application-layer features of industrial protocols. An additional challenge is the shortage of labeled data typical of real industrial environments, which complicates the training of robust attack detection models.

Objective. To develop and evaluate a method for representing industrial traffic as tokenized sequences suitable for transformer models in small Industrial Ethernet networks.

Methods. Flow- and session-level packet aggregation, semantic tokenization of Modbus/TCP and OPC UA protocol fields, embeddings with positional encoding, self-supervised pre-training, and subsequent fine-tuning of the model for session classification were used. The approach was evaluated on the Malware in Smart Factory dataset.

Results. A protocol-aware representation of network sessions that preserves communication context and the specific features of industrial protocols was developed. On the Malware in Smart Factory dataset, the transformer model achieved an F1-score of 99.3 % in attack detection and outperformed LSTM and CNN models; pre-training further improved classification performance.

Theoretical and practical significance. The proposed approach provides a unified input format for analyzing Modbus/TCP and OPC UA traffic and can be used in anomaly and intrusion detection systems for small industrial networks.

Keywords: Industrial Ethernet, Modbus/TCP, OPC UA, transformer models, network traffic, tokenization, positional encoding, intrusion detection, self-supervised learning

For citation: Bui N.T., Pashchenko F.F., To P.T. Representation of Industrial Traffic and Encoding of Protocol Semantics for Transformer Models in Small Industrial Ethernet Networks. *Proceedings of Telecommunication Universities*. 2026;12(2):74–91. (in Russ.) DOI:10.31854/1813-324X-2026-12-2-74-91. EDN:GEUZWL

1. Введение

Индустриальные системы управления (ICS, аббр. от англ. Industrial Control Systems) и сети Industrial Ethernet широко используются для соединения контроллеров, датчиков и исполнительных механизмов на производственных объектах. В таких се-

тях передаются данные технологического процесса, команды управления, телеметрия и другие критически важные сообщения. Надежность и безопасность обмена информацией являются приоритетными, однако многие промышленно-коммуникационные протоколы исторически разрабатыва-

лись без учета современных требований к кибербезопасности [1]. Например, протокол Modbus изначально не предусматривал аутентификацию или шифрование, из-за чего трафик уязвим для перехвата и подмены. С другой стороны, более новые протоколы, такие как OPC UA, проектировались с встроенными механизмами безопасности, но их сложность выше. В последние годы значительно возросло число кибератак на ICS, включая случаи вредоносного вмешательства (Stuxnet, Triton и др.), что подтверждает актуальность разработки эффективных средств мониторинга и обнаружения аномалий в промышленном трафике [1].

Классическим средством защиты сетей являются системы обнаружения вторжений (IDS, аббр. от англ. Intrusion Detection System), анализирующие сетевой трафик на наличие известных сигнатур или аномальных паттернов. Однако стандартные IDS с сигнатурным анализом плохо справляются с ранее неизвестными атаками и зашифрованным трафиком. Машинное обучение и, в особенности глубокие нейронные сети, рассматриваются как перспективный инструмент для построения IDS нового поколения, способных выявлять скрытые закономерности в сетевых данных [1]. Ранее в задачах классификации трафика применялись статистические методы и алгоритмы машинного обучения (деревья решений, SVM и др.), а также модели на основе сверточных (CNN, аббр. от англ. Convolutional Neural Network) и рекуррентных (RNN / LSTM, аббр. от англ. Recurrent Neural Network / Long Short-Term Memory) нейронных сетей [2, 3]. Эти подходы дали определенные улучшения по сравнению с ручными правилами, позволяя автоматически извлекать признаки из исходных данных. Тем не менее, они обладают рядом ограничений. В частности, RNN склонны терять долгосрочные зависимости при обработке длинных последовательностей и могут требовать тщательной настройки для разных протоколов, тогда как CNN эффективны для локальных шаблонов, но с трудом улавливают глобальный контекст без значительного увеличения глубины сети.

Трансформерная архитектура, изначально разработанная для обработки последовательностей в задачах естественного языка, в последние годы привлекла внимание исследователей сетевой безопасности. Трансформеры способны параллельно обрабатывать длинные последовательности и учитывать отношения между любыми позициями посредством механизма самовнимания. Это создает предпосылки для лучшего учета контекста в сетевых коммуникациях, где важны взаимосвязи между событиями, удаленными в последовательности по времени. Кроме того, трансформерные модели позволяют эффективно использовать self-supervised обучение, то есть обучение на больших объемах

неразмеченных данных, что особенно актуально для промышленных сетей, где размеченные данные об атаках ограничены или получение их затруднительно [2, 4]. Уже появляются работы, демонстрирующие успешное применение трансформеров для сетевого трафика: например, предложены гибридные модели, сочетающие трансформер и LSTM для классификации вредоносного трафика с высокой точностью (около 99 %) [2], а также специализированные сетевые трансформеры для аномалий в ICS, способные выявлять нештатное поведение устройств без ручной разметки благодаря самообучению на структуре сетевого обмена [4].

Анализ предметной области и обзор литературы

Промышленные сети Industrial Ethernet характеризуются высокой долей периодического и детерминированного обмена, что позволяет строить IDS как на основе правил и спецификаций, так и на основе данных. В литературе подходы к обнаружению атак в ICS условно можно разделить на несколько классов. Общие обзоры методов IDS и применения машинного обучения для кибербезопасности и киберфизических систем приведены, например, в [5, 6]. Для ICS и операционных технологий (OT, аббр. от англ. Operational Technology) важны также экспериментальные стенды и наборы данных, обеспечивающие воспроизводимость исследований; среди наиболее известных – SWaT и WADI [7, 8], а примеры IDS, ориентированные на реальные промышленные сети, представлены в [9].

Сигнатурные и правило-ориентированные IDS (например, на базе Snort / Suricata и протокольных правил) хорошо выявляют известные шаблоны, но требуют постоянного ручного сопровождения и обычно слабо применимы к ранее неизвестным атакам. В промышленной среде дополнительно используются «белые списки» допустимых команд и коммуникационных сценариев согласно рекомендациям по ОТ-безопасности [10].

Модельно-ориентированные и протоколно-специфичные методы используют особенности промышленных протоколов и типовые сценарии взаимодействия между операторскими интерфейсами и программируемыми логическими контроллерами. Так, для Modbus/TCP предлагается строить формальные модели допустимых последовательностей сообщений, вплоть до детерминированных автоматов, что повышает интерпретируемость и точность в стабильных средах [11]. Ограничением является необходимость настройки под конкретный объект и протокол, а также чувствительность к изменениям конфигурации.

Методы машинного и глубокого обучения, применяемые к агрегированным потокам и сессиям, предполагают предварительное сворачивание пакетов в потоки и расчет статистических признаков (размеры, интервалы, направления, количество пакетов

и т. д.), после чего используются классические модели (RF, SVM) или нейросети. Такой подход обеспечивает компактность и удобство обучения, однако часть прикладной семантики протокола может теряться [1, 12]. Для онлайн-обнаружения аномалий по потоковым признакам также применяются ансамбли автокодировщиков, например Kitsune [13].

Глубокое обучение на низкоуровневых представлениях (байты, первые N байт пакета, «карты» байтов) позволяет автоматически извлекать признаки и широко используется для классификации сетевого трафика, в том числе в сценариях с ограниченной видимостью содержимого [14]. Однако в контексте ICS такой подход сталкивается с разнородностью протоколов и риском переобучения на артефакты конкретного стенда или реализации стека.

Трансформерные и self-supervised модели опираются на способность трансформера учитывать дальние зависимости в последовательностях [15], а также на идеи предобучения, характерные для моделей семейства BERT, в частности на маскированное моделирование токенов (MTM, аббр. от англ. Masked Token Modeling) [16]. Для сетевого трафика предлагаются решения семейства BERT (TSFN [2], ET-BERT [17], NetGPT [18]) и специализированные архитектуры Network Transformer для обнаружения аномального поведения устройств и сервисов [4]. Их ключевое достоинство – возможность использовать большие объемы неразмеченного трафика и переносить знания на задачи с малым числом меток.

Для Industrial Ethernet принципиально важно сохранить протокольный смысл сообщений. Поэтому перспективным направлением является протоколно-осмысленная токенизация, при которой признаки выбираются так, чтобы отражать роли сообщений и их контекст (коды функций, типы сервисов, размеры, параметры, временные характеристики). Предлагаемый в статье подход следует этой линии: формируются семантические токены на основе полей Modbus/TCP и OPC UA, затем они агрегируются в сессионные последовательности, после чего обучаются трансформерные модели для выявления аномалий в малых промышленных сетях.

Цель данной статьи – описать метод представления промышленного сетевого трафика и кодирования протокольной семантики, который позволяет эффективно применять трансформерные модели в малых сетях Industrial Ethernet. Под «малыми» понимаются локальные промышленные сегменты с ограниченным числом узлов и предсказуемым профилем трафика, например сеть цехового контроллера и нескольких программируемых логических контроллеров. В таких условиях особенно важна способность модели учитывать регулярные шаблоны обмена и выявлять даже скрытые отклонения.

В работе рассматриваются:

– способы потокового и сессионного агрегирования пакетов (раздел 2);

– характеристики промышленных протоколов Modbus/TCP и OPC UA и вытекающие из них требования к представлению данных (раздел 3);

– метод семантического кодирования трафика и формализация представления пакетов и их последовательностей как токенизированных данных (разделы 4–5);

– варианты реализации эмбедингов признаков и позиционного кодирования (раздел 6);

– аргументы в пользу использования трансформерной архитектуры по сравнению с RNN / CNN для анализа сетей управления (раздел 7);

– подходы self-supervised обучения на сетевых данных (раздел 8);

– экспериментальная оценка предложенного подхода на реальном наборе данных промышленного трафика (раздел 9).

2. Агрегирование трафика по потокам и сессиям

Промышленный сетевой трафик представляет собой последовательность отдельных пакетов, обмен которыми происходит между множеством устройств (контроллеры, датчики, приводы и т. д.). Для корректного анализа поведения системы важно не рассматривать каждый пакет изолированно, а агрегировать пакеты в логические группы, соответствующие коммуникационным потокам или сессиям. Под потоком обычно понимается совокупность пакетов с общими атрибутами низкого уровня, например одинаковыми IP-адресами отправителя и получателя, номерами портов и т. п., в пределах определенного тайм-аута. Сессия может включать один или несколько потоков и отражать законченную операцию или диалог между узлами, например выполнение команды чтения регистра программируемого логического контроллера и передачу ответа. Агрегирование на уровне сессий позволяет восстановить контекст: какие запросы и ответы составляют единый обмен, каков временной интервал между ними и какова последовательность действий.

В промышленных сетях часто наблюдается цикличность и периодичность трафика: например, контроллер опрашивает датчики через фиксированные интервалы, обновляет значения, записывает команды на исполнительные устройства. Такой регулярный фон образует нормальный профиль сессий. При агрегации пакетов в потоки эти повторяющиеся коммуникации проявляются как схожие последовательности пакетов. Любое отклонение (новая последовательность команд, изменение длины или частоты пакетов, незапланированный запрос) может указывать на аномалию или атаку. Таким образом, потоково-сессионное представление повышает информативность данных для алгоритма обнаружения – вместо разрозненных пакетов модель видит целостную картину взаимодействия.

Формирование сессий из исходных дампов трафика осуществлялось с помощью анализа заголовков и временных меток пакетов. В данной работе под сессией понимается одно устойчивое соединение между парой узлов по определенному протоколу верхнего уровня. Для протокола Modbus/TCP это может соответствовать одному TCP-соединению, в рамках которого происходит серия запросов и ответов Modbus. Для OPC UA сессия начинается с установки соединения и инициализации сессионного контекста, включающей обмен HEL / ACK и открытие сессии, после чего следуют операционные сообщения, а завершается она закрытием сессии. Если соединение является длительным и включает множество циклов обмена, его можно разбить на подсессии по дополнительным признакам, например по временному окну или по окончанию логической транзакции. В общем случае критерии агрегирования задаются так, чтобы каждая сессия представляла логически завершенную последовательность взаимодействий.

Подход с использованием потоков и сессий принят в ряде современных наборов данных для IDS ICS. Например, в наборе данных ICS-Flow представлены как исходные сетевые пакеты, так и агрегированные записи потоков, что облегчает применение методов машинного обучения [1]. В разделе 9 используется именно такое агрегирование: исходные файлы трасс сетевого трафика в формате PCAP предварительно обработаны утилитами, выделяющими отдельные сессии обмена. Это позволяет далее токенизировать каждую сессию как самостоятельную последовательность (см. раздел 5). Таким образом, агрегирование трафика служит первым этапом предварительной обработки данных перед подачей в модель на этапе обучения или обнаружения.

3. Протокольные особенности Modbus/TCP и OPC UA

Промышленные сети используют специализированные протоколы прикладного уровня, которые накладывают определенную структуру на передаваемые данные. Рассмотрим два характерных протокола: Modbus/TCP – один из традиционных промышленных протоколов полевого уровня, и OPC UA – современный стандарт интеграции в промышленных системах.

Modbus/TCP

Modbus – один из старейших и наиболее распространенных промышленных протоколов, изначально разработанный в 1979 г. для связи между контроллерами, датчиками и исполнительными устройствами. Изначально данный протокол работал поверх последовательных линий (Modbus RTU по RS-485), однако вариант Modbus/TCP инкапсулирует Modbus PDU (*аббр. от англ. Protocol Data*

Unit) в пакеты TCP/IP, что позволяет использовать его в Ethernet-сетях без изменений в логике протокола. Modbus имеет простую структуру «запрос–ответ»: ведущий узел, например программируемый логический контроллер или SCADA-система, отправляет запрос ведомому устройству с указанием кода функции (чтение, запись и т. п.) и необходимых параметров (адрес регистра, число регистров, данные), на что получает ответ, содержащий запрошенные данные или подтверждение записи. Каждый Modbus PDU имеет фиксированный формат: заголовок (транзакционный идентификатор, идентификатор протокола, длина, адрес устройства) и последующий блок данных приложения с кодом функции и байтами данных. Пример: запрос чтения регистра 0×1000 содержит функцию 0×03 (Read Holding Registers), адрес 0×1000 и количество 1; ответ содержит функцию 0×03, байт количества данных и затем значение регистра.

Особенности Modbus-трафика состоят в том, что это короткие сообщения фиксированного формата, которые чаще всего передаются в открытом виде, без шифрования. Modbus не имеет встроенных механизмов аутентификации, шифрования или контроля целостности, поэтому любой узел в сети, получивший пакет, может его прочитать или имитировать ответ. В нормальном режиме Modbus-трафик весьма регулярен: одни и те же запросы выполняются циклически, а значения могут повторяться или меняться плавно. Объем данных в одном пакете, как правило, невелик – несколько байтов полезной нагрузки. Временные задержки между запросом и ответом минимальны; для проводного Ethernet время реакции обычно составляет порядка 20 мс. С точки зрения представления для методов машинного обучения каждый Modbus-пакет несет семантически значимые поля – код функции, адрес, значение (для ответов), длину и др., – которые желательно использовать при токенизации (см. раздел 4). В то же время отсутствие шифрования позволяет анализировать содержимое напрямую, что упрощает анализ, но создает риски безопасности.

OPC UA

OPC Unified Architecture – современный протокол, разработанный OPC Foundation в 2006 г. как преемник устаревшего OPC DA (OLE for Process Control на базе DCOM). OPC UA спроектирован как универсальный механизм обмена данными между разнородными компонентами промышленной системы – от полевого уровня до облачных сервисов – с поддержкой богатой семантической модели данных и встроенных средств безопасности. OPC UA не привязан к конкретному транспорту: спецификация определяет несколько вариантов привязки к транспортной среде (TCP, HTTP/HTTPS, WebSockets и др.), однако наиболее распространен бинарный

протокол OPC UA поверх TCP. Коммуникационная модель OPC UA является клиент-серверной: клиенты, например SCADA-система или приложение мониторинга, устанавливают сессии с серверами, например OPC-сервером на контроллере, запрашивают или обновляют данные либо подписываются на изменения.

Особенности OPC UA-трафика состоят в следующем: сессия OPC UA начинается с рукопожатия, при котором клиент отправляет сообщение HELLO (HEL) с предложением параметров соединения, а сервер отвечает ACKNOWLEDGE (ACK). Далее происходит открытие безопасного канала – обмен сообщениями OPEN (OPN) с установлением уникального идентификатора сессии и согласованием режима безопасности (шифрование / подпись). После этого клиент создает сессию верхнего уровня в логике OPC UA, получает SessionID и может выполнять сервисные запросы: чтение атрибутов узлов адресного пространства, запись значений, подписку на изменения, вызов методов и т. п. Каждый такой запрос оформляется как сообщение MESSAGE (MSG) определенного типа сервиса, может включать сложно структурированные данные, например массивы и структуры, и получает соответствующий отклик MSG. В конце взаимодействия сессия завершается обменом CLOSE (CLO) и разрывом соединения.

Важное достоинство OPC UA заключается в наличии встроенных средств безопасности. При установлении сессии могут использоваться сертификаты для аутентификации клиента и сервера, а весь трафик сессии шифруется, если выбран режим SecureChannel с шифрованием. Это усложняет анализ содержимого трафика, так как полезная нагрузка становится недоступной для инспекции без ключей. Тем не менее даже зашифрованный OPC UA-трафик имеет отличительные характеристики на уровне метаданных: типы сообщений (HEL, OPN, MSG и др.), размеры пакетов, временные интервалы между сообщениями и последовательность запросов. Эти характеристики можно использовать для построения последовательности токенов, представляющих сессию, не раскрывая содержимого, например токенизировать типы сообщений и размеры. Если же трафик доступен в открытом виде, например в тестовой среде без шифрования или на уровне до шифрования, семантическое содержание OPC UA становится еще богаче: можно извлекать сами запрашиваемые узлы, методы и значения. В данной работе предполагается возможность токенизации как открытых протокольных полей для Modbus, так и ряда характеристик OPC UA. Например, учитывается последовательность типов сообщений OPC UA в сессии (HEL → OPN → несколько MSG → CLO), типы сервисов (Read, Write,

Subscribe и т. д.) и базовые показатели объема данных.

Сравнивая Modbus/TCP и OPC UA, отметим следующие моменты, влияющие на выбор представления данных:

- длина и сложность сообщений (Modbus-пакеты короткие и фиксированные, пакеты OPC UA могут быть значительно больше и вариативнее, особенно при передаче массивов данных или сертификатов);
- структурированность данных (Modbus несет простые регистровые данные, например, числа, флаги, OPC UA – объектно-ориентированные: узлы адресного пространства с атрибутами, разными типами);
- частота обмена (опросы Modbus часто являются периодическими и высокочастотными, тогда как OPC UA может работать как в режиме опроса, так и по подписке, что формирует иной паттерн трафика);
- безопасность (OPC UA с шифрованием требует ориентироваться на метаданные – тайминги, размеры, вместо содержимого при анализе; Modbus – прямой анализ содержимого, но с риском подмены);
- транспорт (оба работают поверх TCP (в рассматриваемом в данной работе случае), потому с точки зрения сетевых потоков поведение сходно – есть установленное соединение, обмен парными сообщениями «запрос-ответ», хотя OPC UA может отправлять и асинхронные уведомления по подписке).

Учет всех этих особенностей диктует, что единый подход к кодированию трафика должен быть достаточно гибким, чтобы извлечь полезные признаки из различных протоколов. В следующем разделе описывается общий метод семантического представления, который может быть специализирован под конкретные протоколы путем настройки набора токенов и признаков.

4. Семантическое кодирование сетевого трафика

Непосредственное использование трафика в форме последовательности байтов, поступающей на сетевой интерфейс, неэффективно для обучения нейронных сетей, особенно таких сложных моделей, как трансформеры. Необходимо преобразовать исходные данные в форму, которая, с одной стороны, пригодна для подачи в модель в виде числовых векторов, а с другой – сохраняет как можно больше семантически значимой информации о передаваемых сообщениях. Семантическое кодирование трафика – это процесс преобразования пакетов и потоков в последовательности токенов, отражающих структурные поля и значения, существенные для конкретного протокола.

Подход, предлагаемый в данной работе, опирается на аналогию с обработкой естественного языка: сетевой обмен рассматривается как «текст», состоящий не из слов, а из токенов протокольных

сообщений. Каждый токен соответствует либо определенному полю / значению в пакете, либо определенному атрибуту коммуникации. Ключевая идея – вместо неструктурированной совокупности числовых признаков (длина, байты и пр.) предоставить трансформеру дискретные метки, которые имеют прямой смысл в контексте протокола. Такой подход успешно применялся, например, для трафика интернета вещей (IoT, *аббр. от англ.* Internet of Things): разработаны процедуры кастомной токенизации, которые превращают статистические характеристики потока в семантически осмысленные пары «признак–значение» [3]. Авторы адаптируют эту идею к пакетно-сессионному представлению.

Принципы выделения токенов

Идентификаторы протоколов и сессий. Каждый поток принадлежит определенному протоколу (Modbus или OPC UA в рассматриваемом случае). Имеет смысл вводить токен, обозначающий начало сессии конкретного протокола, например <MBUS> для начала сессии Modbus, <OPCUA> – для OPC UA. Это играет роль своеобразного «тега контекста».

Тип сообщения. Для Modbus все операционные сообщения – запросы или ответы, определяются кодом функции. Поэтому для Modbus целесообразно вводить токены вида Func:XX (где XX – номер функции). Например, Func:03 для функции чтения регистров, Func:16 для записи нескольких регистров и т. д. В OPC UA – токены для типов верхнеуровневых сервисов: Service:Read, Service:Write, Service:Subscribe и т. п., а также отдельные токены для системных сообщений HEL, OPN, CLO. Благодаря этому последовательность токенов сразу фиксирует последовательность выполняемых операций.

Параметры запросов. В Modbus ключевым параметром является адрес регистра и количество регистров, а для записи – еще и сами данные. Можно кодировать адрес как отдельный токен, например Addr:4096 (для десятичного адреса 4096) или категоризовать диапазоны адресов (например, <CoilAddr>, <InputRegAddr> и т. д.), если известно назначение диапазона. Данные (значения регистров) могут быть любыми числами; их можно дискретизовать (например, токены <Value:0>, <Value:NonZero> для флагов, или диапазонные для аналоговых величин) либо вообще опускать из токенизации, если численные значения малоинформативны сами по себе для выявления аномалий. В OPC UA параметры разнообразны – идентификаторы узлов, запрашиваемые атрибуты, значения. Можно использовать упрощенный подход: токенизировать, к примеру, размер передаваемых данных (<Payload:Small> или <Payload:Large>), или тип узла (Node:Temperature, Node:Pressure и т. п. по семантике тега, если известна). В общем случае, для OPC UA с зашифрованным содержимым, внутренняя семантика скрыта –

тогда токенизируются лишь внешние признаки (тип сервиса, длина).

Структурные метки. Помимо собственно содержимого сообщений, важны структурные моменты: начало и конец сессии. Можно вводить специальные токены, аналогичные start-of-sequence и end-of-sequence в NLP. Например, <START_SESSION> и <END_SESSION>. Они помогут модели понимать границы последовательности, особенно если модель обучается на непрерывном потоковом трафике без заранее заданных границ сессий.

Временные интервалы. Хотя трансформер в базовой версии не учитывает неравномерность времени между токенами, поскольку оперирует позициями в последовательности, для сетевого трафика это существенный фактор. Можно встраивать дополнительные токены или метки, отражающие задержки, например <Delay:100ms>, если между пакетами i и $i + 1$ прошло 100 мс. В нашем случае промышленные сети малого масштаба обычно характеризуются регулярным опросом, поэтому значительные отклонения по времени сами по себе могут указывать на проблему. Однако для упрощения базовой модели авторы не стали явно токенизировать время: оно вводится косвенно через позиционное кодирование (см. раздел 6) или может быть добавлено отдельным каналом признаков.

Таким образом, сессия преобразуется в последовательность токенов различного типа. Упрощенно Modbus-сессия чтения нескольких регистров может быть представлена как: <START_SESSION>, <MBUS>, Func:03, Addr:4096, <RESP>, Func:03, DataLen:2, <END_SESSION>. Здесь <RESP> – условный токен, обозначающий переход от запроса к ответу (можно иначе кодировать, например, включая признак «ответ» в токен функции). Этот пример иллюстрирует, что даже короткий диалог разбивается на ряд осмысленных элементов. Для OPC UA начальная сессия может выглядеть следующим образом: <START_SESSION>, <OPCUA>, HEL, ACK, OPN, OPN, Service:Read, Service:Read, Service:Write, CLO, <END_SESSION> (представлена последовательность этапов и операций).

Подход семантической токенизации позволяет вручную инкорпорировать знания о протоколе в представление данных, что облегчает последующему обучению выявление аномалий. Например, если злоумышленник выполняет нетипичную команду (скажем, функцию 0×05 Force Coil в Modbus, редко используемую в нормальной работе), то в последовательности появится соответствующий токен Func:05 – модель может распознать его как аномальный токен последовательности. Или если атака приводит к рассинхронизации сессии OPC UA (неправильная последовательность OPN / HEL сообщений), то возникнет нелогичный порядок токенов, который трансформер также может уловить, обучаясь на нормальных данных.

Важно подчеркнуть, что семантическое кодирование должно быть разработано для каждого протокола (или класса протоколов) с учетом специфики. Данное исследование сфокусировано на Modbus и OPC UA, но общая методика применима и к другим промышленным протоколам (DNP3, IEC-104, S7 и др.) – для каждого можно определить набор значимых токенов (команды, коды функций, типы сообщений, параметры), которые затем включить в общий «словарь» модели.

5. Формализация представления пакетов и потоков как токенизированных последовательностей

Для строгого описания процедуры кодирования введем некоторый формализм. Пусть имеется сессия S , состоящая из последовательности n сетевых сообщений (пакетов) $S = \{P_1, P_2, \dots, P_n\}$, упорядоченных по времени. Каждый пакет P_i характеризуется набором атрибутов или полей $\{x_{i1}, x_{i2}, \dots, x_{im}\}$, где m зависит от протокола и уровня детализации (например, для Modbus можно выделить поля: тип сообщения («запрос-ответ»), код функции, адрес, значение, длина и т. д., для OPC UA: тип сервиса, флаг безопасности, размер полезных данных и т. п.).

Для целей токенизации определим функцию T , отображающую значение поля в дискретный токен:

$$T(x_{ij}) = t_{ij}, t_{ij} \in V,$$

где V – множество всех токенов (словарь) для рассматриваемых протоколов; например, если x_{ij} – это код функции Modbus со значением 3, то:

$$T(x_{ij}) = t_{\text{Func:03}},$$

а если x_{ij} – индикатор типа сообщения OPC UA со значением «HEL», то:

$$T(x_{ij}) = t_{\text{HEL}}.$$

Каждому пакету P_i может соответствовать не один, а последовательность токенов. В общем случае определим набор токенов для пакета P_i как объединение токенов по всем его полям:

$$t(P_i) = [T(x_{i1}), T(x_{i2}), \dots, T(x_{im})],$$

где квадратные скобки обозначают упорядоченный список токенов.

Таким образом, пакет с множеством полей разворачивается в последовательность токенов. Если пакет содержит поля, которые решено не токенизировать, например числовое значение регистра, обрабатываемое как непрерывный признак, то такие поля можно исключить из $t(P_i)$ либо заменить обобщенным токеном, например <VAL> без указания конкретного значения.

После этого для всей сессии S получим суммарную последовательность токенов путем конкатенации токенов всех пакетов по порядку:

$$T(S) = [t(P_1); t(P_2); \dots; t(P_n)],$$

где точка с запятой обозначает конкатенацию списков; $T(S)$ – это конечное дискретное представление всей сессии, готовое для подачи в трансформерную модель.

Длиной получившейся последовательности будет сумма токенов по всем пакетам. В худшем случае (если каждый пакет порождает несколько токенов) длина может быть значительно больше числа пакетов. Чтобы ограничить размер последовательностей (что важно для уменьшения вычислительной сложности трансформера), можно использовать два подхода:

- 1) сократить максимальное число пакетов n в сессии, например рассматривать не более 100 сообщений, усекая или объединяя слишком длинные сессии;
- 2) ограничить число токенов на пакет (например, токенизировать только ключевые поля, игнорируя второстепенные).

В ходе проведения экспериментов был установлен предел длины последовательности в 128 токенов, который оказался достаточным для типичных сессий в целевом наборе данных. Длинные последовательности, встречавшиеся редко, в основном при длительных OPC UA-сессиях укорачивались: авторы отбрасывали самые ранние части сессии, полагая, что для обнаружения актуальных атак важнее последние события. Альтернативно такие последовательности можно было бы разбивать на несколько сессий.

После токенизации необходимо преобразовать дискретные токены $T(S)$ в числовые векторы для подачи в нейросеть.

Введем отображение E (эмбединг токенов):

$$V \rightarrow \mathbb{R}^d,$$

которое каждому токеноу $t \in V$ сопоставляет обучаемый вектор:

$$z = E(t) \in \mathbb{R}^d,$$

где d – размерность эмбединга.

В дальнейшем t_l обозначает l -й токен в последовательности $T(S)$, а z_l – его векторное представление.

Применяя эмбединг ко всей последовательности $T(S)$, получаем последовательность векторов:

$$E(T(S)) = [z_1, z_2, \dots, z_L],$$

где $L = |T(S)|$ – длина токенизированной последовательности (число токенов в сессии); z_l – d -мерный эмбединг l -го токена:

$$z_l = E(t_l) \in \mathbb{R}^d.$$

Полученная последовательность $Z = [z_1, z_2, \dots, z_L]$ далее дополняется позиционной информацией и подается на вход трансформерной модели.

Дополнительно введем функцию позиционного кодирования $P: \{1, \dots, L\} \rightarrow \mathbb{R}^d$, которая каждой позиции l в последовательности сопоставляет позиционный вектор $p_l = P(l)$.

Тогда окончательный вход в трансформер формируется как сумма эмбединга токена и позиционного вектора:

$$h_l = z_l + P(l), \\ H = [h_1, h_2, \dots, h_L],$$

где h_l – итоговый вектор на позиции l (эмбединг токена с учетом позиции).

Последовательность $H = [h_1, h_2, \dots, h_L]$ используется как вход для последующих слоев трансформера.

6. Эмбединги и позиционное кодирование

6.1. Эмбединг токенов

После токенизации (раздел 4) и формальной укладки последовательности (раздел 5) каждый токен представляет собой некоторый символ из словаря V . Используется стандартный подход, принятый в NLP-моделях: каждому уникальному токenu соответствует обучаемый параметр-вектор размерности d (гиперпараметр модели). Матрица эмбедингов E является частью параметров модели, инициализируется случайно (либо заимствуется из предварительно обученной модели, если таковая имеется) и оптимизируется в процессе обучения трансформера.

Размерность эмбединга d выбирается с учетом баланса между выразительностью и сложностью модели. В проведенных экспериментах использовались значения $d = 64$ или $d = 128$; этого достаточно для кодирования нескольких десятков или сотен возможных токенов без перенасыщения модели избыточными параметрами. Некоторые токены могут не нести большого объема информации, например `<START_SESSION>` или `<END_SESSION>`, однако они все равно получают собственные векторы, которые модель может настроить по необходимости.

Особенность промышленного трафика – потенциально ограниченный размер словаря токенов по сравнению, скажем, с естественным языком. Если рассматривать только Modbus и OPC UA, общее число видов токенов обычно составляет порядка сотен: все возможные коды функций Modbus (чуть более 10 часто используемых), несколько категорий адресов / данных; типы сервисов OPC UA (десятки) и системные сообщения; а также технические токены начала / конца сессии, задержек и т. п.

В итоге $|V|$ может оказаться < 200 . Это означает, что даже случайная инициализация эмбедингов при обучении достаточно быстро начнет разделять токены. В отличие от слов естественного языка, где смысл часто проявляется в геометрии пространства эмбедингов, здесь скорее ожидается, что модель сама соотнесет токены с их полезностью для задачи (например, токен опасной команды может получить вектор, который повышает отклик на выходном слое «аномалия»). Теоретически можно заложить дополнительную структуру, используя раздельные подэмбединги для разных типов токенов (например, отдельные параметры для Func:XX и для адресов / параметров), а затем объединять их в общий вектор. Однако такой ручной дизайн усложняет модель и не всегда дает выигрыш. Поэтому в данной работе используется единое пространство эмбедингов для всех токенов, полагаясь на способность модели различать их тип по контексту (например, токены Func:XX обычно встречаются в характерных позициях сообщения, что может быть уловлено позиционным кодированием).

6.2 Позиционное кодирование

В архитектуре трансформера базовый механизм самовнимания сам по себе не учитывает порядок следования токенов, если явно не добавить позиционную информацию [3]. Классический подход состоит в добавлении к каждому эмбедингу токена позиционного вектора $p_l = P(l)$, где $l = 1, \dots, L$ – индекс позиции токена в последовательности, а L – ее длина. Вариант $P(l)$ может быть детерминированным, например задаваться синусоидальными функциями разной частоты, как в работе [15], либо обучаемым. Для сетевых данных позиция в последовательности соответствует порядку поступления пакетов по времени; интервалы могут быть неравномерными, однако в рамках выбранного представления кодируется именно порядковый номер появления токена.

Использовались два варианта позиционных кодов.

Вариант 1. Синусоидальное позиционное кодирование задается формулами:

$$P(\text{pos}, 2i) = \sin\left(\frac{\text{pos}}{10000^{\frac{2i}{d}}}\right), \\ P(\text{pos}, 2i + 1) = \cos\left(\frac{\text{pos}}{10000^{\frac{2i}{d}}}\right).$$

где pos – позиция токена в последовательности ($\text{pos} = 1, \dots, L$), $i = 0, \dots, d/2 - 1$; d – размерность эмбединга (d_{model}).

Вариант 2. Обучаемое кодирование позиций: в этом варианте вектор позиции, также как и эмбединг токена, является обучаемым параметром.

Для этого вводится таблица позиций:

$$P \in \mathbb{R}^{L_{\max} \times d},$$

где $L_{\max}$ – максимальная длина последовательности.

Вектор p_l берется из этой таблицы и оптимизируется в ходе обучения. Подход более гибок, однако требует осторожности при генерализации на длины, превышающие $L_{\max}$ или редко встречающиеся при обучении.

В реализации было выбрано обучаемое позиционное кодирование, так как максимальная длина известна и относительно невелика: $L_{\max} = 128$. Это позволило модели потенциально скорректировать представление позиций с учетом специфики данных (например, токены запроса и ответа часто идут парой – тогда позиционные векторы для позиций l и $l + 1$ могут быть сделаны более похожими). При этом классическое синусоидальное кодирование также показало работоспособность в наших экспериментах, но с небольшим снижением качества. Таким образом, выбор способа позиционного кодирования не является критичным, однако обучаемый вариант дал немного лучшие результаты, поэтому далее используем именно его.

Отметим интересный момент: комбинирование нескольких признаков пространств. Можно было бы вводить позиционные признаки не только для линейной позиции, но и, например, кодировать номер пакета в сессии отдельно от позиции токена внутри пакета. Или ввести двумерное кодирование: индекс сообщения и индекс токена в сообщении. В теории это повысило бы способность модели различать начало нового пакета даже без специального токена <RESP> или аналогичных. Были проведены эксперименты с такими схемами, однако существенного выигрыша не было получено – вероятно потому, что модель и так распознавала паттерны вроде «токен функции всегда первый после <MBUS>». Тем не менее, для более сложных протоколов или смешанного трафика этот прием может пригодиться.

Итак, на выходе этапа эмбединга и позиционного кодирования для каждой сессии формируется последовательность векторов $H = [h_1, h_2, \dots, h_L]$, готовая к подаче в блоки трансформера. Далее опишем архитектуру модели и обоснуем, почему именно она предпочтительна в рассматриваемом случае.

7. Преимущества применения трансформерной архитектуры для анализа трафика

Трансформерная архитектура в последние годы стала фактическим стандартом в задачах обработки последовательностей, во многих случаях вытеснив RNN. Рассмотрим, какие преимущества она дает применительно к анализу промышленного трафика и почему в данной работе сделан выбор в ее пользу.

1) Учет долгосрочных зависимостей. В промышленных протоколах аномальное поведение может проявляться как последовательность нескольких событий, разнесенных по времени. Например, атака повторного воспроизведения сначала перехватывает пакеты, а спустя некоторое время повторяет их; кроме того, нарушитель может последовательно посылать несколько команд в нетипичном для штатной работы порядке. RNN с ограниченной памятью может «забыть» о первом событии, когда дойдет до пятого, особенно если между ними прошли десятки других пакетов. Трансформер благодаря механизму самовнимания способен напрямую связывать представления удаленных по позиции токенов: каждый выходной признак рассчитывается как взвешенная сумма всех входных, а веса отражают внимание к соответствующим токенам [4].

Формально самовнимание в одной голове механизма внимания вычисляется как:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{Q K^T}{\sqrt{d_k}}\right)V,$$

где Q, K, V – матрицы запросов (Queries), ключей (Keys) и значений (Values), полученные линейными преобразованиями исходных эмбедингов; d_k – размерность пространства запросов и ключей.

Вышепредставленное выражение показывает, что для каждого токена формируется взвешенная сумма всех входных векторов V , а веса вычисляются (через Softmax) на основе сходства соответствующих запросов и ключей $Q K^T$, нормированного на $\sqrt{d_k}$. Таким образом, если, например, текущий токен – Func:05 (команда принудительной установки выхода), модель может «обратить внимание» на предыдущие токены (например, Func:03 – штатный опрос регистра), сравнить контекст и обнаружить несоответствие последовательности. В RNN подобные зависимости также могут учитываться через скрытое состояние, но с ростом расстояния по последовательности влияние ранних событий обычно ослабевает. В трансформере связь между любыми позициями поддерживается напрямую механизмом внимания.

2) Параллелизм и эффективность обучения. При обучении на больших объемах данных трансформеры могут обрабатывать последовательность целиком параллельно в одном временном шаге (self-attention вычисляется матрично), в то время как RNN приходится прогонять пакет за пакетом последовательно. Хотя скорость работы не всегда критична для офлайн-анализа, при тонкой настройке модели на большом датасете (например, гигабайты трафика) параллельное обучение позволяет эффективнее использовать GPU. Кроме того, параллельность облегчает реализацию предварительно обученных методов типа MLM (аббр. от англ.

Masked Language Modeling), где можно одновременно маскировать и предсказывать несколько позиций.

3) Гибкость архитектуры и возможность дообучения. Трансформерные модели легко масштабируются: можно изменять количество слоев, количество голов самовнимания и размер эмбедингов, не меняя принципиально структуру данных. Они также хорошо подходят для переноса знаний: модель, обученная на одном наборе трафика, может быть дообучена на другом или использоваться как инициализация. В контексте ICS это особенно ценно, поскольку трафик одного промышленного объекта может служить основой для предварительного обучения модели, которая затем адаптируется к сети другого предприятия. RNN и CNN также допускают дообучение, однако в трансформерах особенно просто реализовать механизмы дополнительного дообучения последних слоев, заморозки части параметров и т. д. с учетом модульности модели (энкодер / декодер или только энкодер). В наших экспериментах предварительно обучался трансформерный энкодер на задаче реконструкции пакетов, то есть режиме self-supervised обучения (см. раздел 8), а затем изменялся выходной слой и выполнялось дообучение на задаче классификации. При аналогичном подходе модели RNN демонстрировали более медленную и менее устойчивую сходимость.

4) Обобщающая способность. Благодаря своему устройству трансформер способен выявлять более общие закономерности, а не подгоняться под частный порядок событий. Например, CNN хорошо распознает паттерн вида «*A*, затем *B*», если он всегда находится в окне из пары пакетов, однако, если между *A* и *B* иногда появляются другие события, паттерн разрушается. RNN, как отмечалось выше, склонны забывать далекие зависимости. Трансформер же может обрабатывать и последовательности вида «*A*, ..., *B*», и «*A*, *B*» сходным образом, если он выучил, что важна связь между *A* и *B* независимо от расстояния. Это обеспечивает лучшую переносимость между разными режимами работы сети. Практические исследования показывают, что трансформеры способны эффективнее обобщать ранее не наблюдавшиеся варианты трафика после достаточного предобучения [3]. В частности, отмечается, что сочетание трансформера с продуманной токенизацией позволяет достигать высокой точности на различных наборах IoT-трафика без переобучения под каждый из них, что указывает на более высокую обобщающую способность по сравнению с классическими методами.

5) Интерпретируемость (относительная). Несмотря на репутацию «черных ящиков», трансформеры предоставляют механизм внимания, который можно визуализировать для интерпретации ре-

зультатов. В области сетевой безопасности появились работы, где на основе матриц внимания выясняется, на какие именно токены, то есть поля пакетов, модель ориентируется при выявлении аномалии [4]. Это может помочь эксперту понять природу обнаруженного отклонения: например, если внимание сосредоточено на токене функции Modbus, подозрительной является именно команда. RNN и CNN в этом отношении менее удобны, поскольку в них сложнее выделить, какие именно части последовательности повлияли на решение. Хотя интерпретация трансформеров остается активной областью исследований и не всегда дает однозначные выводы, ее потенциал в IDS заключается в том, что модель может не только сигнализировать о срабатывании, но и указывать, какие сообщения вызвали подозрение.

6) Совместимость с self-supervised обучением. Об этом подробнее сказано в следующем разделе; здесь лишь отметим, что трансформеры, особенно энкодеры семейства BERT, изначально разрабатывались с расчетом на предварительное обучение на маскированном моделировании. Это делает их естественным выбором, если требуется использовать неразмеченные дампы трафика для обучения представлений. RNN тоже можно предварительно обучать, однако, например, маскирование части входов при последовательном чтении оказывается менее эффективным, чем в трансформере, где модель оценивает маскированные позиции на основе контекста с двух сторон. Таким образом, архитектура трансформера функционально лучше приспособлена к таким приемам.

Разумеется, у трансформеров есть и недостатки – главным образом, требовательность к вычислительным ресурсам при росте длины последовательности L , поскольку вычислительная сложность механизма самовнимания составляет $O(L^2)$. В контексте малых сетей Industrial Ethernet эта проблема не столь остра: как правило, сессии не содержат тысяч сообщений, а предлагаемый подход с агрегацией и ограничением длины последовательности удерживает L в пределах сотен. Кроме того, можно применять техники ограничения внимания, например локальное внимание или иерархическое агрегирование, как показано в работе [22]. В предлагаемом прототипе модели без таких усложнений удалось обойтись.

В итоге, сопоставив возможности различных архитектур, авторы пришли к выводу, что трансформер предпочтительнее RNN / CNN для рассматриваемой задачи, поскольку способен обрабатывать семантически токенизированный трафик с учетом произвольных зависимостей и эффективно использовать предварительное обучение. Далее описывается, как именно применяется self-supervised обучение для нашей модели и какие результаты оно дает.

8. Self-supervised обучение на трафике

Как уже отмечалось, большая проблема применения глубокого обучения в промышленных сетях состоит в дефиците размеченных данных об атаках. Сбор полноценного набора данных, включающего разнообразные кибератаки на реальное промышленное оборудование, затруднен по соображениям безопасности и конфиденциальности, поэтому крайне важно уметь использовать имеющиеся большие объемы неразмеченного трафика нормальной работы для обучения модели. Self-supervised обучение предоставляет такой подход: формируется вспомогательная задача, которую модель учится решать, и тем самым вырабатывает полезные представления.

Авторы реализовали два варианта self-supervised обучения трансформерного энкодера на промышленном трафике: маскированное моделирование токенов (MTM) и предсказание следующего события (Next Event Prediction).

MTM представляет собой аналог метода MLM из NLP. Последовательность токенов $T(S)$ берется из нормального трафика; случайным образом выбирается часть токенов, например 15 %, которые заменяются специальным маркером <MASK>. Модель, то есть только энкодер, обучается предсказывать исходные токены в позициях <MASK> по контексту остальных элементов. Иными словами, выход трансформера проходит через линейный классификатор на словарь и обучается с функцией потерь кросс-энтропии по истинному токеноу на замаскированных позициях. Такой подход заставляет модель использовать контекст с обеих сторон маски. Например, если в нормальной последовательности присутствовали Func:03, Addr:4096, <RESP>, Func:03 и был замаскирован элемент Addr:4096, то модель должна по окружающим токенам предположить, что именно этот адрес наиболее вероятен, поскольку функция 03 в рассматриваемом наборе данных обычно обращается к определенным адресам. Обучаясь на множестве таких примеров, трансформерный энкодер формирует представление, в котором токены со сходным контекстом имеют близкие эмбединги и сходные активации.

Режим предсказания следующего события предполагает, что модель получает на вход последовательность токенов сессии без последнего пакета и должна предсказать следующий пакет или следующие токены. Этот режим похож на обучение языковой модели слева направо. Однако авторы сочли его менее удобным для решения поставленной задачи, поскольку последовательности обычно циклически и предсказание точного следующего сообщения затруднительно: несколько вариантов могут быть нормальными. Тем не менее была принята попытка обучать трансформер предсказывать не конкретный токен, а совокупность призна-

ков, например тип следующего сообщения – команда или ответ. Качество таких предсказаний также улучшалось со временем, однако в итоге авторы сосредоточились на MTM как на более мощном методе.

После предварительного обучения на большом наборе исходных сессий, где использовался только нормальный трафик без атак, чтобы модель усвоила профиль штатной работы, трансформерный энкодер оказался способен генерировать обобщенные скрытые представления токенов. Эти представления были использованы для решения целевой задачи – обнаружения аномалий или классификации атак – с минимальным дополнительным обучением. В экспериментальной части авторы выполнили тонкую настройку: добавили выходной слой, например бинарный классификатор классов «атака» и «норма», поверх усредненных по времени активаций энкодера и провели дообучение на размеченной выборке, которая существенно меньше по объему, чем данные для предварительного обучения. Выяснилось, что такая стратегия заметно повышает качество: модель, прошедшая self-supervised этап, достигает примерно на 5–10 % более высокого значения F1, чем аналогичная архитектура, обученная без предварительного обучения.

Применение self-supervised обучения к сетевому трафику ICS уже опробовано в других работах. Например, в [4] предложен специальный Network Transformer, в котором использовались структурированные данные о топологии сети, а модель обучалась выявлять аномалии без меток, то есть, по сути, изучала норму и сигнализировала об отклонениях. Настоящее исследование ближе к подходам из NLP: используется маскирование токенов. Это позволяет интегрировать решение в существующие фреймворки трансформеров семейства BERT. Более того, потенциально можно использовать готовые веса крупных моделей, предварительно обученных на разных сетевых данных. В настоящее время появляются работы по большим языковым моделям для сетей; например, модель NetGPT [18] представляет собой трансформер, предварительно обученный генерировать сетевые пакеты. Потенциально такие модели можно адаптировать к конкретным протоколам ICS.

В рамках данной статьи авторы ограничились собственным предварительным обучением на промышленном наборе данных. Технически обучение MTM было реализовано следующим образом: использовался корпус из примерно 100 тыс. сессий нормального трафика (из набора данных, см. раздел 9); случайным образом маскировалось 15 % токенов, после чего трансформерный энкодер с шестью слоями, четырьмя головами механизма внимания и размерностью $d = 128$ обучался предсказывать эти токены. Обучение продолжалось около пяти эпох до сходимости, когда значение кросс-эн-

тропии на масках переставало улучшаться. Затем веса энкодера использовались как инициализация для модели обнаружения атак.

Еще один вариант self-supervised обучения связан с автокодировщиком аномалий, когда модель, например трансформерный автокодировщик, учится воспроизводить входную последовательность на выходе, минимизируя ошибку реконструкции, а аномалии определяются по росту этой ошибки. В настоящем исследовании этот вариант не был реализован полностью, однако в ходе экспериментов простой трансформерный декодер обучался восстанавливать исходную последовательность токенов в задаче тождественного отображения. Эксперимент показал, что такая схема быстро переобучается, то есть фактически копирует вход на выход, поэтому для практической пользы необходимо ограничение пропускной способности скрытого пространства, чтобы нормальный трафик реконструировался хорошо, а аномальный – хуже. Эта проблема будет изучена авторами в дальнейшем.

Подводя итог, отметим, что self-supervised обучение заметно повышает качество и надежность трансформерной модели при анализе промышленного трафика. Оно позволяет использовать большие объемы нормальных данных для формирования представления о протокольной семантике и закономерностях без ручной разметки. В разделе 9 приводится сравнение моделей с предварительным обучением и без него, демонстрирующее выигрыш. В реальных условиях эксплуатации IDS для ICS это означает, что систему можно сначала обучить на обычном рабочем трафике предприятия, например на данных, собранных за несколько дней или недель, а затем использовать в режиме обнаружения с более высокой чувствительностью к отклонениям.

9. Экспериментальное исследование

Для экспериментальной оценки разработанных методов использовался набор данных реального промышленного трафика, включающий как штатное взаимодействие устройств, так и инсценированные атаки. В частности, был задействован общедоступный набор Malware in Smart Factory [19, 20]. Этот набор данных содержит около 173 ГБ исходных файлов сетевого трафика в формате PCAP, записанных за 16 дней работы небольшой промышленной сети с оборудованием числового программного управления, датчиками и программируемыми логическими контроллерами. В сети использовались современные промышленные протоколы, основными из которых были MQTT, OPC UA и Modbus/TCP. Часть трафика соответствует нормальной работе (около 14 дней), а в течение двух дней осуществлялись контролируемые атаки двух типов: агрессивные, то есть легко заметные, и скрытные, характеризующиеся

минимальным отклонением от нормального поведения. Каждая сессия в наборе данных снабжена меткой, указывающей, относится ли она к какому-либо сценарию атаки или является легитимной. Тем самым данный набор хорошо подходит для обучения и тестирования рассматриваемых моделей, поскольку содержит как примеры нормальной работы, так и разнообразные атаки на промышленные протоколы.

Предобработка данных

Из всего объема данных авторы сосредоточились на сессиях, относящихся к протоколам Modbus/TCP и OPC UA, поскольку именно они являлись объектом семантического кодирования. MQTT-трафик, представлявший телеметрию, в данном исследовании не рассматривался. Для извлечения сессий использовался скрипт на основе Zeek [21]: сессия определялась по уникальной паре IP-адресов и портов, а также по протоколу. Для OPC UA дополнительно контролировались сообщения HEL / ACK как признак начала новой сессии. Всего было выделено порядка 1,2 млн нормальных сессий Modbus и около 50 тыс. нормальных сессий OPC UA, а также несколько тысяч сессий с атаками различных типов. Выборка оказалась несбалансированной: нормальных сессий значительно больше. Для предварительного обучения в self-supervised режиме была сформирована обучающая выборка, включавшая около 100 тыс. случайно выбранных нормальных сессий, примерно 90 тыс. Modbus и 10 тыс. OPC UA, что отражало частоту их появления. Отдельно был подготовлен набор для обучения детектора атак: 5000 нормальных сессий и все доступные 3500 сессий с атаками; около 20 % этого набора было отложено в тестовую выборку.

Реализация модели

Трансформерная модель была реализована на PyTorch с использованием архитектуры энкодера семейства BERT в базовой конфигурации, уменьшенной по числу параметров:

- количество слоев (блоков самовнимания) – 4;
- количество голов механизма внимания – 8;
- размер эмбединга и скрытого представления $d = 128$;
- размер промежуточного полносвязного слоя – 256;
- словарь токенов – 150 единиц (разработанный вручную набор для Modbus/TCP и OPC UA, как описано в разделе 4);
- максимальная длина последовательности токенов – 128 (большие сессии укорачивались).

Выходной слой для задачи обнаружения атак представлял собой полносвязный классификатор с функцией Softmax для двух классов: «атака» и «норма». Для обучения использовался оптимизатор AdamW; начальная скорость обучения составляла 5×10^{-4} с линейным спадом.

Режимы обучения:

– предварительное обучение: на 100 тыс. нормальных сессий решалась задача МТМ (маскировалось 15 % случайных токенов; обучение продолжалось пять эпох, после чего значение перплексии на масках стабилизировалось примерно на уровне 1,2; то есть модель почти безошибочно предсказывала замаскированные токены);

– тонкая настройка: продолжение обучения модели, инициализированной весами после предварительного обучения, на задаче классификации с метками «атака» и «норма»; обучение выполнялось 10 эпох с контролем по F1-мере на валидационной выборке и остановкой при отсутствии улучшения в течение двух эпох;

– обучение без предварительного обучения: для сравнения та же архитектура обучалась на задаче классификации без предшествующего этапа self-supervised обучения.

Дополнительно для сопоставления были обучены более простые модели: LSTM и одномерная CNN. LSTM-модель имела два слоя по 128 нейронов; на вход ей подавались те же эмбединги токенов. Иными словами, для корректности сравнения LSTM обучалась на том же токенизированном представлении, а не на исходных признаках. CNN-модель включала четыре одномерных сверточных слоя с размером ядра 3 и 64 выходных канала в каждом слое; сеть обрабатывала последовательность токенов и выявляла локальные n-граммы. Эти модели обучались только на задаче классификации; предварительное обучение для них не проводилось, хотя, например, автокодировщик на основе LSTM теоретически возможен.

Результаты

На тестовой выборке, включавшей ранее не наблюдавшиеся сессии, в том числе 700 атакованных и 1000 нормальных, были получены показатели качества обнаружения атак, приведенные в таблице 1 (положительным считался класс «атака»).

Как видно из таблицы 1, предложенная трансформерная модель, предварительно обученная в self-supervised режиме, показывает наилучший результат: F1 = 99,3 %, что соответствует практически безошибочному выявлению атакующих сессий при минимуме ложных срабатываний. Для сравнения та же модель без предварительного обучения дает F1 около 96,3 %, то есть несколько уступает, хотя все равно превосходит RNN и CNN. Модели LSTM удалось достичь F1 около 93,4 %; ее основная потеря связана с более низкой точностью, то есть с большим числом ложных срабатываний, когда нормальные сессии ошибочно маркируются как атаки. CNN показала еще более низкий результат — F1 около 88,7 %; при этом чаще наблюдались и пропуски атак (полнота 88 %). Такой разрыв между

трансформером и классическими моделями глубокого обучения согласуется с ожиданиями, поскольку трансформер лучше учитывает сложные последовательные шаблоны атак, в том числе скрытых.

Отдельно следует отметить пользу предварительного обучения: разница между 99,3 и 96,3 % в абсолютных величинах выглядит небольшой, однако на практике это означает, что трансформер без предварительного обучения либо пропускал около 3 % атак, либо давал дополнительные ложные тревоги, тогда как предварительно обученная модель допускала менее 1 % таких ошибок. В условиях промышленной эксплуатации IDS трехкратное снижение ложных тревог особенно важно, поскольку уменьшает нагрузку на оператора; вместе с тем пропуск даже нескольких процентов атак может оказаться критичным. Следовательно, self-supervised этап позволил модели тоньше различать норму и аномалию.

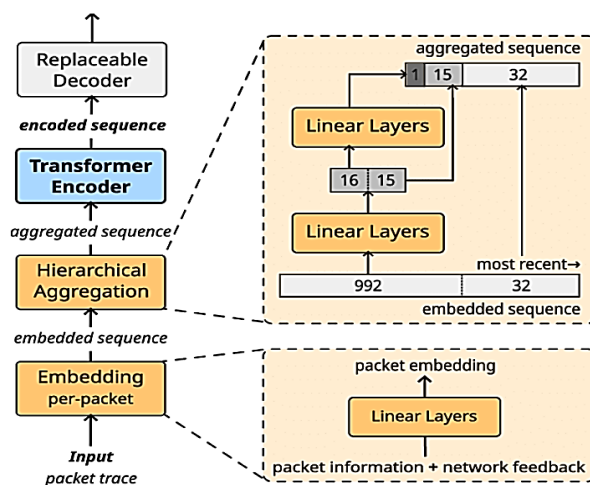
ТАБЛИЦА 1. Сравнение моделей по качеству классификации сессий («атака / норма»), %

TABLE 1. Comparison of Models by Session Classification Quality (Attack / Normal), %

Модель	Accuracy	Precision	Recall	F1-score
Трансформер (предобучение)	99,3	98,9	99,7	99,3
Трансформер (без предобучения)	96,8	95,5	97,1	96,3
LSTM	94,6	91,0	96,0	93,4
CNN	91,2	89,5	88,0	88,7

Примечание: Accuracy – точность; Precision – точность по атаке; Recall – полнота по атаке; F1-score – F1 по атаке

Для наглядности на рисунке 1 приведена схема трансформерной архитектуры для моделирования сетевого трафика (Network Traffic Transformer) [21].



The Network Traffic Transformer (NTT) contains three main stages: an embedding layer, an aggregation layer, and a transformer encoder. It outputs a context-rich encoded sequence that is fed into a task-specific decoder. The aggregation layer aggregates 1024 packet embeddings into 48 in two stages. The most recent packets are kept as-is, less recent packets are aggregated once, and the least recent twice.

Рис. 1. Архитектура Network Traffic Transformer [21, рис. 3]

Fig. 1. Network Traffic Transformer architecture [22, Fig. 3]

В экспериментальной сети присутствовали узлы, работающие по протоколам Modbus и OPC UA, журналы которых анализировались. Конфигурация включала систему числового программного управления, выступавшую OPC UA-сервером, программируемый логический контроллер с поддержкой Modbus, датчики энергии и атакующий узел, локально подключенный для проведения тестов [19]. Эта схема (см. рисунок 1) помогла при интерпретации ряда атак: например, атака типа «man-in-the-middle» на Modbus моделировалась внедрением пакетов от имени программируемого логического контроллера, и трансформер успешно идентифицировал нарушение последовательности «запрос-ответ», в частности появление некорректного токена ответа после запроса и искажение временных меток.

Помимо приведенных результатов, также оценивалась обобщающая способность модели. Авторы пробовали дообучить трансформер, обученный на данном датасете, на другом наборе (для разнообразия – синтетическом датасете ICS-Flow [1], содержащем имитацию атак на водоканале). Модель быстро адаптировалась, показав, что предварительно выученные эмбединги Modbus довольно универсальны – она успешно нашла атаки, даже специфические, например, посылку некорректных команд управления насосом (что проявлялось в токене функции, выходящем за нормальный диапазон). RNN при переносе потребовал почти полного переобучения заново, т. е. его внутренние состояния оказались менее переносимыми.

Обсуждение результатов

Высокая точность трансформерного подхода объясняется несколькими факторами:

- семантическая токенизация предоставляет модели информацию в «сконцентрированном» виде; ей не пришлось самой выучивать, что, скажем, байт 0x03 в определенном контексте значит команду чтения – авторы это обозначили токеном; поэтому большая часть потенциала модели пошла непосредственно на выявление зависимостей между токенами (внимание между функциями, адресами, последовательностями сообщений);

- трансформер эффективно использовал контекст сессии: скрытые атаки, не изменявшие мгновенные статистики, например размер пакета, но изменявшие команду, выявлялись по несоответствию общей последовательности; ложных срабатываний почти не наблюдалось, что свидетельствует о хорошем усвоении профиля нормального трафика;

- LSTM, даже работая с токенами, вероятно теряла часть контекста: у нее имеется фиксированный вектор состояния, тогда как у трансформера несколько голов механизма внимания могли параллельно отслеживать разные аспекты — последовательность функций, временные интервалы и шаблон «запрос-ответ»;

- CNN в данной задаче показывает более слабые результаты, поскольку шаблоны аномалий не сводятся к коротким локальным подпоследовательностям и могут охватывать всю сессию.

Стоит упомянуть и вычислительную эффективность: на этапе применения на одной сессии трансформер с четырьмя слоями и восемью головами механизма внимания обрабатывал около 100 токенов за доли миллисекунды на современном CPU, что позволяет работать в онлайн-режиме с небольшой задержкой даже при сотнях сессий в секунду. LSTM была несколько быстрее на коротких последовательностях, но на длинных проигрывала из-за необходимости итеративной обработки по шагам.

Таким образом, эксперименты подтвердили, что предложенное представление промышленного трафика в виде токенов + трансформерная модель позволяет достичь высоких результатов в задаче обнаружения аномалий и атак. В заключении суммируем выводы и наметим направления дальнейших работ.

10. Заключение

В статье изложен подход к аналитической обработке трафика промышленных сетей с применением трансформерных моделей глубокого обучения.

Основные результаты и выводы работы можно сформулировать следующим образом.

1) Представление трафика

Разработан метод семантического представления индустриального трафика: пакеты агрегируются в сессии, после чего превращаются в последовательности токенов, отражающих протокольную семантику (коды команд, типы сообщений, параметры и др.). Формально определено отображение пакетов в токены и задан процесс эмбединга и позиционного кодирования последовательностей. Показано, что такой метод позволяет унифицировать разнородные протоколы (на примере Modbus/TCP и OPC UA) в одном формате входных данных для нейросети.

2) Трансформерная архитектура в сравнении с RNN / CNN

Обосновано преимущество трансформерной архитектуры для задач анализа сетевого трафика по сравнению с традиционными RNN и CNN: трансформер эффективно улавливает долгосрочные зависимости и сложные шаблоны в последовательностях токенов, что критично для выявления атак, маскирующихся под нормальное поведение. Экспериментально подтверждено, что трансформерная модель дает более высокую точность обнаружения аномалий (F1-мера до 99 %) и более низкий уровень ложных срабатываний, чем LSTM или CNN, при анализе тех же данных. Это соответствует мировой тенденции применения трансформеров в кибербезопасности [2, 3].

3) Self-supervised обучение

Впервые для рассматриваемой предметной области продемонстрировано успешное применение self-supervised обучения трансформера на исходных данных промышленного трафика. Маскирование токенов и их предсказание по контексту позволили модели выучить нормальные шаблоны протоколов без учителя. Дальнейшая тонкая настройка на небольшом количестве размеченных примеров атак дала существенный прирост качества. Такой подход особенно ценен для малых сетей Industrial Ethernet, где аномальные ситуации редки и трудно собрать статистически значимую обучающую выборку: модель может обучиться на большом объеме штатных данных, а затем распознавать редкие отклонения за счет накопленных знаний о норме [4].

4) Практическая применимость

Рассмотренный набор данных и сценарии атак близки к реальным условиям, поскольку атаки проводились в действующей тестовой промышленной сети с реальным оборудованием [19, 20]. Высокая точность, достигнутая моделью, позволяет предполагать, что подобные методы могут быть внедрены в системы кибербезопасности промышленных предприятий, например в виде модуля интеллектуального анализа трафика на промышленных шлюзах или в системах IDS, мониторящих технологические сети. Малые сети Industrial Ethernet, характеризующиеся относительно стабильным трафиком, особенно подходят для таких решений, поскольку модель может четко выучить профиль нормальной работы и выявлять даже минимальные отклонения.

5) Ограничения и дальнейшие работы

Настоящее исследование сфокусировано на двух протоколах и относительно небольшом количестве типов атак. В дальнейшем планируется расширить словарь токенов и методику на другие распространенные протоколы (DNP3, IEC 60870-5-104, PROFINET и др.), а также учесть смешанный трафик, в котором в одной сети присутствуют несколько разных протоколов; это потребует более сложной токенизации и, возможно, мультипротокольного обучения. Кроме того, перспективным направлением является интеграция графовой информации о топологии сети в модель – по аналогии с подходом в работе [4] к интерпретируемости, –

когда в трансформер добавляются специальные токены устройств или соединений. Это может помочь локализовать источник аномалии, то есть конкретный узел или сегмент сети. Еще одно направление связано с применением генеративных моделей, например NetGPT, для синтеза аномального трафика и тестирования системы на новых типах атак.

Отдельно отметим класс атак нулевой динамики на киберфизические системы, при которых нарушитель подбирает воздействие так, чтобы выходные измерения оставались согласованными с нормальным режимом, а изменения происходили во внутренней динамике объекта. Такие атаки рассматриваются как особый тип скрытных атак внедрения данных [23]. Для их выявления часто требуется изменить структуру наблюдения или использовать дополнительные источники информации, например независимые измерения и модельно-ориентированные методы контроля [24, 25]. Поскольку предложенный в статье подход опирается на анализ сетевого трафика и протокольной семантики, гарантированное обнаружение атак нулевой динамики в общем случае невозможно, если злоумышленник использует легитимные последовательности команд и сохраняет нормальный профиль обмена. Тем не менее, в малых детерминированных сетях даже скрытные атаки нередко приводят к сдвигам в последовательностях, временных интервалах или наборе операций, что потенциально может быть уловлено трансформерной моделью. Авторы рассматривают интеграцию сетевых токенов с данными технологического процесса и мультимодальные модели, например трансформеры для многомерных временных рядов, как одно из ключевых направлений продолжения работы [26].

В заключение отметим, что комбинация семантически осмысленного представления данных и мощных архитектур глубокого обучения открывает новые возможности для обеспечения безопасности промышленных информационно-управляющих систем. Предложенный способ кодирования трафика для трансформеров может послужить основой для создания универсальных моделей мониторинга Industrial Ethernet, способных адаптироваться к различным протоколам и своевременно выявлять угрозы в автоматическом режиме.

Список источников

1. Dehlaghi-Ghadim A., Moghadam H.M., Balador A., Hansson H. Anomaly Detection Dataset for Industrial Control Systems // IEEE Access. 2023. Vol. 11. PP. 107982–107996. DOI:10.1109/ACCESS.2023.3320928. EDN:ZMFCCE
2. Shi Z., Luktarhan N., Song Y., Yin H. TSFN: A Novel Malicious Traffic Classification Method Using BERT and LSTM // Entropy. 2023. Vol. 25. Iss. 5. P. 821. DOI:10.3390/e25050821. EDN:HNRTUY
3. Afifi F., Zaki F., Hanif H., Aqil N., Anuar N.B. Transformer-based tokenization for IoT traffic classification across diverse network environments // PeerJ Computer Science. 2025. Vol. 11. P. e3126. DOI:10.7717/peerj-cs.3126. EDN:LFNGZM
4. Marino D.L., Wickramasinghe C.S., Rieger C., Manic M. Self-Supervised and Interpretable Anomaly Detection Using Network Transformers // IEEE Transactions on Industrial Informatics. 2025. Vol. 21. Iss. 5. PP. 4252–4261. DOI:10.1109/TII.2025.3534443. EDN:GDCTHE
5. Buczak A.L., Guven E. A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection // IEEE Communications Surveys & Tutorials. 2016. Vol. 18. Iss. 2. PP. 1153–1176. DOI:10.1109/COMST.2015.2494502
6. Mitchell R., Chen I.-R. A survey of intrusion detection techniques for cyber-physical systems // ACM Computing Surveys. 2014. Vol. 46. Iss. 4. P. 55. DOI:10.1145/2542049. EDN:QKQUJE

7. Mathur A.P., Tippenhauer N.O. SWaT: a water treatment testbed for research and training on ICS security // *Proceedings of the International Workshop on Cyber-Physical Systems for Smart Water Networks (CySWater, Vienna, Austria, 11 April 2016)*. IEEE, 2016. PP. 31–36. DOI:10.1109/CySWater.2016.7469060
8. Ahmed C.M., Palleti V.R., Mathur A.P. WADI: a water distribution testbed for research in the design of secure cyber physical systems // *Proceedings of the 3rd International Workshop on Cyber-Physical Systems for Smart Water Networks (CySWater, Pittsburgh, USA, 21 April 2017)*. New York: ACM, 2017. PP. 25–28. DOI:10.1145/3055366.3055375
9. Anthi E., Williams L., Burnap P., Jones K. A three-tiered intrusion detection system for industrial control systems // *Journal of Cybersecurity*. 2021. Vol. 7. Iss. 1. P. tyab006. DOI:10.1093/cybsec/tyab006. EDN:PZNPOU
10. Stouffer K., Pease M., Tang C., Zimmerman T., Pillitteri V., Lightman S., et al. NIST SP 800-82 Rev. 3. Guide to Operational Technology (OT) Security. 2023. DOI:10.6028/NIST.SP.800-82r3
11. Goldenberg I., Wool A. Accurate modeling of Modbus/TCP for intrusion detection in SCADA systems // *International Journal of Critical Infrastructure Protection*. 2013. Vol. 6. Iss. 2. PP. 63–75. DOI:10.1016/j.ijcip.2013.05.001
12. Azab A., Khasawneh M., Alrabaee S., Choo K.K.R., Sarsour M. Network traffic classification: Techniques, datasets, and challenges // *Digital Communications and Networks*. 2024. Vol. 10. Iss. 3. PP. 676–692. DOI:10.1016/j.dcan.2022.09.009. EDN:BWGLKV
13. Mirsky Y., Doitshman T., Elovici Y., Shabtai A. Kitsune: An Ensemble of Autoencoders for Online Network Intrusion Detection // *Network and Distributed System Security Symposium (NDSS, San Diego, USA, 18–21 February 2018)*. 2018. DOI:10.14722/ndss.2018.23204
14. Lotfollahi M., Siavoshani M.J., Zade R.S.H., Saberian M. Deep packet: a novel approach for encrypted traffic classification using deep learning // *Soft Computing*. 2020. Vol. 24. P. 1999–2012. DOI:10.1007/s00500-019-04030-2. EDN:XWSLQV
15. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., et al. Attention Is All You Need // *Advances in Neural Information Processing Systems*. 2017. DOI:10.48550/arXiv.1706.03762
16. Devlin J., Chang M.-W., Lee K., Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics (NAACL, Minneapolis, USA, 2–7 June 2019)*. 2019. DOI:10.48550/arXiv.1810.04805
17. Lin X., Xiong G., Gou G., Li Z., Shi J., Yu J. ET-BERT: A Contextualized Datagram Representation with Pre-training Transformers for Encrypted Traffic Classification // *arXiv2202.06335*. 2022. DOI:10.48550/arXiv.2202.06335
18. Meng X., Lin C., Wang Y., Zhang Y. NetGPT: Generative Pretrained Transformer for Network Traffic // *arXiv:2304.09513*. 2023. DOI:10.48550/arXiv.2304.09513
19. Brenner P., Fabini J., Offermanns M., Semper S., Zseby T. Malware communication in smart factories: A network traffic data set // *Computer Networks*. 2024. Vol. 255. P. 110804. DOI:10.1016/j.comnet.2024.110804. EDN:FZKJWH
20. Malware in Smart Factory // *TU Wien Research Data Repository*. 2024. DOI:10.48436/2sv8c-ykc69
21. Paxson V. Bro: a system for detecting network intruders in real-time // *Computer Networks*. 1999. Vol. 31. Iss. 23–24. PP. 2435–2463. DOI:10.1016/S1389-1286(99)00112-7
22. Dietmüller A., Ray S., Jacob R., Vanbever L. A new hope for network model generalization // *Proceedings of the 21st ACM Workshop on Hot Topics in Networks (HotNets '22, Austin, USA, 14–15 November 2022)*. New York: Association for Computing Machinery, 2022. PP. 152–159. DOI:10.1145/3563766.3564104
23. Teixeira A.M.H., Shames I., Sandberg H., Johansson K.H. Revealing stealthy attacks in control systems // *Proceedings of the 50th Annual Allerton Conference on Communication, Control, and Computing (Monticello, USA, 01–05 October 2012)*. IEEE, 2012. PP. 1806–1813. DOI:10.1109/Allerton.2012.6483441
24. Teixeira A., Shames I., Sandberg H., Johansson K.H. A secure control framework for resource-limited adversaries // *Automatica*. 2015. Vol. 51. PP. 135–148. DOI:10.1016/j.automatica.2014.10.067
25. Kim J., Back J., Park G., Lee C., Shim H., Voulgaris P.G. Neutralizing zero dynamics attack on sampled-data systems via generalized holds // *Automatica*. 2020. Vol. 113. P. 108778. DOI:10.1016/j.automatica.2019.108778. EDN:GKLFBB
26. Tuli S., Casale G., Jennings N.R. TranAD: Deep Transformer Networks for Anomaly Detection in Multivariate Time Series Data // *Proceedings of the VLDB Endowment*. 2022. Vol. 15. Iss. 6. PP. 1201–1214. DOI:10.14778/3514061.3514067. EDN:LPTVUQ

References

1. Dehlaghi-Ghadim A., Moghadam H.M., Balador A., Hansson H. Anomaly Detection Dataset for Industrial Control Systems. *IEEE Access*. 2023;11:107982–107996. DOI:10.1109/ACCESS.2023.3320928. EDN:ZMFCCE
2. Shi Z., Luktarhan N., Song Y., Yin H. TSFN: A Novel Malicious Traffic Classification Method Using BERT and LSTM. *Entropy*. 2023;25(5):821. DOI:10.3390/e25050821. EDN:HNRTUY
3. Afifi F., Zaki F., Hanif H., Aqil N., Anuar N.B. Transformer-based tokenization for IoT traffic classification across diverse network environments. *PeerJ Computer Science*. 2025;11(e3126). DOI:10.7717/peerj-cs.3126. EDN:LFGZM
4. Marino D.L., Wickramasinghe C.S., Rieger C., Manic M. Self-Supervised and Interpretable Anomaly Detection Using Network Transformers. *IEEE Transactions on Industrial Informatics*. 2025;21(5):4252–4261. DOI:10.1109/TII.2025.3534443. EDN:GDCTHE
5. Buczak A.L., Guven E. A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection. *IEEE Communications Surveys & Tutorials*. 2016. Vol. 18. Iss. 2. PP. 1153–1176. DOI:10.1109/COMST.2015.2494502
6. Mitchell R., Chen I.-R. A Survey of Intrusion Detection Techniques for Cyber-Physical Systems. *ACM Computing Surveys*. 2014;46(4):55. DOI:10.1145/2542049. EDN:OQKUJE
7. Mathur A.P., Tippenhauer N.O. SWaT: a water treatment testbed for research and training on ICS security. *Proceedings of the International Workshop on Cyber-Physical Systems for Smart Water Networks, CySWater, 11 April 2016, Vienna, Austria*. IEEE; 2016. p.31–36. DOI:10.1109/CySWater.2016.7469060
8. Ahmed C.M., Palleti V.R., Mathur A.P. WADI: a water distribution testbed for research in the design of secure cyber physical systems. *Proceedings of the 3rd International Workshop on Cyber-Physical Systems for Smart Water Networks, CySWater, 21 April 2017, Pittsburgh, USA*. New York: ACM; 2017. p.25–28. DOI:10.1145/3055366.3055375

9. Anthi E., Williams L., Burnap P., Jones K. A three-tiered intrusion detection system for industrial control systems. *Journal of Cybersecurity*. 2021;7(1):tyab006. DOI:10.1093/cybersec/tyab006. EDN:PZNPOU
10. Stouffer K., Pease M., Tang C., Zimmerman T., Pillitteri V., Lightman S., et al. *NIST SP 800-82 Rev. 3. Guide to Operational Technology (OT) Security*. 2023. DOI:10.6028/NIST.SP.800-82r3
11. Goldenberg I., Wool A. Accurate modeling of Modbus/TCP for intrusion detection in SCADA systems. *International Journal of Critical Infrastructure Protection*. 2013;6(2):63–75. DOI:10.1016/j.ijcip.2013.05.001
12. Azab A., Khasawneh M., Alrabaa S., Choo K.K.R., Sarsour M. Network traffic classification: Techniques, datasets, and challenges. *Digital Communications and Networks*. 2024;10(3):676–692. DOI:10.1016/j.dcan.2022.09.009. EDN:BWGLKV
13. Mirsky Y., Doitshman T., Elovici Y., Shabtai A. Kitsune: An Ensemble of Autoencoders for Online Network Intrusion Detection. *Network and Distributed System Security Symposium, NDSS, 18–21 February 2018, San Diego, USA*. 2018. DOI:10.14722/ndss.2018.23204
14. Lotfollahi M., Siavoshani M.J., Zade R.S.H., Saberian M. Deep packet: a novel approach for encrypted traffic classification using deep learning. *Soft Computing*. 2020;24:1999–2012. DOI:10.1007/s00500-019-04030-2. EDN:XWSLQV
15. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., et al. Attention Is All You Need. *Advances in Neural Information Processing Systems*. 2017. DOI:10.48550/arXiv.1706.03762
16. Devlin J., Chang M.-W., Lee K., Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics, NAACL, 2–7 June 2019, Minneapolis, USA*. 2019. DOI:10.48550/arXiv.1810.04805
17. Lin X., Xiong G., Gou G., Li Z., Shi J., Yu J. ET-BERT: A Contextualized Datagram Representation with Pre-training Transformers for Encrypted Traffic Classification. *arXiv2202.06335*. 2022. DOI:10.48550/arXiv.2202.06335
18. Meng X., Lin C., Wang Y., Zhang Y. NetGPT: Generative Pretrained Transformer for Network Traffic. *arXiv:2304.09513*. 2023. DOI:10.48550/arXiv.2304.09513
19. Brenner P., Fabini J., Offermanns M., Semper S., Zseby T. Malware communication in smart factories: A network traffic data set. *Computer Networks*. 2024;255:110804. DOI:10.1016/j.comnet.2024.110804. EDN:FZKJWH.
20. *Malware in Smart Factory*. TU Wien Research Data Repository. 2024. DOI:10.48436/2sv8c-ykc69
21. Paxson V. Bro: a system for detecting network intruders in real-time // *Computer Networks*. 1999. Vol. 31. Iss. 23–24. PP. 2435–2463. DOI:10.1016/S1389-1286(99)00112-7
22. Dietmüller A., Ray S., Jacob R., Vanbever L. A new hope for network model generalization. *Proceedings of the 21st ACM Workshop on Hot Topics in Networks, HotNets '22, 14–15 November 2022, Austin, USA*. New York: Association for Computing Machinery; 2022. p.152–159. DOI:10.1145/3563766.3564104
23. Teixeira A.M.H., Shames I., Sandberg H., Johansson K.H. Revealing Stealthy Attacks in Control Systems. *Proceedings of the 50th Annual Allerton Conference on Communication, Control, and Computing, 01–05 October 2012, Monticello, USA*. IEEE; 2012. p.1806–1813. DOI:10.1109/Allerton.2012.6483441
24. Teixeira A., Shames I., Sandberg H., Johansson K.H. A secure control framework for resource-limited adversaries. *Automatica*. 2015;51:135–148. DOI:10.1016/j.automatica.2014.10.067
25. Kim J., Back J., Park G., Lee C., Shim H., Voulgaris P.G. Neutralizing zero dynamics attack on sampled-data systems via generalized holds. *Automatica*. 2020;113:108778. DOI:10.1016/j.automatica.2019.108778. EDN:GKLFBB
26. Tuli S., Casale G., Jennings N.R. TranAD: Deep Transformer Networks for Anomaly Detection in Multivariate Time Series Data. *Proceedings of the VLDB Endowment*. 2022;15(6):1201–1214. DOI:10.14778/3514061.3514067. EDN:LPTVUQ

Статья поступила в редакцию 04.01.2026; одобрена после рецензирования 24.02.2026; принята к публикации 05.03.2026

The article was submitted 04.01.2026; approved after reviewing 24.02.2026; accepted for publication 05.03.2026

Информация об авторах:

**БУИ
Нгуен Чунг**

аспирант кафедры инфокоммуникационных систем и сетей Московского физико-технического института (национального исследовательского университета)
 <https://orcid.org/0009-0004-1489-927X>

**ПАЩЕНКО
Федор Федорович**

доктор технических наук, профессор, главный научный сотрудник Лаборатории № 40 «Интеллектуальных систем управления и моделирования» Института проблем управления им. В.А. Трапезникова РАН, профессор кафедры инфокоммуникационных систем и сетей Московского физико-технического института (национального исследовательского университета)
 <https://orcid.org/0000-0001-8898-2720>

**ТО
Фи Тхань**

аспирант кафедры инфокоммуникационных систем и сетей Московского физико-технического института (национального исследовательского университета)
 <https://orcid.org/0009-0006-6742-5398>

Авторы сообщают об отсутствии конфликтов интересов.

The authors declare no conflicts of interests.

Научная статья

УДК 621.39

<https://doi.org/10.31854/1813-324X-2026-12-2-92-101>

EDN:LWPUOU



Анализ последствий атак на элементы системы синхронизации, функционирующие по алгоритму ВТСА на основе протокола точного времени

✉ Евгений Валерьевич Опарин¹, onapuh@mail.ru✉ Федор Алексеевич Прошин² ✉, fedorproshin@gmail.com¹ЗАО «Институт телекоммуникаций»,

Санкт-Петербург, 194100, Российская Федерация

²Петербургский государственный университет путей сообщения Императора Александра I,

Санкт-Петербург, 190031, Российская Федерация

Аннотация

Протокол точного времени получил широкое распространение во многих сферах, включая такие критически важные отрасли, как энергетический и промышленный комплексы, экономика и связь – где точность синхронизации является необходимым условием функционирования. Несвоевременность получения данных, вызванная возникновением отклонения шкалы времени, может привести к неисправности, что недопустимо. Общедоступность данного протокола делает его привлекательным для потенциального нарушителя, который может вмешаться в работу устройств. Следовательно, все большую актуальность приобретает задача выявления подобных воздействий на начальном этапе с целью недопущения дальнейшего рассогласования системы.

Цель: исследовать вероятные стратегии атак на элементы системы синхронизации, функционирующие по алгоритму поиска лучшего источника; на основе сформированных стратегий атак разработать имитационные модели функционирования элементов системы синхронизации в условиях воздействия атак.

Методы: сбор, систематизация и анализ научно-технической информации, имитационное моделирование с применением агентного подхода.

Результаты: разработан комплекс имитационных моделей процесса функционирования элементов системы синхронизации, функционирующих по алгоритму поиска лучшего источника в условиях воздействия информационных атак.

Теоретическая значимость: на основе комплекса разработанных моделей возможно отслеживать состояния портов устройств протокола точного времени, определять время сходимости алгоритма поиска лучшего источника и в целом отслеживать процесс функционирования элементов системы синхронизации, исходя из атрибутов портов устройств, характеристик каналов связи и различных стратегий проведения атак.

Практическая значимость. Полученные результаты могут быть использованы на реальных сетях связи для выявления скомпрометированных устройств системы синхронизации при наличии информации о характеристиках узлов связи. Имея в наличии статистику оценки времени сходимости алгоритма поиска лучшего источника и сведения о состоянии портов устройств, можно сделать вывод о вероятных объектах атаки и предпринять шаги по ликвидации деструктивных воздействий.

Ключевые слова: сетевая синхронизация, ВТСА, РТР, протокол точного времени, злоумышленник, атака

Ссылка для цитирования: Опарин Е.В., Прошин Ф.А. Анализ последствий атак на элементы системы синхронизации, функционирующие по алгоритму ВТСА на основе протокола точного времени // Труды учебных заведений связи. 2026. Т. 12. № 2. С. 92–101. DOI:10.31854/1813-324X-2026-12-2-92-101. EDN:LWPUOU

Original research

<https://doi.org/10.31854/1813-324X-2026-12-2-92-101>

EDN:LWPUOU

An Analysis of a Sequence to Attacks on Synchronization System Elements Operating According to BTCA Based on Precision Time Protocol

Evgeniy V. Oparin¹, onapuh@mail.ru

Fedor A. Proshin² ✉, fedorproshin@gmail.com

¹Institute of Telecommunications JSC,
St. Petersburg, 194100, Russian Federation

²Emperor Alexander I St. Petersburg State Transport University,
St. Petersburg, 190031, Russian Federation

Annotation

Precision time protocol is widely common in different areas that include critically significant industries such as power generation, manufacturing, economics and communication where high accuracy is the necessary provision for its working. Time offset can causes delays in data receiving that potentially leads to faults. Precision time protocol is a quite attractive for malicious user to interfere in equipment functioning because of its accessibility. Therefore, preventable detecting of such actions is one of the actual goals for defending from faults.

The objective is to research possible attack strategies on the elements of the synchronization system operating according to the best time transmitter clock algorithm, and to form simulation models of the functioning of the elements of the synchronization system under the influence of attacks based on the formed attack strategies.

Methods include collection, systematization and analysis of scientific and technical information, simulation using an agent-based approach.

Results. A complex of simulation models of the functioning of synchronization system elements operating according to the best time transmitter clock algorithm under the influence of information attacks are developed.

The theoretical significance is that a complex of developed models makes it possible to monitor the port status of precision time protocol's devices to determine the convergence time of the best time transmitter clock algorithm, and generally, to monitor the functioning of synchronization system elements based on the attributes of device ports, characteristics of communication channels, and various attack strategies.

The practical significance is that the obtained results can be used for real communication networks to identify anomaly synchronization system devices in the presence of information about the characteristics of communication nodes. When we have the statistics on the estimation of the convergence time of the best time transmitter clock algorithm and information on the status of device ports, it is possible to draw a conclusion about the objects for attacking and to take steps for eliminating destructive influences.

Keywords: network synchronization, BTCA, PTP, precision time protocol, attacker, attack

For citation: Oparin E.V., Proshin F.A. An Analysis of a Sequence to Attacks on Synchronization System Elements Operating According to BTCA Based on Precision Time Protocol. *Proceedings of Telecommunication Universities*. 2026;12(2):92–101. (in Russ.) DOI:10.31854/1813-324X-2026-12-2-92-101. EDN:LWPUOU

Введение

Протокол точного времени (ПТР, аббр. от англ. Precision Time Protocol) представляет собой пакетно-ориентированный протокол, спроектиро-

ванный для точной синхронизации распределенных часов с разрешением менее микросекунды и работающий, в основном, по Ethernet или IP. Согласно стандарту IEEE Std 1588-2019 [1], протокол

РТР обеспечивает точность синхронизации в сети выше 1 мкс (до 10 нс). Процесс функционирования РТР основан на обмене сообщениями, передающими необходимую информацию в виде меток времени.

Функционирование протокола РТР может быть концептуально разделено на два этапа. На первом этапе устройства РТР логически самоорганизуются в иерархическое дерево синхронизации с использованием алгоритма поиска лучшего источника (ВТСА, аббр. от англ. Best Time Transmitter Clock Algorithm). В сети связи устройства РТР постоянно обмениваются друг с другом качественными характеристиками своих внутренних часов. Устройство РТР с часами самого высокого качества в системе в конечном итоге берет на себя роль Грандмастера (GM, аббр. от англ. Grandmaster) и выполняет функции источника времени для всей сети связи. На втором этапе осуществляется непрерывная передача меток времени от GM вниз по иерархии устройств сети связи [1–3].

Полная иерархия дерева синхронизации целиком формируется отдельными узлами РТР, которые решают, в каком состоянии должен находиться каждый порт, просматривая сообщения Announce, которые узел получает от других портов РТР, находящихся в состоянии «Master».

Алгоритм ВТСА состоит из двух независимых алгоритмов [1, 4]:

- алгоритм сравнения набора данных (от англ. data set comparison algorithm);
- алгоритм определения состояния (от англ. the state decision algorithm).

Алгоритм сравнения набора данных основывается на попарном сравнении атрибутов портов узлов со следующими приоритетами [1, 4]:

- *priority1*: атрибут, задаваемый пользователем, который описывает, что генератор находится в перечне генераторов, из которых выбирается ведущий (Master);
- *clockClass*: атрибут, определяющий отслеживание часами международного атомного времени (TAI, аббр. от англ. International Atomic Time);
- *clockAccuracy*: атрибут, определяющий точность часов;
- *offsetScaledLogVariance*: атрибут, определяющий стабильность часов;
- *priority2*: атрибут, задаваемый пользователем, который описывает более тонкое упорядочивание среди одинаковых часов;
- *clockIdentity*: дополнительный параметр.

Алгоритм определения состояния вычисляет, будет ли следующее состояние порта установлено как «Master», «Slave», или «Passive», основываясь на результатах алгоритма сравнения набора данных. Цель ВТСА – определить состояние каждого порта

системы РТР. Это определение основано на информации, содержащейся в сообщениях синхронизации, полученных портами, и на заданных по умолчанию значениях данных для узлов системы РТР.

Алгоритм ВТСА работает независимо для каждого часа в системе РТР, то есть часы не согласовывают, какие из них должны быть ведущими, а какие ведомыми. Вместо этого каждый вычисляет только свое собственное состояние, при этом ВТСА избегает конфигураций с двумя ведущими, или без них [1, 4, 5].

Результат работы ВТСА существенно зависит от информации, находящейся в сообщениях Announce, и отсутствие координации между узлами во время выбора ведущих устройств делают участников РТР уязвимыми для атак, связанных с процессом функционирования данного алгоритма.

При атаке на элементы системы синхронизации, функционирующие по ВТСА на основе протокола РТР, злоумышленник, завладев отдельным узлом РТР, может вмешаться в процесс выбора ведущего устройства, заявляя преувеличенные и неверные характеристики скомпрометированного узла, и, став ведущим, манипулировать синхронизацией всех ведомых устройств [6–8].

Манипулирование содержимым сообщения Announce при организации процесса функционирования алгоритма ВТСА

Полная иерархия дерева синхронизации полностью формируется отдельными узлами РТР, которые решают, в каком состоянии должен находиться каждый порт, просматривая сообщения Announce, которые узел получает от других портов РТР, находящихся в состоянии «Master».

Для моделирования атаки на элементы системы синхронизации, функционирующие по ВТСА на основе протокола РТР, рассмотрим простую топологию, где исходные данные для моделирования представлены в виде сети связи (рисунок 1).

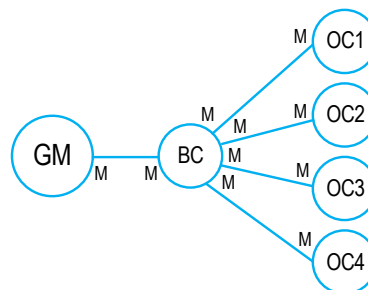


Рис. 1. Сеть РТР, находящаяся в исходном состоянии

Fig. 1. PTP Network in the Initial State

Сеть связи на рисунке 1 содержит одно устройство уровня GM. Дополнительно она состоит из граничных часов (BC, аббр. от англ. Boundary Clock), а также из обычных часов (OC, аббр. от англ. Ordinary

Clock), которые выступают потребителями сигналов точного времени. В качестве исходных данных для моделирования также выступают состояния портов устройств. Условно считаем, что изначально порты всех устройств находятся в состоянии «Master» и осуществляют рассылку сообщений Announce [1, 9, 10].

Главным итогом работы ВТСА должно быть такое состояние портов устройств, чтобы обеспечить создание дерева синхронизации и передачу меток времени потребителям. Любой порт РТР в состоянии «Master» периодически отправляет сообщения Announce на общий адрес групповой рассылки РТР. Поля в этом сообщении содержат атрибуты портов и устройств, используемые алгоритмом ВТСА для определения состояния портов. Указанные атрибуты портов и устройств представлены в таблице 1 [1, 11].

ТАБЛИЦА 1. Атрибуты портов и устройств, используемые алгоритмом ВТСА для определения состояния портов

TABLE 1. Per-Port and Per-System Attributes That are Used in BTCA for Port Initialization

Поле	Значение
priority1	0–255
clockClass	6–255
clockAccuracy	17–31
clockVariance	000016–FFFE16
priority2	0–255
clockIdentity	глобально уникальный номер в 8 октетах
portNumber	1–255
stepsRemoved	0–255

Определение состояния портов осуществляется путем поэтапного сравнения атрибутов. Проведем моделирование процесса функционирования реализации ВТСА для примера сети связи, показанного на рисунке 1, при нормативных условиях, а также в условиях атаки манипулирования содержимым сообщения Announce.

Для проведения имитационного моделирования зададим изначально значения атрибутов портов устройств (таблица 2). Отметим в качестве замечания, что значения атрибутов «clockVariance» и «clockIdentity» приведены к десятичному виду. В качестве критерия функционирования ВТСА примем время сходимости. Под временем сходимости ВТСА будем понимать время, когда определено состояние последнего порта рассматриваемой сети. В итоге имитационная модель для оценки процесса функционирования алгоритма ВТСА примет вид, показанный на рисунке 2 [1, 11–15]. Дополнительно заданы переменные «Port_N_N», в которых будет отражено значение порта по окончании процесса функционирования ВТСА.

ТАБЛИЦА 2. Значения атрибутов портов устройств рассматриваемой сети

TABLE 2. Per-Port Attribute Values for the Given Network

Поле	GM	BC	OC1	OC2	OC3	OC4
priority1	3	16	32	32	32	32
clockClass	6	20	32	40	44	44
clockAccuracy	17	25	50	53	56	56
clockVariance	0	25	50	58	62	62
priority2	0	25	50	64	70	70
clockIdentity	0	16	50	52	54	54
portNumber	1	1	1	1	1	1
stepsRemoved	0	1	2	2	2	2

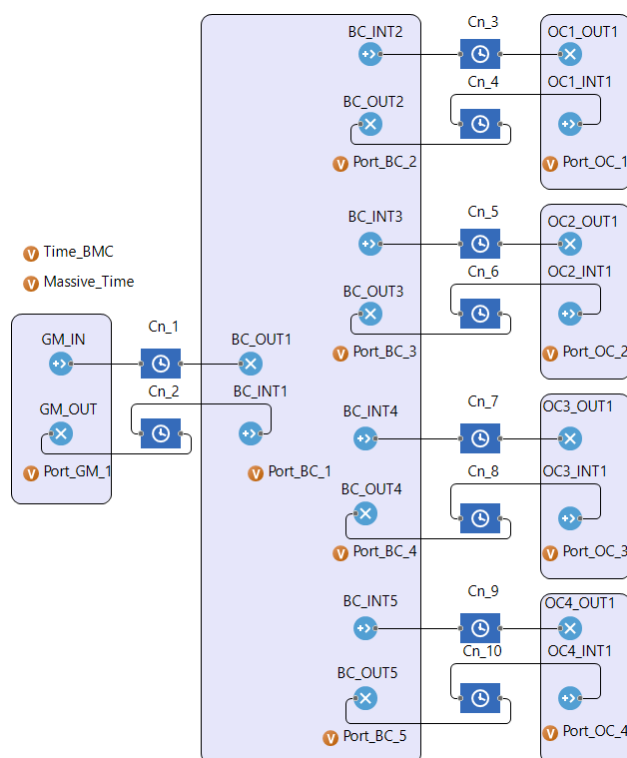


Рис. 2. Сетевая модель работы домена сети РТР при выполнении алгоритма ВТСА

Fig. 2. The Network Model of PTP Domain Functioning in Accordance with BTCA

Построенная модель также позволяет производить оценку процесса функционирования ВТСА в зависимости от характеристик каналов передачи данных, которые заданы соответствующими блоками задержек. Переменная Massive_Time содержит массив времен определения состояния портов по каждому каналу, переменная Time_BMC представляет собой время сходимости алгоритма ВТСА. Атрибуты портов устройств рассматриваемой сети заданы как параметры агентов и при проведении процесса моделирования записываются в соответствующие переменные [1, 12, 15]. При передаче сообщения Announce, т. е. при прохождении агента по

каналу происходит обработка алгоритма сравнения атрибутов портов устройств. После обработки алгоритма сравнения атрибутов портов устройств происходит определение состояния каждого порта. Время определения состояния соответствующего порта записывается в массив *Massive_Time*, после обработки которого происходит определение времени сходимости алгоритма BTCA. При отсутствии атаки на элементы системы синхронизации и корректном функционировании алгоритма BTCA состояния портов устройств примут вид, как показано на рисунке 3. По результатам моделирования определены состояния портов устройств в соответствии с заданными атрибутами устройств, а также определено время сходимости алгоритма BTCA, которое составило 6,807 мкс.

В случае, если злоумышленник контролирует какое-либо из устройств РТР, он в состоянии подменить содержимое сообщений Announce, а значит вмешаться в процесс работы BTCA. Рассмотрим случай, когда злоумышленник контролирует узел ОС4.

В данном случае он в состоянии подменить значения атрибутов портов устройства ОС4, например, присвоить атрибуту «*priority1*» значение 2, которое станет минимальным из всех устройств. Результат моделирования указанного случая показан на рисунке 4, из которого видно, что изменились состояния портов устройств РТР, а значит изменились маршруты доставки сигналов точного времени. Также следует отметить, что увеличилось время сходимости алгоритма BTCA (12,998 мкс), данный факт связан с тем, что потребовалась дополнительная отправка сообщения Announce в направлении устройства GM.

Построенная модель позволяет отслеживать состояния портов устройств РТР, определять время сходимости алгоритма BTCA и в целом отслеживать процесс его функционирования, исходя из атрибутов портов устройств и характеристик каналов связи, что может являться основой для фиксации и выявления возможных атак на домены РТР.

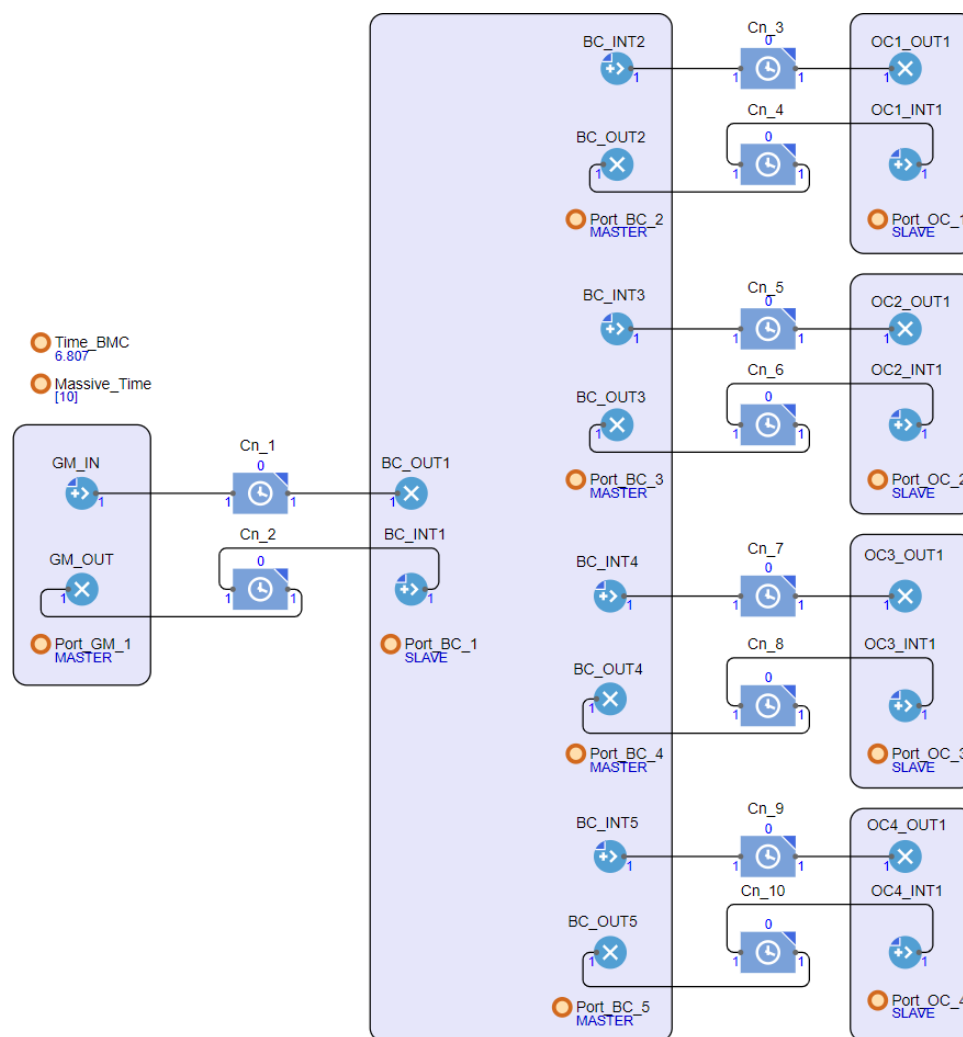


Рис. 3. Результат моделирования процесса функционирования алгоритма BTCA при отсутствии атаки на элементы системы синхронизации

Fig. 3. The Results of BTCA Functioning Model without Attack on Network Elements

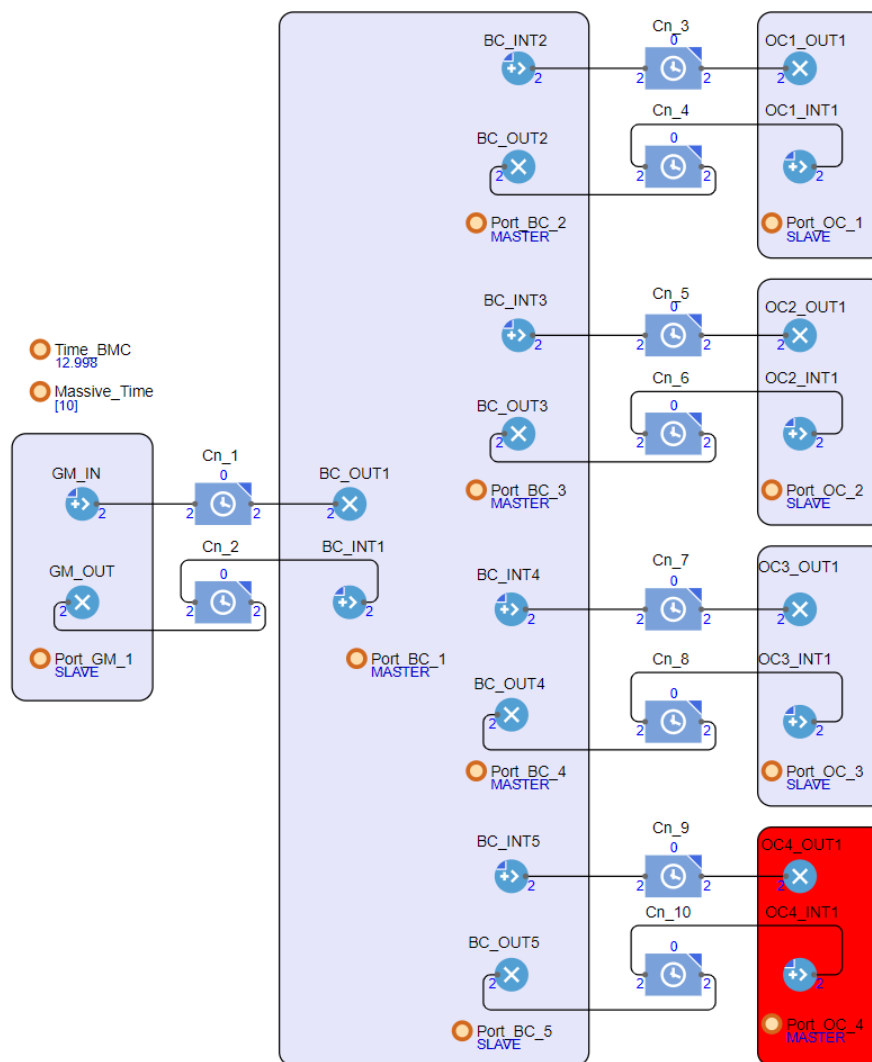


Рис. 4. Результат моделирования процесса функционирования алгоритма BTCA в случае контроля злоумышленником узла OC4

Fig. 4. The Results of BTCA Functioning Model with Node OC4 Controlled by Attacker

Манипулирование задержкой сообщения Announce при организации процесса функционирования алгоритма BTCA

Сообщение Announce является общим сообщением при функционировании протокола RTP и используется для распространения информации о характеристиках и состоянии портов устройств RTP, а также данных о качестве синхронизации и приоритетах. Злоумышленник в случае контроля отдельных устройств RTP способен вносить задержки в процесс обработки сообщений Announce, влияя таким образом на работу алгоритма BTCA, повышая его время сходимости, в результате чего могут видоизменяться маршруты доставки сигналов единого времени в отдельные моменты времени.

Рассмотрим случай, когда злоумышленник контролирует узел BC. Для этого добавлены дополнительные элементы задержки «Delay», которые имитируют воздействие на пакеты RTP при контроле злоумышленником узла BC на каждый канал пере-

дачи сообщения Announce, в результате чего имитационная модель для оценки процесса функционирования BTCA примет вид, показанный на рисунке 5 [1, 11, 15].

Построенная модель позволяет оценивать время сходимости BTCA при различных видах внесения задержек на каждый канал (как постоянных, так и блуждающих). Для примера оценим время сходимости алгоритма при дополнительной задержке 2 мкс на каждый канал передачи и обработки сообщения Announce при исходных данных согласно таблице 2. В результате проведенного моделирования получены следующие результаты. Алгоритм отработал корректно, однако наблюдается значительное увеличение времени сходимости, равное 13,117 мкс. В случае сетей связи значительного размера и масштаба увеличение времени сходимости алгоритма из-за внесения дополнительных задержек в передачу сообщения Announce может повлиять на точность доставки сигналов единого времени и маршруты их передачи.

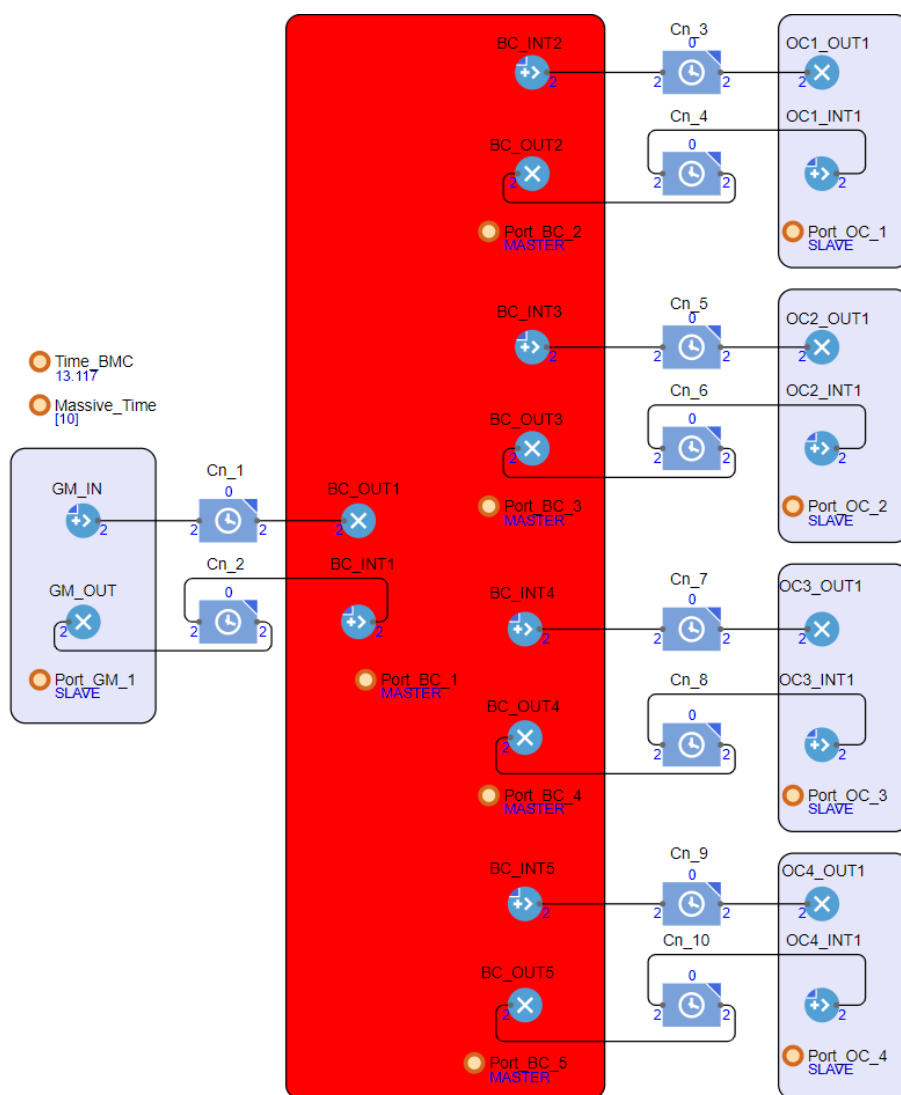


Рис. 5. Результат моделирования процесса функционирования алгоритма ВТСА при атаке манипулирования задержкой сообщения Announce в случае контроля злоумышленником узла BC

Fig. 5. The Results of BTCA Functioning Model When Attacker is Manipulating on Announce Messages from Controlled Node BC

С помощью построенной модели возможно производить оценку времени сходимости ВТСА с учетом возможных атак на различные устройства РТР и в зависимости от величины возможных задержек. Имея в наличии сведения по оценке времени сходимости алгоритма, можно сделать вывод о вероятных атаках на различные устройства РТР и принять меры по их устранению.

Удаление сообщений Announce при организации процесса функционирования алгоритма ВТСА

При атаке с удалением сообщений Announce замедляется процесс построения дерева синхронизации устройств РТР в результате снижения времени сходимости ВТСА за счет повторной отправки сообщений. Подобные виды атак являются труднообна-

руживаемыми и труднофиксируемыми, так как зачастую на потерю сообщений в сетях связи влияет значительное число дестабилизирующих факторов, таких как качество каналов связи, электромагнитная совместимость, качество электропитания устройств и т. д. Тем не менее, замедление процесса функционирования алгоритма влияет на качество синхронизации, а значит – на качество предоставляемых услуг связи и сервисов реального времени конечным потребителям.

Рассмотрим случай, когда злоумышленник также контролирует узел BC и способен удалять сообщения Announce, в результате имитационная модель для оценки процесса функционирования ВТСА примет вид, показанный на рисунке 6 [1, 7, 15]. Исходная модель процесса функционирования алгоритма дополнена блоком «SelectOutput», позволяющим задавать вероятность возникновения ошибок

при передаче сообщения Announce. При возникновении ошибки осуществляется повторная передача сообщения.

Для примера оценим время его сходимости при вероятности удаления сообщения, равной 0,3 на каждый канал передачи и обработки сообщения Announce при исходных данных согласно таблице 2. В результате проведенного моделирования получены следующие результаты. Алгоритм корректно определил состояния портов устройств исходя из их атрибутов, однако с учетом удаления отдельных сообщений Announce значительно возросло время

сходимости алгоритма, которое в данном случае составляет 15,126 мкс в результате дополнительной отправки сообщений.

При помощи разработанной модели возможно производить оценку времени сходимости ВТСА с учетом вероятных атак на сетевые узлы протокола точного времени, когда злоумышленник, завладев узлом РТР, осуществляет умышленное удаление сообщений Announce. Имея в наличии статистику оценки времени сходимости алгоритма, можно сделать вывод о вероятных объектах атаки и предпринять шаги по ликвидации атаки.

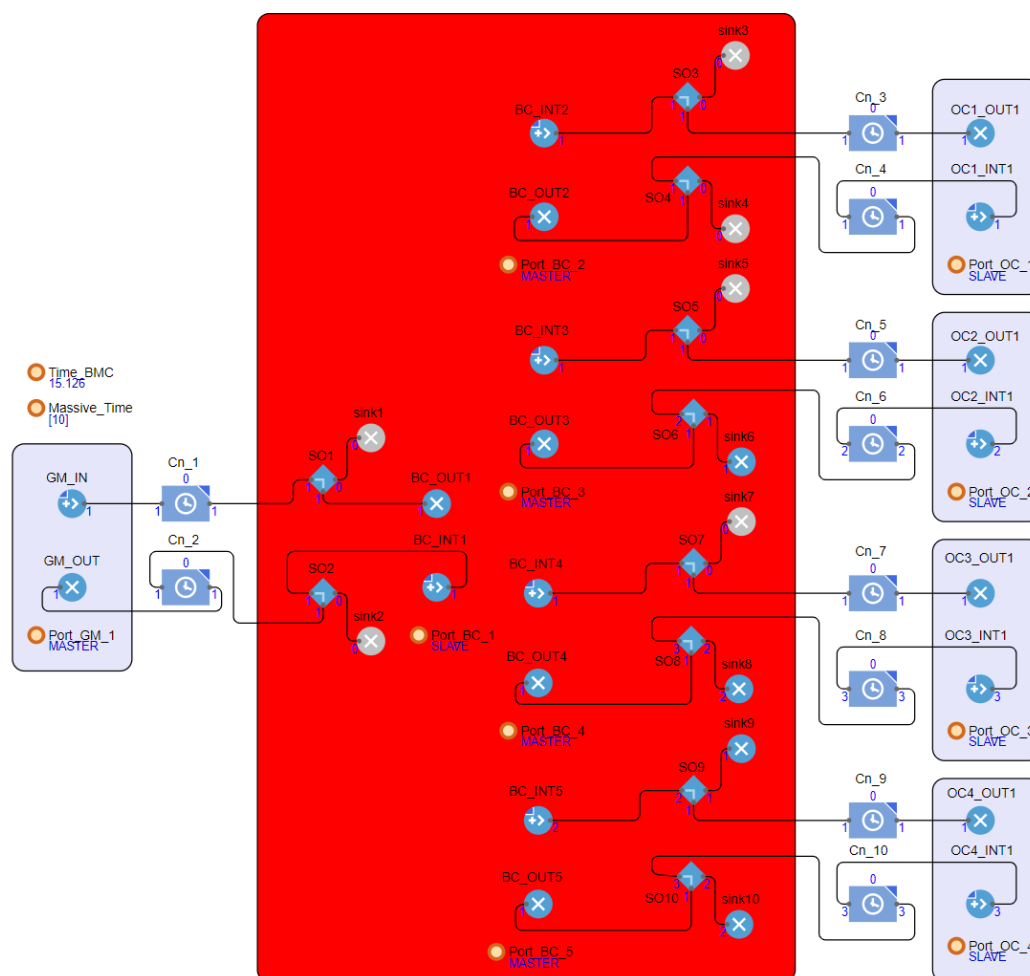


Рис. 6. Результат моделирования процесса функционирования алгоритма ВТСА при атаке удаления сообщений Announce в случае контроля злоумышленником узла BC

Fig. 6. The Results of BTCA Functioning When Attacker is Deleting Announce Messages from Controlled Node BC

Заключение

Алгоритм ВТСА является ключевым компонентом протокола РТР, который обеспечивает синхронизацию времени в пакетно-ориентированных сетях связи. Реализация ВТСА обеспечивает определение устройств РТР с наилучшим качеством, исходя из значений атрибутов портов устройств и характеристик самих устройств. От оперативности и корректности его работы зависит оперативность,

точность и стабильность синхронизации всех узлов в сети связи, а значит и качество предоставляемых услуг, поэтому атаки на элементы системы синхронизации, функционирующие по данному алгоритму, являются одними из вероятных сценариев атак. В статье исследованы наиболее возможные стратегии атак на элементы системы синхронизации, функционирующие по ВТСА, такие как манипулирование содержимым сообщения Announce,

манипулирование задержкой сообщения Announce и удаление сообщения Announce. Разработанный комплекс моделей позволяет отслеживать процесс функционирования алгоритма ВТСА при указанных стратегиях атак, отслеживать состояния портов устройств, оценить качество и достоверность

построения дерева синхронизации и время сходимости алгоритма, что может являться основой для выявления возможных атак. Построенные модели могут быть применены на реальных сетях связи для выявления скомпрометированных устройств при наличии информации о характеристиках узлов связи.

Список источников

1. IEEE 1588-2019 Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems. DOI:10.1109/IEEESTD.2020.9120376
2. Рыжков А.В. Частотно-временное обеспечение в сетях электросвязи: учебное пособие для вузов. М.: Горячая линия – Телеком, 2021. 270 с.
3. Лоховин В.А., Шварц М.Л., Рыжков А.В. Перспективные направления развития систем связи и синхронизации сложных инфраструктурных объектов // Т-Comm: Телекоммуникации и транспорт. 2024. Т. 18. № 11. С. 30–37. DOI:10.36724/2072-8735-2024-18-11-30-37. EDN:UZQBNC
4. Мазуренко Д.К. Аспекты построения системы частотно-временной сетевой синхронизации сигналов // Т-Comm: Телекоммуникации и транспорт. 2017. Т. 11. № 8. С. 4–8. EDN:ZGFBST
5. Прошин Ф.А., Сторожук М.Н., Сторожук Н.Л. Методы синхронизации в сетях связи // Первая миля. 2024. № 2(118). С. 62–69. DOI:10.22184/2070-8963.2024.118.2.62.69. EDN:BQBPFX
6. Alghamdi W., Schukat M. Cyber Attacks on Precision Time Protocol Networks – A Case Study // Electronics. 2020. Vol. 9(9). P. 1398. DOI:10.3390/electronics9091398. EDN:BVQULX
7. Alghamdi W., Schukat M. Precision time protocol attack strategies and their resistance to existing security extensions // Cybersecurity. 2021. Vol. 4(1). P. 12. DOI:10.1186/s42400-021-00080-y. EDN:WBOCZY
8. Moradi M., Jahangir A.H. A Petri net model for Time-Delay Attack detection in Precision Time Protocol-based networks // IET Cyber-Physical Systems: Theory & Applications. 2024. Vol. 9. Iss. 4. PP. 407–423. DOI:10.1049/cps2.12088
9. Moussa B., Kassouf M., Hadjidj R., Debbabi M., Assi C. An Extension to the Precision Time Protocol (PTP) to Enable the Detection of Cyber Attacks // IEEE Transactions on Industrial Informatics. 2019. Vol. 16. Iss. 1. PP. 18–27. DOI:10.1109/TII.2019.2943913
10. Мазуренко Д.К., Пальцин Д.А., Фень А.С., Федоров А.В., Шварц М.Л. Угрозы безопасности для протоколов сетевого времени в сетях с коммутацией пакетов // Электросвязь. 2024. № 10. С. 69–74. DOI:10.34832/ELSV.2024.59.10.010. EDN:LOHBJL
11. Опарин Е.В. Функционирование системы частотно-временного обеспечения железнодорожного транспорта в условиях воздействия дестабилизирующих факторов. М.: ООО «Издательский центр РИОР», 2025. 248 с. DOI:10.29039/02173-6. EDN:VVRPWS
12. Саенко И.Б., Лаута О.С., Карпов М.А., Крибель А.М. Модель угроз ресурсам ИТКС как ключевому активу критически важного объекта инфраструктуры // Электросвязь. 2021. № 1. С. 36–44. DOI:10.34832/ELSV.2021.14.1.004. EDN:OYYZGF
13. Карпов М.А., Митрофанов М.В., Лаута О.С., Пальцин Д.А. Методика управления защитой информационно-телекоммуникационной сети // Электросвязь. 2021. № 12. С. 49–57. DOI:10.34832/ELSV.2021.25.12.007 EDN:CYNHOX
14. Канаев А.К., Опарин Е.В., Горбач А.Н. Моделирование атаки на систему управления сетью синхронизации // Вестник Рязанского государственного радиотехнического университета. 2020. № 72. С. 35–47. DOI:10.21667/1995-4565-2020-72-35-47. EDN:HYKAKY
15. Бунцев И.А. Создание и реализация имитационных моделей в программной среде AnyLogic. М.: Горячая Линия – Телеком, 2016. 154 с.

References

1. IEEE 1588-2019 Standard for A Precision Clock Synchronization Protocol for Networked Measurement and Control Systems. DOI:10.1109/IEEESTD.2020.9120376
2. Ryzhkov A.V. *Frequency-Time Support in Telecommunication Networks*. Moscow: Hotline – Telecom Publ.; 2021. 270 p. (in Russ.)
3. Lokhovin V.A., Schwartz M.L., Ryzhkov A.V. Development trends regarding communication and synchronization systems of complex infrastructure facilities. *T-Comm*. 2024;18(11):30–37. (in Russ.) DOI:10.36724/2072-8735-2024-18-11-30-37. EDN:UZQBNC
4. Mazurenko D.K. Aspects of the building a system of network synchronization signals in frequency, phase shift and time. *T-Comm*. 2017;11(8):4–8. (in Russ.) EDN:ZGFBST
5. Proshin F.A., Storozhuk M.N., Storozhuk N.L. Methods of synchronization communication networks. *Last Mile*. 2024;2(118):62–69. (in Russ.) DOI:10.22184/2070-8963.2024.118.2.62.69. EDN:BQBPFX
6. Alghamdi W., Schukat M. Cyber Attacks on Precision Time Protocol Networks – A Case Study. *Electronics*. 2020;9(9):1398. DOI:10.3390/electronics9091398. EDN:BVQULX

7. Alghamdi W., Schukat M. Precision time protocol attack strategies and their resistance to existing security extensions. *Cybersecurity*. 2021;4(1):12. DOI:10.1186/s42400-021-00080-y. EDN:WBOCZY
8. Moradi M., Jahangir A.H. A Petri net model for Time-Delay Attack detection in Precision Time Protocol-based networks. *IET Cyber-Physical Systems: Theory & Applications*. 2024;9(4):407–423. DOI:10.1049/cps2.12088
9. Moussa B., Kassouf M., Hadjidj R., Debbabi M., Assi C. An Extension to the Precision Time Protocol (PTP) to Enable the Detection of Cyber Attacks. *IEEE Transactions on Industrial Informatics*. 2019;16(1):18–27. DOI:10.1109/TII.2019.2943913
10. Mazurenko D.K., Paltsin D.A., Fen A.S., Fedorov A.V., Schwartz M.L. Security threats of time protocols in packet switched networks. *Electrosvyaz*. 2024;10:69–74. (in Russ.) DOI:10.34832/ELSV.2024.59.10.010. EDN:LOHBJL
11. Oparin E. *Functioning of the frequency-time support system for railway transport under the influence of destabilizing factors*. Moscow: RIOR Publ.; 2025. 248 p. (in Russ.) DOI:10.29039/02173-6. EDN:VVRPWS
12. Sayenko I.B., Lauts O.S., Karpov M.A., Kribel A.M. Model of threats to information and telecommunication network resources as a key asset of critical infrastructure. *Electrosvyaz*. 2021;1:36–44. (in Russ.) DOI:10.34832/ELSV.2021.14.1.004. EDN:OYYZGF
13. Lauts O.S., Karpov M.A., Mitrofanov M.V., Paltsin D.A. Information and telecommunication network security management methodology. *Electrosvyaz*. 2021;12:49–57. (in Russ.) DOI:10.34832/ELSV.2021.25.12.007. EDN:CYNHOX
14. Kanaev A.K., Oparin E.V., Gorbach A.N. Simulation of attack on synchronization network control system. *Vestnik of Ryazan State Radio Engineering University*. 2020;72:35–47. (in Russ.) DOI:10.21667/1995-4565-2020-72-35-47. EDN:HYKAKY
15. Buntsev I.A. *Creation and Implementation of Simulation Models in the AnyLogic Software Environment*. Moscow: Hotline – Telecom Publ.; 2016. 154 p. (in Russ.)

Статья поступила в редакцию 22.12.2025; одобрена после рецензирования 15.01.2026; принята к публикации 26.01.2026

The article was submitted 22.12.2025; approved after reviewing 15.01.2026; accepted for publication 26.01.2026

Информация об авторах:

ОПАРИН
Евгений Валерьевич

кандидат технических наук, доцент, ведущий специалист отдела разработки систем связи ЗАО «Институт телекоммуникаций»
 <https://orcid.org/0000-0003-4622-9161>

ПРОШИН
Федор Алексеевич

кандидат технических наук, ассистент кафедры «Электрическая связь» Петербургского государственного университета путей сообщения Императора Александра I
 <https://orcid.org/0009-0009-5513-9302>

Авторы сообщают об отсутствии конфликтов интересов.

The authors declare no conflicts of interests.

Научная статья

УДК 004.056.53

<https://doi.org/10.31854/1813-324X-2026-12-2-102-112>

EDN:MNHGTS



Модель процесса мониторинга сетевой безопасности сети передачи данных в условиях многоэтапных атак

✉ Александр Александрович Шевченко¹, shevchenko.aa@sut.ru

✉ Вадим Александрович Задбоев² ✉, zadboev89@mail.ru

✉ Валерий Алексеевич Липатников², lipatnikovanl@mail.ru

✉ Кирилл Витальевич Мелехов², kirill_melehov@bk.ru

✉ Павел Игоревич Кузин³, kuzik78@mail.ru

¹Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича, Санкт-Петербург, 193232, Российская Федерация

²Военная академия связи им. Маршала Советского Союза С.М. Буденного, Санкт-Петербург, 194064, Российская Федерация

³Санкт-Петербургский государственный лесотехнический университет им. С.М. Кирова, Санкт-Петербург, 194021, Российская Федерация

Аннотация

Постановка задачи. Развитие современных технологий приводит к тому, что в процессе реализации многоэтапных атак злоумышленники используют интеллектуальные и высокотехнологические инструменты для сокрытия своего воздействия на сетевую инфраструктуру. Противостояние данным атакам является одной из основных задач мониторинга информационной безопасности сети передачи данных. На основе существующей проблемы постоянно требуются новые инструменты противодействия или же оптимизации работы существующих систем обнаружения аномалий, которые позволяют не только эффективнее получать необходимую информацию, но и на ее основе улучшить прогнозирование потенциальных кибератак в будущем.

Цель исследования: определить взаимосвязь вероятностно-временных характеристик процесса мониторинга сетевой безопасности.

Результаты. Предложена структура средства мониторинга сетевой безопасности, в которой используются потоковые двухслойные рекуррентные нейронные сети с управляемыми синапсами для классификации и прогнозирования многоэтапных атак. Проведено моделирование процесса мониторинга сетевой безопасности сети передачи данных с целью определения влияния на него различных факторов. Создано программное обеспечение для расчета вероятностно-временных характеристик процесса мониторинга сетевой безопасности сети передачи данных в условиях воздействия злоумышленника.

Теоретическая значимость: использование предлагаемой модели и программного обеспечения позволяет сформировать требования к различным подпроцессам мониторинга сетевой безопасности сети передачи данных.

Практическая значимость: применение предлагаемой модели возможно в качестве основы при разработке систем по предотвращению многоэтапных атак злоумышленника.

Ключевые слова: моделирование, рекуррентные нейронные сети, киберугроза, управляемые синапсы, информационная безопасность, сеть передачи данных, многоэтапная атака, раннее обнаружение

Ссылка для цитирования: Шевченко А.А., Задбоев В.А., Липатников В.А., Мелехов К.В., Кузин П.И. Модель процесса мониторинга сетевой безопасности сети передачи данных в условиях многоэтапных атак // Труды учебных заведений связи. 2026. Т. 12. № 2. С. 102–112. DOI:10.31854/1813-324X-2026-12-2-102-112. EDN:MNHGTS

Original research

<https://doi.org/10.31854/1813-324X-2026-12-2-102-112>

EDN:MNHGTS

A Model for the Network Security Monitoring Process in Data Transmission Networks Under Multi-Stage Attack Conditions

 **Aleksandr A. Shevchenko**¹, shevchenko.aa@sut.ru

 **Vadim A. Zadboev**²✉, zadboev89@mail.ru

 **Valery A. Lipatnikov**², lipatnikovanl@mail.ru

 **Kirill V. Melekhov**², kirill_melehov@bk.ru

 **Pavel I. Kuzin**³, kuzik78@mail.ru

¹The Bonch-Bruевич Saint Petersburg State University of Telecommunications,
St. Petersburg, 193232, Russian Federation

²Telecommunications Military Academy,
St. Petersburg, 194064, Russian Federation

³Saint-Petersburg State Forest Technical University,
St. Petersburg, 194021, Russian Federation

Annotation

Statement of the problem. The advancement of modern technologies has enabled adversaries to employ intelligent and sophisticated tools during the execution of multi-stage attacks to conceal their activities within the network infrastructure. Countering such attacks constitutes one of the primary objectives of information security monitoring for data transmission networks. Given this persistent challenge, there is a continuous demand for the development of novel countermeasures or the optimization of existing anomaly detection systems. These systems must not only facilitate more efficient acquisition of relevant information but also leverage this information to enhance the prediction of potential cyberattacks.

Purpose: to determine the relationship between the probabilistic and temporal characteristics of the network security monitoring process.

Results. A network security monitoring tool structure is proposed that utilizes streaming two-layer recurrent neural networks with controlled synapses to classify and predict multi-stage attacks. The network security monitoring process for a data transmission network was modeled to determine the impact of various factors. Software was developed to calculate the probabilistic and temporal characteristics of the network security monitoring process for a data transmission network under attacker attack.

Theoretical significance. The model and software used allow for the formulation of requirements for various sub-processes of network security monitoring for a data transmission network.

Practical Significance: The proposed model can serve as a foundational framework for the development of systems designed to prevent multi-stage attacks.

Keywords: modeling, recurrent neural networks, cyber threat, controlled synapses, information security, data network, multi-stage attack, early detection

For citation: Shevchenko A.A., Zadboev V.A., Lipatnikov V.A., Melekhov K.V., Kuzin P.I. A Model for the Network Security Monitoring Process in Data Transmission Networks Under Multi-Stage Attack Conditions. *Proceedings of Telecommunication Universities*. 2026;12(2):102–112. (in Russ.) DOI:10.31854/1813-324X-2026-12-2-102-112. EDN:MNHGTS

Введение

Современные сети передачи данных (СПД) часто могут стать объектами воздействия злоумышленников, реализующегося в виде многоэтапных атак

(МЭА). МЭА, в отличие от одноэтапных атак, представляют собой последовательно взаимосвязанные этапы, разнесенные во времени и логически зависящие друг от друга образующие единый жизненный

цикл. Жизненный цикл МЭА в основном включает в себя пять ключевых этапов: разведку (OSINT), создание инструментов и выбор методов атаки, проникновение в систему через уязвимости, повышение прав в системе и замещение следов для сокрытия своей активности. Благодаря такой особенности МЭА длятся дольше, но имеют высокую степень скрытности и адаптации к защитным мерам.

В процессе реализации МЭА злоумышленники эксплуатируют различные уязвимости в программном, аппаратном и организационном обеспечении сетевой инфраструктуры. Для этого применяются различные инструментариумы в зависимости от текущей задачи. На этапах разведки широко используются сканеры портов и уязвимостей, такие как *Nmap*, *Masscan* и *OpenVAS*. Для проникновения в систему применяются фреймворки *Metasploit*, *Cobalt Strike*, *PowerShell Empire* и аналогичные средства. Для закрепления в системе и повышения прав используются троянские программы, бэкдоры, ботнет-клиенты и средства скрытого удаленного доступа. Для сокрытия своей деятельности применяются методы шифрования трафика, прокси-серверы и VPN.

МЭА характеризуются рядом наблюдаемых признаков. На ранних этапах возможна аномальная сетевая активность в виде нетипичных сканирований или попыток подключения к закрытым портам. На более поздних этапах отмечаются множественные неудачные попытки аутентификации, появление неизвестных процессов и служб на узлах, изменения в системных файлах и реестре или аномальные обращения к ресурсам, не соответствующие профилю пользователя.

В противодействие данным атакам требуются инструменты активной защиты, эффективно выявляющие и предотвращающие деятельность злоумышленников в СПД путем вмешательства в этапы жизненного цикла МЭА. Активная защита предусматривает мониторинг информационной безопасности (ИБ) СПД, который включает в себя своевременный анализ киберугроз, определение геолокации источника кибератаки в совокупности с планированием и принятием мер противодействия злоумышленнику путем его нейтрализации. Ввиду того, что геополитическая обстановка в мире остается напряженной, информационная инфраструктура Российской Федерации, которая включает в себя совокупность СПД органов власти и госкорпораций, ежедневно подвергается колоссальному кибервлиянию, поэтому задача организации мониторинга защищенности СПД в условиях МЭА остается актуальной. Для обеспечения защиты СПД от МЭА необходима разработка средств мониторинга сетевой безопасности (СМСБ), которые позволят своевременно обнаруживать, определять параметры вторжений и геолокацию злоумышленника.

В настоящее время решением вышеуказанной задачи занимаются ведущие вендоры программных и аппаратных систем ИБ, основывающие свои разработки на результатах исследований в данной области. В статье [1] предложена концептуальная модель процесса функционирования системы для предупреждения атак, которая обеспечивает сравнение эталонных профилей с текущими профилями процесса функционирования с учетом динамики изменения состояния системы. В [2] авторы предлагают метод повышения защищенности сети, основанный на использовании средств определения геолокации злоумышленника, который позволяет ускорить расследование инцидентов ИБ. Разрабатываются способы анализа и прогнозирования категорий уязвимостей в конфигурациях устройств с помощью методов искусственного интеллекта, один из таких предложен в [3–4]. Способ позволяет прогнозировать категории CVE на основе CPE URI с высокой точностью. В статье [5] рассматривается вопрос разработки моделей оценки процесса подготовки и реализации вторжений в сетях IP-телефонии, который обеспечивает прогнозирование, управление устойчивостью и защищенностью сети VoIP, позволяющий эффективно выбирать необходимые способы нейтрализации атак.

В данных источниках описываются различные модели угроз, способы их прогнозирования и устранения, однако не рассматриваются подходы к разработке СМСБ.

Целью статьи является определение взаимосвязи вероятностно-временных характеристик процесса мониторинга сетевой безопасности.

В интересах достижения цели необходимо решить следующие задачи:

- разработать структуру СМСБ, в основе которого используется потоковая двухслойная рекуррентная нейронная сеть (РНС) с управляемыми сигналами, которая будет способна выявлять аномалии на ранних стадиях, а также классифицировать распознанные МЭА;
- провести моделирование процесса мониторинга сетевой безопасности СПД при МЭА с целью определения влияния на него различных факторов.

Разработка структуры средства мониторинга сетевой безопасности

В качестве решения поставленной задачи была разработана структура СМСБ (рисунок 1), способная выявлять аномалии на ранних стадиях, классифицировать и прогнозировать распознанные МЭА.

Далее представлены описания каждого блока СМСБ.

Блок обработки цифрового потока (ЦП) отвечает за аналитическую обработку потока данных, осуществляя над ними различные операции, такие как

классификация, регрессия, сегментация и другие [6–7]. Он адаптирует нейросетевую архитектуру под конкретные задачи обработки входящих данных, обеспечивая их корректное восприятие системой.

Блок формирования синтаксической конструкции ЦП и протокола отвечает за организацию и структурирование информации для последующей передачи, кодируя данные согласно установленным правилам [8–9]. Синтаксическая конструкция определяет формат данных, их структуру и правила обмена информацией, включая заголовки, поля данных, контрольные суммы и другие элементы. Этот блок обеспечивает корректное представление информации как внутри системы, так и при внешней передаче, гарантируя совместимость и безопасность данных.

Блок нормализации и кодирования выполняет предварительную обработку данных, приводя их к единому формату для оптимального восприятия РНС [10]. Он отвечает за нормализацию числовых значений, кодирование категориальных данных с помощью метода one-hot-кодирования, обработку текстовых данных через токены, удаление стоп-слов и преобразование слов в числовые векторы с использованием технологий типа Word2Vec или TF-IDF, а также масштабирование данных для улучшения сходимости модели при обучении.

Блок задержки используется для работы с временными рядами, сохраняющие информацию о предыдущих состояниях системы и передачи ее на

следующий временной шаг [11–12]. Этот механизм позволяет учитывать долгосрочные зависимости в данных, что особенно важно при решении задач, связанных с анализом речи, видео или других последовательных данных. Блок может быть реализован с помощью различных архитектур, таких как LSTM или GRU.

Блок декодирования генерирует выходные сигналы на основе внутреннего состояния сети и входной информации [13–14]. Этот блок состоит из нескольких слоев, включая рекуррентные слои для передачи информации о предыдущих состояниях и выходные слои для формирования конечного результата. Для повышения качества генерации выходных данных могут использоваться механизмы внимания (*attention*), которые позволяют сфокусироваться на наиболее важных частях входных данных.

Блок управления (БУ) координирует работу всех компонентов системы, организует процессы обучения нейросети, принимает решения на основе полученных результатов и управляет вычислительными ресурсами [15–16]. Он может использовать различные алгоритмы и стратегии, такие как обратное распространение ошибки, градиентный спуск или регуляризация, для оптимизации производительности модели. Помимо этого, БУ анализирует результаты работы системы и отправляет команды другим системам или устройствам для выполнения конкретных действий.

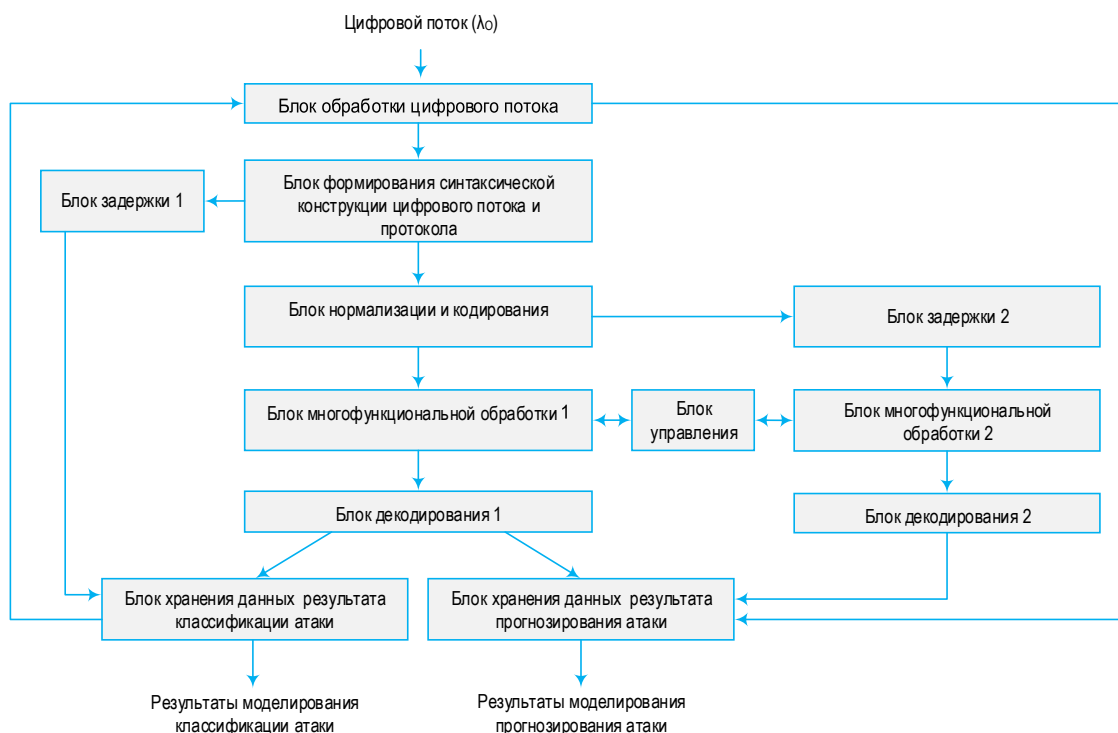


Рис. 1. Структура средства мониторинга сетевой безопасности

Fig. 1. Structure of the Network Security Monitoring Tool

Блок многофункциональной обработки отвечает за обработку последовательных данных. В представленных на рисунках 2 и 3 схемах работает два разных блока, один из которых использует технологии самоорганизующихся карт Кохонена и древовидные модели классификации для анализа зависимостей между элементами данных. Карты Ко-

хонена позволяют выполнять кластеризацию данных и находить скрытые закономерности, а древовидные модели эффективно классифицируют объекты на основе множества признаков. Эти технологии особенно полезны при работе с большими объемами данных, где требуется выявить сложные паттерны и связи.

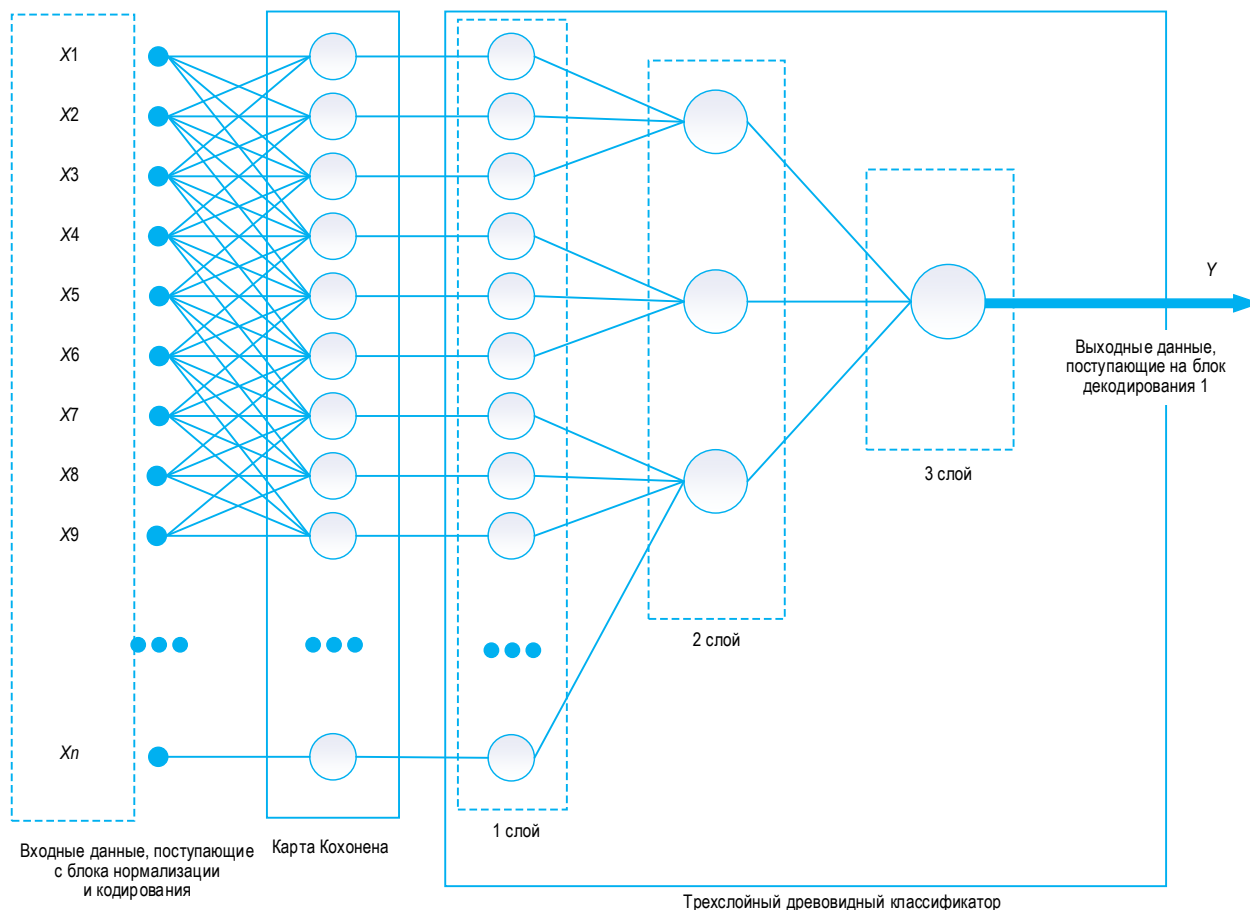


Рис. 2. Схема блока многофункциональной обработки 1

Fig. 2. Multifunctional Processing Unit Diagram 1

Данный блок отвечает за группировку данных ($X_1, X_2, \dots, X_n$), включающих в себя результаты анализа сетевого трафика, параметры клиентских приложений и накопленную информацию. Также он отвечает за фильтрацию результатов и выделение целевого класса, характеризующего состояние СПД и обработку значений с помощью древовидного нейронного классификатора, который определяет принадлежность входных значений к заранее заданному классу, описывающему текущее состояние СПД. На основе входных данных модель позволяет однозначно интерпретировать выходное значение как оценку текущего состояния.

На этапе выявления первичных признаков вторжений идентифицируются структурные, неизменные и корреляционные характеристики проводимой атаки, далее обновляются базы данных и уточняются сценарии атак, а после разрабатываются

оптимальные меры противодействия и нейтрализации возникшей угрозы.

При выявлении вторичных признаков вторжений проводится анализ корреляционных связей между признаками, далее обновляется база данных со вторичными признаками, уточняется сценарий атаки и на основе обработанной информации разрабатываются эффективные меры противодействия.

Второй блок многофункциональной обработки потока предназначен для прогнозирования МЭА и контроля принятия решений первого блока. Он основан на использовании РНС, имеющие обратную связь между выходным и входным слоями, то есть результат предыдущего шага в потоке является входным сигналом для следующего и так далее [17].

Суть работы второго блока заключается в параллельной обработке разработанных мер противодействия на первом блоке и поиске оптимального решения из предложенных на основе прогнозирования поведения атаки. Результаты прогнозирования предоставляются обратно на первый блок для дальнейшей классификации атаки и принятия решения для ее противодействия.

Так как при прогнозировании МЭА необходимо учитывать состав и последовательность исторических данных для расчета будущих значений, предлагается использовать сети Джордана в виде трехслойной РНС с контекстным слоем (рисунок 3). В сети Джордана контекстные нейроны получают сигнал из выходного слоя и подаются обратно во времени на вход сети. Это позволяет модели использовать контекст прошлых состояний для прогнозирования будущих состояний [18].

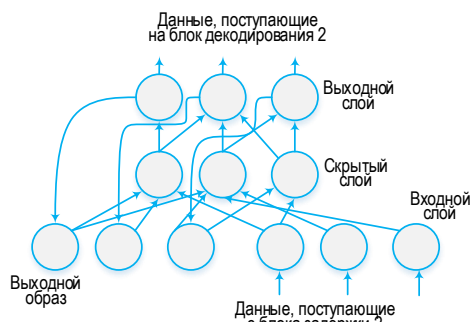


Рис. 3. Схема блока многофункциональной обработки 2

Fig. 3. Multifunctional Processing Unit Diagram 2

Сеть Джордана в виде трехслойной РНС с контекстным слоем имеет полносвязные слои с рекуррентной обратной связью от выходного слоя, позволяющая использовать выходные сигналы на предыдущем шаге для прогнозирования и последующего обучения. Однако следует учитывать одну из ключевых проблем сетей Джордана, связанную с затуханием или взрывом градиентов, что может осложнять процесс обучения и негативно сказываться на производительности модели, особенно при работе с длинными временными последовательностями данных.

По итогу текущего цикла работы СМСБ при параллельной работе двух блоков многофункциональной обработки выводится результат классификации атаки с мерами защиты против нее и прогноз ее возможного поведения. По окончании цикла работы результаты отправляются на входные данные СМСБ для дальнейшего обучения.

Моделирование процесса мониторинга сетевой безопасности передачи данных при многоэтапных атаках

Для корректного описания процесса мониторинга сетевой безопасности СПД при МЭА прове-

дено его преобразование в математическую модель, которая бы учитывала ключевые особенности СМСБ. В качестве возможного подхода используется представление СМСБ как система с переменной структурой, где ее поведение на случайных временных интервалах описывается разными режимами работы: наблюдением, обороной и восстановлением; а переходы между этими структурами зависят от текущих параметров системы, например фазовых координат, и подчиняются вероятностным законам. В отличие от точностных характеристик, отражающих корректность классификации сетевых атак, или временных показателей, определяющих быстродействие системы обнаружения, вероятностно-временные характеристики позволяют оценить вероятность выявления вторжения за заданный промежуток времени. Данные показатели наиболее полно отражают эффективность СМСБ в условиях МЭА. Для такого подхода используется теория марковских процессов, где следующее состояние системы определяется только текущим состоянием, а для анализа динамики вероятностных распределений применяется уравнение Фоккера – Планка – Колмогорова, позволяющее прогнозировать эволюцию системы, минимизировать риски и оптимизировать ее работу в условиях неопределенности, как это показано в [19]. На рисунке 4 изображен граф состояний СМСБ в процессе мониторинга сетевой безопасности СПД при МЭА, где P_i обозначает вероятность нахождения системы в i -м состоянии. В таблице 1 приведены параметры, характеризующие данный процесс.

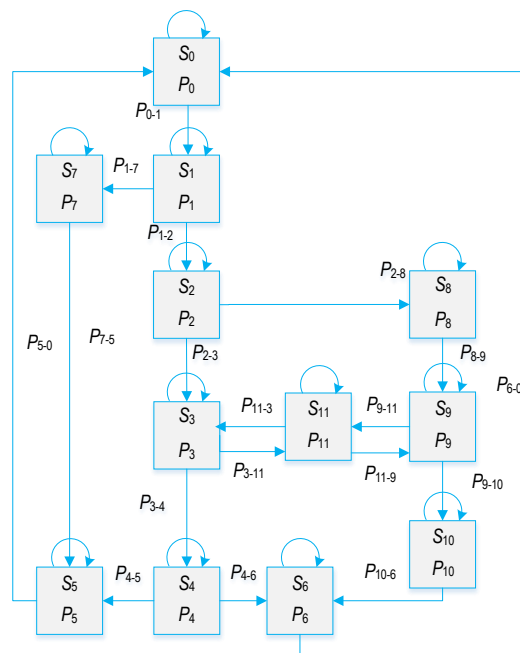


Рис. 4. Граф состояний средства мониторинга сетевой безопасности СПД при МЭА

Fig. 4. State Graph of a Network Security Monitoring Tool for Data Transmission Networks Under a multi-stage attack

ТАБЛИЦА 1. Параметры процесса реализации успешного выявления аномалий

TABLE 1. Parameters of the Process for Implementing Successful Anomaly Detection

S_n	Описание состояния S_n	Интенсивность перехода в состояние S_m	Параметр	Описание параметра	Значение параметра, с
S_0	СМСБ ожидает входной ЦП данных для начала обработки	$\lambda_{0-1} = \frac{1}{\bar{t}_0}$	$\bar{t}_0$	Среднее время обработки входного ЦП данных	2000
S_1	СМСБ производит кодирование и структурирование информации	$\lambda_{1-2/1-7} = \frac{1}{\bar{t}_{\text{ФС}}}$	$\bar{t}_{\text{ФС}}$	Среднее время формирования синтаксической конструкции цифрового потока и протокола	10 20 30
S_2	СМСБ обрабатывает входные данные перед их подачей на РНС	$\lambda_{2-3/2-8} = \frac{1}{\bar{t}_{\text{ОД}}}$	$\bar{t}_{\text{ОД}}$	Среднее время обработки входных данных перед их подачей на РНС	500
$S_{3(9)}$	СМСБ анализирует и моделирует зависимости между элементами последовательности	$\lambda_{3-\frac{4}{3}-11(9-\frac{10}{9}-11)} = \frac{1}{\bar{t}_{\text{AM1(AM2)}}}$	$\bar{t}_{\text{AM1(AM2)}}$	Среднее время анализа и моделирования зависимости между элементами последовательности	110 (115)
$S_{4(10)}$	СМСБ генерирует выходной сигнал или последовательность данных на основе предыдущих скрытых состояний и входных данных	$\lambda_{4-5/4-6(10-6)} = \frac{1}{\bar{t}_{\Gamma 1(\Gamma 2)}}$	$\bar{t}_{\Gamma 1(\Gamma 2)}$	Среднее время генерации выходного сигнала или последовательности данных на основе предыдущих скрытых состояний и входных данных	1 (130)
S_5	СМСБ классифицирует МЭА	$\lambda_{5-0} = \frac{1}{\bar{t}_{\text{СК}}}$	$\bar{t}_{\text{СК}}$	Среднее время классификации МЭА	100
S_6	СМСБ прогнозирует МЭА	$\lambda_{6-0} = \frac{1}{\bar{t}_{\text{СП}}}$	$\bar{t}_{\text{СП}}$	Среднее время прогнозирования МЭА	300
$S_{7(8)}$	СМСБ сохраняет информацию о предыдущих состояниях и передает ее на следующий временной шаг	$\lambda_{7-5(8-9)} = \frac{1}{\bar{t}_{\text{СИ1(СИ2)}}}$	$\bar{t}_{\text{СИ1(СИ2)}}$	Среднее время сохранения информации о предыдущих состояниях и передачи ее на следующий временной шаг	100 (120)
S_{11}	СМСБ обучает и использует РНС	$\lambda_{11-3/11-9} = \frac{1}{\bar{t}_{\text{ОП}}}$	$\bar{t}_{\text{ОП}}$	Среднее время обучения и использования РНС	1000

Усл. обозначения: S_n – состояние системы

На основе полученного графа состояний составим систему дифференциальных уравнений:

$$\begin{cases} \frac{dP_0(t)}{dt} = \lambda_{5-0}P_5(t) + \lambda_{6-0}P_6(t) - \lambda_{0-1}P_0(t); \\ \frac{dP_1(t)}{dt} = \lambda_{0-1}P_0(t) - (\lambda_{1-2} + \lambda_{1-7})P_1(t); \\ \frac{dP_2(t)}{dt} = \lambda_{1-2}P_1(t) - (\lambda_{2-3} + \lambda_{2-8})P_2(t); \\ \frac{dP_3(t)}{dt} = \lambda_{2-3}P_2(t) + \lambda_{11-3}P_{11}(t) - (\lambda_{3-4} + \lambda_{3-11})P_3(t); \\ \frac{dP_4(t)}{dt} = \lambda_{3-4}P_3(t) - (\lambda_{4-5} + \lambda_{4-6})P_4(t); \\ \frac{dP_5(t)}{dt} = \lambda_{4-5}P_4(t) + \lambda_{7-5}P_7(t) - \lambda_{5-0}P_5(t); \\ \frac{dP_6(t)}{dt} = \lambda_{4-6}P_4(t) + \lambda_{10-6}P_{10}(t) - \lambda_{6-0}P_6(t); \\ \frac{dP_7(t)}{dt} = \lambda_{1-7}P_1(t) - \lambda_{7-5}P_7(t); \\ \frac{dP_8(t)}{dt} = \lambda_{2-8}P_2(t) - \lambda_{8-9}P_8(t); \\ \frac{dP_9(t)}{dt} = \lambda_{8-9}P_8(t) + \lambda_{11-9}P_{11}(t) - (\lambda_{9-10} + \lambda_{9-11})P_9(t); \\ \frac{dP_{10}(t)}{dt} = \lambda_{9-10}P_9(t) - \lambda_{10-6}P_{10}(t); \\ \frac{dP_{11}(t)}{dt} = \lambda_{3-11}P_3(t) + \lambda_{9-11}P_9(t) - (\lambda_{11-9} + \lambda_{11-3})P_{11}(t). \end{cases}$$

Решение системы уравнений представлено выражением:

$$\begin{cases} P_0 = \alpha; \\ P_1 = a\alpha; \\ P_2 = b\alpha; \\ P_3 = c\alpha; \\ P_4 = d\alpha; \\ P_5 = e\alpha; \\ P_6 = f\alpha; \\ P_7 = g\alpha; \\ P_8 = h\alpha; \\ P_9 = j\alpha; \\ P_{10} = k\alpha; \\ P_{11} = l\alpha, \end{cases}$$

где:

$$\alpha = \frac{1}{1 + a + b + c + d + e + f + g + h + i + j + k},$$

$$a = \frac{\lambda_{0-1}}{\lambda_{1-2} + \lambda_{1-7}},$$

$$b = \frac{\lambda_{0-1}\lambda_{1-2}}{(\lambda_{1-2} + \lambda_{1-7})(\lambda_{2-3} + \lambda_{2-8})},$$

$$c = \frac{\lambda_{0-1}\lambda_{1-2}(\lambda_{2-3} + \lambda_{11-3}b)}{(\lambda_{1-2} + \lambda_{1-7})(\lambda_{2-3} + \lambda_{2-8})(\lambda_{3-4} + \lambda_{3-11})},$$

$$d = \frac{\lambda_{0-1}\lambda_{1-2}\lambda_{3-4}(\lambda_{2-3} + \lambda_{11-3}b)}{(\lambda_{1-2} + \lambda_{1-7})(\lambda_{2-3} + \lambda_{2-8})(\lambda_{3-4} + \lambda_{3-11})(\lambda_{4-5} + \lambda_{4-6})},$$

$$e = \frac{\lambda_{0-1}}{\lambda_{5-0}(\lambda_{1-2} + \lambda_{1-7})} \times$$

$$\times \left(\frac{\lambda_{1-2}\lambda_{3-4}\lambda_{4-5}(\lambda_{2-3} + \lambda_{11-3}b)}{(\lambda_{2-3} + \lambda_{2-8})(\lambda_{3-4} + \lambda_{3-11})(\lambda_{4-5} + \lambda_{4-6})} + \lambda_{1-7} \right),$$

$$f = \frac{\lambda_{0-1}\lambda_{1-2}}{\lambda_{6-0}(\lambda_{1-2} + \lambda_{1-7})(\lambda_{2-3} + \lambda_{2-8})} \times$$

$$\times \left(\frac{\lambda_{3-4}(\lambda_{2-3} + \lambda_{11-3}c)}{(\lambda_{3-4} + \lambda_{3-11})(\lambda_{4-5} + \lambda_{4-6})} + \frac{\lambda_{9-10}(\lambda_{2-8} + \lambda_{11-9}b)}{\lambda_{9-10} + \lambda_{9-11}} \right),$$

$$g = \frac{\lambda_{0-1}\lambda_{1-7}}{\lambda_{7-5}(\lambda_{1-2} + \lambda_{1-7})},$$

$$h = \frac{\lambda_{0-1}\lambda_{1-2}\lambda_{2-8}}{\lambda_{8-9}(\lambda_{1-2} + \lambda_{1-7})(\lambda_{2-3} + \lambda_{2-8})},$$

$$i = \frac{\lambda_{0-1}\lambda_{1-2}(\lambda_{2-8} + \lambda_{11-9}b)}{(\lambda_{1-2} + \lambda_{1-7})(\lambda_{2-3} + \lambda_{2-8})(\lambda_{9-10} + \lambda_{9-11})},$$

$$j = \frac{\lambda_{0-1}\lambda_{1-2}\lambda_{9-10}(\lambda_{2-8} + \lambda_{11-9}b)}{\lambda_{10-6}(\lambda_{1-2} + \lambda_{1-7})(\lambda_{2-3} + \lambda_{2-8})(\lambda_{9-10} + \lambda_{9-11})},$$

$$k = \frac{\lambda_{0-1}\lambda_{1-2}b}{(\lambda_{1-2} + \lambda_{1-7})(\lambda_{2-3} + \lambda_{2-8})}.$$

Успешное выявление аномалий (P_A) может быть в двух состояниях, а именно при реализации классификации (P_5) или при прогнозировании кибератаки (P_6), где в конечном итоге процесс переходит к выявлению последующей аномалии, то есть система должна быть эффективной в независимости от режима работы или других состояний СМСБ.

С учетом вышесказанного вероятность успешного выявления аномалии примет вид:

$$\begin{cases} P_A = \alpha(1 + e), \text{ если кибератака неизвестна;} \\ P_A = \alpha(1 + f), \text{ если кибератака известна.} \end{cases}$$

На основе системы уравнений с исходными данными из таблицы 1 была проведена верификация разработанной модели. В целях автоматизации будущих расчетов зависимостей было разработано программное обеспечение [20–21], интерфейсы которого представлены на рисунке 5. Программное обеспечение в консольном окне (рисунок 5а) получает на вход все требуемые значения параметров и проводит расчет вероятностей на основе заранее подготовленных формул. По окончании вычислений для наглядного представления выводятся графики (рисунок 5б) полученных зависимостей.

В результате успешных расчетов получены зависимости вероятности успешного выявления аномалий (P_A) от времени обработки входных данных ($t_{од}$)

в виде закодированной и структурированной информации перед их подачей на РНС в случаях, когда обнаружена:

- 1) неизвестная кибератака и необходима ее дальнейшая классификация (рисунок 6а);
- 2) известная кибератака и необходимо прогнозировать ход ее возможной реализации (рисунок 6б).

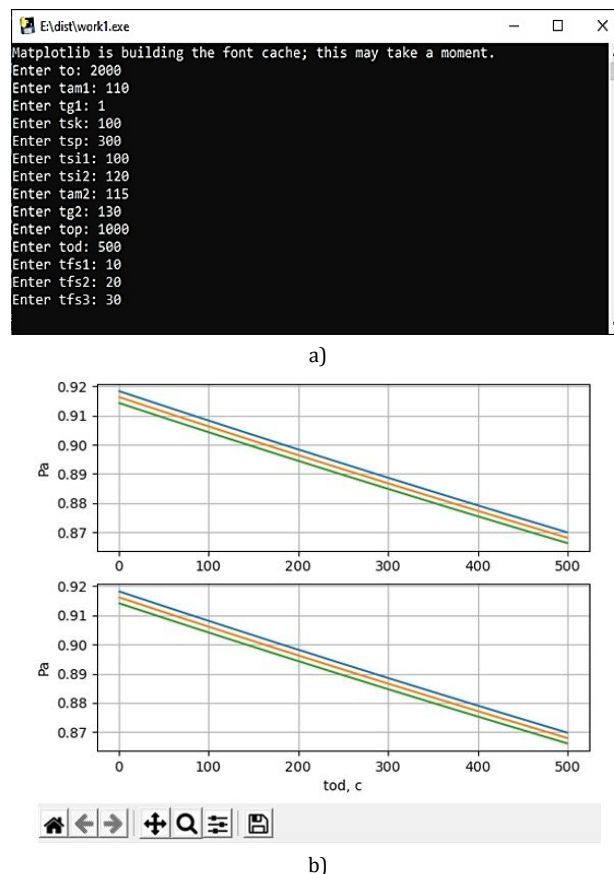


Рис. 5. Интерфейс программного продукта «Программа расчета вероятностно-временных характеристик средств сетевого контроля в условиях многоэтапных атак»:
а) консоль ввода исходных данных; б) поле вывода графиков

Fig. 5. Interface of the Software Product «Program for Calculating the Probabilistic-Temporal Characteristics of Network Control Tools Under Conditions of Multi-Stage Attacks»: a) Console for Inputting Initial Data; b) Field for Outputting Graphs

Анализ зависимостей, представленных на рисунке 6, позволяет сделать следующие выводы:

- вероятность выявления аномалии (P_A) зависит от времени кодирования и структурирования информации ($t_{фс}$), в связи с этим при разработке СМСБ необходимо использовать оптимальные по времени методы реализации данного подпроцесса;
- на обработку данных процесса выявления аномалий в случае реализации злоумышленником неизвестной МЭА тратится больше на 5 %, чем при реализации известной МЭА, следовательно, необходимо при разработке СМСБ также использовать оптимальный по времени метод кодирования и структурирования информации, либо наиболее быстродействующие методы обработки данных.

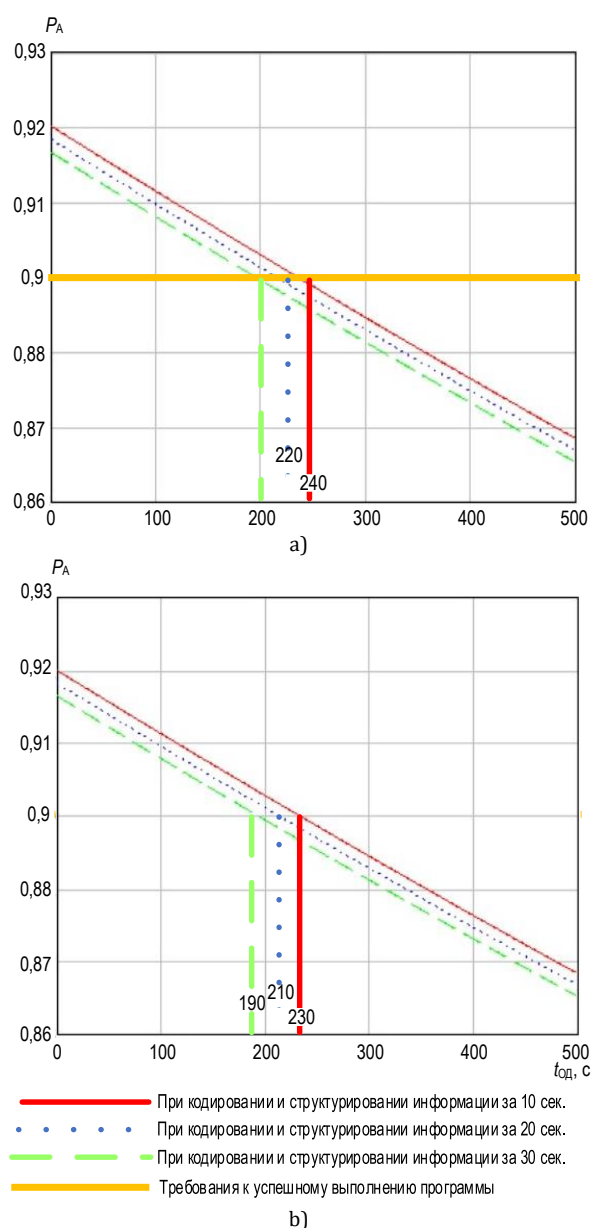


Рис. 6. Изменение вероятности P_A успешного выявления аномалий в случае реализации злоумышленником:
а) неизвестной МЭА; б) известной МЭА

Fig. 6. Change in the Probability of P_A Successful Detection of Anomalies:
a) in the Case of an Unknown Multi-Stage Attack by an Intruder;
b) in the Case of an Attacker Implementing a Known Multi-Stage Attack

Список источников

1. Котенко И.В., Саенко И.Б., Захарченко Р.И., Величко Д.В. Динамическая модель контроля функционирования для предупреждения компьютерных атак // Правовая информатика. 2024. № 2. С. 35–43. DOI:10.21681/1994-1404-2024-2-35-43. EDN:OGOTEF
2. Липатников В.А., Задбоев В.А., Мелехов К.В., Шевченко А.А. Метод повышения защищенности информационно-телекоммуникационной сети с учетом использования средств определения геолокации нарушителя // Труды учебных заведений связи. 2023. Т. 9. № 4. С. 86–96. DOI:10.31854/1813-324X-2023-9-4-86-96. EDN:FWQHUC
3. Левшун Д.С., Веснин Д.В., Котенко И.В. Прогнозирование категорий уязвимостей в конфигурациях устройств с помощью методов искусственного интеллекта // Вопросы кибербезопасности. 2024. № 3(61). С. 33–39. DOI:10.21681/2311-3456-2024-3-33-39. EDN:FTORLR
4. Kim S., Park K.J., Lu C.A Survey on Network Security for Cyber-Physical Systems: From Threats to Resilient Design // IEEE Communications Surveys & Tutorials. 2022. Vol. 24. Iss. 3. PP. 1534–1573. DOI:10.1109/COMST.2022.3187531. EDN:ZEUEHY
5. Липатников В.А., Косолапов В.С., Шевченко А.А., Сокол Д.С. Модель оценки процесса подготовки и реализации

Заключение

В рамках исследования была предложена структура СМСБ, в которой используются потоковые двухслойные РНС с управляемыми синапсами для классификации и прогнозирования МЭА.

Основной функцией предлагаемого СМСБ является мониторинг сетевой безопасности, который является сложным технологическим процессом, зависящим от множества внутренних и внешних факторов. В связи с этим проведено моделирование данного процесса при МЭА с целью определения влияния на него различных факторов. На основе разработанной модели создано программное обеспечение для расчета вероятностно-временных характеристик процесса мониторинга сетевой безопасности СПД в условиях воздействия злоумышленника. Использование модели и программного обеспечения позволяет сформировать требования к различным подпроцессам мониторинга сетевой безопасности СПД, которые возможно будет использовать как критерии выбора технологий, методов и способов при разработке СМСБ.

Результаты моделирования процесса мониторинга сетевой безопасности СПД при МЭА позволяют сформулировать логичные корректные выводы по взаимосвязи вероятностно-временных характеристик данного процесса, что говорит об адекватности предлагаемых решений.

Научная новизна приведенных исследований заключается в том, что представленный подход к мониторингу и прогнозированию сетевой безопасности СПД при МЭА отличается использованием двухслойных РНС с управляемыми синапсами.

Теоретическая значимость заключается в использовании основных положений теорий массового обслуживания, вероятности, марковских цепей для развития методологических положений теории управления защищенностью систем.

Практическая значимость определяется возможностью использования предложенной модели в качестве основы при разработке систем по предотвращению МЭА злоумышленника.

вторжений в сетях IP-телефонии // Информация и космос. 2021. № 4. С. 55–69. EDN:FRPUPC

6. Zhou P., Zhou G., Wu D., Fei M. Detecting multi-stage attacks using sequence-to-sequence model // Computers & Security. 2021. Vol. 105. P. 102203. DOI:10.1016/j.cose.2021.102203. EDN:KUKUOO

7. Mishra S., Alotaibi W.B., Alshehri M., Saxena S. Cyber-attacks visualisation and prediction in complex multi-stage network // International Journal of Computer Applications in Technology. 2022. Vol. 68. Iss. 4. PP. 345–356. DOI:10.1504/IJCAT.2022.125180. EDN:CQPVYR

8. Weerakody P.B., Wong K.W., Wang G., Ela W. A review of irregular time series data handling with gated recurrent neural networks // Neurocomputing. 2021. Vol. 441. PP. 161–178. DOI:10.1016/j.neucom.2021.02.046. EDN:JKFTCG

9. Olszewski D. A data-scattering-preserving adaptive self-organizing map // Engineering Applications of Artificial Intelligence. 2021. Vol. 105. P. 104420. DOI:10.1016/j.engappai.2021.104420. EDN:EGGQEI

10. Осипов В.Ю., Никифоров В.В. Кодирование и устойчивость обработки сигналов в потоковых рекуррентных нейронных сетях // Информационно-управляющие системы. 2021. № 3(112). С. 9–18. DOI:10.31799/1684-8853-2021-3-9-18. EDN:TMUFQK

11. Al-Turaiki I., Altwaijry N. A Convolutional Neural Network for Improved Anomaly-Based Network Intrusion Detection // Big Data. 2021. Vol. 9. Iss. 3. PP. 233–252. DOI:10.1089/big.2020.0263. EDN:BINXTC

12. Sarker I.H. Deep Cybersecurity: A Comprehensive Overview from Neural Network and Deep Learning Perspective // SN Computer Science. 2021. Vol. 2. Iss. 3. P. 154. DOI:10.1007/s42979-021-00535-6. EDN:EKIHP

13. Gong J. Network Information Security Pipeline Based on Grey Relational Cluster and Neural Networks // Proceedings of the 5th International Conference on Computing Methodologies and Communication (Erode, India, 08–10 April 2021). IEEE, 2021. PP. 971–975. DOI:10.1109/ICCMCS1019.2021.9418311

14. Khan M.A. HCRNNIDS: Hybrid Convolutional Recurrent Neural Network-Based Network Intrusion Detection System // Processes. 2021. Vol. 9. Iss. 5. P. 834. DOI:10.3390/pr9050834. EDN:BBATMC

15. Долгачев М.В., Москвичев А.Д., Москвичева К.С. Обнаружение атак на веб-приложение с помощью самоорганизующихся карт Кохонена // Вопросы кибербезопасности. 2024. № 1(59). С. 38–44. DOI:10.21681/2311-3456-2024-1-38-44. EDN:KHTKXR

16. Плетенкова А.Д., Соколов А.Н. Применение двухэтапного метода кластеризации на основе самоорганизующейся карты Кохонена для обнаружения аномалий в синтетических наборах данных // Вестник УрФО. Безопасность в информационной сфере. 2024. Т. 4(54). С. 49–60. DOI:10.14529/secur240406. EDN:ZLGTJQ

17. Pinto A., Herrera L.-C., Donoso Y., Gutierrez J.A. Survey on Intrusion Detection Systems Based on Machine Learning Techniques for the Protection of Critical Infrastructure // Sensors. 2023. Vol. 23. Iss. 5. P. 2415. DOI:10.3390/s23052415. EDN:GBGCLV

18. Yang H., Li X., Qiang W., Zhao Y., Zhang W., Tang C. A network traffic forecasting method based on SA optimized ARIMA-BP neural network // Computer Networks. 2021. Vol. 193. P. 108102. DOI:10.1016/j.comnet.2021.108102. EDN:NKGPOO

19. Липатников В.А., Шевченко А.А. Математическая модель процесса управления информационной безопасностью распределенной информационной системы в условиях несанкционированного воздействия злоумышленника // Информационные системы и технологии. 2022. № 3(131). С. 121–130. EDN:KSBCGK

20. Робак В.А., Липатников В.А., Парфилов В.А., Задбоев В.А., Шевченко А.А., Петренко М.И. и др. Программа расчета вероятностно-временных характеристик средств сетевого контроля в условиях многоэтапных атак. Свидетельство о регистрации программы для ЭВМ № RU 2024661259 от 11.04.24. Опубл. 16.05.24. EDN:WTNJLT

21. Савина А.Г., Малявкина Л.И., Герасимова Ю.Я., Жилина Д.Е. Язык программирования Python в научных вычислениях // Национальная научно-практическая конференция «Инфраструктура цифрового развития образования и бизнеса» (Орел, Российская Федерация, 01–30 апреля 2021 г.). Орёл: Орловский государственный университет экономики и торговли, 2021. С. 64–69. EDN:CPPPGG

References

1. Kotenko I., Saenko I., Zakharchenko R., Velichko D. A Dynamic Functioning Control Model for Preventing Computer Attacks. *Legal Informatics*. 2024;2:35–43. (in Russ.) DOI:10.21681/1994-1404-2024-2-35-43. EDN:OGOTEF

2. Lipatnikov V., Zadboev V., Melekhov K., Shevchenko A. A Method of Improving the Security of Information and Telecommunications Network Using the Means of Determining Intruder's Geolocation. *Proceedings Telecommunication Universities*. 2023;9(4):86–96. (in Russ.) DOI:10.31854/1813-324X-2023-9-4-86-96. EDN:FWQHUC

3. Levshun D.S., Vesnin D.V., Kotenko I.V. Prediction of Vulnerability Categories in Configurations of Devices Using Artificial Intelligence Methods. *Cybersecurity Issues*. 2024;3(61):33–39. (in Russ.) DOI:10.21681/2311-3456-2024-3-33-39. EDN:FTORLR

4. Kim S., Park K.J., Lu C. A Survey on Network Security for Cyber-Physical Systems: From Threats to Resilient Design. *IEEE Communications Surveys & Tutorials*. 2022;24(3):1534–1573. DOI:10.1109/COMST.2022.3187531. EDN:ZEUEHY

5. Lipatnikov V.A., Kosolapov V.S., Shevchenko A.A., Sokol D.S. A Model for Evaluating the Process of Preparing and Implementing Intrusions in IP Telephony Networks. *Information and Space*. 2021;4:55–69. (in Russ.) EDN:FRPUPC

6. Zhou P., Zhou G., Wu D., Fei M. Detecting multi-stage attacks using sequence-to-sequence model. *Computers & Security*. 2021;105:102203. DOI:10.1016/j.cose.2021.102203. EDN:KUKUOO

7. Mishra S., Alotaibi W.B., Alshehri M., Saxena S. Cyber-attacks visualisation and prediction in complex multi-stage network. *International Journal of Computer Applications in Technology*. 2022;68(4):345–356. DOI:10.1504/IJCAT.2022.125180. EDN:CQPVYR

8. Weerakody P.B., Wong K.W., Wang G., Ela W. A review of irregular time series data handling with gated recurrent neural networks. *Neurocomputing*. 2021;441:161–178. DOI:10.1016/j.neucom.2021.02.046. EDN:JKFTCG

9. Olszewski D. A data-scattering-preserving adaptive self-organizing map. *Engineering Applications of Artificial Intelligence*. 2021;105:104420. DOI:10.1016/j.engappai.2021.104420. EDN:EGGQEI

10. Osipov V.Yu., Nikiforov V.V. Coding and stability of signal processing in streaming recurrent neural networks. *Information and Control Systems*. 2021;3(112):9–18. (in Russ.) DOI:10.31799/1684-8853-2021-3-9-18. EDN:TMUFQK

11. Al-Turaiki I., Altwaijry N. A convolutional neural network for improved anomaly-based network intrusion detection. *Big Data*. 2021;9(3):233–252. DOI:10.1089/big.2020.0263. EDN:BINXTC
12. Sarker I.H. Deep Cybersecurity: A Comprehensive Overview from Neural Network and Deep Learning Perspective. *SN Computer Science*. 2021;2(3):154. DOI:10.1007/s42979-021-00535-6. EDN:EKIIP
13. Gong J. Network Information Security Pipeline Based on Grey Relational Cluster and Neural Networks // Proceedings of the 5th International Conference on Computing Methodologies and Communication, 08–10 April 2021, Erode, India. IEEE, 2021. p.971–975. DOI:10.1109/ICCMC51019.2021.9418311
14. Khan M.A. HCRNNIDS: Hybrid convolutional recurrent neural network-based network intrusion detection system. *Processes*. 2021;9(5):834. DOI:10.3390/pr9050834. EDN:BBATMC
15. Dolgachev M.V., Moskvichev A.D., Moskvicheva K.S. Detection of attacks on a web application using self-organizing Kohonen maps. *Cybersecurity Issues*. 2024;1(59):38–44. (in Russ.) DOI:10.21681/2311-3456-2024-1-38-44. EDN:KHTKXR
16. Pletenkova A.D., Sokolov A.N. Application of a two-stage clustering method based on the Kohonen self-organizing map for detecting anomalies in synthetic data sets. *Journal of the Ural Federal District. Information security*. 2024;4(54):49–60. (in Russ.) DOI:10.14529/secur240406. EDN:ZLGTJQ
17. Pinto A., Herrera L.-C., Donoso Y., Gutierrez J.A. Survey on Intrusion Detection Systems Based on Machine Learning Techniques for the Protection of Critical Infrastructure. *Sensors*. 2023;23(5):2415. DOI:10.3390/s23052415. EDN:GBGCLV
18. Yang H., Li X., Qiang W., Zhao Y., Zhang W., Tang C. A network traffic forecasting method based on SA optimized ARIMA-BP neural network. *Computer Networks*. 2021;193:108102. DOI:10.1016/j.comnet.2021.108102. EDN:NKGPOO
19. Lipatnikov V.A., Shevchenko A.A. Mathematical Model of Information Security Management Process for a Distributed Information System Under Conditions of Unauthorized Attacker Impact. *Information systems and technologies*. 2022;3(131):121–130. (in Russ.) EDN:KSBCGK
20. Robak V.A., Lipatnikov V.A., Parfirov V.A., Zadboev V.A., Shevchenko A.A., Petrenko M.I., et al. *Program for Calculating Probabilistic-Temporal Characteristics of Network Control Tools Under Conditions of Multi-Stage Attacks*. Patent RF, no. 2024661259, 11.04.2024. (in Russ.) EDN:WTNJT
21. Savina A.G., Malyavkina L.I., Gerasimova Yu.Ya., Zhilina D.E. Python Programming Language in Scientific Computing. *Proceedings of the National Scientific and Practical Conference on Infrastructure for Digital Development of Education and Business, 1–30 April 2021, Orel, Russian Federation*. Orel: Oryol State University of Economics and Trade Publ.; 2021. p.64–69 (in Russ.) EDN:CPPPGG

Статья поступила в редакцию 08.12.2025; одобрена после рецензирования 03.02.2026; принята к публикации 09.02.2026

The article was submitted 08.12.2025; approved after reviewing 03.02.2026; accepted for publication 09.02.2026

Информация об авторах:

ШЕВЧЕНКО
Александр Александрович

кандидат технических наук, доцент кафедры защищенных систем связи Санкт-Петербургского государственного университета телекоммуникаций им. проф. М.А. Бонч-Бруевича
 <https://orcid.org/0000-0001-9113-1089>

ЗАДБОВЕ
Вадим Александрович

младший научный сотрудник научно-исследовательского центра Военной академии связи им. Маршала Советского Союза С.М. Буденного
 <https://orcid.org/0009-0003-9362-1307>

ЛИПАТНИКОВ
Валерий Алексеевич

доктор технических наук, профессор, старший научный сотрудник научно-исследовательского центра Военной академии связи им. Маршала Советского Союза С.М. Буденного
 <https://orcid.org/0000-0002-3736-4743>

МЕЛЕХОВ
Кирилл Витальевич

кандидат технических наук, преподаватель кафедры организации боевой подготовки и повседневной деятельности войск связи Военной академии связи им. Маршала Советского Союза С.М. Буденного
 <https://orcid.org/0009-0007-3474-412X>

КУЗИН
Павел Игоревич

кандидат технических наук, доцент, доцент кафедры информационных систем и технологий Санкт-Петербургского государственного лесотехнического университета им. С.М. Кирова
 <https://orcid.org/0000-0003-0880-6204>

Авторы сообщают об отсутствии конфликтов интересов.

The authors declare no conflicts of interests.

Научная статья
УДК 004.732.056
<https://doi.org/10.31854/1813-324X-2026-12-2-113-120>
EDN:OETOU1

Синтез реалистичных синтетических данных с помощью диффузионной модели TabDDPM в задачах обнаружения сетевых атак

Олег Иванович Шелухин, sheluhin@mail.ru
Фёдор Андреевич Маторин✉, f.a.matorin@mtuci.ru

Московский технический университет связи и информатики,
Москва, 111024, Российская Федерация

Аннотация

Актуальность. Для исследований и тестирования средств обнаружения атак требуются репрезентативные наборы сетевого трафика, однако их сбор и разметка трудозатратны, а распространение ограничено требованиями конфиденциальности и рисками утечек. Синтетические данные позволяют увеличить объем выборки и моделировать редкие сценарии и сценарии «нулевого дня» при сохранении статистических свойств сетевого трафика.

Цель: повышение качества и воспроизводимости формирования синтетических табличных признаков сетевого трафика на примере Android-приложений за счет применения диффузионной модели Tabular Denoising Diffusion (TabDDPM) и комплексной валидации результатов генерации на согласованном наборе метрик.

Методы. Использована диффузионная генеративная модель TabDDPM, применимая к произвольным табличным наборам данных. Эффективность генерации оценивается методами статистического анализа: сравнение распределений и зависимостей признаков, проверка полезности в прикладной задаче, а также оценка отличия синтетических данных от реальных.

Результат. Выполнен комплексный анализ качества TabDDPM при генерации табличных признаков сетевого трафика атак или нежелательных приложений. Показана возможность формирования контролируемых синтезированных наборов данных, сохраняющих характерные паттерны трафика и обеспечивающих масштабирование обучающих выборок без прямого копирования исходных записей.

Новизна. Предложен согласованный протокол постгенерационной валидации синтетического трафика, сочетающий метрики реалистичности, полезности и неотличимости, что снижает риск некорректных выводов при разрозненной оценке. Для количественной оценки качества генерации синтезируемых данных предложен интегральный показатель качества на основе партиципальных метрик.

Теоретическая значимость состоит в развитии методического подхода к верификации табличных диффузионных моделей для задач кибербезопасности.

Практическая значимость заключается в возможности применения получаемых синтетических наборов данных для моделирования компьютерных атак, сценариев «нулевого дня», стресс-тестирования и обучения / проверки систем обнаружения вторжений.

Ключевые слова: машинное обучение, генеративно-состязательные сети, TabDDPM, диффузионные модели, синтезированный трафик, Android-приложения, табличные данные, метрики, обнаружение вторжений

Ссылка для цитирования: Шелухин О.И., Маторин Ф.А. Синтез реалистичных синтетических данных с помощью диффузионной модели TabDDPM в задачах обнаружения сетевых атак // Труды учебных заведений связи. 2026. Т. 12. № 2. С. 113–120. DOI:10.31854/1813-324X-2026-12-2-113-120. EDN:OETOU1

Original research

<https://doi.org/10.31854/1813-324X-2026-12-2-113-120>

EDN:OETOU

Realistic Synthetic Data Generation Using the TabDDPM Diffusion Model for Network Attack Detection

Oleg I. Sheluhin, sheluhin@mail.ru

Fedor A. Matorin✉, f.a.matorin@mtuci.ru

Moscow Technical University of Communications and Informatics,
Moscow, 111024, Russian Federation

Annotation

Relevance. Representative network-traffic datasets are required for research and testing of attack detection tools; however, their collection and annotation are labor-intensive, and data sharing is constrained by confidentiality requirements and the risk of leakage. Synthetic data can increase sample sizes and enable modeling of rare and zero-day scenarios while preserving the statistical properties of network traffic.

Objective. To improve the quality and reproducibility of generating synthetic tabular features of network traffic, using Android applications as a case study, by applying the Tabular Denoising Diffusion model (TabDDPM) and performing comprehensive validation of the generated data using a consistent set of metrics.

Methods. We employ the TabDDPM diffusion-based generative model, which is applicable to arbitrary tabular datasets. Generation performance is assessed via statistical analysis methods, including comparisons of feature distributions and inter-feature dependencies, evaluation of utility in a downstream task, and estimation of the discrepancy between synthetic and real data.

Results. A comprehensive quality assessment of TabDDPM is conducted for generating tabular features of network traffic associated with attacks or unwanted applications. The results demonstrate the feasibility of producing controlled synthetic datasets that preserve characteristic traffic patterns and enable scaling of training samples without directly copying the original records.

Novelty. We propose a unified post-generation validation protocol for synthetic traffic that integrates realism, utility, and indistinguishability metrics, thereby reducing the risk of misleading conclusions arising from fragmented evaluation. In addition, an integral quality indicator is introduced to quantify generation performance by aggregating partial metrics.

The theoretical significance lies in advancing a methodological framework for verifying tabular diffusion models in cybersecurity applications.

The practical significance is the ability to use the resulting synthetic datasets to model cyberattacks and zero-day scenarios, perform stress testing, and train and / or evaluate intrusion detection systems.

Keywords: machine learning, generative adversarial network; TabDDPM, diffusion models, synthetic traffic, Android applications, tabular data, metrics, intrusion detection

For citation: Sheluhin O.I., Matorin F.A. Realistic Synthetic Data Generation Using the TabDDPM Diffusion Model for Network Attack Detection. *Proceedings of Telecommunication Universities*. 2026;12(2):113–120. (in Russ.) DOI:10.31854/1813-324X-2026-12-2-113-120. EDN:OETOU

Введение

В условиях, когда получение большого объема реального сетевого трафика и сетевых атак сопряжено с юридическими, техническими и этическими трудностями, особую актуальность приобретают методы, позволяющие искусственно расширить

обучающее множество без ухудшения его качества и репрезентативности. Генерация синтезированных данных является мощным и гибким подходом и заключается в создании новых, ранее не существовавших примеров на основе генеративных моделей.

Наибольшую известность получили несколько методов создания синтетических данных. Под «синтетическими» понимаются искусственно созданные данные, имитирующие реальные наблюдения и используемые для подготовки моделей машинного обучения, в условиях, когда получение реальных данных невозможно из-за сложности или дороговизны. Так, для увеличения объема обучающих данных и повышения эффективности машинного обучения используется метод аугментации данных [1], направленный на создание синтезированных данных на основе существующих наборов данных [2]. При использовании этого метода увеличивается как количество, так и разнообразие данных, доступных для обучения и тестирования моделей, устраняя необходимость в сборе новых данных. Увеличение объема данных может быть достигнуто либо за счет их генерации «с нуля» путем обучения генератора (например, с помощью генеративно-сопоставительных нейронных сетей (GAN, аббр. от англ. Generative Adversarial Nets) [3]), либо путем применения к имеющимся выборкам заранее определенного набора преобразований [4]. Оба подхода повышают эффективность моделей глубокого обучения за счет предоставления более разнообразного и обширного набора данных.

В контексте аугментации табличных данных в работах [5] с помощью GAN [6] и вариационных автокодировщиков (VAE, аббр. от англ. Variational Autoencoder) [7] используются разные методологии. Так, GAN хорошо справляются с формированием сложных распределений, что делает их высокоэффективными для создания реалистичных табличных данных. Однако процесс их обучения часто нестабилен и требует тщательной сопоставительной настройки. Напротив, VAE обеспечивают стабильное обучение с помощью своей вероятностной структуры и полезны при изучении скрытых представлений, позволяя интерполировать и исследовать данные. Несмотря на эти преимущества, как правило VAE генерируют менее четкие или реалистичные данные по сравнению с GAN, а балансировка возникающих потерь при реконструкции с регуляризацией остается сложной задачей.

В последние годы разработаны диффузионные модели [8], которые стали мощным инструментом для создания синтетических данных в условиях их дефицита. Диффузионные модели в задачах обнаружения атак (в т. ч. кибератак) применяются как генеративные модели для выявления аномалий. Их ключевая идея – научиться «восстанавливать» нормальные данные и фиксировать значительные отклонения при попытке реконструкции аномальных паттернов. В отличие от GAN и VAE диффузионные модели, также известные как вероятностные диффузионные модели или генеративные мо-

дели, основаны на оценке. Они опираются на теоретическую базу для стабильного обучения и позволяют генерировать высококачественные данные, устраняя некоторые ограничения генеративно-сопоставительных и вариативно-генеративно-сопоставительных моделей. Однако это преимущество достигается за счет более высоких требований к вычислительным ресурсам.

Табличную диффузионную вероятностную модель шумоподавления (TabDDPM, аббр. от англ. Tabular Denoising Diffusion Probabilistic Model) [9] можно универсально применять к любому табличному набору данных, работающему со всеми типами признаков. Она использует полиномиальную диффузию для категориальных и бинарных, а также гауссову диффузию для числовых признаков. Модель TabDDPM эффективно работает со смешанными типами данных и стабильно генерирует высококачественные синтетические данные.

Целью работы является генерация реалистичного синтетического сетевого трафика Android-приложений, представленного в виде нормализованных числовых признаков, с использованием генеративной диффузионной модели TabDDPM, адаптированной для табличных данных и комплексный анализ качества синтезированного трафика с помощью согласованного набора метрик.

Для достижения цели необходимо решить следующие задачи:

- выбрать исходный набор данных сетевого трафика Android-приложений и выполнить его предварительную обработку для обучения модели;
- обучить генеративную диффузионную модель TabDDPM на подготовленных данных и подобрать оптимальные гиперпараметры для улучшения качества синтетических выборок;
- оценить степень сходства между распределениями сгенерированного и реального трафика.

Обеспечение корректности и контролируемости генерации

Одной из важнейших задач при разработке генеративных моделей, предназначенных для синтеза сетевого трафика, является обеспечение корректности и воспроизводимости результатов генерации. Корректность означает, что выходные данные соответствуют логике и структуре реального трафика, не содержат аномалий, противоречий или артефактов. Под контролируемостью понимается, что генерация управляется заранее заданными параметрами и может быть повторена в идентичных условиях с предсказуемыми результатами.

Перед генерацией модель обучается на очищенном и строго структурированном датасете, в котором каждый признак имеет четко определенный тип, диапазон значений и статистическую значи-

мость. На этапе вывода модель использует идентичную структуру признаков, что гарантирует соответствие синтезированных записей формату исходных данных. Взаимно зависимые признаки (например, протокол и порт, флаги и тип соединения) обрабатываются с учетом их совместного распределения вероятности, что позволяет избежать появления комбинаций, не соответствующих установленным в системе правилам обработки.

Последующая валидация результатов генерации – сгенерированный пример проверяется по нескольким критериям:

- тип данных каждого признака должен соответствовать ожидаемому (например, целочисленные порты, категориальные метки протокола, вещественные интервалы);
- далее контролируется попадание значений в допустимые диапазоны, зафиксированные при анализе обучающего датасета;
- признаки, обладающие логической связью (например, TCP-флаг SYN и поле *init_win_bytes_forward*), проверяются на предмет совместимости, что позволяет исключить аномальные комбинации;
- отсутствие пропущенных или поврежденных значений данных (NaN, *аббр. от англ.* Not a Number) гарантируется за счет применения маскирования и восстановления их после нормализации.

Каждая процедура генерации характеризуется набором параметров, которые могут быть зафиксированы заранее или переданы через конфигурационный файл. Ключевыми из них являются: размер латентного вектора; объем выходных данных; класс генерируемого трафика (например, «Normal», «DDoS», «Botnet»); версия модели (номер эпохи, весовой чекпоинт); фиксированное значение Random Seed, обеспечивающее детерминированность результата.

Это дает возможность не только гибко управлять характеристиками выходных данных, но и гарантирует воспроизводимость эксперимента – важную для научных исследований и тестирования систем обнаружения вторжений.

Структура TabDDPM

Табличные наборы данных часто ограничены в объеме из-за проблем с конфиденциальностью при их сборе. TabDDPM может создавать новые синтетические данные без ущерба для конфиденциальности. Общая структура модели представлена на рисунке 1. Числовые и категориальные данные обычно представляются двумя ветвями: квантильным преобразователем для числовых данных и однократным кодированием для категориальных данных. Эти новые представления данных затем передаются на вход диффузионной вероятностной

модели шумоподавления, реализованной на основе многослойных перцептронов (MLP, *аббр. от англ.* Multilayer Perceptron) для минимизации двух типов потерь L_{num} и L_{cat} с помощью функции Softmax. Блок One Hot кодирования осуществляет преобразование категориальных данных в числовой формат, при котором каждая уникальная категория становится отдельным бинарным признаком, принимающим значение 1 или 0.

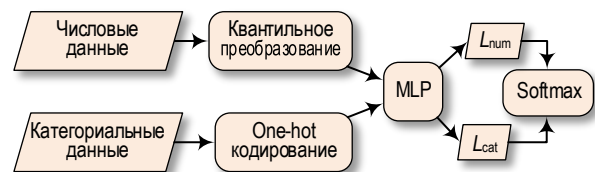


Рис. 1. Общая структура TabDDPM

Рис. 1. General Structure of TabDDPM

Диффузионная модель TabDDPM состоит из двух основных компонентов: прямого и обратного процесса [9]. Прямой процесс (Forward Process) заключается в постепенном добавлении шума удовлетворяющего распределению $q(x_{1:T}|x_0)$ к исходному образцу x_0 из распределения данных $q(x_0)$:

$$q(x_{1:T}|x_0) = \prod_{t=1}^T q(x_t|x_{t-1}).$$

Шум выбирается из заранее заданных распределений $q(x_t|x_{t-1})$ с дисперсиями $\{\beta_1, \dots, \beta_T\}$ до тех пор, пока итоговое распределение не станет неотличимым от случайного шума.

Обратный процесс (Backward Process), удовлетворяющий распределению $p(x_{0:T})$, постепенно устраняет шум в скрытой переменной $x_T \sim q(x_T)$ и позволяет порождать новые выборки данных из $q(x_0)$:

$$p(x_{0:T}) = \prod_{t=1}^T p(x_{t-1}|x_t).$$

Начиная с чистого шума, модель постепенно устраняет шумовую компоненту и тем самым восстанавливает по скрытым переменным $x_1, \dots, x_T$ исходные «чистые» $x_0 \sim q(x_0)$.

Совместное распределение $p_\theta(x_{0:T})$ называется обратным процессом и задается как цепь Маркова с обучаемыми гауссовыми переходами, начинающаяся с $p(x_T) = \mathcal{N}(x_T; 0, I)$:

$$p_\theta(x_{0:T}) = p(x_T) \prod_{t=1}^T p_\theta(x_{t-1}|x_t). \quad (1)$$

Отличие диффузионных моделей от других типов со скрытыми переменными состоит в том, что приближенное апостериорное распределение $q(x_{1:T}|x_0)$, называемое прямым процессом или диффузионным процессом, фиксируется как цепь Маркова, которая

постепенно добавляет к данным гауссов шум согласно расписанию дисперсий $\beta_1, \dots, \beta_T$:

$$q(x_{1:T} | x_0) = \prod_{t=1}^T q(x_t | x_{t-1}). \quad (2)$$

Гауссовские диффузионные модели работают в непрерывных пространствах $(x_t | \mathbb{R}^n)$, где прямой и обратный процессы задаются нормальными распределениями:

$$q(x_t | x_{t-1}) = N(x_t; \sqrt{1 - \beta_t} x_{t-1}, \beta_t I), \quad (3)$$

$$q(x_T) = N(x_T; 0, I), \quad (4)$$

$$p_\theta(x_{t-1} | x_t) = N(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)). \quad (5)$$

Необходимо выбрать дисперсии β_t прямого процесса, а также архитектуру модели и параметризацию гауссовского распределения для обратного процесса. Дисперсии β_t могут быть получены либо с помощью повторной параметризации [10], либо зафиксированы как гиперпараметры. В [11] предлагают использовать диагональную матрицу $\Sigma_\theta(x_t, t)$ с постоянной β_t и вычислять среднее значение $\mu_\theta(x_t, t)$ как функцию от x_t и $\epsilon_\theta(x_t, t)$:

$$\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{\beta_t}{\sqrt{1 - \alpha_t}} \epsilon_\theta(x_t, t) \right), \quad (6)$$

где $\alpha_t = 1 - \beta_t$, $\bar{\alpha}_t = \prod_{i \leq t} \alpha_i$, а $\epsilon_\theta(x_t, t)$ предсказывает «истинную» шумовую компоненту ϵ для зашумленного образца данных x_t .

На практике целевая функция может быть представлена в виде суммы среднеквадратичных ошибок между предсказанным шумом $\epsilon_\theta(x_t, t)$ и истинным шумом ϵ по всем временным шагам t :

$$L_t^{\text{simple}} = \mathbb{E}_{x_0, \epsilon, t} \| \epsilon - \epsilon_\theta(x_t, t) \|_2^2. \quad (7)$$

Исходный набор данных сетевого трафика Android-приложений

Генерация реалистичного сетевого трафика Android-приложений проводилась на примере экспериментальных данных мобильных приложений, приведенных в работах [12, 13]. Для формирования обучающей и тестовой выборок на мобильных устройствах под управлением ОС Android осуществлялся сбор необработанных данных сетевого трафика в виде IP-пакетов. Трафик различных типов мобильных приложений характеризовался набором из $M = 23$ атрибутов. Общее число экспериментально измеренных потоков каждого приложения составляло $K = 5000$.

В исследовании использовался подготовленный датасет сетевого трафика Android-приложений [12, 13], содержащий $N = 90777$ записей (сетевых сессий) и $d = 21$ числовых признаков. Каждый образец был помечен идентификатором приложения $y = \text{app_id}$, сгенерировавшего данный трафик. Для

оценки качества генерации синтетического трафика рассматривались три из 18 уникальных приложений из датасета: *Mail.ru* ($\text{app_id} = 9$), *SberMobile* ($\text{app_id} = 12$) и *4PDA* ($\text{app_id} = 2$). Для каждого выбранного приложения формировалась модель TabDDPM с отдельной обучающей выборкой, аппроксимирующей эмпирическое распределение признаков данного класса.

На первом этапе к исходным данным применялась нормализация признаков, поскольку значения исходных числовых атрибутов имеют разную природу (размеры и количества пакетов, байтов, соотношения и т. д.) и характеризуются сильной вариативностью и асимметрией распределений.

На втором этапе производилось ограничение выбросов (*clip*). С этой целью для числовых признаков с наибольшим размахом вычисляются пороги $P_{0.99}(j)$ – 99-й перцентиль по обучающей выборке. Все значения, превышающие $P_{0.99}(j)$, заменялись пороговым значением: $x_j := \min(x_j, P_{0.99}(j))$.

На третьем этапе производилось масштабирование с помощью квантильного нормализатора $QT: \mathbb{R} \rightarrow [0, 1]$, переводящего распределение каждого признака в равномерное на отрезке $[0, 1]$. Формально для каждого признака j значение x_j заменялось на $u_j = F_j(x_j)$, где F_j – эмпирическая функция распределения данного признака на обучающей выборке.

Для построения и оценки модели датасет разбивался на обучающую, валидационную и тестовую выборки. Разделение производилось по классам для сохранения пропорций редких и частых классов. В проведенных экспериментах для обучения выделялось 80 % наблюдений, 10 % для подбора гиперпараметров и 10 % для финального тестирования генератора.

На основе особенностей используемых числовых данных и метода их предобработки структурная схема модели TabDDPM может быть представлена в виде, представленном на рисунке 2.

Модель обучается путем минимизации суммы среднеквадратичной ошибки для гауссовой диффузионной части:

$$L_{\text{MSE}} = \frac{1}{M} \sum_{i=1}^M \| \hat{\epsilon}_\theta(x_{t_i}^{(i)} | t_i) - \epsilon^{(i)} \|^2. \quad (8)$$

На вход модели подается вектор $\{x_1, x_2, \dots, x_n\}$ числовых признаков сессии одного из приложений. Блок Quantile Transformer по каждому признаку выполняет масштабирование к диапазону $[-1, 1]$. На выходе блока получаем нормализованный тензор x_{norm} размерности $\text{batch} \times \text{features}$. После указанного преобразования данные попадают в скоринговую сеть диффузионной модели TabDDPM, в качестве которой используется MLP, формирующий оценку добавленного шума ϵ той же размерности.

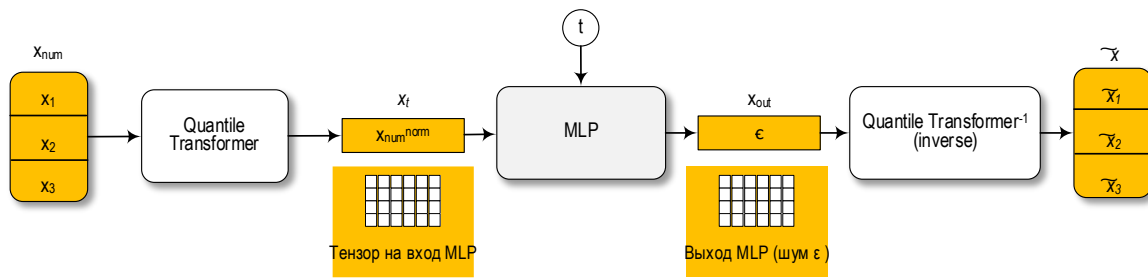


Рис. 2. Структурная схема модели TabDDPM

Puc. 2. Structural Diagram of the TabDDPM

На заключительном этапе выполняется обратное к исходному масштабу преобразование в блоке Quantile Transformer⁻¹. В результате на выходе модели формируется вектор синтезированных данных $\{\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n\}$, сопоставимый по масштабу с реальными данными.

Оценка качества синтезированных данных

Для количественной оценки качества сходства синтезированных выборок, порожденных генеративной моделью TabDDPM и выбранного приложения реального трафика, использовался набор метрик, представленный в таблице 1.

ТАБЛИЦА 1. Метрики оценки сходства распределений

ТАБЛИЦА 1. Metrics for Assessing Distribution Similarity

№ п/п	Метрика	Формула
1	Kolmogorov-Smirnov (KS)	$D_j = \sup_x F_j(x) - \hat{F}_j(x) $ $KS = \frac{1}{d} \sum D_j$ (9)
2	First-Order Wasserstein Distance (WD)	$WD = \frac{1}{d} \sum_{j=1}^d \int_0^1 F_j^{-1}(u) - \hat{F}_j^{-1}(u) du$ (10)
3	Reference Range Coverage (Coverage)	$\text{cov}_j(W_k) = \frac{1}{ W_k } \sum_{x \in W_k} 1 \{F_j^{-1}(0,01) \leq x^{(j)} \leq F_j^{-1}(0,99)\}$ (11)
		$\text{Coverage}(W_k) = \frac{1}{d} \sum_{j=1}^d \text{cov}_j(W_k)$ (12)
4	Correlation Consistency (ΔCorr)	$\Delta\text{Corr} = \frac{2}{d(d-1)} \sum_{1 \leq j < k \leq d} \rho_{jk}^{(\text{real})} - \rho_{jk}^{(\text{synth})} $ (13)
5	Area Under the ROC Curve (AUC)	$AUC = \int_0^1 TPR(FPR^{-1}(t)) dt$ (14)

В формулах, представленных в таблице 1, обозначениями $F_j(x)$ и $\hat{F}_j(x)$ представлены кумулятивные функции распределения реальных и синтезированных данных:

$F_j^{-1}(u)$ и $\hat{F}_j^{-1}(u)$ – функции распределения обратные к $F_j(x)$ и $\hat{F}_j(x)$;

$\rho_{jk}^{(\text{real})}$ и $\rho_{jk}^{(\text{synth})}$ – коэффициенты корреляции Пирсона между j -м и k -м признаками для реальных и синтетических данных;

В ходе экспериментов с генеративной моделью TabDDPM для всех метрик, представленных в таблице 1, и каждого рассматриваемого приложения были получены конкретные числовые значения, приведенные в таблице 2. Представленные численные значения метрик подтверждают, что генера-

ция синтезированных с помощью модели TabDDPM данных воспроизводит как масштаб и форму распределений, так и существенную часть межпризнаковых зависимостей. Это позволяет применять синтез выбранных приложений для дальнейшей настройки и валидации детекторов, ориентированных на появление неизвестных классов.

ТАБЛИЦА 2. Сводная таблица метрик по приложениям

TABLE 2. Summary of Metrics by Application

app_id	AUC	KS	WD	ΔCorr	Coverage
2	0,6778	0,0520	0,0287	0,0485	1
9	0,6141	0,0501	0,0242	0,0433	1
12	0,5914	0,0658	0,0329	0,0679	0,9997

Каждая из представленных метрик покрывает «слепые зоны» других метрик. Так, метрики KS / WD служат для оценки сходства одномерных распреде-

лений и сдвиги массы, а ΔCorr оценивает многомерные связи. Метрика Coverage оценивает охват значений, а AUC – практическую различимость. Каждая из применяемых метрик AUC, Coverage, ΔCorr , KS и WD характеризует свой индивидуальный, принципиально отличный аспект качества генерации, синтезируемых данных с помощью модели TabDDPM и потому не является взаимозаменяемым. Анализ результатов моделирования показывает, что синтезированные данные почти полностью совпадают с реальными, поскольку центры и масштабы кластеров близки, а локальные окрестности идентичны. Незначительные отличия касаются главным образом глобального положения центров у приложений SberMobile и частично у 4PDA. При этом локальная структура соседств остается на уровне 95–98 %, что подтверждает корректную генерацию характерных паттернов трафика с помощью модели TabDDPM.

Учитывая разнонаправленный характер изменения величины указанных метрик, для характеристики индивидуальных особенностей перечисленных критериев предлагается ввести интегральный показатель качества синтезированных данных на основе модифицированных метрик, приведенных в таблице 3. Количественно эффективность моделирования может быть оценена показателем $\sum K_i$.

ТАБЛИЦА 3. Сводная таблица модифицированных метрик по приложениям

TABLE 3. Summary Table of Modified Metrics by Application

app_id	K1	K2	K3	K4	K5	$\sum K_i$
2	0,6444	0,948	0,971	0,951	1	4,192
9	0,7718	0,949	0,976	0,957	1	4,268
12	0,8172	0,934	0,967	0,922	0,999	4,232

Примечание: $K2 = 1 - KS$; $K3 = 1 - WD$; $K4 = 1 - \Delta\text{Corr}$; $K5 = \text{Coverage}$

Для визуализации введенных метрик, характеризующих качественные показатели данных для различных типов анализируемых приложений и разработанного метода моделирования синтезированных данных, использовалась лепестковая диаграмма, представленной на рисунке 3. Для удобства метрика AUC дискриминатора, для которого целевым является значение 0,5, была преобразована в «индекс неотличимости», принимающий значение, равное 1 при $AUC = 0,5$, и убывающий к 0 по мере роста различимости: $K1 = 1 - 2 | AUC - 0,5 |$.

Список источников

1. Ding J., Li X., Kang X., Gudivada V.N. A Case Study of the Augmentation and Evaluation of Training Data for Deep Learning // Journal of Data and Information Quality. 2019. Vol. 11. Iss. 4. PP. 1–22. DOI:10.1145/3317573
2. Bansal A., Sharma R., Kathuria M. A Systematic Review on Data Scarcity Problem in Deep Learning: Solution and Applications // ACM Computing Surveys. 2022. Vol. 54. Iss. 10s. PP. 1–29. DOI:10.1145/3502287
3. Alqahtani H., Kavakli-Thorne M., Kumar G. Applications of Generative Adversarial Networks (GANs): An Updated Review // Archives of Computational Methods in Engineering. 2021. Vol. 28. Iss. 2. PP. 525–552. DOI:10.1007/s11831-019-09388-y. EDN:FTPNOL
4. Cubuk E.D., Zoph B., Mané D., Vasudevan V., Le Q.V. AutoAugment: Learning Augmentation Strategies From Data // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR, Long Beach, USA, 2019). IEEE, 15–20 June 2019. PP. 113–123. DOI:10.1109/CVPR.2019.00020

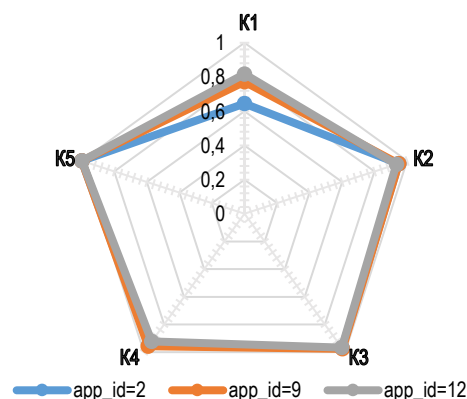


Рис. 3. Лепестковая диаграмма метрик для исследуемых приложений

Fig. 3. Radar Chart of Metrics for the Applications under Study

Заключение

Достоинством диффузионных моделей является возможность детализированной реконструкции данных и способности выявлять редкие аномалии. Представленные метрики подтверждают, что генерация синтезированных данных с помощью модели TabDDPM воспроизводит как масштаб и форму распределений, так и существенную часть межпризнаковых зависимостей. Это позволяет применять синтез выбранных приложений для дальнейшей настройки и валидации детекторов, ориентированных на появление неизвестных классов.

Каждая из представленных метрик покрывает «слепые зоны» других метрик. Так, метрики KS / WD служат для оценки сходства одномерных распределений и сдвиги массы, а ΔCorr оценивает многомерные связи. Метрика Coverage оценивает охват значений, а AUC – практическую различимость. Каждая из применяемых метрик AUC, Coverage, ΔCorr , KS и WD характеризует свой индивидуальный, принципиально отличный аспект качества генерации, синтезируемых данных с помощью модели TabDDPM и потому не являются взаимозаменяемыми. Для количественной оценки качества генерации синтезируемых данных предложено ввести в рассмотрение интегральный показатель качества на основе парциальных метрик.

Показано, что с помощью лепестковых диаграмм можно наглядно визуализировать интегральный показатель качества синтезируемых данных с помощью модели TabDDPM.

5. Cui L., Li H., Chen K., Shou L., Chen G. Tabular data augmentation for machine learning: Progress and prospects of embracing generative AI // arXiv preprint arXiv:2407.21523. 2024. DOI:10.48550/arXiv.2407.21523
6. Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., et al. Generative adversarial networks // Communications of the ACM. 2020. Vol. 63. Iss. 11. PP. 139–144. DOI:10.1145/3422622. EDN:SESCXD
7. Kingma D.P., Welling M. An Introduction to Variational Autoencoders // Foundations and Trends in Machine Learning. 2019. Vol. 12. Iss. 4. PP. 307–392. DOI:10.1561/22000000056
8. Yang L., Zhang Z., Song Y., Hong S., Xu R., Zhao Y., et al. Diffusion Models: A Comprehensive Survey of Methods and Applications // ACM Computing Surveys. 2024. Vol. 56. Iss. 4. PP. 1–39. DOI:10.1145/3626235
9. Kotelnikov A., Baranchuk D., Rubachev I., Babenko A. TabDDPM: Modelling tabular data with diffusion models // arXiv preprint arXiv:2209.15421. 2022. DOI:10.48550/arXiv.2209.15421
10. Kingma D.P., Welling M. Auto-encoding variational Bayes // arXiv preprint arXiv:1312.6114. 2022. DOI:10.48550/arXiv.1312.6114
11. Ho J., Jain A., Abbeel P. Denoising Diffusion Probabilistic Models // Advances in Neural Information Processing Systems 33 (NeurIPS). 2020.
12. Шелухин О.И., Маторин Ф.А. Снижение размерности массивов данных с помощью многослойных автокодировщиков в задаче классификации мобильных приложений // Труды учебных заведений связи. 2024. Т. 10. № 6. С. 111–120. DOI:10.31854/1813-324X-2024-10-6-111-120. EDN:TOPDUA
13. Шелухин О.И., Маторин Ф.А., Ванюшина А.В. Оценка свойств многослойных автокодировщиков в задаче обнаружения и классификации мобильных приложений // Научные технологии в космических исследованиях Земли. 2024. Т. 16. № 6. С. 12–20. DOI:10.36724/2409-5419-2024-16-6-12-20. EDN:CRPLOR

References

1. Ding J., Li X., Kang X., Gudivada V.N. A Case Study of the Augmentation and Evaluation of Training Data for Deep Learning. *Journal of Data and Information Quality*. 2019;11(4):1–22. DOI:10.1145/3317573
2. Bansal A., Sharma R., Kathuria M. A Systematic Review on Data Scarcity Problem in Deep Learning: Solution and Applications. *ACM Computing Surveys*. 2022;54(10s):1–29. DOI:10.1145/3502287
3. Alqahtani H., Kavakli-Thorne M., Kumar G. Applications of Generative Adversarial Networks (GANs): An Updated Review. *Archives of Computational Methods in Engineering*. 2021;28(2):525–552. DOI:10.1007/s11831-019-09388-y. EDN:FTPNOL
4. Cubuk E.D., Zoph B., Mané D., Vasudevan V., Le Q.V. AutoAugment: Learning Augmentation Strategies From Data. *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 15–20 June 2019, Long Beach, USA. IEEE*; 2019. p.113–123. DOI:10.1109/CVPR.2019.00020
5. Cui L., Li H., Chen K., Shou L., Chen G. Tabular data augmentation for machine learning: Progress and prospects of embracing generative AI. *arXiv preprint arXiv:2407.21523*. 2024. DOI:10.48550/arXiv.2407.21523
6. Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., et al. Generative adversarial networks. *Communications of the ACM*. 2020;63(11):139–144. DOI:10.1145/3422622. EDN:SESCXD
7. Kingma D.P., Welling M. An Introduction to Variational Autoencoders. *Foundations and Trends in Machine Learning*. 2019;12(4):307–392. DOI:10.1561/22000000056
8. Yang L., Zhang Z., Song Y., Hong S., Xu R., Zhao Y., et al. Diffusion Models: A Comprehensive Survey of Methods and Applications. *ACM Computing Surveys*. 2024;56(4):1–39. DOI:10.1145/3626235
9. Kotelnikov A., Baranchuk D., Rubachev I., Babenko A. TabDDPM: Modelling tabular data with diffusion models. *arXiv preprint arXiv:2209.15421*. 2022. DOI:10.48550/arXiv.2209.15421
10. Kingma D.P., Welling M. Auto-encoding variational Bayes. *arXiv preprint arXiv:1312.6114*. 2022. DOI:10.48550/arXiv.1312.6114
11. Ho J., Jain A., Abbeel P. Denoising Diffusion Probabilistic Models. *Advances in Neural Information Processing Systems 33 (NeurIPS)*. 2020.
12. Sheluhin O.I., Matorin F.A. Reducing the Dimensionality of Data Arrays Using Multi-Layer Autoencoders in the Task of Classifying Mobile Applications. *Proceedings of Telecommunication Universities*. 2024;10(6):111–120. (in Russ.) DOI:10.31854/1813-324X-2024-10-6-111-120. EDN:TOPDUA
13. Sheluhin O.I., Matorin F.A., Vanyushina A.V. Evaluation of the properties of multilayer autoencoders in the task of detecting and classifying mobile applications. *H&ES Research*. 2024;16(6):12–20. (in Russ.) DOI:10.36724/2409-5419-2024-16-6-12-20. EDN:CRPLOR

Статья поступила в редакцию 30.01.2026; одобрена после рецензирования 19.02.2026; принята к публикации 24.02.2026

The article was submitted 30.01.2026; approved after reviewing 19.02.2026; accepted for publication 24.02.2026

Информация об авторах:

ШЕЛУХИН
Олег Иванович

доктор технических наук, профессор, заведующий кафедрой «Информационная безопасность» Московского технического университета связи и информатики
 <https://orcid.org/0000-0001-7564-6744>

МАТОРИН
Фёдор Андреевич

аспирант кафедры «Информационная безопасность» Московского технического университета связи и информатики
 <https://orcid.org/0009-0002-4897-2338>

Авторы сообщают об отсутствии конфликтов интересов.

The authors declare no conflicts of interests.

Санкт-Петербургский государственный университет
телекоммуникаций им. проф. М. А. Бонч-Бруевича

СПбГУТ)))

15th INTERNATIONAL CONFERENCE ON ADVANCED INFOTELECOMMUNICATIONS ICAIT
Международная научно-техническая и научно-методическая конференция
«Актуальные проблемы инфотелекоммуникаций в науке и образовании»



ОСНОВНЫЕ НАУЧНЫЕ НАПРАВЛЕНИЯ:

-  Инфокоммуникационные сети и системы
-  Радиоэлектронные системы и робототехника
-  Информационные технологии и программная инженерия
-  Кибербезопасность
-  Социальные технологии и экономика данных
-  Сети связи специального назначения
-  Инновации для Арктики



Круглый стол
партнеров СПбГУТ

12–15
МАЯ 2026

Подробности
о конференции
на сайте:
API.NO.SUT.RU



План издания научной литературы 2026 г., п. 12

Усл.-печ. л.
15,25

Формат
60×84_{1/8}

Заказ
№ 1650

Учредитель и издатель:

Федеральное государственное бюджетное образовательное учреждение
высшего образования "Санкт-Петербургский государственный университет
телекоммуникаций им. проф. М.А. Бонч-Бруевича"

E-mail: tuzs@sut.ru Web: tuzs.sut.ru VK: vk.com/spbtuzs

