

# Программные продукты и системы

DOI 10.15827/0236-235X

## С Новым 2025 годом!



№ 4

ТОМ 37

2024

---

Научно-исследовательский институт  
«Центрпрограммсистем»

---

# Программные продукты и системы

НАУЧНЫЙ ЖУРНАЛ

2024, том 37, № 4  
(год издания тридцать седьмой)

**Главный редактор**  
**Г.И. САВИН**, академик РАН

## SOFTWARE & SYSTEMS

Research Journal

2024, vol. 37, no. 4

**Editor-in-Chief**  
**G.I. SAVIN**, Academician of the Russian Academy of Sciences

---

Research Institute CENTERPROGRAMSYSTEM

---

## © ПРОГРАММНЫЕ ПРОДУКТЫ И СИСТЕМЫ

Научный журнал

2024. Т. 37. № 4

DOI: 10.15827/0236-235X.148

### Главный редактор

**Г.И. САВИН**, академик РАН

### Научный редактор номера:

**А.Н. СОТНИКОВ**, д.ф.-м.н., профессор

Издатель НИИ «Центрпрограммсистем»

(г. Тверь, Россия)

Учредитель **В.П. Куприянов**

Журнал зарегистрирован в Роскомнадзоре

3 марта 2020 г.

Регистрационное свидетельство ПИ № ФС 77-77843

Подписной индекс в каталоге

Урал-Пресс **70799**

ISSN 0236-235X (печатн.)

ISSN 2311-2735 (онлайн)

## РЕДАКЦИОННАЯ КОЛЛЕГИЯ

**Семенов Н.А.** – заместитель главного редактора, д.т.н., профессор Тверского государственного технического университета (г. Тверь, Россия)

**Сотников А.Н.** – заместитель главного редактора, д.ф.-м.н., профессор, заместитель директора Межведомственного суперкомпьютерного центра РАН (г. Москва, Россия)

**Афанасьев А.П.** – д.ф.-м.н., профессор Московского физико-технического института (технического университета), заведующий Центром распределенных вычислений Института проблем передачи информации РАН (г. Москва, Россия)

**Баламетов А.Б.** – д.т.н., профессор Азербайджанского научно-исследовательского и проектно-изыскательского института энергетики (г. Баку, Азербайджан)

**Борисов В.В.** – д.т.н., профессор филиала Национального исследовательского университета «МЭИ» в г. Смоленске (г. Смоленск, Россия)

**Голенков В.В.** – д.т.н., профессор Белорусского государственного университета информатики и радиоэлектроники (г. Минск, Беларусь)

**Елизаров А.М.** – д.ф.-м.н., профессор Института математики и механики им. Н.И. Лобачевского Казанского федерального университета (г. Казань, Россия)

**Еремеев А.П.** – д.т.н., профессор Национального исследовательского университета «МЭИ» (г. Москва, Россия)

**Кузнецов О.П.** – д.т.н., профессор Института проблем управления РАН (г. Москва, Россия)

**Мамросенко К.А.** – к.т.н., доцент Московского авиационного института (Национального исследовательского университета), руководитель Центра визуализации и спутниковых информационных технологий НИИСИ РАН (г. Москва, Россия)

**Палюх Б.В.** – д.т.н., профессор Тверского государственного технического университета (г. Тверь, Россия)

**Сулейманов Д.Ш.** – академик АН Республики Татарстан, д.т.н., профессор Казанского государственного технического университета (г. Казань, Россия)

**Татарникова Т.М.** – д.т.н., доцент, профессор Санкт-Петербургского государственного электротехнического университета «ЛЭТИ» им. В.И. Ульянова (Ленина) (г. Санкт-Петербург, Россия)

**Ульянов С.В.** – д.ф.-м.н., профессор, ведущий научный сотрудник Объединенного института ядерных исследований (г. Дубна, Россия)

**Хорошевский В.Ф.** – д.т.н., профессор Московского физико-технического института (технического университета) (г. Москва, Россия)

**Шабанов Б.М.** – д.т.н., чл.-корр. РАН, директор Межведомственного суперкомпьютерного центра РАН (г. Москва, Россия)

**Язенин А.В.** – д.ф.-м.н., профессор Тверского государственного университета (г. Тверь, Россия)

## АССОЦИИРОВАННЫЕ ЧЛЕНЫ РЕДАКЦИИ

*Национальный исследовательский университет «МЭИ», г. Москва, Россия*

*Технологический институт Южного федерального университета, г. Таганрог, Россия*

*Тверской государственный технический университет, г. Тверь, Россия*

### АДРЕС ИЗДАТЕЛЯ И РЕДАКЦИИ

г. Тверь, просп. Николая Корыткова, д. 3а,  
170024, Россия

Телефон: (482-2) 39-91-49

Факс: (482-2) 39-91-00

E-mail: [red@cps.tver.ru](mailto:red@cps.tver.ru)

Сайт: [www.swsys.ru](http://www.swsys.ru)

Дата выхода в свет 16.12.2024 г.

Отпечатано ИПШ «Фактор и К»

г. Тверь, ул. Крылова, д. 26, 170100, Россия

Выпускается один раз в квартал

Год издания тридцать седьмой

Формат 60×84 1/8. Объем 184 стр.

Заказ № 12. Тираж 1000 экз. Цена 550,00 руб.

© SOFTWARE & SYSTEMS

Research Journal

2024, vol. 37, no. 4

DOI: 10.15827/0236-235X.148

Editor-in-chief

G.I. SAVIN, Academician of RAS

Science editor of the issue:

A.N. SOTNIKOV, Dr.Sci. (Physics and Mathematics), Professor

Publisher Research Institute

CENTERPROGRAMSYSTEM (Tver, Russian Federation)

Founder V.P. Kupriyanov

The journal is registered with the Federal Service  
for Supervision of Communications, Information Technology  
and Mass Communications (Roskomnadzor)  
March 3<sup>rd</sup>, 2020

Registration certificate ПИИ № ФС 77-77843

ISSN 0236-235X (print)

ISSN 2311-2735 (online)

## EDITORIAL BOARD

**Semenov N.A.** – Deputy Editor-in-Chief, Dr.Sci. (Engineering), Professor of the Tver State Technical University (Tver, Russian Federation)

**Sotnikov A.N.** – Deputy Editor-in-Chief, Dr.Sci. (Physics and Mathematics), Professor, Deputy Director of the Joint Supercomputer Center of the Russian Academy of Sciences (Moscow, Russian Federation)

**Afanasiev A.P.** – Dr.Sci. (Physics and Mathematics), Professor of the Moscow Institute of Physics and Technology, Head of Centre for Distributed Computing of Institute for Information Transmission Problems (Moscow, Russian Federation)

**Balametov A.B.** – Dr.Sci. (Engineering), Professor of the Azerbaijan Scientific-Research & Design-Prospecting Power Engineering Institute (Baku, Azerbaijan)

**BorISOV V.V.** – Dr.Sci. (Engineering), Professor of the MPEI Branch in Smolensk (Smolensk, Russian Federation)

**Golenkov V.V.** – Dr.Sci. (Engineering), Professor of the Belarusian State University of Informatics and Radioelectronics (Minsk, Republic of Belarus)

**Elizarov A.M.** – Dr.Sci. (Physics and Mathematics), Professor of the N.I. Lobachevsky Institute of Mathematics and Mechanics of the Kazan Federal University (Kazan, Russian Federation)

**Eremeev A.P.** – Dr.Sci. (Engineering), Professor of the National Research University "Moscow Power Engineering Institute" (Moscow, Russian Federation)

**Kuznetsov O.P.** – Dr.Sci. (Engineering), Professor of the Institute of Control Sciences of the Russian Academy of Sciences (Moscow, Russian Federation)

**Mamrosenko K.A.** – Ph.D. (Engineering), Associate Professor of the Moscow Aviation Institute (National Research University), Head of the Center of Visualization and Satellite Information Technologies SRISA RAS (Moscow, Russian Federation)

**Palyukh B.V.** – Dr.Sci. (Engineering), Professor of the Tver State Technical University (Tver, Russian Federation)

**Suleimanov D.Sh.** – Academician of TAS, Dr.Sci. (Engineering), Professor of the Kazan State Technical University (Kazan, Russian Federation)

**Tatarnikova T.M.** – Dr.Sci. (Engineering), Associate Professor, Professor of the St. Petersburg Electrotechnical University "LETI" (St. Petersburg, Russian Federation)

**Ulyanov S.V.** – Dr.Sci. (Physics and Mathematics), Professor of the Dubna International University for Nature, Society and Man (Dubna, Russian Federation)

**Khoroshevsky V.F.** – Dr.Sci. (Engineering), Professor of the Moscow Institute of Physics and Technology (Moscow, Russian Federation)

**Shabanov B.M.** – Dr.Sci. (Engineering), Corresponding Member of the RAS, Director of the Joint Supercomputer Center of the Russian Academy of Sciences (Moscow, Russian Federation)

**Yazenin A.V.** – Dr.Sci. (Physics and Mathematics), Professor of the Tver State University (Tver, Russian Federation)

## ASSOCIATED EDITORIAL BOARD MEMBERS

*National Research University "Moscow Power Engineering Institute", Moscow, Russian Federation*  
*Technology Institute at Southern Federal University, Taganrog, Russian Federation*  
*Tver State Technical University, Tver, Russian Federation*

## EDITORIAL BOARD AND PUBLISHER OFFICE ADDRESS

Nikolay Korytkov Ave, 3a, Tver, 170024, Russian Federation

Phone: (482-2) 39-91-49 Fax: (482-2) 39-91-00

E-mail: red@cps.tver.ru

Website: www.swsys.ru

Release date 16.12.2024

Printed in printing-office "Faktor i K"

Krylova St. 26, Tver, 170100, Russian Federation

Published quarterly. 37th year of publication

Format 60×84 1/8. Wordage 184 pages

Prod. order № 12. Circulation 1000 copies. Price 550,00 rub.

## Вниманию авторов

Журнал «Программные продукты и системы» публикует материалы научного и научно-практического характера по новым информационным технологиям, результаты академических и отраслевых исследований в области использования средств вычислительной техники. Практикуются выпуски тематических номеров по искусственному интеллекту, системам автоматизированного проектирования, по технологиям разработки программных средств и системам защиты, а также специализированные выпуски, посвященные научным исследованиям и разработкам отдельных вузов, НИИ, научных организаций.

Журнал «Программные продукты и системы» внесен в Перечень ведущих рецензируемых научных журналов и изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученых степеней кандидата и доктора наук.

Информация об опубликованных статьях по установленной форме регулярно предоставляется в систему РИНЦ, в CrossRef и в другие базы и электронные библиотеки.

Журнал «Программные продукты и системы» включен в ядро коллекции РИНЦ, размещенное на платформе Web of Science в виде базы данных RSCI.

Автор статьи отвечает за подбор, оригинальность и точность приводимого фактического материала. При перепечатке ссылка на журнал обязательна. Статьи публикуются бесплатно.

### Условия публикации

К рассмотрению принимаются оригинальные материалы, отвечающие редакционным требованиям и соответствующие тематике журнала. Группы научных специальностей:

#### **1.2. Компьютерные науки и информатика**

1.2.1. Искусственный интеллект и машинное обучение (физико-математические науки).

1.2.2. Математическое моделирование, численные методы и комплексы программ (физико-математические науки, технические науки).

#### **2.3. Информационные технологии и телекоммуникации**

2.3.1. Системный анализ, управление и обработка информации, статистика (технические науки, физико-математические науки).

2.3.2. Вычислительные системы и их элементы (технические науки).

2.3.3. Автоматизация и управление технологическими процессами и производствами (технические науки).

2.3.5. Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей (технические науки, физико-математические науки).

2.3.6. Методы и системы защиты информации (технические науки, физико-математические науки).

2.3.7. Компьютерное моделирование и автоматизация (технические науки, физико-математические науки).

2.3.8. Информатика и информационные процессы (технические науки).

Работа представляется в электронном виде в формате Word. Объем статьи вместе с иллюстрациями – не менее 10 000 знаков. Диаграммы, схемы, графики должны быть доступными для редактирования (Word, Visio, Excel). Заголовок должен быть информативным; сокращения, а также терминологию узкой тематики желательно в нем не использовать. Количество авторов на одну статью – не более четырех, количество статей одного автора в номере, включая соавторство, – не более двух. Список литературы, наличие которого обязательно, должен включать не менее 10 пунктов.

Необходимы также содержательная структурированная аннотация (не менее 200 слов), ключевые слова (7–10) и индекс УДК. Название статьи, аннотация и ключевые слова должны быть переведены на английский язык (машинный перевод недопустим), а фамилии авторов, названия и юридические адреса организаций (если нет официального перевода) – транслитерированы по стандарту BGN/PCGN.

Вместе со статьей следует прислать экспертное заключение о возможности открытого опубликования материала и авторскую справку. Обязательно соблюдение автором договора (публичной оферты).

### Порядок рецензирования

Все статьи, поступающие в редакцию (соответствующие тематике и оформленные согласно требованиям к публикации), подлежат двойному слепому рецензированию в течение месяца с момента поступления, рецензия отправляется авторам.

В редакции сформирован устоявшийся коллектив рецензентов, среди которых члены редколлегии журнала, эксперты из числа крупных специалистов в области информатики и вычислительной техники ведущих вузов страны, а также ученые и специалисты НИИСИ РАН, МСЦ РАН (г. Москва) и НИИ «Центрпрограмм-систем» (г. Тверь).

Редакция журнала «Программные продукты и системы» в своей работе руководствуется сводом правил Кодекса этики научных публикаций, разработанным и утвержденным Комитетом по этике научных публикаций (Committee on Publication Ethics – COPE).

## Управление пользовательскими заданиями в сети суперкомпьютерных центров с применением федеративной аутентификации

А.В. Баранов<sup>1,2</sup>✉, Е.Е. Кузнецов<sup>2</sup>

<sup>1</sup> Межведомственный суперкомпьютерный центр РАН, г. Москва, 119334, Россия

<sup>2</sup> Национальный исследовательский центр «Курчатовский институт», г. Москва, 123182, Россия

### Ссылка для цитирования

Баранов А.В., Кузнецов Е.Е. Управление пользовательскими заданиями в сети суперкомпьютерных центров с применением федеративной аутентификации // Программные продукты и системы. 2024. Т. 37. № 4. С. 461–471. doi: 10.15827/0236-235X.148.461-471

### Информация о статье

Группа специальностей ВАК: 2.3.6

Поступила в редакцию: 20.07.2024

После доработки: 22.08.2024

Принята к публикации: 30.08.2024

**Аннотация.** Предметом представленного в статье исследования является управление пользовательскими заданиями в распределенной сети научных суперкомпьютерных центров (СКЦ) коллективного пользования. Сеть СКЦ объединяет высокопроизводительные вычислительные системы разной архитектуры, принадлежащие различным СКЦ. Каждый центр самостоятельно определяет политику безопасности и поддерживает собственную базу учетных записей пользователей, что усложняет управление пользовательскими заданиями, в частности, затрудняет оперативное перераспределение заданий между вычислительными системами разных СКЦ. Методология исследования базируется на совмещении двухуровневой иерархической системы управления заданиями и федеративного управления идентификацией, в частности, федеративной аутентификации. В работе предложен новый метод управления пользовательскими заданиями в распределенной сети СКЦ, основанный на федеративной аутентификации. Верхний уровень иерархии управления представлен глобальной очередью, из которой задания распределяются по вычислительным системам распределенной сети СКЦ. Локальные очереди этих вычислительных систем образуют нижний уровень иерархии управления заданиями. Аутентификация и авторизация для каждого задания должны производиться дважды: при постановке в глобальную очередь и при распределении в одну из локальных очередей. Предлагаемый метод учитывает, что за время нахождения в глобальной очереди задание с точки зрения информационной безопасности превращается из объекта в субъект, который заново должен быть авторизован в локальной очереди. Как показано в статье, применение федеративной аутентификации при авторизации пользователей и их заданий позволяет построить простую и безопасную схему управления заданиями в сети СКЦ. Практическую значимость исследования составляют представленный в статье порядок функционирования системы управления заданиями в распределенной сети СКЦ и анализ безопасности такой системы.

**Ключевые слова:** сеть суперкомпьютерных центров, управление заданиями, очередь заданий, федеративная аутентификация, поставщик услуг, поставщик идентификационных данных

**Благодарности.** Работа выполнена в МСЦ РАН и НИЦ «Курчатовский институт» в рамках госзадания по теме FNEF-2024-0014

**Введение.** Формированию единого научно-образовательного пространства информационных технологий в стране придается большое значение. Так, в Межведомственном суперкомпьютерном центре РАН ведутся работы по созданию прикладной цифровой платформы, объединяющей в единую сеть вычислительные ресурсы территориально распределенных суперкомпьютерных центров (СКЦ) коллективного пользования в интересах организаций науки, высшего образования и промышленности Российской Федерации [1]. Представленный в [2] метод управления заданиями пользователей в распределенной сети СКЦ основан на двухуровневой организации системы управления. Единицей вычислительной работы в такой системе является задание, включающее параллельную программу для прикладных расчетов, требования к ресурсам и входные данные. Единицей оборудования в составе распределенной сети СКЦ является высокопроизводительная

вычислительная система (ВС). Для управления отдельной ВС используется локальная система управления ресурсами (ЛСУР), действия всех ВС в сети координирует глобальная система управления ресурсами (ГСУР). В качестве ЛСУР может выступать любая система управления заданиями (СУПЗ, SLURM, PBS), поступающие в ЛСУР задания могут выполняться только на вычислительных ресурсах локальной ВС. Задания могут быть направлены в ГСУР, которая ведет глобальную очередь. Задания из глобальной очереди допускают обработку на вычислительных ресурсах любой ВС сети. Для распределения заданий из глобальной очереди в очереди ЛСУР могут быть применены различные алгоритмы, в том числе основанные на экономических (аукционных) методах.

Вне зависимости от способа распределения заданий глобальной очереди пользователь (субъект) оставляет после себя объекты – зада-

ния, которые продолжительное время (от нескольких минут до нескольких суток) проводят в глобальной очереди. Прошедшее глобальную очередь задание будет претендовать на вычислительные ресурсы назначенной ему локальной ВС. Однако на момент распределения задания из глобальной очереди в локальную пользователь может потерять право доступа к ресурсам назначенной ВС или исчерпать квоту. Фактически за время нахождения в глобальной очереди задание из информационного объекта превращается в субъект, который в том числе должен быть авторизован в СКЦ на уровне планирования локальных ресурсов.

В работе [2] была предложена децентрализованная автоматизированная система управления заданиями и ресурсами в распределенной сети СКЦ. Авторы решили проблему повторной проверки прав доступа путем ввода дополнительных сущностей – диспетчера СКЦ и диспетчера ВС. В предложенной схеме диспетчер СКЦ служит точкой доступа к ГСУР, а коллектив равноправных диспетчеров ВС предназначен для распределения заданий из глобальной очереди в локальные. Недостаток такого подхода в том, что диспетчеры являются узлами отдельной защищенной распределенной сети. Масштабируемость подобного решения ограничена, так как при росте числа СКЦ и ВС в сети увеличиваются технические и организационные сложности поддержки функционирования защищенной сети.

Другим важным направлением исследований в области построения распределенной сети СКЦ является упрощение доступа пользователей к вычислительным ресурсам сети. С этой целью была предложена федеративная схема аутентификации и авторизации, позволяющая пользователям использовать одну учетную запись для доступа ко всем суперкомпьютерным ресурсам сети [3]. Распространенным решением для реализации механизмов аутентификации и использования идентификационной информации в нескольких организациях является федеративное управление идентификацией (*Federated Identity Management – FIM*). В федерации формируются трехсторонние отношения между поставщиком услуг (*Service Provider – SP*), поставщиком идентификационных данных (*Identification Provider – IdP*) и пользователем, которые позволяют пользователям получать доступ к SP с помощью одного набора учетных данных, таких как подписанные доверенным IdP токены и утверждения. Например, пользователь после аутентификации

у поставщика идентификационных данных получает утверждение (билет) SAML (*Security Assertion Markup Language – стандарт управления федеративными идентификационными данными*), и этот IdP-билет позволяет ему авторизоваться на стороне поставщика услуг.

В [3] рассмотрен подход к организации авторизации пользователей в распределенной сети СКЦ, который дает возможность сочетать механизм федеративной аутентификации с традиционными методами доступа к суперкомпьютерам, основанными на протоколе ssh. В работе [4] авторы уточнили процедуру информационного обмена и состава информации (метаданных), предоставляемых организацией-клиентом (IdP) СКЦ при постановке заданий в очередь и передаваемых при перераспределении задания внутри сети СКЦ. Однако в исследованиях [3, 4] не учитывается двухуровневая организация управления заданиями пользователей. Рассмотрен только процесс взаимодействия пользователя с организацией, которая в рамках федерации сети СКЦ не является домашней.

В настоящей работе предложен метод управления заданиям пользователей на основе федеративного подхода организации сети СКЦ. Авторы рассматривают ГСУР как SP в федеративной сети СКЦ.

#### Актуальные исследования в области федеративной организации сети СКЦ

Для реализации федеративного управления идентификацией используются стандартные протоколы аутентификации и авторизации, такие как OAuth 2.0 (<https://datatracker.ietf.org/doc/html/rfc6749>), OIDC ([https://openid.net/specs/openid-connect-core-1\\_0.html](https://openid.net/specs/openid-connect-core-1_0.html)) или SAML (<https://www.oasis-open.org/committees/download.php/13525/sstc-saml-exec-overview-2.0-cd-01-2col.pdf>). Существует достаточно много исследований, посвященных изучению и сравнению этих протоколов, например, работа [5], в которой определен ряд критериев сравнения и дана оценка каждому протоколу по каждому из критериев.

Рассмотрим основные направления исследований в области безопасности федеративного управления идентификацией.

1. *Поиск уязвимостей в различных реализациях протоколов.* В работах [6, 7] проводится сравнение различных реализаций протоколов федеративного управления идентификацией, таких как OAuth/OIDC и SAML. Анализиру-

ются их сильные и слабые стороны, уязвимости и общая безопасность. Например, в [7] исследуется практическая реализация OAuth/OIDC в платформозависимых приложениях из Google Play Store, выявлены типовые нарушения общепринятых рекомендаций.

2. *Моделирование работы протоколов для обнаружения уязвимостей.* В таких исследованиях моделируются различные сценарии использования протоколов для выявления потенциальных уязвимостей. В [8] моделируются потоки авторизации OAuth 2.0 и представляются решения для ослабления известных уязвимостей. В работе [9] авторы предлагают улучшение протокола OAuth 2.0 путем добавления криптографических схем для сохранения конфиденциальности данных пользователя.

3. *Улучшение подотчетности и прозрачности в системах федеративного управления идентификацией.* В работе [10] представлена система TicketT, в [11] – система T-FIM. Оба подхода используют общедоступные журналы для хранения информации о выданных билетах, что позволяет внешним участникам проверять их подлинность, обеспечивая при этом анонимность пользователей.

4. *Внедрение решений по управлению идентификацией на основе блокчейна, позволяющие пользователю взять на себя контроль над своей собственной идентификацией (самосуверенной идентификацией, Self-Sovereign Identity – SSI).* Самосуверенная идентификация – возможность пользователя контролировать свою цифровую идентичность. В работе [12] представлен наиболее полный обзор публикаций и предложений на рынке, касающихся применимости решений SSI на базе технологии блокчейн, а также обсуждаются основные компоненты архитектуры, анализируется безопасность, приводится сравнение систем идентификации на основе SSI с FIM и традиционными централизованными системами управления идентификацией.

В [13] авторы предлагают подход к оценке потенциальных атак на систему SSI и рисков безопасности, используя комбинацию модели дерева атак и матрицы рисков для оценки потенциальных атак и рисков безопасности. В работе [14] сравниваются SSI и FIM с точки зрения внешних угроз. Дана классификация 23 общих угроз для систем FIM по семи основным целям. Кроме этого, 20 общих угроз SSI классифицированы аналогично угрозам FIM для облегчения сравнения. Авторы делают вывод, что в целом система SSI менее подвержена атакам, несет меньше рисков и по сравнению

с системой FIM не так полно раскрывает пользовательские данные. Исследование проводилось на теоретических моделях FIM и SSI, а не на конкретных реализациях этих подходов.

5. *Аутентификация пользователей СКЦ.* В работе [15] представлен набор программных компонентов, интегрированных для создания масштабируемого, готового к внедрению решения многофакторной аутентификации для систем высокопроизводительных вычислений с большим числом пользователей. Предлагаемое решение протестировано на системе, поддерживающей более 10 000 учетных записей пользователей.

Используемые инструменты и процесс интеграции портала MIT SuperCloud Portal для федеративной аутентификации с федерацией InCommon и инфраструктурой открытых ключей (PKI) правительства США обсуждаются в работе [16]. Авторы рассматривают ПО и методы, необходимые для настройки их системы на прием учетных данных, полученных от этих двух поставщиков идентификационных данных. В работе рассмотрены улучшения в области безопасности и удобства использования, наиболее заметным из которых является возможность использовать надежные системы многофакторной аутентификации, развернутые домашними организациями пользователей. В статье уделено внимание различным методам веб-доступа к суперкомпьютерным мощностям и процессу самостоятельной регистрации и проверки ключей протокола SSH.

Работы по интеграции с системами федеративной аутентификации научных организаций проводились в рамках инфраструктуры XSEDE (США) [17]. В ходе проекта AARC [18] разрабатывались проекты федеративной аутентификации для европейской инфраструктуры суперкомпьютерных приложений PRACE [19].

Актуальные публикации демонстрируют активность исследовательских работ в области создания федераций и протоколов федеративной аутентификации. Однако авторы данной статьи не обнаружили решений, связывающих иерархическое управление заданиями и вычислительными ресурсами в распределенной сети СКЦ с ее федеративной организацией.

#### **Метод управления пользовательскими заданиями и вычислительными ресурсами на основе федеративной организации распределенной сети СКЦ коллективного пользования**

Для управления заданиями в иерархически организованной распределенной сети СКЦ пред-

лагается метод, основанный на федеративной аутентификации. Рассмотрим базовые положения метода.

1. В основе метода лежит двухуровневая организация управления в сети СКЦ [1, 2], при которой выделяются глобальный (на уровне сети СКЦ) и локальный (на уровне отдельной ВС из состава сети СКЦ) уровни управления. Для каждой ВС из состава сети выделяется локальная система управления ресурсами, в которой ведется локальная очередь заданий. На уровне распределенной сети глобальной системой управления ресурсами ведется глобальная очередь заданий.

2. Сеть СКЦ объединяет вычислительные системы нескольких СКЦ, связанных коммуникационными каналами. Она характеризуется списком ВС, входящих в ее состав, и общей глобальной очередью заданий. Важно отметить, что сеть имеет децентрализованный характер, поскольку все СКЦ коллективного пользования являются независимыми. Ими владеют и управляют различные научные организации в разных ведомствах. Каждый СКЦ обрабатывает персональные данные тех пользователей, для которых он является домашней организацией, и, таким образом, выступает как провайдер вычислительных ресурсов и одновременно является поставщиком идентификационных данных.

3. Планирование заданий глобальной очереди заключается в их распределении по локальным очередям. Алгоритм распределения может быть произвольным: например, очередное задание может направляться в наименее загруженную локальную очередь. Алгоритм распределения в предлагаемом методе рассматривается в качестве специализированного сервиса, определяющего для каждого задания глобальной очереди ВС, в которой это задание будет выполнено.

4. Представим каждую ЛСУР и соответствующую ей локальную очередь в виде поставщика услуг (SP), глобальную очередь и ГСУР – в качестве отдельного SP. При наличии в составе сети СКЦ  $N$  ВС можно ввести следующие обозначения:  $SP_0$  – глобальная очередь заданий,  $SP_1, SP_2, \dots, SP_N$  – локальные очереди. Пользователь имеет возможность обратиться к любому поставщику услуг из множества  $\{SP_0, SP_1, \dots, SP_N\}$ . Задание, поступившее в локальную очередь  $SP_i$ , может быть выполнено только на ресурсах  $i$ -й ВС. Задание, поступившее в глобальную очередь  $SP_0$ , может быть выполнено на любой ВС из состава сети СКЦ или на их заданном подмножестве.

5. Введем понятие глобального прокси-IdP, который выступает в качестве брокера федерации, функционируя и как IdP, и как SP. Прокси-IdP позволяет упростить дальнейшую интеграцию новых ВС в сеть СКЦ и предоставляет общий интерфейс для аутентификации на глобальной очереди (рис. 1).

6. Примем следующий общий порядок распределения заданий глобальной очереди.

6.1. Пользователь проходит идентификацию, аутентификацию и авторизацию в одном из СКЦ распределенной сети.

6.2. Авторизованный пользователь направляет задание в глобальную очередь  $SP_0$  как информационный объект.

6.3. Задание, прошедшее глобальную очередь и распределенное в некую локальную очередь  $SP_i$ , преобразуется из объекта в субъект и авторизуется в  $i$ -й ВС.

Рассмотрим положения предлагаемого метода более подробно. Отметим, что в контексте организации сети СКЦ каждая ВС имеет уникальное имя в рамках СКЦ и собственную локальную очередь пользовательских заданий. Каждый СКЦ включает одну или несколько ВС с различными именами и характеризуется

- уникальным именем;
- организационной принадлежностью и территориальным расположением;
- списком ВС, входящих в его состав;
- отдельной системой хранения данных, содержащей проектные каталоги, исходные данные и результаты расчетов всех пользователей;
- отдельной системой авторизации пользователей, хранящей учетные записи всех пользователей СКЦ.

Попытаемся сохранить иерархическую архитектуру сети, однако глобальную очередь рассматриваем как поставщика услуг в рамках федерации. СКЦ будут рассмотрены как поставщики вычислительных ресурсов (поставщики услуг) и поставщики идентификационных данных пользователей, для которых они являются домашними организациями. Другими словами, каждый СКЦ должен предоставлять идентификационные данные пользователей через собственный IdP, поддерживающий стандартный протокол идентификации SAML,

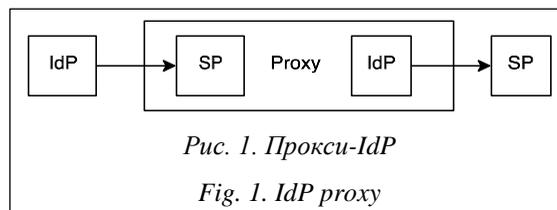


Рис. 1. Проксу-IdP

Fig. 1. IdP proxy

соответствующий определенному набору атрибутов (рис. 2).

Следует отметить, что можно обойтись и без прокси-IdP, однако тогда усложняется интеграция СКЦ в федерацию. Каждый СКЦ будет обязан синхронизировать свои метаданные с другим СКЦ и договариваться о формате передачи данных (рис. 3). Поскольку в данном случае федерация статическая, то есть создана на уровне администратора и связана юридическим договором с использованием определенного набора административных процедур, процесс интеграции нового СКЦ в федерацию будет занимать продолжительное время.

Использование прокси-IdP позволяет СКЦ синхронизировать метаданные только с глобальным прокси-IdP. Введение прокси-IdP дает возможность определить набор необходимых атрибутов, который должны передавать провайдеры каждого СКЦ. Прокси-IdP может сам приводить атрибуты к общему формату или дополнять их (например, для авторизации) и передавать поставщику вычислительных ресурсов. Подобная организация федерации значительно упрощает процесс интеграции в нее новых СКЦ (рис. 4). Однако у такого решения есть очевидный минус – появляется точка централизации.

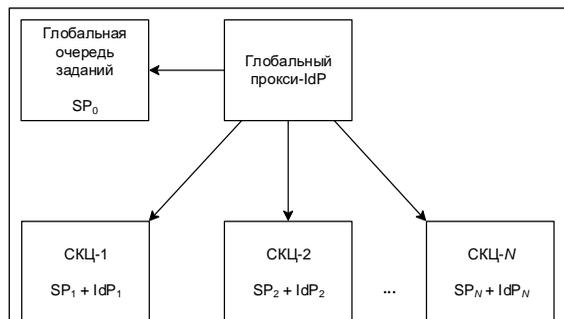


Рис. 2. Общая схема федерации СКЦ

Fig. 2. General scheme of SCC federation

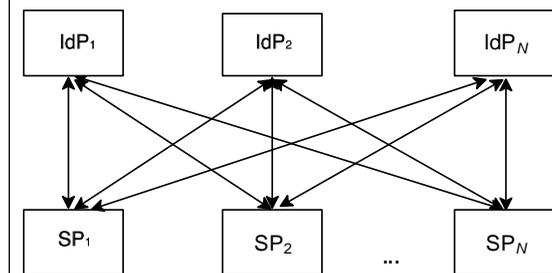


Рис. 3. Организация федерации без прокси-IdP

Fig. 3. Federation without IdP proxy

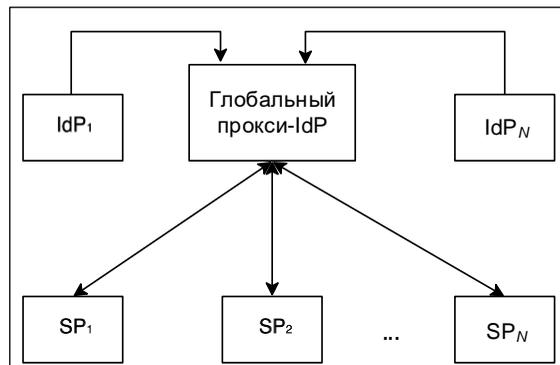


Рис. 4. Организация федерации с прокси-IdP

Fig. 4. Federation with IdP proxy

### Порядок функционирования глобальной системы управления заданиями и вычислительными ресурсами распределенной сети СКЦ

Управление заданиями и вычислительными ресурсами основано на обращении пользователя для постановки задания в глобальную очередь и последующей авторизации задания в ЛСУР (рис. 5). Сама организация глобальной очереди инвариантна и рассматривается авторами как специализированный сервис в виде некоторого черного ящика. Пример одной из возможных схем организации глобальной очереди приведен в работе [2].

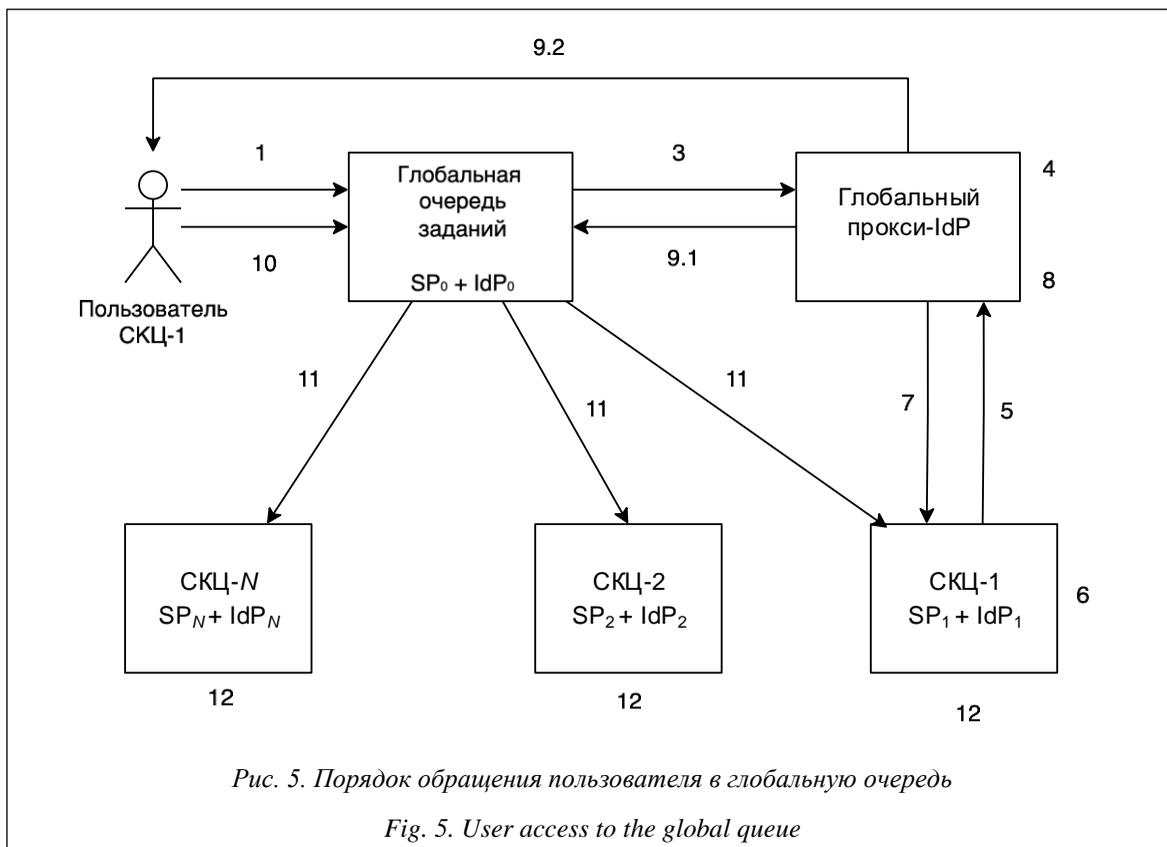
Глобальная очередь должна также предоставлять пользователю удобный интерфейс для управления и мониторинга заданий в очереди, то есть позволять

- добавлять новые глобальные задания;
- отслеживать состояния глобальных заданий;
- получать результаты выполнения глобальных заданий;
- просматривать конфигурацию сети СКЦ;
- проверять конфигурацию пользовательских настроек.

Добавление СКЦ (BC) в состав сети СКЦ и удаление СКЦ (BC) из состава сети осуществляются администраторами СКЦ и администраторами глобальной очереди путем синхронизации метаданных.

Задание может иметь один из статусов:

- «в глобальной очереди» – задание находится в глобальной очереди, ожидает распределения в BC какого-либо СКЦ сети;
- «в локальной очереди СКЦ X» – задание ожидает в локальной очереди освобождения ресурсов BC;



– «выполняется» – задание выполняется на вычислительных ресурсах;

– «завершено» – задание выполнилось и освободило ресурсы;

– «ошибка выполнения» – произошла ошибка на любом из этапов обработки задания.

Рассмотрим порядок обращения пользователя к глобальной очереди (рис. 5).

1. Пользователь СКЦ-1 пытается получить доступ к глобальной очереди заданий.

2. Пользователь выбирает глобальный прокси-IdP в качестве поставщика идентификационных данных.

3. Перенаправление на глобальный прокси-IdP.

4. Пользователь выбирает СКЦ-1 как IdP<sub>1</sub>.

5. Перенаправление на IdP<sub>1</sub>.

6. Аутентификация.

7. Передача атрибутов глобальному прокси-IdP.

8. Пользователь дает разрешение на передачу атрибутов глобальному прокси-IdP.

9.1. Передача утверждения от прокси-IdP SP<sub>0</sub> глобальной очереди.

9.2. Передача ссылки на утверждение пользователю.

10. Получение доступа к глобальной очереди, постановка задания в глобальную очередь.

11. Передача задания в ЛСУР одного из СКЦ и утверждения от IdP<sub>0</sub> глобальной очереди.

12. Проверка уровня доверия к процессу аутентификации утверждения SP<sub>x</sub> СКЦ X.

12.1. Если уровень доверия к процессу аутентификации соответствует политике СКЦ X, то авторизация задания и предоставление доступа к ЛСУР, уведомление пользователя о постановке задания в локальной очереди СКЦ X.

12.2. Если уровень доверия к процессу аутентификации не соответствует политике СКЦ X, то отказ в доступе к ЛСУР, уведомление пользователя с требованием повторного прохождения процедуры аутентификации с соответствующим уровнем доверия, например, с помощью двухфакторной аутентификации.

Присоединение СКЦ к сети не должно сказываться на его локальных пользователях. Соответственно, при распределении нового задания из глобальной очереди должны соблюдаться ограничения, установленные администратором СКЦ, например, по числу глобальных заданий в локальной очереди или требуемых ресурсов для глобального задания. Администратор СКЦ также может выделить отдельный раздел ВС для обработки заданий из ГСУР.

Присоединение СКЦ к сети СКЦ не сказывается на процессе получения доступа пользо-

вателей к вычислительным ресурсам домашней организации. Пользователи могут подключаться к ЛСУР напрямую, например, по SSH. При этом появится возможность подключения к ЛСУР с помощью глобального IdP через SSH-соединение, например, как рассмотрено в [4]. Отметим, что существуют и другие решения, в частности, предложенные в [20].

Процесс получения пользователем доступа к ЛСУР посредством глобального прокси-IdP аналогичен процессу доступа к глобальной очереди (рис. 6).

1. Пользователь СКЦ-2 пытается получить доступ к ресурсам СКЦ-1.
2. Пользователь выбирает глобальный прокси-IdP как поставщика идентификационных данных.
3. Перенаправление на глобальный прокси-IdP.
4. Пользователь выбирает СКЦ-2 как IdP.
5. Перенаправление на IdP<sub>2</sub>.
6. Аутентификация.
7. Передача атрибутов глобальному прокси-IdP.
8. Пользователь дает разрешение на передачу атрибутов глобальному прокси-IdP.
9. Предоставление доступа, пользователь может ставить задания в локальную очередь СКЦ-1.

### Безопасность данных пользователей при федеративной организации распределенной сети

Современный характер научных исследований, основанный на кооперации ученых, обу-

словил создание интегрированных исследовательских инфраструктур, например, сети СКЦ. Подобные инфраструктуры имеют распределенный характер и охватывают несколько административных доменов. Инфраструктуры выступают в качестве поставщиков услуг, предоставляемых независимыми организациями. Эти организации должны работать согласованно, обмениваться данными между собой, чтобы предоставлять услуги общим пользователям. Для этого необходимо хранить, обрабатывать и передавать внутри инфраструктуры персональные данные пользователей. Чтобы обеспечить конфиденциальность персональных данных, а также соблюдение регуляторных требований, передаваемые и обрабатываемые данные должны быть соответствующим образом защищены. Соответственно, необходимо разработать политику не только предоставления атрибутов поставщиками идентификационных данных поставщикам услуг, но и обработки предоставленных персональных данных. Эти вопросы довольно сложны и выходят за рамки статьи, в работе рассматриваются лишь общие риски безопасности информации в объединенной распределенной инфраструктуре.

Здесь и далее под данными будут подразумеваться только те, которые необходимы для учета, мониторинга и взаимодействия, но не исследовательские наборы данных, порой содержащие конфиденциальную информацию. Кроме этого, не планируется подробно исследовать вопрос раскрытия атрибутов пользователя (данных, передаваемых от поставщика идентификационных данных поставщику услуг), а рассмотрены лишь ситуации, когда атрибуты

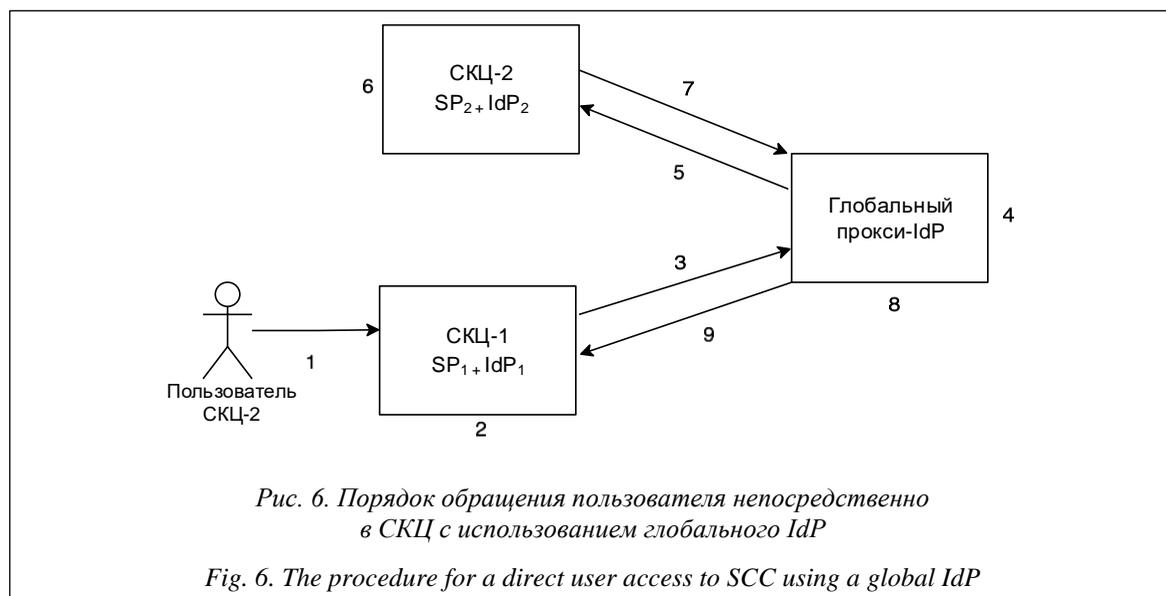


Рис. 6. Порядок обращения пользователя непосредственно в СКЦ с использованием глобального IdP

Fig. 6. The procedure for a direct user access to SCC using a global IdP

передаются с целью учета, мониторинга и организации взаимодействия.

В стандарте X.1250 федерация определяется как ассоциация, объединяющая любое количество поставщиков услуг и поставщиков идентификационных данных. Тот факт, что различные поставщики сформировали ассоциацию, означает, что они должны иметь определенный уровень доверия, достаточный для обмена сообщениями друг с другом. Когда эти сообщения содержат данные аутентификации и авторизации пользователей, позволяющие пользователям иметь доступ к ресурсам объединенной инфраструктуры, получается федеративное управление идентификацией. Федерация позволяет IdP предоставлять атрибуты аутентификации и (необязательно) атрибуты авторизации нескольким отдельно управляемым поставщикам услуг с помощью протоколов и утверждений федерации. Аналогично поставщики услуг могут использовать более одного поставщика идентификационных данных в качестве источников идентификационных данных. В FIM пользователь может использовать свои учетные данные от одного или нескольких поставщиков идентификационных данных для получения доступа к другим поставщикам услуг в рамках федерации. Другими словами, некоторый SP принимает утверждение, предоставленное некоторым IdP, и использует его для принятия решения об аутентификации и авторизации исходя из доверия к федерации, процессу регистрации и текущему событию аутентификации.

Федеративное управление позволяет минимизировать объем передаваемых атрибутов пользователя, что обеспечивает лучшую защищенность данных по сравнению с традиционными методами. Это достигается техническими методами (могут запрашиваться конкретные атрибуты пользователя) или политикой, которая будет определять категории передаваемых атрибутов, например, только электронная почта организации и логин пользователя. У федерации есть еще одно существенное преимущество: учетные данные пользователей (пароли, ключи доступа) хранятся в одном месте (в домашней организации пользователя), что способствует ограничению распространения информации между организациями.

При оценке рисков подразумеваются риски для пользователей (или субъектов данных), получающих доступ к ресурсам и использующих их. Хотя эти риски отличаются от рисков поставщиков услуг, их можно рассматривать вме-

сте в ситуациях, когда риски одного из них влияют на другого. Например, при проникновении в сервисы может произойти утечка данных, содержащих личную информацию пользователей.

Информация о пользователях раскрывается только при доступе к некоторому сервису, причем лишь та информация, которая требуется этому сервису. В работе [4] рассмотрены соответствующие необходимые атрибуты. Такой сценарий выгоден всем участникам федерации, в том числе организациям, предоставляющим услуги, поскольку он обеспечивает гарантии в отношении информации о пользователях и позволяет поставщикам услуг идентифицировать пользователя и взаимодействовать с ним. В схеме также участвуют доверенные третьи стороны для посредничества между домашней организацией пользователя и поставщиком услуг либо для передачи дополнительных атрибутов. Все это способствует ограничению распространения информации о пользователях.

Дополнительные данные, которые обычно могут быть собраны, – это данные о подключении, такие как IP-адреса, журналы событий и др. Они, хотя и остаются персональными, не являются конфиденциальными. Нет оснований утверждать, что из-за распределенного характера инфраструктуры возрастет уровень угрозы по сравнению с традиционным подходом, например, при сценарии, где один СКЦ столкнется с утечкой пользовательских данных. Во-первых, не все пользователи получают доступ к распределенной инфраструктуре, а во-вторых, последствия все равно будут менее значительными, чем при сценарии, когда глобальный поставщик услуг сталкивается с утечкой данных. Следует учитывать склонность пользователей к повторному применению учетных данных на разных сервисах, что делает федеративный подход более предпочтительным, а последствия утечки данных одного поставщика услуг менее серьезными.

Таким образом, с точки зрения рисков безопасности рассматриваемый метод управления заданиями в распределенной сети СКЦ является более безопасным и предпочтительным для защиты информации, в том числе персональных данных пользователей.

## Заключение

В статье предложен метод управления пользовательскими заданиями в распределенной сети СКЦ коллективного пользования. Метод

основан на применении двухуровневой схемы управления в сочетании с федеративной организацией распределенной сети СКЦ. Двухуровневая иерархия подразумевает наличие уровня глобальной очереди заданий, из которой задания распределяются в локальные очереди ВС из состава сети СКЦ. Предлагаемый метод учитывает то, что за время нахождения в глобальной очереди задание с точки зрения безопасности превращается из объекта в субъект, который заново должен быть авторизован в локальной очереди суперкомпьютера. Суще-

ность метода заключается в том, что глобальная очередь заданий рассматривается как еще один член федерации. Глобальная очередь выступает в качестве поставщика услуг для пользователей федерации и одновременно является поставщиком идентификационных данных для авторизации заданий, распределяемых в локальные очереди СКЦ. Рассмотренные в статье сценарии применения метода показывают, что он позволяет повысить надежность и безопасность обрабатываемых данных, в том числе персональных данных пользователей.

### Список литературы

1. Шабанов Б.М., Овсянников А.П., Баранов А.В., Лещев С.А., Долгов Б.В., Дербышев Д.Ю. Проект распределенной сети суперкомпьютерных центров коллективного пользования // Программные системы: теория и приложения. 2017. № 4. С. 245–262. doi: 10.25209/2079-3316-2017-8-4-245-262.
2. Шабанов Б.М., Телегин П.Н., Овсянников А.П., Баранов А.В., Тихомиров А.И., Ляховец Д.С. Система управления заданиями распределенной сети суперкомпьютерных центров коллективного пользования // Тр. НИИСИ РАН. 2018. Т. 8. № 6. С. 65–73.
3. Баранов А.В., Овсянников А.П., Шабанов Б.М. Федеративная аутентификация в распределенной инфраструктуре суперкомпьютерных центров // Тр. НИИСИ РАН. 2018. Т. 8. № 6. С. 79–83.
4. Гончар А.А., Морин Ю.Н., Овсянников А.П. Некоторые вопросы федеративной аутентификации в распределенной сети суперкомпьютерных центров // Тр. НИИСИ РАН. 2020. Т. 10. № 5-6. С. 13–20.
5. Naik N., Jenkins P. Securing digital identities in the cloud by selecting an apposite federated identity management from SAML, OAuth and OpenID Connect. Proc. Int. Conf. RCIS, 2017, pp. 163–174. doi: 10.1109/RCIS.2017.7956534.
6. Aldosary M., Norah A. A survey on federated identity management systems limitation and solutions. IJNSA, 2021, vol. 13, no. 3, pp. 43–59.
7. Sharif A., Carbone R., Sciarretta G., Ranise S. Best current practices for OAuth/OIDC Native Apps: A study of their adoption in popular providers and top-ranked Android clients. JISA, 2022, vol. 65, art. 103097. doi: 10.1016/j.jisa.2021.103097.
8. Singh J., Chaudhary N. OAuth 2.0: Architectural design augmentation for mitigation of common security vulnerabilities. JISA, 2022, vol. 65, 103091. doi: 10.1016/j.jisa.2021.103091.
9. Sucasas V., Mantas G., Althunibat S., Oliveira L., Antonopoulos A., Otung I., Rodriguez J. A privacy-enhanced OAuth 2.0 based protocol for Smart City mobile applications. Computers & Security, 2018, vol. 74, no. C, pp. 258–274. doi: 10.1016/j.cose.2018.01.014.
10. Chu D., Lin J., Li F., Zhang X., Wang Q., Liu G. Ticket transparency: Accountable single sign-on with privacy-preserving public logs. In: LNICST. Proc. SecureComm, 2019, vol. 304, pp. 511–531. doi: 10.1007/978-3-030-37228-6\_25.
11. Xu B., Zhang Z., Sun A. et al. T-FIM: Transparency in federated identity management for decentralized trust and forensics investigation. Electronics, 2023, vol. 12, no. 17, art. 3591. doi: 10.3390/electronics12173591.
12. Ahmed M.R., Islam A.M., Shatabda S., Islam S. Blockchain-based identity management system and self-sovereign identity ecosystem: a comprehensive survey. IEEE Access, 2022, vol. 10, pp. 113436–113481. doi: 10.1109/ACCESS.2022.3216643.
13. Naik N., Grace P., Jenkins P., Naik K., Song J. An evaluation of potential attack surfaces based on attack tree modelling and risk matrix applied to self-sovereign identity. Computers & Security, 2022, vol. 120, art. 102808. doi: 10.1016/j.cose.2022.102808.
14. Le A., Epiphaniou G., Maple C. A comparative cyber risk analysis between federated and self-sovereign identity management systems. Data & Policy, 2023, vol. 5, art. e38. doi: 10.1017/dap.2023.41.
15. Cyrus W.C., Storm P., Hanlon M.R., Mendoza N. Securing HPC: Development of a low cost, open source multi-factor authentication infrastructure. Proc. Int. Conf. SC, 2017, art. 37. doi: 10.1145/3126908.3126957.
16. Prout A., Klein A., Michaleas P. et al. Securing HPC using Federated Authentication. Proc. IEEE HPEC, 2019, pp. 1–7. doi: 10.1109/HPEC.2019.8916255.
17. Towns J., Cockerill T., Dahan M. et al. XSEDE: Accelerating scientific discovery. Computing in Science & Engineering, 2014, vol. 16, no. 5, pp. 62–74. doi: 10.1109/MCSE.2014.80.
18. Liampotis N. AARC Blueprint Architecture. AARC, 2019, no. AARC-G045, pp. 1–12. URL: <https://core.ac.uk/download/pdf/289271206.pdf> (дата обращения: 10.08.2024).
19. Oorsprong M., O'Neill H. PRACE Annual Report 2022. PRACE, 2022, pp. 1–42. URL: [https://insightm.co.uk/wp-content/uploads/2023/06/PRACE\\_AP\\_22\\_Single\\_Hi\\_res.pdf](https://insightm.co.uk/wp-content/uploads/2023/06/PRACE_AP_22_Single_Hi_res.pdf) (дата обращения: 24.07.2024).
20. Simmel D., Filus S. Flexible enforcement of multi factor authentication with SSH via Linux-PAM for federated identity users. Proc. PEARC, 2017, no. 10, pp. 1–9. doi: 10.1145/3093338.3093392.

**The user job managing in a HPC network using federated authentication**Anton V. Baranov<sup>1,2</sup>✉, Egor E. Kuznetsov<sup>2</sup><sup>1</sup> Joint Supercomputer Center of RAS, Moscow, 119334, Russian Federation<sup>2</sup> National Research Centre “Kurchatov Institute”,  
Moscow, 123182, Russian Federation**For citation**Baranov, A.V., Kuznetsov, E.E. (2024) ‘The user job managing in a HPC network using federated authentication’, *Software & Systems*, 37(4), pp. 461–471 (in Russ.). doi: 10.15827/0236-235X.148.461-471**Article info**

Received: 20.07.2024

After revision: 22.08.2024

Accepted: 30.08.2024

**Abstract.** The paper presents a research on user task management in a distributed network of scientific HPC centers. The HPC network unites supercomputers of different architecture belonging to different centers. Each center independently determines its own security policy and maintains its own user accounting database. This complicates the management of user jobs; in particular, it complicates the operational job redistribution between supercomputers of different centers. The authors base their research methodology on combining two-level hierarchical job management and federated identity management, in particular federated authentication. The paper proposes a new method for managing user jobs in a distributed HPC network based on federated authentication. The upper level of the management hierarchy is a global job queue; the global queue jobs are distributed to the supercomputers of the distributed HPC network. The local queues of these supercomputers form the lower level of the job management hierarchy. Each job should be authenticated and authorized twice: when it is placed in the global queue and when it is allocated to a local queue. The proposed method takes into account that a job, when being in the global queue, from the point of view of information security, turns from an object into a subject, which must be authorized again in a local queue. The paper shows that the application of federated authentication in authorization of users and their jobs allows building a simple and secure scheme of job management in the HPC network. The practical significance of the research is the operation of the job management system in the distributed HPC network and the security analysis of such solution.

**Keywords:** SCC, task management, federated authentication, job queue, service provider, identification provider**Acknowledgements.** The paper was carried out under the government assignment, project no. FNEF-2024-0014**References**

1. Shabanov, B.M., Ovsianikov, A.P., Baranov, A.V., Leshchev, S.A., Dolgov, B.V., Derbyshev, D.Yu. (2017) ‘The distributed network of the supercomputer centers for collaborative research’, *Program Systems: Theory and Applications*, (4), pp. 245–262 (in Russ.). doi: 10.25209/2079-3316-2017-8-4-245-262.
2. Shabanov, B.M., Telegin, P.N., Ovsianikov, A.P., Baranov, A.V., Tikhomirov, A.I., Lyakhovets, D.S. (2018) ‘Task management system for a distributed network of supercomputer centres for collective use’, *Proc. of NIISI RAS*, 8(6), pp. 65–73 (in Russ.).
3. Baranov, A.V., Ovsianikov, A.P., Shabanov, B.M. (2018) ‘Federated authentication in distributed infrastructure of supercomputer centres’, *Proc. of NIISI RAS*, 8(6), pp. 79–83 (in Russ.).
4. Gonchar, A.A., Morin, Yu.N., Ovsianikov, A.P. (2020) ‘Some issues of federated authentication in a distributed network of supercomputer centres’, *Proc. of NIISI RAS*, 10(5-6), pp. 13–20 (in Russ.).
5. Naik, N., Jenkins, P. (2017) ‘Securing digital identities in the cloud by selecting an apposite federated identity management from SAML, OAuth and OpenID Connect’, *Proc. Int. Conf. RCIS*, pp. 163–174. doi: 10.1109/RCIS.2017.7956534.
6. Aldosary, M., Norah, A. (2021) ‘A survey on federated identity management systems limitation and solutions’, *IJNSA*, 13(3), pp. 43–59.
7. Sharif, A., Carbone, R., Sciarretta, G., Ranise, S. (2022) ‘Best current practices for OAuth/OIDC Native Apps: A study of their adoption in popular providers and top-ranked Android clients’, *JISA*, 65, art. 103097. doi: 10.1016/j.jisa.2021.103097.
8. Singh, J., Chaudhary, N. (2022) ‘OAuth 2.0: Architectural design augmentation for mitigation of common security vulnerabilities’, *JISA*, 65, art. 103091. doi: 10.1016/j.jisa.2021.103091.
9. Sucasas, V., Mantas, G., Althunibat, S., Oliveira, L., Antonopoulos, A., Otung, I., Rodriguez, J. (2018) ‘A privacy-enhanced OAuth 2.0 based protocol for Smart City mobile applications’, *Computers & Security*, 74(C), pp. 258–274. doi: 10.1016/j.cose.2018.01.014.
10. Chu, D., Lin, J., Li, F., Zhang, X., Wang, Q., Liu, G. (2019) ‘Ticket transparency: Accountable single sign-on with privacy-preserving public logs’, in *LNICST. Proc. SecureComm*, pp. 511–531. doi: 10.1007/978-3-030-37228-6\_25.
11. Xu, B., Zhang, Z., Sun, A. et al. (2023) ‘T-FIM: transparency in federated identity management for decentralized trust and forensics investigation’, *Electronics*, 12(17), art. 3591. doi: 10.3390/electronics12173591.
12. Ahmed, M.R., Islam, A.M., Shatabda, S., Islam, S. (2022) ‘Blockchain-based identity management system and self-sovereign identity ecosystem: a comprehensive survey’, *IEEE Access*, 10, pp. 113436–113481. doi: 10.1109/ACCESS.2022.3216643.
13. Naik, N., Grace, P., Jenkins, P., Naik, K., Song, J. (2022) ‘An evaluation of potential attack surfaces based on attack tree modelling and risk matrix applied to self-sovereign identity’, *Computers & Security*, 120, art. 102808. doi: 10.1016/j.cose.2022.102808.

14. Le, A., Epiphaniou, G., Maple, C. (2023) 'A comparative cyber risk analysis between federated and self-sovereign identity management systems', *Data & Policy*, 5, art. e38. doi: 10.1017/dap.2023.41.
15. Cyrus, W.C., Storm, P., Hanlon, M.R., Mendoza, N. (2017) 'Securing HPC: Development of a low cost, open source multifactor authentication infrastructure', *Proc. Int. Conf. SC*, art. 37. doi: 10.1145/3126908.3126957.
16. Prout, A., Klein, A., Michaleas, P. et al. (2019) 'Securing HPC using Federated Authentication', *Proc. IEEE HPEC*, pp. 1–7. doi: 10.1109/HPEC.2019.8916255.
17. Towns, J., Cockerill, T., Dahan, M. et al. (2014) 'XSEDE: Accelerating scientific discovery', *Computing in Science & Engineering*, 16(5), pp. 62–74. doi: 10.1109/MCSE.2014.80.
18. Liampotis, N. (2019) 'AARC Blueprint Architecture', *AARC*, (AARC-G045), pp. 1–12, available at: <https://core.ac.uk/download/pdf/289271206.pdf> (accessed August 10, 2024).
19. Oorsprong, M., O'Neill, H. (2022) 'PRACE Annual Report 2022', *PRACE*, pp. 1–42, available at: [https://insightm.co.uk/wp-content/uploads/2023/06/PRACE\\_AP\\_22\\_Single\\_Hi\\_res.pdf](https://insightm.co.uk/wp-content/uploads/2023/06/PRACE_AP_22_Single_Hi_res.pdf) (accessed July 24, 2024).
20. Simmel, D., Filus, S. (2017) 'Flexible enforcement of multi factor authentication with SSH via Linux-PAM for federated identity users', *Proc. PEARC*, (10), pp. 1–9. doi: 10.1145/3093338.3093392.

**Авторы**

**Баранов Антон Викторович**<sup>1,2</sup>, к.т.н.,  
доцент, ведущий научный сотрудник,  
abaranov@jscc.ru  
**Кузнецов Егор Евгеньевич**<sup>2</sup>,  
инженер-исследователь, egor57k@jscc.ru

<sup>1</sup> Межведомственный суперкомпьютерный  
центр РАН, г. Москва, 119334, Россия

<sup>2</sup> Национальный исследовательский центр  
«Курчатовский институт», г. Москва, 123182, Россия

**Authors**

**Anton V. Baranov**<sup>1,2</sup>, Cand. of Sci. (Engineering),  
Associate Professor, Leading Researcher,  
abaranov@jscc.ru  
**Egor E. Kuznetsov**<sup>2</sup>,  
Engineer-Researcher, egor57k@jscc.ru

<sup>1</sup> Joint Supercomputer Center of RAS,  
Moscow, 119334, Russian Federation

<sup>2</sup> National Research Centre "Kurchatov Institute",  
Moscow, 123182, Russian Federation

## Ожидания от производительности вычислительного кластера при выборе параллельной файловой системы

О.С. Аладышев<sup>1,2</sup>✉, А.В. Захарченко<sup>1,2</sup>, В.Ф. Огарышев<sup>1,2</sup>, Б.М. Шабанов<sup>1,2</sup>

<sup>1</sup> Межведомственный суперкомпьютерный центр РАН, г. Москва, 119334, Россия

<sup>2</sup> Национальный исследовательский центр «Курчатовский институт», г. Москва, 123182, Россия

### Ссылка для цитирования

Аладышев О.С., Захарченко А.В., Огарышев В.Ф., Шабанов Б.М. Ожидания от производительности вычислительного кластера при выборе параллельной файловой системы // Программные продукты и системы. 2024. Т. 37. № 4. С. 472–486. doi: 10.15827/0236-235X.148.472-486

### Информация о статье

Группа специальностей ВАК: 2.3.5

Поступила в редакцию: 20.06.2024

После доработки: 27.08.2024

Принята к публикации: 30.08.2024

**Аннотация.** При создании высокопроизводительного вычислительного кластера одной из наиболее актуальных задач является обеспечение производительной внешней системы хранения данных под будущую рабочую нагрузку. В статье исследуются принципы работы параллельных файловых систем, которые могут определять их производительность для различных рабочих нагрузок, и предлагается метод определения пределов производительности внешних систем хранения данных. Основным преимуществом предлагаемого метода является поэтапность выявления пределов производительности. Сначала определяются пределы аппаратной части (инфраструктуры) на основе теоретических расчетов. Потом пределы уточняются с помощью тестов и/или показаний системы мониторинга аппаратной части системы хранения данных. Вместе с выбором файловой системы и конфигурированием аппаратной части формируются программные факторы, которые могут влиять на производительность файловой системы для требуемой рабочей нагрузки. С помощью различных предметных тестов или моделей пределы продолжают уточняться. В конечном итоге выявленные пределы проверяются в предварительно настроенной под требуемую рабочую нагрузку системе хранения данных. Предлагаемый подход к выбору параллельной файловой системы для высокопроизводительного вычислительного кластера и к настройке системы хранения данных для определенного спектра параллельных суперкомпьютерных приложений позволяет обойтись без применения сложных моделей и необходимости анализа больших объемов результатов тестирования. Кроме того, он помогает лучше понять характеристики создаваемой системы хранения данных. Авторский метод поэтапной оценки производительности параллельных файловых систем позволяет упростить и ускорить процесс разработки системы хранения данных с параллельной файловой системой. Метод так же хорошо может работать и с современными специализированными файловыми системами, динамически создаваемыми для суперкомпьютерного приложения. **Ключевые слова:** высокопроизводительный вычислительный кластер, параллельные файловые системы, внешние системы хранения данных, суперкомпьютеры

**Благодарности.** Работа выполнена в МСЦ РАН и НИЦ «Курчатовский институт» по теме FNEF-2024-0016

**Введение.** С момента появления суперкомпьютеров количество процессоров, вычислительных устройств и узлов постоянно росло. Еще в начале 90-х годов основные позиции занимали суперкомпьютеры, состоящие из единого многопроцессорного узла на общей памяти (*Symmetric Multiprocessing – SMP*), а требования к системам хранения данных основывались на принципе централизации. Сейчас в рейтинге TOP500 не найти SMP-системы, да и требования к системам хранения данных существенно изменились. Количество вычислительных узлов в суперкомпьютере многократно увеличилось, а системы хранения стали внешними по отношению к вычислительным узлам. Появились и специализированные *файловые системы* (ФС) ad-hoc [1], создаваемые только на время работы суперкомпьютерного приложения.

Современные *параллельные ФС* (ПФС) состоят из нескольких серверов, обслуживающих единое адресное пространство данных. Каж-

дый сервер может предоставлять доступ одновременно к данным и метаданным, а также только к данным или только к метаданным. Непосредственные носители данных обычно распределяются по серверам данных; программная часть системы хранения разделилась на серверы и клиенты, образуя кластер серверов ФС и вычислительное поле. Требования к таким системам постоянно меняются в зависимости от потребностей приложений, однако основные принципы работы остаются неизменными. Сегодня наиболее актуальным остается принцип масштабируемости. Возможность масштабирования основана в первую очередь на методах разделения данных между параллельными процессами. При увеличении пределов масштабирования увеличиваются и пределы производительности.

В настоящей статье рассматриваются принципы построения ФС для высокопроизводительных вычислительных кластеров, которые

тем или иным способом влияют на время выполнения операций ввода-вывода в приложении, то есть на производительность системы хранения.

### Общие требования к ПФС

ПФС в первую очередь должны соответствовать всем основным правилам POSIX. Их нарушение обычно приводит к значительному уменьшению совместимости, и тем самым обрекает ФС на ограниченное использование. Но для больших вычислительных кластеров некоторыми принципами поступаются в угоду производительности [2], например, штампом времени доступа к файлу (*atime*), который для ПФС не имеет смысла.

На втором месте стоит требование масштабируемости ПФС. Оно относится и к носителям данных, и к механизмам разделения доступа к самим данным. Масштабируемость напрямую влияет на предел производительности.

Следующими требованиями к ПФС являются независимая работа вычислительных узлов над операциями чтения и записи с непересекающимися областями общего файла, а также конкурентные операции записи и чтения различных файлов. Существует множество методов решения этих задач, и выбор наилучшего остается препятствием для команд разработчиков. Основной проблемой является выбор механизма блокировки.

Традиционно суперкомпьютерные приложения, работающие в кластере, требуют параллельного доступа с различных узлов как к общим, так и к отдельным файлам [3]. Другие приложения, например, обработка и поиск информации в больших данных, характеризуются параллельным доступом к отдельным файлам, и преимущественно для чтения. В этом классе приложений общий доступ к отдельным файлам не требуется, но, когда эти файлы расположены в общих каталогах или на общем диске, параллельный доступ к системной информации все равно необходим. Отсюда требование: ПФС должна поддерживать параллельный доступ не только к данным, но и к метаданным. В кластерной системе административная работа, такая как добавление диска в ПФС или оптимизация расположения данных на дисках, является непростой задачей. Эти операции могут влиять на производительность ввода-вывода, даже если они выполняются в фоновом режиме. Увеличение степени

их параллелизма может компенсировать накладные расходы в ПФС и тем самым снизить влияние на производительность.

Закладываемые в ПФС правила по выполнению общих требований и обычно не изменяемые после инсталляции, являются неотъемлемой частью системы хранения и имеют основное влияние на пределы производительности. Поэтому можно считать, что при выборе ПФС для указанной аппаратной архитектуры системы хранения данных определяется и уровень снижения пределов производительности.

### Типы ПФС

Для лучшего понимания пределов производительности разделим системы хранения данных высокопроизводительного вычислительного кластера на несколько типов. Из внешних систем хранения остановимся на двух основных типах ФС:

- централизованные, в которых все запросы на метаданные проходят через единую точку;
- распределенные, в которых метаданные, как и данные, распределены между множеством точек доступа к данным.

Разнообразие распределенных систем хранения определяется разнообразием типов рабочих нагрузок в вычислительной системе. Под традиционную нагрузку (последовательное чтение и запись большими порциями данных) разработаны GPFS (*General Parallel File System*), Lustre (*Linux Cluster*), CXFS (*Clustered External File System*), GFS (*Global File System*).

ПФС Lustre [4], как и многие другие системы этого класса, появилась в суперкомпьютерах кластерного типа для обеспечения контрольной точки. Разделение функций хранения данных и управления метаданными является основной концепцией этой ПФС. Хотя данные и распределены по серверам данных (*Object Storage Servers – OSS*), метаданные плохо распределяются по разным точкам доступа и в ПФС используется только один сервер метаданных, продублированный для надежности, а не для производительности. ПФС Lustre очень хорошо подходит для процессов массовой параллельной записи/чтения в разные файлы/объекты или в разные области одного файла/объекта.

Сетевые ФС, основанные на принципе pNFS [5], как и ПФС Lustre, распределяют данные по различным серверам. Метаданные, а также структура данных для серверов данных могут кэшироваться на узлах, считывающих

данные. Если метаданные устареют, сервер метаданных попросит клиента перечитать карту распределения данных. Расширение pNFS разделяет метаданные и данные между узлами ввода-вывода. Сервер метаданных централизован, как и в Lustre.

Распределенная ФС HDFS (*Hadoop Distributed File System*) [6] предназначена для обработки больших данных с использованием MapReduce. Технология MapReduce не ориентирована только на HDFS и может использоваться в качестве ФС Lustre, SS3, KFS и т.п. Подобно Lustre и pNFS HDFS состоит из двух типов узлов ФС – метаданных и данных. HDFS хранит блоки данных в трех или более копиях на разных узлах данных и позволяет выполнять запись только в конец файла, оставляя исходные данные неизменными. Парадигма обработки больших данных подразумевает перенос выполняющегося программного кода к хранимым данным, что делает такие ФС непригодными для традиционных суперкомпьютерных нагрузок.

В этой работе новые ФС, такие как SuperFS, OceanFS, DAO, ParaStor, flashfs, рассматриваются бегло. Более специализированные ФС, предназначенные для узкого спектра прикладной нагрузки, такие как Ceph, PufferScale, HEP-OS, UnifyFS, Hercules, являются наложенными на базовую архитектуру какой-либо системы хранения, и предел производительности прежде всего зависит от нее. Эти ФС по-прежнему остаются статичными и конфигурируются перед запуском. С появлением более производительных вычислительных систем, которые могут гибко управлять доступными ресурсами для удовлетворения требований различных рабочих нагрузок, системы ad-hoc также должны стать гибкими [1].

Таким образом, каждый тип ПФС имеет свой характер влияния на пределы производительности. Например, ПФС, предназначенные для обработки больших данных, в разы снижают пределы для операций вывода. Это влияние имеет смысл оценивать отдельно от других факторов и после того, как будет оценено влияние основных принципов ПФС.

### Работы по оценке производительности

Производительность ПФС можно оценивать исходя из наблюдений за ее поведением при работе различных приложений. Такого рода мониторинг не должен ухудшать ни основную характеристику суперкомпьютера –

производительность, ни сбалансированность для традиционных суперкомпьютеров, где параллельные процессы синхронизированы. Так, в работе [3] с помощью системы мониторинга произведен анализ профилей выполнения пользовательских программ и показана степень влияния характеристик внешней системы хранения суперкомпьютера кластерного типа на время выполнения суперкомпьютерных приложений.

Оценку производительности ПФС с учетом рабочей нагрузки или без нее проводят и с помощью различных моделей систем хранения. Так, в работах [7, 8] моделируется среднее время отклика жесткого диска как элемента системы хранения, в [9] с помощью аналитического моделирования прогнозируется пропускная способность дискового массива. Исследование [10] посвящено использованию статистической модели для прогнозирования времени отклика систем хранения как черного ящика. В [11] прогнозируются производительность устройств хранения как черных ящиков с помощью CART-моделей, обученных на время отклика на запрос, и характеристики рабочей нагрузки. В работе [12] исследуется использование табличных моделей для измерения производительности дискового массива в других точках. Рассмотренные в указанных работах подходы применимы и к более сложным системам хранения, где используется иерархия носителей информации.

Оценкой факторов, таких как количество серверов данных, журналирование изменений, наличие кэша, степень гранулирования записываемых данных и других, влияющих на производительность ввода-вывода для конкретной ПФС, занимаются давно. Например, в работе [13] показано, что избыточное чередование данных в ФС может снижать производительность ПФС Lustre. В [14] рассматривается иерархическая конфигурация системы хранения, оценивается производительность отдельных компонентов системы хранения и предлагается улучшение масштабируемости суперкомпьютерного приложения путем эмпирического анализа и изменения схемы распределения файлов. Авторы предлагают объединять объекты (файлы) таким образом, чтобы снизить накладные расходы, связанные с большой шириной полосы пропускания устройств хранения данных. Работа [15] посвящена активным системам хранения, в которых простаивающие ресурсы серверов ПФС используются для некоторых операций с рядом лежащими данными. Тем самым

предлагается сократить объем перемещаемых между хранилищем и вычислительными узлами данных. Авторы [16, 17] воспользовались тестом IOR (от LLNL) для имитации рабочей нагрузки ключевых приложений с интенсивным вводом-выводом, чтобы оценить характеристики ПФС. В [18] описывается механизм группового (секционного) объединения запросов от параллельных процессов суперкомпьютерного приложения для уменьшения накладных расходов на синхронизацию ввода-вывода в ПФС Lustre и для увеличения утилизации пропускной способности системы хранения. Авторы исследуют влияние предлагаемого механизма на производительность системы хранения и показывают существенное уменьшение накладных расходов. В [19] показано, что механизмы разделения доступа к данным между параллельными процессами в ПФС Lustre и ROMIO – реализация функций ввода-вывода MPI для приложений – не согласуются между собой. Также установлено, что максимальная производительность ПФС Lustre достигается при одновременной записи параллельными процессами в независимые файлы.

В работе [20] проведен подробный анализ литературы и подобран внушительный список факторов, влияющих на производительность ПФС Lustre при различных конфигурациях системы хранения данных на ее базе. Рассматриваются следующие факторы (настройки или конфигурации) систем хранения данных: *факторы серверов метаданных* (количество серверов данных, наличие кэша на чтение/запись, тактика распределения блоков по дискам и серверам, количество параллельных нитей на сервере, количество памяти, стратегия управления метаданными, факторы дисковой подсистемы метаданных); *факторы дисковой подсистемы* (журналирование, типы носителей, шаблоны распределения данных по носителям, структура дисковой подсистемы, размер и число индексных дескрипторов файла); *факторы серверов данных* (количество дисков и дисковых подсистем, количество параллельных нитей, обслуживающих потоки данных, пропускная способность шин сервера, кэш); *факторы сетевой среды* (тип сети, количество каналов на сервер данных, пропускная способность); *факторы клиентов файловой системы* (кэш, шаблоны доступа к данным: последовательное или случайное чтение/запись, большие или маленькие порции, разделение или совмещение данных, характер запросов ввода-вывода); *факторы клиентской сетевой среды*

(тип сети и пропускная способность, агрегирование или сегментирование каналов, агрегирование или разделение запросов ввода-вывода). Авторы исследования разработали модель, с помощью которой можно определить производительность различных систем хранения с ПФС Lustre при разных настройках и понять влияние всевозможных факторов. При этом не предлагается сравнивать различные ПФС с разными механизмами разделения доступа к данным и т.п. Авторы не рассматривают реализацию ввода-вывода в приложении, которая может существенно менять характер рабочей нагрузки. Но сам предложенный подход может быть полезен и в этих случаях. Авторами также отмечено, что настройки системы хранения системно зависят друг от друга и провести чистый эксперимент только с одним фактором не представляется возможным. Таким образом, в качестве теста производительности той или иной конфигурационной настройки модель не годится, к тому же она сложна в применении.

В работе [21] для выяснения характеристик производительности системы хранения применен более прагматичный подход – тестирование производительности на различных схемах ввода-вывода, которые могут быть использованы в суперкомпьютерных приложениях: запись/чтение данных параллельными процессами большими порциями, каждый в свой файл, каждый из своего файла; запись в различные системы хранения или ФС с различными конфигурациями или чтение в этих системах; создание/удаление данных. Авторы разделяют вопросы аппаратных конфигураций и настройки ФС для того, чтобы лучше понимать возможные пределы производительности системы хранения. Подобный подход к выяснению характеристик производительности систем хранения был применен и в работе [22]. Авторы тестировали ФС, глобально распределенные через WAN, на всевозможных операциях ввода-вывода.

Анализируя существующие подходы разных лет, можно выявить тенденцию смещения акцента от аппаратных решений систем хранения к программным реализациям ФС под определенный спектр рабочей нагрузки. В некоторых работах, например в [20], в модель производительности стремятся включить все факторы, влияющие на производительность, – от аппаратных до программных на самом верхнем уровне, прикладном.

В данном исследовании авторы постарались разделить оценку производительности по уров-

ням и использовать принцип, по которому пределы производительности более низкого уровня (аппаратного) на более высоком уровне можно только уменьшить. Хотя не исключается тот факт, что на более высоком уровне допустимо применять механизмы повторного использования данных (например, кэширования), масштабирования и тиражирования данных (удвоение, утроение информации), а также использовать дополнительные вычислительные ресурсы для сокращения издержек на передачу данных (например, дедубликацию), при которых пределы производительности на самом верхнем уровне могут быть увеличены. Такие методы оптимизации потоков информации в представленной работе не рассматриваются, поскольку считаются их частными случаями.

### Аппаратная зависимость пределов производительности

Для оценки максимальной производительности ФС прежде всего необходимо понять, какие пределы есть у аппаратной части системы хранения. Предполагается, что эти пределы превысить невозможно.

### Предел производительности системы хранения

Рассмотрим общую схему для всех систем хранения данных. В архитектуру системы хранения данных входят следующие компоненты (рис. 1):

- клиенты ФС (узлы, на которых не размещается значительная ФС);
- узлы ФС (узлы, на которых расположена значительная часть ФС, но носители данных не подключены к ним непосредственно);
- узлы ввода-вывода (серверы ФС, на которых работает серверная часть ФС и к которым носители данных подключены напрямую или через сетевую инфраструктуру; узел ввода-вывода может обслуживать только метаданные, только данные или метаданные и данные одновременно);
- носители данных (различные устройства и системы для непосредственного хранения данных; диск, дисковая полка, если она напрямую или через сеть хранения данных подключена к узлам ввода-вывода; хранилище данных с одним, двумя или более серверами резервирования, обеспечивающими эффективное объединение носителей данных более низкого порядка);
- транспортная сеть (сеть доставки данных между клиентами, узлами ФС и узлами ввода-вывода);
- сервисная сеть (сеть для обмена служебной информацией, включая блокировочные токены);
- сеть хранения данных (SAN, сеть обмена данными на уровне блоков).

Именно максимальная пропускная способность сетевой фабрики между параллельными процессами приложения и данными на носителях будет недостижимым пределом производительности.

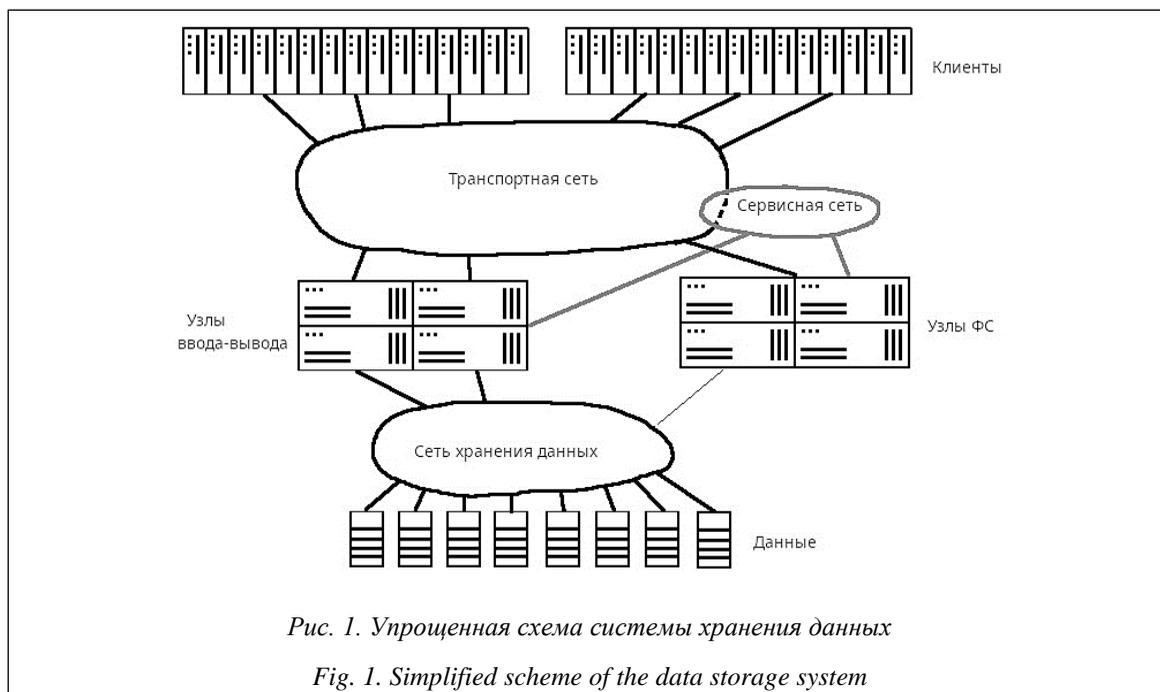


Рис. 1. Упрощенная схема системы хранения данных

Fig. 1. Simplified scheme of the data storage system

### **Факторы, влияющие на снижение предела производительности**

На пределы производительности влияют накладные расходы в организации программно-аппаратной части системы хранения данных.

Уменьшить пределы производительности могут задержка времени отклика, а также дробление сетевых каналов на несколько частей из-за физических нюансов аппаратуры и протоколов связи. При этом клиенты ФС вынуждены подстраиваться под такую схему, если система хранения сама динамически не подстраивается под характер приложения.

Узкие места, такие как несогласованность сетевой и агрегированной производительности носителей информации (учитывая схемы распределения данных и влияние механизмов обеспечения надежности), тоже снижают пределы. Если подобные нюансы имеются в системе, это, скорее, недостатки проектирования, поэтому в представленной модели учитываться не будут.

В ПФС существует глубокая иерархия блоков данных [23]. Это связано с наличием различных типов носителей информации (дисков, памяти узлов ФС и т.д.) и с организацией единого адресного доступа к данным. То есть данные через различные сети могут распределяться блоками между узлами, задействованными в работе ФС, и только потом непосредственно между носителями. Таким образом, различные схемы распределения данных по-разному могут влиять на предел производительности.

На производительность могут повлиять и пределы производительности сервисной сети, по которой ФС осуществляет обмен служебной информацией, такой как блокировочные токены, синхронизация метаданных и т.п. Если транспортная сеть системы хранения объединена с сервисной, то служебная составляющая трафика будет уменьшать пределы производительности всей системы хранения. Служебная составляющая содержит постоянную и дополнительную компоненту. Постоянная компонента – непрерывный трафик служебной информации, не зависящий от рабочей нагрузки ФС. Дополнительная компонента зависит от характера рабочей нагрузки всей системы хранения.

### **Расчет пределов производительности**

Первый и основной предел производительности составляет аппаратная часть системы хранения. Для его оценки рассчитываются следующие компоненты.

1. Пропускная способность клиентов ФС:

$$B_c = N_c C,$$

где  $C$  – пропускная способность канала, выделенного для одного клиентского процесса;  $N_c$  – количество клиентских процессов.

В более общем плане пропускная способность клиентов

$$B_C = \sum_i C_i. \quad (1)$$

2. Пропускная способность  $N_s$  узлов ввода-вывода в транспортной сети, где  $S$  – пропускная способность каналов одного узла:

$$B_s = N_s S. \quad (2)$$

3. Пропускная способность сетевой инфраструктуры между клиентами и серверами ввода-вывода:

$$B_{net}. \quad (3)$$

4. Максимальная производительность хранилища:

$$B_d. \quad (4)$$

### **Программная зависимость пределов производительности**

В работе [20] аппаратные и программные факторы учитываются в одной модели. То, что эти виды факторов не разделяются, во-первых, существенно усложняет процесс выяснения характеристик системы хранения, а во-вторых, вносит неопределенность в понимание степени влияния конкретного фактора.

В настоящем исследовании авторы полностью разделяют эти два вида факторов, что позволяет пошагово оптимизировать систему хранения данных для определенного класса нагрузки. При этом к аппаратным факторам относят такие, как количество носителей и серверов данных, сетевая инфраструктура всей системы хранения, аппаратное распределение блоков данных по носителям. К программным факторам, влияющим на производительность ФС, относят механизмы блокирования данных в ФС, методы организации единого адресного пространства в иерархии данных, механизмы консолидации и распределения блоков данных непосредственно по носителям, кроме аппаратных. Аппаратные настройки обычно устанавливаются статически при начальной инсталляции или настройке системы хранения, в дальнейшем они остаются неизменными. В современных динамически создаваемых ФС эти настройки могут быть изменены, но в очень узких границах. Вклад таких настроек в пределы производительности системы хранения требует дополнительных исследований.

К программным факторам относится также и программное распределение данных по раз-

личным по скорости доступа носителям информации (диски, память узлов ФС и т.п.).

### **Программные факторы**

Чтобы достичь максимальной производительности при выполнении операций чтения и записи в один большой файл, система хранения распределяет блоки данных по всем дискам и дисковым контроллерам. Количество дисков, подключенных к одному узлу ввода-вывода, может быть большим, особенно при наличии сети хранения данных. Например, GPFS делит большие файлы на блоки данных одинаковой длины и последовательно размещает их на разных дисках: первый блок на первом диске, второй – на втором и так далее, блок  $N$  – на диске  $N$ , блок  $N + 1$  – на первом диске и т.д. Чтобы свести к минимуму относительно длительную операцию установки жесткого магнитного диска на нужное место, используются блоки подходящей длины, согласующейся с аппаратными характеристиками дисков, RAID-контроллеров и т.п. Например, когда у дисков наименьший размер блока равен 4 Кб, RAID имеет соответствующий размер блока, равный произведению количества дисков (без избыточности) и размера блока диска, размер блока ФС можно варьировать в зависимости от ее назначения и количества серверов ввода-вывода. Если больших файлов много, то большой размер блока будет иметь преимущество и за одну операцию ввода-вывода возможно прочитать или записать больше данных без накладных расходов. Небольшие файлы и концы больших файлов могут храниться в небольших блоках фиксированной длины, называемых подблоками ФС, например,  $1/32$  от размера блока ФС, как в первых версиях GPFS. Этот метод экономит место на диске и не должен приводить к существенному снижению производительности. Размер блока ФС – основной программный фактор, влияющий на пределы производительности.

Механизмы опережающего чтения и кэширования записи – второй существенный программный фактор, влияющий на пределы производительности. Однако для большинства рабочих нагрузок именно для высокопроизводительного вычислительного кластера опережающее чтение или кэширование записи только уменьшает накладные расходы непосредственно при работе с носителями информации.

Существуют различные способы работы ПФС с большими каталогами. Если в одном каталоге может быть очень много файлов, то механизмы распределения метаданных по точкам

доступа (серверам) существенно влияют на архитектуру системы хранения и непосредственно на производительность ФС.

В большой ФС невозможно выполнять проверку целостности каждый раз, когда ФС монтируется или происходит сбой узла кластера. Одним из способов решения проблемы является ведение специального протокола изменений метаданных каждым узлом ФС отдельно, который обязательно записывает протокол на диск, перед тем как производить сами изменения метаданных на диске. Механизмы регистрации изменений и восстановления при сбоях могут существенно влиять на производительность ФС.

Производительность кластерной ФС может значительно превышать производительность некластерной ФС, если используются параллельные операции чтения и записи на нескольких узлах ФС. С другой стороны, существуют требования к целостности данных и ограничению параллелизма в POSIX. ПФС как минимум должна гарантировать поддержку требований POSIX на одном узле для выполнения операций во всем кластере.

Указанные программные факторы и подобные им являются настраиваемыми. Оптимизацией настроек целесообразно заниматься после определения пределов производительности на более низком уровне иерархии в системе хранения данных.

### **Механизмы блокировки**

Релевантность механизмов блокировки рабочей нагрузки существенно влияет на производительность ФС. В GPFS, например, реализованы два способа POSIX-синхронизации данных:

- *распределенная блокировка*: каждая операция ввода-вывода должна получить блокировку на чтение или запись, чтобы синхронизировать конфликтующую операцию на другом узле перед чтением или записью данных или метаданных;

- *централизованная синхронизация*: все конфликтующие операции ввода-вывода перенаправляются на центральный или указанный узел ввода-вывода, который выполняет запрошенную операцию чтения или записи.

От применяемых способов блокировки существенно зависит величина накладных расходов по параллельному доступу к данным.

Механизм распределенной блокировки обеспечивает больший параллелизм, чем централизованная синхронизация, до тех пор, пока разные

узлы работают с разными частями данных или метаданных с учетом независимости блоков данных в иерархической структуре. С другой стороны, централизованный подход может быть более эффективным при частом чтении или записи данных или метаданных с разных узлов. Когда конфликты блокировок начинают возникать часто, накладные расходы в механизме распределенной блокировки могут превысить стоимость пересылки запросов на центральный узел при централизованной синхронизации. Степень детализации блокировок сильно влияет и на производительность механизма. Так, при высокой степени детализации блокировка небольших блоков данных может привести к большому количеству запросов на блокировку, а при низкой степени детализации блокировка больших блоков данных с большей вероятностью приведет к возникновению конфликтующих запросов. На рисунке 2 изображена зависимость скорости записи данных от вида применяемого механизма их блокировки.

Влияние механизмов блокировки на пределы производительности происходит на двух уровнях. Первый – это внутренние механизмы, не зависящие от характера ввода-вывода в приложении. Здесь механизмы выбираются статично. Например, если размер операции ввода-вывода меньше размера блока ФС, то данные перенаправляются к узлу, который обслуживает весь блок данных, а если больше или равно, то ФС старается распределить нагрузку равномерно между узлами ввода-вывода. Второй уровень – прикладная задача. На этом уровне можно распознавать или статически за-

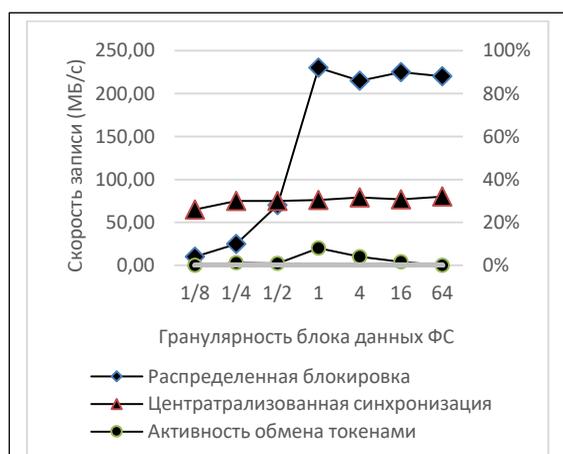


Рис. 2. Зависимость скорости записи данных от механизма блокировки

Fig. 2. Dependence of data writing speed on the locking mechanism

давать шаблоны распределения нагрузки в соответствии с необходимостью руководствоваться POSIX или с отсутствием такой необходимости.

Для эффективной работы с широким спектром вычислительных задач недостаточно применять только один подход к процессу блокировки. Применяемые в ПФС механизмы блокировки обычно специфичны для нее. Релевантность этих механизмов рабочей нагрузке в существенной мере влияет на производительность ФС. Выявление степени влияния применяемых механизмов лучше делать отдельно, от нижнего к верхнему уровню.

### Методика оценки пределов производительности ПФС

Оценку пределов производительности системы хранения данных, а именно аппаратной и программной составляющих, целесообразно проводить поэтапно, начиная с аппаратной составляющей и заканчивая наблюдением за рабочей нагрузкой конкретного приложения.

Для оценки пределов производительности аппаратной части можно воспользоваться несколькими способами.

1. Расчет пределов производительности исходя из технических характеристик сетевых составляющих по модели (1)–(4). В этом случае минимум будет основным пределом. Если постоянный сервисный трафик системы хранения включается в транспортную сеть, то вместо (3) оценивается (3) за вычетом той производительности, которую берет на себя сервисный трафик.

2. Выполнение различных тестов по оценке максимальной производительности сетевых составляющих, причем каждой в отдельности. Вместе с первым способом это позволит выявить несоответствия документированных и реальных технических характеристик аппаратуры, которые могут вызывать, например используемые драйверы устройств и протоколы обмена.

3. Создание рабочей нагрузки, которая будет идеальной для исследуемой системы хранения, и замер параметров производительности. Этот способ является менее точным. В данном случае предел будет выявляться автоматически с учетом всех накладных расходов, в том числе и на сервисный трафик.

4. Мониторинг производительности при идеальной и неидеальной для системы хранения рабочих нагрузках. Рабочая нагрузка может создаваться специальными имитационными те-

стами или приложениями. В данном случае оценивается максимальная производительность сетевых составляющих и далее рассчитывается по модели (1)–(4).

Для лучшего понимания точности оценки пределов производительности желательно использовать все четыре способа.

После оценки пределов производительности аппаратной части можно оценивать программную зависимость и определять эффективность выбранной ФС с различными настройками и в условиях всевозможных рабочих настроек.

Пусть результат теста производительности ФС равен  $B$ . Очевидно, что он не может превышать ни одного из пределов аппаратной части (1), (2), (3), (3) минус сервисный трафик, (4). Если он превышает минимум, значит, для теста не были созданы правильные условия. Возможно, механизм кэширования данных оказал существенное влияние на результаты теста.

Принимая во внимание соотношение

$$\eta = B/\min(B_c, B_s, B_{net}, B_d), \quad (5)$$

получаем эффективность  $\eta$  ФС, то есть ее реальную производительность в процентах от теоретического предела. Соотношение

$$P = \min(B_s, B_{net}, B_d)/C \quad (6)$$

даст предел распараллеливания операций ввода-вывода. Здесь предполагается, что  $B_c$  больше  $B_s$  и не влияет на значения (5) и (6), поскольку количество клиентских узлов обычно не меньше количества узлов ввода-вывода.

Производительность ФС должна увеличиваться линейно с увеличением количества клиентских процессов, пока (5) не достигнет значения «1». Дальнейшее добавление параллельных клиентских процессов не приведет к увеличению общей производительности, но может вызвать ее снижение из-за роста накладных расходов.

### Эксперименты и результаты

Стало очевидным, что для тестирования пределов программной составляющей системы хранения данных необходимо оценить пределы производительности аппаратной части. Это можно сделать теоретически, рассчитав  $B_c$ ,  $B_s$ ,  $B_{net}$  как сумму пропускных способностей сетевых каналов для соответствующего уровня и экспериментально оценив, является ли  $B_d$  пределом. Можно проверить правильность теоретической оценки, выполнив сетевые тесты и/или тесты параллельного ввода-вывода. Для этого необходимо уменьшить влияние ограничений, устранить зависимость операций ввода-

вывода, то есть минимизировать количество блокировок данных во время теста. Например, массово и максимально параллельно записывать новые данные в различные файлы клиентскими процессами. ПФС Lustre, в которой есть единый сервер метаданных, распределяет получаемые непосредственно с клиентов выводимые данные по всем узлам ввода-вывода и по всем носителям данных, если операции ввода-вывода проводятся последовательно и большими блоками. Таким образом, Lustre является идеальной ПФС для тестирования аппаратной составляющей системы хранения.

Разработанная экспериментальная система хранения данных состоит из одного узла ввода-вывода метаданных, пяти узлов ввода-вывода данных и большого количества клиентских узлов (более 64). Узлы соединены сетью Intel OmniPath пропускной способностью 100 Гбит/с на интерфейс каждого узла. Носители данных подключаются к узлам ввода-вывода через те же интерфейсы Intel OmniPath. Носители данных находятся на клиентских узлах. Сетевой интерфейс является двунаправленным, поэтому влияние на производительность сетевого интерфейса при одновременном выводе с клиента и записи на носители данных минимально. При тестировании аппаратной части выполняются массовые параллельные операции последовательной записи большими блоками случайных данных множеством параллельных процессов с клиентов в новые файлы в одном каталоге. Наличие зависимости выводимых данных от единого каталога компенсируется наличием только одной точки синхронизации метаданных в Lustre – одного централизованного сервера ФС, сервера метаданных. Чтобы компенсировать возможное дробление каналов связи, на каждом клиентском узле запускается несколько параллельных процессов. Число процессов обычно делают кратным количеству процессоров на клиентском узле ФС.

Тестовая конфигурация:

–  $M$  узлов,  $M \in \{1, 2, 4, 8, 16, 32, 64\}$ ;

–  $N$  параллельных процессов ввода-вывода на узел,  $N \in \{1, 2, 4, 8, 16, 32, 64\}$ ;

– количество клиентских процессов  $N_s = N \times M$ ;

– пропускная способность канала одного клиентского узла – 100 Гбит/с;

– пропускная способность канала, выделенного

для одного клиентского процесса,  $C = \frac{100 \text{ Гб/с}}{N}$ ;

– пропускная способность клиентов ФС  $B_c = N_c C = M \times 100 \text{ Гбит/с}$ ;

– пропускная способность узла ввода-вывода в транспортной сети  $S = 100 \text{ Гбит/с} = 12\,500 \text{ Мбит/с}$ ;  
 – количество узлов ввода-вывода  $N_s = 5$ ;  
 –  $B_c = N_c S = 5 \times 100 \text{ Гбит/с} = 500 \text{ Гбит/с} = 62\,500 \text{ Мбит/с}$ .

Поскольку сеть является полносвязной, пропускная способность сетевой инфраструктуры транспортной сети  $B_{net} = \min(B_c, B_s)$ .

В представленном примере транспортная сеть объединена с сетью хранения, так как носители данных находятся на клиентских узлах. Сетевые каналы имеют пропускную способность 100 Гбит/с как на ввод, так и на вывод, тем самым общая пропускная способность сети хранения данных  $B_d = B_s$ .

Предполагается, что распределение данных по носителям в сети хранения данных является идеальным.

Intel OmniPath (4xEDR) использует кодировку 64b/66b, накладные расходы составляют около 3 %. Максимальная производительность пяти интерфейсов узлов ввода-вывода составляет 60 606,06 Мб/с. Тем самым теоретический предел производительности экспериментальной системы хранения равен 60,6 Гб/с.

Для уточнения аппаратных пределов производительности воспользуемся еще двумя описанными способами.

Модель рабочей нагрузки в тесте производительности будет наиболее приближена к тому, для чего создавалась ПФС Lustre – массовый параллельный вывод данных, имитация записи контрольной точки. Чтобы обеспечить независимость по данным операций вывода, записываем данные в различные независимые файлы. Файлы будут располагаться в одной и той же директории, но это не должно вызвать

существенных дополнительных накладных расходов в ПФС Lustre, так как управление метаданными централизуется на одном сервере метаданных (узле ФС). Для оценки влияния процесса дробления каналов связи со стороны транспортной сети и сети хранения данных и его компенсации создадим несколько независимых параллельных процессов на одном вычислительном узле. Известно, что сервер метаданных в Lustre распределяет нагрузку по объектам, а не по блокам данных, то есть при операциях ввода-вывода с одного пользовательского процесса передача информации на носители будет происходить только через один узел ввода-вывода. Для второго процесса, даже если он запущен на том же вычислительном узле, будет назначен следующий узел ввода-вывода.

Сетевой трафик отслеживается с помощью специальных средств, что даст понимание полноты использования пропускной полосы в сетевых каналах.

Результаты тестирования с помощью имитации контрольной точки (массовый и продолжительный вывод данных в независимые файлы) приведены в таблице. Из нее видно, что один процесс на вычислительном узле не может полностью утилизировать пропускную полосу сетевых интерфейсов узлов ввода-вывода. Пропускная способность вычислительных узлов в транспортной сети используется полностью, если на одном вычислительном узле выполняются как минимум 8 процессов, на двух узлах – 16 процессов. Когда количество вычислительных узлов становится существенно больше, чем узлов ввода-вывода (в данном случае 32 к 5), для наполнения всей полосы пропускания доста-

**Зависимость производительности операций вывода от количества используемых вычислительных узлов и параллельных процессов вывода на один узел, Мб/с**

**Dependence of output operation performance on the number of used computational nodes and parallel output processes per node, Mb/sec**

		Количество параллельных процессов на узел N						
		1	2	4	8	16	32	64
Количество вычислительных узлов M	1	905,6	1 878,4	3 831,2	6 731,8	6 900,8	6 944,8	6 974,4
	2	1 864,0	3 811,2	7 081,6	12 694,4	13 759,2	14 424,0	14 992,0
	4	3 692,2	8 013,6	15 024,8	25 159,2	34 524,0	37 726,4	33 695,2
	8	7 106,9	16 326,4	28 477,6	40 768,8	43 799,2	42 072,8	43 302,4
	16	14 063,3	31 775,2	45 634,4	50 011,2	50 189,6	50 108,0	49 491,2
	32	23 555,6	47 838,4	50 400,0	49 600,0	50 160,0	50 943,2	48 209,6
	64	26 436,2	49 270,4	50 400,0	49 840,0	49 690,4	49 760,0	50 310,4

точно двух процессов на узел. Эффективная схема распределения параллельных процессов ввода-вывода по вычислительной системе требует отдельного анализа.

Наилучшая производительность системы хранения составляет  $B = 50942,2$  Мб/с и достигается при 32 вычислительных узлах ( $M = 32$ ) и 32 параллельных процессах ввода-вывода на вычислительный узел ( $N = 32$ ). Примерно такая же производительность достигается уже при  $M = 16$  и  $N = 8$  или  $M = 32$  и  $N = 4$ . Это подтверждает теоретические расчеты и то, что самым узким горлом в системе является общая пропускная способность каналов узлов ввода-вывода в транспортной сети (сети хранения данных). При этом пределы производительности  $B_c = 3\,103,03$  Гб/с,  $B_s = 60,60606$  Гб/с и  $B_{net} = B_s$ .

На рисунке 3 изображены графики производительности сетевых интерфейсов узлов ввода-вывода. Во время теста достигалась максимальная скорость передачи данных на узел ввода-вывода – 10,4 Гб/с, что на 16,8 % меньше пропускной полосы.

Эффективность файловой системы (5) составляет  $\frac{50,9422 \text{ Гб/с}}{60,60606 \text{ Гб/с}} \approx 0,84$ .

### Анализ результатов экспериментов

Теоретический предел производительности экспериментальной установки составил 60,6 Гб/с. Влияние программных факторов в проведенных экспериментах по выяснению аппаратных пределов производительности ослабляется различными приемами. Параллельные операции вывода проводятся максимально параллельно и независимо. Для ПФС Lustre с централизованным сервером метаданных не требуется проводить синхронизацию операций

управления метаданными. В отказе от POSIX тоже нет необходимости, поскольку метаданные хранятся в быстрой оперативной памяти централизованного сервера. Механизмы журналирования записи, выделения свободного места для записи, работы с большими каталогами в проведенном эксперименте также в какой-либо значительной степени не влияют на производительность, так как сервер метаданных централизован, файлов немного, записываемые данные независимы. Единый каталог для записываемых файлов, если и влияет на производительность, то тоже в очень незначительной мере из-за наличия единой точки управления метаданными. Влияние эффекта кэширования на уровне ФС (на клиентах или серверах) исключается в схеме ввода-вывода, когда параллельные операции записи проводятся массово и большими блоками. Схема распределения (распараллеливания) потоков данных по носителям данных в ПФС Linux такова, что при большом количестве этих потоков дробление данных и распределение по независимым каналам и носителям данных, перевод блоков данных на нижние уровни иерархии данных проводятся без особых ухищрений и задержек. При выполнении операций записи большими порциями и параллельно при наличии единого сервера метаданных механизмы блокировок, которые могут оказать эффект на скорость записи данных, не задействуются. Таким образом, в проведенных экспериментах выявленная производительность системы хранения данных близка к пределу возможностей ее аппаратной части.

По результатам экспериментов получено снижение пределов производительности на 16,8 %.

Из графиков производительности сетевых интерфейсов узлов ввода-вывода (рис. 3) видно, что основное влияние на предел оказывают

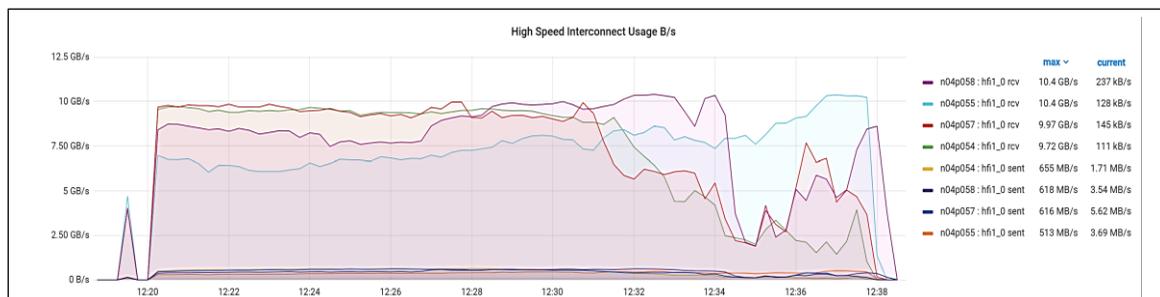


Рис. 3. Максимальная производительность интерфейсов ввода-вывода транспортной сети на узлах ввода-вывода во время работы теста

Fig. 3. Maximum performance of transport network I/O interfaces on I/O nodes during the test operation

факторы, связанные с уменьшением пропускной способности интерфейсов узлов ввода-вывода. Эффект от процесса дробления каналов связи компенсируется количеством клиентских узлов и параллельных процессов на один узел. В данном случае конфликты между параллельными потоками данных не возникают, что можно проверить сетевыми тестами. Авторы теоретически показали, что узким горлом в представленной системе хранения является максимальная пропускная способность интерфейсов узлов ввода-вывода в транспортной сети и сети хранения данных (в данном случае одна и та же сеть). Этот факт подтверждает такую несогласованность и дает повод изменить архитектуру системы хранения данных, увеличить количество узлов ввода-вывода и/или их сетевых интерфейсов.

На снижение пределов производительности для архитектуры системы хранения данных могло повлиять то, что сервер метаданных и серверы обслуживания ФС (узлы ФС) располагаются на узлах ввода-вывода. Данные системы мониторинга показали, что узлы ввода-вывода неравнозначны: два узла ввода-вывода имеют меньшую производительность, чем три других. Такой эффект создается дополнительной нагрузкой на узлы ввода-вывода – наличием сервера метаданных и/или сервера мониторинга системы хранения. Сам эффект можно оценить, если определить разницу в средней производительности узлов ввода-вывода в период тестирования.

В данном случае транспортная сеть и сеть хранения данных объединены и влияют друг на друга в случае, когда для вывода данных возникает обратный трафик. Этот эффект хорошо виден на рисунке 3, где максимальный обратный трафик составляет порядка 5 % от пропускной полосы 12,5 Гб/с. Таким образом, теоретически максимальную производительность интерфейса можно увеличить до 11 Гб/с, если разделить транспортную сеть и сеть хранения данных.

ПФС Lustre устроена так, что трафик по сервисной сети между узлами ввода-вывода между собой и узлами ФС ничтожен, то есть данный фактор несущественен.

Таким образом, выявленный предел производительности аппаратной части экспериментальной системы хранения составляет 51 Гб/с. После этого можно проводить сравнительные тесты для определения эффекта на предел производительности, оказываемого различными программными факторами.

## Выводы

Способы оценки производительности ПФС для высокопроизводительного вычислительного кластера зачастую сложны и недостаточно точны. Предлагаемый способ поэтапной оценки пределов производительности по сравнению с другими способами имеет ряд неоспоримых преимуществ. Во-первых, он позволяет проще и глубже понять основные факторы, влияющие на производительность ПФС для различного рода прикладных приложений. Во-вторых, оценив основной предел производительности системы хранения еще до этапа выбора ПФС, можно понять, какой будет предел производительности вычислительного кластера на операциях ввода-вывода. Если производительности недостаточно для предполагаемой рабочей нагрузки, то необходимо корректировать проект суперкомпьютера. В-третьих, постепенное уточнение пределов производительности позволяет более точно и полно учитывать многочисленные факторы, влияющие на производительность ввода-вывода, а также нюансы архитектур и настроек систем хранения данных и ПФС. Сложная системная зависимость многочисленных факторов друг от друга может не позволить проведение моделирования или тестирования производительности всевозможных вариантов конфигурации системы хранения данных в один шаг. Оценка влияния факторов на разных архитектурных уровнях позволяет поделить все факторы на независимые группы. Влияние малых групп факторов легче оценивать, а при постепенном снижении пределов производительности появляется вероятность избежать ошибок и неточностей при анализе этой зависимости.

Принято считать, что максимальную производительность будущей системы хранения данных можно оценить по архитектурным параметрам системы хранения и аппаратной части с какой-либо ПФС, например, оценить агрегированную пропускную способность сетевых интерфейсов и количественные параметры ресурсов, задействованных в операциях ввода-вывода на кластере с учетом характера ФС. Но на самом деле эти оценки могут быть очень далеки от реальных показателей.

Производители суперкомпьютеров обычно тщательно изучают будущую рабочую нагрузку системы хранения. Имея уже готовые варианты решений, они предлагают самые эффективные. Суперкомпьютерные технологии развиваются очень быстрыми темпами. Все ча-

ще приходится использовать суперкомпьютерные установки для новых видов приложений, в которых характер рабочей нагрузки может сильно отличаться от проектной. Для этого применяют динамически создаваемые системы хранения, которые разворачиваются на оборудовании непосредственно перед запуском задачи. Раньше суперкомпьютеры оснащались одной ПФС, которая могла настраиваться или автоматически настраивалась под характер рабочей нагрузки от разных вычислительных

заданий. Но все чаще в суперкомпьютерах стали применять облачные технологии, когда вычислительные ресурсы динамически выделяются под задачи непосредственно по требованию. Появляются простейшие файловые системы (ad-hoc). Для таких вариантов предложенный метод оценки пределов производительности системы хранения и последующего подбора ресурсов, вида ФС, настроек ее работы подходит намного лучше, чем известные способы оценки.

### Список литературы

1. Vef M.-A., Miranda A., Nou R., Brinkmann A. From static to malleable: improving flexibility and compatibility in burst buffer file systems. In: LNCS, 2023, vol. 13999, pp. 3–15. doi: 10.1007/978-3-031-40843-4\_1.
2. Oeste S., Kluge M., Tschüter R., Nagel W.E. Analyzing parallel applications for unnecessary I/O semantics that inhibit file system performance. In: LNCS, 2023, vol. 13999, pp. 161–176. doi: 10.1007/978-3-031-40843-4\_13.
3. Aladyshov S., Kiselev E.A., Zakharchenko A.V., Shabanov B.M., Savin G.I. Influence of external memory characteristics of supercomputer complexes on parallel programs execution. Lobachevskii J. of Math., 2021, vol. 42, pp. 2493–2502. doi: 10.1134/S1995080221110044.
4. Schwan P. Lustre: Building a filesystem for 1,000-node clusters. Proc. Linux Symposium, 2003, pp. 380–386.
5. Hildebrand D., Honeyman P., Adamson W.A. pNFS and Linux: Working towards a heterogeneous future. CITI Tech. Report, 2007. URL: <https://deepblue.lib.umich.edu/bitstream/handle/2027.42/107906/citi-tr-07-1.pdf?sequence=1&isAllowed=y> (дата обращения: 19.07.2024).
6. Shvachko K., Kuang H., Radia S., Chansler R. The Hadoop distributed file system. Proc. IEEE 26th Symposium MSST, 2010, pp. 1–10. doi: 10.1109/MSST.2010.5496972.
7. Varki E., Merchant A., Xu J., Qiu X. Issues and challenges in the performance analysis of real disk arrays. IEEE Transactions on Parallel and Distributed Systems, 2004, vol. 15, no. 6, pp. 559–574. doi: 10.1109/TPDS.2004.9.
8. Thomasian A., Liu C. Comment on “Issues and challenges in the performance analysis of real disk arrays”. IEEE Transactions on Parallel and Distributed Systems, 2005, vol. 16, no. 11, pp. 1103–1104. doi: 10.1109/TPDS.2005.132.
9. Uysal M., Alvarez G.A., Merchant A. A modular, analytical throughput model for modern disk arrays. Proc. MASCOTS, 2001, pp. 183–192. doi: 10.1109/MASCOT.2001.948868.
10. Kelly T., Cohen I., Goldszmidt M., Keeton K. Inducing models of black-box storage arrays. Tech. Report HPL-2004-108, 2004. URL: [https://www.researchgate.net/publication/244334840\\_Inducing\\_Models\\_of\\_BlackBox\\_Storage\\_Arrays](https://www.researchgate.net/publication/244334840_Inducing_Models_of_BlackBox_Storage_Arrays) (дата обращения: 19.07.2024).
11. Wang M., Au K., Ailamaki A. Storage device performance prediction with CART models. Proc. MASCOTS, 2004, pp. 588–595. doi: 10.1109/MASCOT.2004.1348316.
12. Anderson E. Simple table-based modeling of storage devices. SSP Tech. Report HPL-SSP-2001-4, 2001. URL: [https://www.researchgate.net/publication/2364923\\_HPL--SSP--2001--4\\_Simple\\_table-based\\_modeling\\_of\\_storage\\_devices](https://www.researchgate.net/publication/2364923_HPL--SSP--2001--4_Simple_table-based_modeling_of_storage_devices) (дата обращения: 19.07.2024).
13. Yu W., Vetter J.S., Canon R.S. Exploiting lustre file joining for effective collective IO. Proc. CCGrid, 2007, pp. 267–274. doi: 10.1109/CCGRID.2007.51.
14. Yu W., Oral H.S., Canon R.S. Empirical analysis of a large-scale hierarchical storage system. In: LNTCS. Proc. Euro-Par, 2008, vol. 5168, pp. 130–140. doi: 10.1007/978-3-540-85451-7\_15.
15. Piernas J., Nieplocha J., Felix E.J. Evaluation of active storage strategies for the lustre parallel file system. Proc. Int. Conf. SC, 2007, art. 28. doi: 10.1145/1362622.1362660.
16. Shan H.Z., Shalf J. Using IOR to analyze the I/O performance for HPC platforms. Proc. Cray User Group Conf., 2007. URL: [https://crd.lbl.gov/assets/pubs\\_presos/CDS/ATG/cug07shan.pdf](https://crd.lbl.gov/assets/pubs_presos/CDS/ATG/cug07shan.pdf) (дата обращения: 19.07.2024).
17. Yu W., Oral S., Vetter J. Efficiency evaluation of Cray XT parallel IO stack. Proc. CUG Meeting, 2007. URL: <https://www.cs.fsu.edu/~yuw/pubs/2008-IPDPS-CrayXT-IO.pdf> (дата обращения: 19.07.2024).
18. Yu W., Vetter J. ParColl: Partitioned collective I/O on the Cray XT. Proc. ICPP, 2008, pp. 562–569. doi: 10.1109/ICPP.2008.76.
19. Logan J., Dickens P. Towards an understanding of the performance of MPI-IO in lustre file systems. Proc. IEEE Int. Conf. on Cluster Computing, 2008, pp. 330–335. doi: 10.1109/CLUSTER.2008.4663791.
20. Zhao T., March V., Dong S., See S. Evaluation of a performance model of lustre file system. Proc. Annual ChinaGrid Conf., 2010, pp. 191–196. doi: 10.1109/chinagrid.2010.38.
21. Kluge M. Performance evaluation of the CXFS file system on the HPC/Storage complex for data-intensive computing at the TU Dresden. CMST, 2006, no. 1, pp. 29–32. doi: 10.12921/cmst.2006.SI.01.29-32.
22. Aguilera A., Kluge M., William T., Nagel W.E. HPC file systems in wide area networks: Understanding the performance of lustre over WAN. In: LNTCS. Proc. Euro-Par, 2012, vol. 7484, pp. 65–76. doi: 10.1007/978-3-642-32820-6\_9.

23. Aladyshev O.S., Shabanov B.M., Zakharchenko A.V. Expectations of the high performance computing cluster file system selection. *Lobachevskii J. of Math.*, 2023, vol. 44, pp. 5132–5147. doi: 10.1134/S1995080223120041.

Software &amp; Systems

doi: 10.15827/0236-235X.148.472-486

2024, 37(4), pp. 472–486

### Computing cluster performance expectations when selecting a parallel file system

Oleg S. Aladyshev <sup>1,2</sup>✉, Andrey V. Zakharchenko <sup>1,2</sup>, Vladimir F. Ogaryshev <sup>1,2</sup>, Boris M. Shabanov <sup>1,2</sup>

<sup>1</sup> Joint Supercomputer Center of RAS, Moscow, 119334, Russian Federation

<sup>2</sup> National Research Centre “Kurchatov Institute”, Moscow, 123182, Russian Federation

#### For citation

Aladyshev, O.S., Zakharchenko, A.V., Ogaryshev, V.F., Shabanov, B.M. (2024) ‘Computing cluster performance expectations when selecting a parallel file system’, *Software & Systems*, 37(4), pp. 472–486 (in Russ.). doi: 10.15827/0236-235X.148.472-486

#### Article info

Received: 20.06.2024

After revision: 27.08.2024

Accepted: 30.08.2024

**Abstract.** When creating a high-performance computing cluster, one of the most urgent tasks is to provide a productive external storage system for future workload. The paper studies the principles of parallel file systems' operation, which can determine their performance for different workloads; it proposes a method for determining the performance limits of external storage systems. The main advantage of the proposed method is the phased approach to identifying performance limits. Initially, the authors determine the hardware (infrastructure) limits based on theoretical calculations. Then they refine the limits using tests and/or readings of the storage hardware monitoring system. In addition to selecting an appropriate file system and configuring hardware, the authors form program factors that can affect the file system performance for the required workload. Through various subject tests or models, the limits continue to be refined. Finally, the revealed limits are checked in the storage system preliminarily tuned for the required workload. The proposed approach to selecting a parallel file system for a high-performance computing cluster and to tuning a storage system for a certain range of parallel supercomputer applications allows avoiding using complex models and the need to analyze large amounts of test results. Besides, it helps to understand better the characteristics of the storage system being created. The proposed method of systematic performance evaluation of parallel file systems allows simplifying and speeding up the process of developing a storage system with a parallel file system. The method is also applicable to modern ad-hoc file systems dynamically created for a supercomputer application.

**Keywords:** high-performance computing cluster, parallel file systems, external storage systems, supercomputer

**Acknowledgements.** The paper was carried out under the government assignment, project no. FNEF-2024-0016

#### References

1. Vef, M.-A., Miranda, A., Nou, R., Brinkmann, A. (2023) ‘From static to malleable: improving flexibility and compatibility in burst buffer file systems’, in *LNCS*, 13999, pp. 3–15. doi: 10.1007/978-3-031-40843-4\_1.
2. Oeste, S., Kluge, M., Tschüter, R., Nagel, W.E. (2023) ‘Analyzing parallel applications for unnecessary I/O semantics that inhibit file system performance’, in *LNCS*, 13999, pp. 161–176. doi: 10.1007/978-3-031-40843-4\_13.
3. Aladyshev, S., Kiselev, E.A., Zakharchenko, A.V., Shabanov, B.M., Savin, G.I. (2021) ‘Influence of external memory characteristics of supercomputer complexes on parallel programs execution’, *Lobachevskii J. of Math.*, 42, pp. 2493–2502. doi: 10.1134/S1995080221110044.
4. Schwan, P. (2003) ‘Lustre: Building a filesystem for 1,000-node clusters’, *Proc. Linux Symposium*, pp. 380–386.
5. Hildebrand, D., Honeyman, P., Adamson, W.A. (2007) ‘pNFS and Linux: Working towards a heterogeneous future’, *CITI Tech. Report*, available at: <https://deepblue.lib.umich.edu/bitstream/handle/2027.42/107906/citi-tr-07-1.pdf?sequence=1&isAllowed=y> (accessed July 19, 2024).
6. Shvachko, K., Kuang, H., Radia, S., Chansler, R. (2010) ‘The Hadoop distributed file system’, *Proc. IEEE 26th Symposium MSST*, pp. 1–10. doi: 10.1109/MSST.2010.5496972.
7. Varki, E., Merchant, A., Xu, J., Qiu, X. (2004) ‘Issues and challenges in the performance analysis of real disk arrays’, *IEEE Transactions on Parallel and Distributed Systems*, 15(6), pp. 559–574. doi: 10.1109/TPDS.2004.9.
8. Thomasian, A., Liu, C. (2005) ‘Comment on “Issues and challenges in the performance analysis of real disk arrays”’, *IEEE Transactions on Parallel and Distributed Systems*, 16(11), pp. 1103–1104. doi: 10.1109/TPDS.2005.132.
9. Uysal, M., Alvarez, G.A., Merchant, A. (2001) ‘A modular, analytical throughput model for modern disk arrays’, *Proc. MASCOTS*, pp. 183–192. doi: 10.1109/MASCOT.2001.948868.
10. Kelly, T., Cohen, I., Goldszmidt, M., Keeton, K. (2004) ‘Inducing models of black-box storage arrays’, *Tech. Report HPL-2004-108*, available at: [https://www.researchgate.net/publication/244334840\\_Inducing\\_Models\\_of\\_Black-Box\\_Storage\\_Arrays](https://www.researchgate.net/publication/244334840_Inducing_Models_of_Black-Box_Storage_Arrays) (accessed July 19, 2024).
11. Wang, M., Au, K., Ailamaki, A. (2004) ‘Storage device performance prediction with CART models’, *Proc. MASCOTS*, pp. 588–595. doi: 10.1109/MASCOT.2004.1348316.

12. Anderson, E. (2001) 'Simple table-based modeling of storage devices', *SSP Tech. Report HPL--SSP--2001--4*, available at: [https://www.researchgate.net/publication/2364923\\_HPL--SSP--2001--4\\_Simple\\_table-based\\_modeling\\_of\\_storage\\_devices](https://www.researchgate.net/publication/2364923_HPL--SSP--2001--4_Simple_table-based_modeling_of_storage_devices) (accessed July 19, 2024).
13. Yu, W., Vetter, J.S., Canon, R.S. (2007) 'Exploiting lustre file joining for effective collective IO', *Proc. CCGrid*, pp. 267–274. doi: 10.1109/CCGRID.2007.51.
14. Yu, W., Oral, H.S., Canon, R.S. (2008) 'Empirical analysis of a large-scale hierarchical storage system', in *LNTCS. Proc. Euro-Par*, 5168, pp. 130–140. doi: 10.1007/978-3-540-85451-7\_15.
15. Piernas, J., Nieplocha, J., Felix, E.J. (2007) 'Evaluation of active storage strategies for the lustre parallel file system', *Proc. Int. Conf. SC*, art. 28. doi: 10.1145/1362622.1362660.
16. Shan, H.Z., Shalf, J. (2007) 'Using IOR to analyze the I/O performance for HPC platforms', *Proc. Cray User Group Conf.*, available at: [https://crd.lbl.gov/assets/pubs\\_presos/CDS/ATG/cug07shan.pdf](https://crd.lbl.gov/assets/pubs_presos/CDS/ATG/cug07shan.pdf) (accessed July 19, 2024).
17. Yu, W., Oral, S., Vetter, J. (2007) 'Efficiency evaluation of Cray XT parallel IO stack', *Proc. CUG Meeting*, available at: <https://www.cs.fsu.edu/~yuw/pubs/2008-IPDPS-CrayXT-IO.pdf> (accessed July 19, 2024).
18. Yu, W., Vetter, J. (2008) 'ParColl: Partitioned collective I/O on the Cray XT', *Proc. ICPP*, pp. 562–569. doi: 10.1109/ICPP.2008.76.
19. Logan, J., Dickens, P. (2008) 'Towards an understanding of the performance of MPI-IO in lustre file systems', *Proc. IEEE Int. Conf. on Cluster Computing*, pp. 330–335. doi: 10.1109/CLUSTER.2008.4663791.
20. Zhao, T., March, V., Dong, S., See, S. (2010) 'Evaluation of a performance model of lustre file system', *Proc. Annual ChinaGrid Conf.*, pp. 191–196. doi: 10.1109/chinagrid.2010.38.
21. Kluge, M. (2006) 'Performance evaluation of the CXFS file system on the HPC/Storage complex for data-intensive computing at the TU Dresden', *CMST*, (1), pp. 29–32. doi: 10.12921/cmst.2006.SI.01.29-32.
22. Aguilera, A., Kluge, M., William, T., Nagel, W.E. (2012) 'HPC file systems in wide area networks: Understanding the performance of lustre over WAN', in *LNTCS. Proc. Euro-Par*, 7484, pp. 65–76. doi: 10.1007/978-3-642-32820-6\_9.
23. Aladyshev, O.S., Shabanov, B.M., Zakharchenko, A.V. (2023) 'Expectations of the high performance computing cluster file system selection', *Lobachevskii J. of Math.*, 44, pp. 5132–5147. doi: 10.1134/S1995080223120041.

**Авторы**

**Аладышев Олег Сергеевич**<sup>1,2</sup>, к.т.н.,  
ведущий научный сотрудник, o.aladyshev@jssc.ru  
**Захарченко Андрей Викторович**<sup>1,2</sup>,  
научный сотрудник, avz@jssc.ru  
**Огарышев Владимир Федорович**<sup>1,2</sup>,  
к.ф.-м.н., старший научный сотрудник,  
ogaryshev@jssc.ru  
**Шабанов Борис Михайлович**<sup>1,2</sup>,  
д.т.н., чл.-корр. РАН, зам. директора,  
shabanov@jssc.ru

**Authors**

**Oleg S. Aladyshev**<sup>1,2</sup>, Cand. of Sci. (Engineering),  
Leading Researcher, o.aladyshev@jssc.ru  
**Andrey V. Zakharchenko**<sup>1,2</sup>,  
Research Associate, avz@jssc.ru  
**Vladimir F. Ogaryshev**<sup>1,2</sup>,  
Cand. of Sci. (Physics and Mathematics),  
Senior Researcher, ogaryshev@jssc.ru  
**Boris M. Shabanov**<sup>1,2</sup>, Dr.Sci. (Engineering),  
Corresponding Member of the RAS,  
Deputy Director, shabanov@jssc.ru

<sup>1</sup> Межведомственный суперкомпьютерный центр РАН, г. Москва, 119334, Россия

<sup>2</sup> Национальный исследовательский центр «Курчатовский институт», г. Москва, 123182, Россия

<sup>1</sup> Joint Supercomputer Center of RAS, Moscow, 119334, Russian Federation

<sup>2</sup> National Research Centre "Kurchatov Institute", Moscow, 123182, Russian Federation

## Аналитическая обработка больших массивов данных о событиях кибербезопасности с применением суперкомпьютерных вычислений

И.В. Котенко <sup>1</sup>✉, И.Б. Саенко <sup>1</sup>, И.Б. Паращук <sup>1</sup>,  
В.А. Десницкий <sup>1</sup>, Л.А. Виткова <sup>1</sup>

<sup>1</sup> Санкт-Петербургский федеральный исследовательский центр РАН,  
г. Санкт-Петербург, 199178, Россия

### Ссылка для цитирования

Котенко И.В., Саенко И.Б., Паращук И.Б., Десницкий В.А., Виткова Л.А. Аналитическая обработка больших массивов данных о событиях кибербезопасности с применением суперкомпьютерных вычислений // Программные продукты и системы. 2024. Т. 37. № 4. С. 487–494. doi: 10.15827/0236-235X.148.487-494

### Информация о статье

Группа специальностей ВАК: 2.3.6

Поступила в редакцию: 20.06.2024

После доработки: 27.08.2024

Принята к публикации: 30.08.2024

**Аннотация.** Вопрос кибербезопасности критических инфраструктур осложняется неизбежностью обработки больших объемов данных о событиях безопасности. Это приводит к необходимости разработки информационной технологии, сочетающей аналитическую обработку с суперкомпьютерными вычислениями. Предложены общая схема такой технологии и архитектура реализующей ее системы. В системе выделены компоненты для обнаружения в реальном времени компьютерных атак, аномальной активности и нарушений политик безопасности. Кроме того, компоненты системы позволяют оперативно оценивать защищенность сетевых ресурсов, анализировать риски, принимать решения по защите сетевых ресурсов, расследовать компьютерные инциденты, визуализировать большие массивы данных о событиях кибербезопасности и взаимодействовать с суперкомпьютерным центром. При выборе решения использовались принципы датацентричности, открытой сервис-ориентированной архитектуры и платформенности. Представлено высоко- и низкоуровневое описание архитектуры системы. Продемонстрированы экспериментальные результаты, полученные в суперкомпьютерном центре «Политехнический». Оценка разработанной техники выполнялась с использованием набора данных ИАИ, собранного на испытательном стенде промышленной системы управления паровыми турбинами. Решена задача прогнозирования будущих состояний на основании предыдущих, полученных путем кластеризации системных событий. Реализованный метод прогнозирования показал, что точность зависит от количества учитываемых предыдущих состояний и дальности предсказания. Эти результаты подтвердили эффективность предложенной информационной технологии и продемонстрировали ее высокую производительность.

**Ключевые слова:** информационная технология, кибербезопасность, большие данные, событие кибербезопасности, суперкомпьютерные вычисления

**Благодарности.** Исследование поддержано РФФИ, грант № 21-71-20078, в СПб ФИЦ РАН

**Введение.** Расширение масштабов задач по мониторингу и контролю кибербезопасности, рост объемов собираемых данных о событиях кибербезопасности, а также развитие средств и методов их надежного хранения привели к повышению актуальности разработки новых методов и алгоритмов анализа и обработки больших массивов данных в системах кибербезопасности [1–3]. В современных критических инфраструктурах эта задача требует привлечения технологий интеллектуальной аналитической обработки данных, инновационных методов оценки смыслового содержания информации об угрозах, а также методов и средств реализации высокопроизводительных вычислений, включая суперкомпьютерные [4–6].

Аналитическая обработка больших массивов данных нужна для оперативной и достоверной оценки состояния защищаемой системы, поддержки принятия решений и расследования

компьютерных инцидентов. При этом исследование информации о событиях кибербезопасности, идентификация рисков, а также выработка мер по противодействию угрозам зачастую осуществляются в условиях неопределенности [7]. В современных критических инфраструктурах этот процесс представляет собой технологию целенаправленного поиска информации в массивах гетерогенных данных о подобных событиях. Эта технология подразумевает использование статистических, оптимизационных и других математических алгоритмов, позволяющих находить взаимозависимости (корреляция, классификация и т.п.) и синтезировать дедуктивную информацию [8]. Кроме того, аналитическая обработка должна осуществляться с использованием современных когнитивных методов и алгоритмов, таких как нечеткие, нейросетевые и нейро-нечеткие методы, биоинспирированные алгоритмы оптимизации, ме-

тоды распознавания образов, алгоритмы визуализации данных и проч. Учитывая, что обработке подлежат большие объемы данных, применение суперкомпьютерных вычислений становится необходимым условием достижения требуемой эффективности.

Целью статьи является изложение результатов разработки инновационной информационной технологии аналитической обработки больших массивов данных о событиях кибербезопасности, основанной на применении суперкомпьютерных вычислений.

### **Обзор релевантных работ**

Многие авторы рассматривают различные подходы к применению алгоритмов и отдельных средств для обработки больших массивов гетерогенных данных, к анализу и оценке состояния политик безопасности на основе результатов такой обработки, а также к оценке их защищенности. Однако практическое применение этих подходов продолжает оставаться затруднительным.

В значительной степени это обусловлено необходимостью учета переходных процессов, протекающих в подобных инфраструктурах (например, в энергетике, железнодорожном транспорте, инфраструктуре управления большим городом). Они имеют многокритериальный характер требований, предъявляемых к кибербезопасности, и обуславливают постановку не только линейных, но и нелинейных нестатистических задач анализа и обработки больших массивов данных [9, 10]. Подходы к решению этих задач в рамках существующих методик не рассматривались.

В работах [11, 12] рассматриваются методики сбора текущей статистики и предобработки большого количества собранных гетерогенных данных о событиях кибербезопасности. Они влекут большие временные затраты, что негативно влияет на общее время обработки и оперативность оценки состояния, поддержки принятия решений и расследования компьютерных инцидентов.

Предложенные в работах [5–7] частные методики обработки и оценки защищенности данных работают с большими массивами информации и ориентированы на условия неопределенности, но не учитывают применение суперкомпьютерных вычислений.

Методы поиска и обработки информации, использующие анализ взаимозависимостей параметров кибербезопасности с точки зрения их

корреляции, рассмотрены в работах [13, 14]. Подобные методы сложны, поскольку применение алгоритмов статистической обработки для корреляции событий безопасности связано с обеспечением соответствия исходных данных требованиям по их однородности.

Особого внимания заслуживают исследования, касающиеся суперкомпьютерных вычислений, в которых рассматриваются вопросы обеспечения кибербезопасности самого суперкомпьютера [15–17], а не использования его вычислительных мощностей для решения задач безопасности. В некоторых работах исследуется роль суперкомпьютера для обеспечения национальной безопасности или совершенствования вооружения [18, 19]. Выявлено, что вопрос применения суперкомпьютера для обеспечения кибербезопасности не получил широкого обсуждения в научной литературе.

Таким образом, анализ релевантных работ позволяет говорить не только об актуальности, но и об объективной необходимости формирования информационной технологии, позволяющей реализовать интеллектуальные подходы к аналитической обработке, применяя при этом суперкомпьютерные вычисления. Данная технология должна охватывать оценку состояния, поддержку принятия решений и расследование компьютерных инцидентов. Основными областями применения этой технологии являются критические информационные инфраструктуры, отличающиеся повышенными требованиями к кибербезопасности.

### **Содержание информационной технологии аналитической обработки больших массивов данных о событиях кибербезопасности**

В широком смысле технология – это совокупность методов, процессов и материалов, используемых в какой-либо отрасли деятельности, а также научное описание способов производства. В узком смысле слова технология – это комплекс организационных мер, операций и приемов, направленных на изготовление, обслуживание, ремонт и/или эксплуатацию изделия с номинальным качеством и оптимальными затратами, обусловленных текущим уровнем развития науки, техники и общества в целом [20]. При этом процесс понимается как совокупность действий, направленных на достижение поставленной цели.

Существует несколько определений, поясняющих современную сущность информаци-

онной технологии. Наиболее близким к задачам аналитической обработки больших массивов данных о событиях кибербезопасности с помощью суперкомпьютера является следующее трактование: совокупность методов, производственных процессов и программно-технических средств, объединенных в технологическую цепочку, которые обеспечивают сбор, хранение, обработку, вывод и распространение информации для снижения трудоемкости процессов использования информационных ресурсов, для повышения их надежности и оперативности.

Различают три класса информационных технологий, ориентированных на различные предметные области: *глобальный*, включающий модели, методы и средства, формализующие и позволяющие использовать информационные ресурсы общества в целом; *базовый*, предназначенный для определенной области применения; *конкретный*, реализующий обработку определенных данных при решении конкретных функциональных задач пользователя (планирование, учет, анализ и проч.).

Информационная технология предусматривает технические, коммуникационные средства, организационно-методическое обеспечение и стандартизацию.

Требования, предъявляемые к информационной технологии: высокая степень разделения процесса обработки информации на этапы, включение всего набора элементов для достижения поставленной цели. Кроме того, необходимо наличие регулярного характера – этапы технологического процесса должны быть стандартизированы и унифицированы для более эффективного управления информационными процессами.

К свойствам информационной технологии относятся целесообразность, наличие компонентов и структуры, взаимодействие с внешней средой, целостность, развитие во времени.

Современные и перспективные критически важные инфраструктуры являются киберфизическими системами. Для них характерны: большой парк электронных устройств, огромные объемы данных о событиях безопасности, собираемые для последующего анализа, возможное наложение ограничений на коммуникационно-вычислительные ресурсы этих устройств, большое число пользователей, имеющих доступ к этим устройствам. В результате чего они оказываются подвержены атакам известных и новых видов, нередко целевого назначения.

Для выявления атак и принятия адекватных мер противодействия необходимо проводить

сбор и анализ больших объемов разнородной информации по кибербезопасности в кратчайшие сроки, соответствующие реальному масштабу времени или близкому к нему. Эти функции реализуют системы управления информацией и событиями безопасности (*Security Information and Event Management, SIEM*) [21, 22].

Как правило, в SIEM-системах выделяют три уровня построения. На первом, нижнем уровне осуществляются сбор и предварительная обработка данных о событиях безопасности. На втором уровне реализуется поддержка хранилища данных. На третьем, верхнем уровне выполняются окончательный анализ всей собранной информации по кибербезопасности и выработка мер противодействия. Аналитическая обработка больших массивов данных предполагает реализацию функций второго и третьего уровней SIEM-системы.

В существующих и разрабатываемых перспективных SIEM-системах эти функции включают оценку состояния или текущей ситуации по безопасности – обеспечение осведомленности о безопасности, выработку и выбор вариантов мер противодействия атакам, расследование последствий и причин реализации атак.

В свою очередь, в осведомленность о безопасности входит следующее:

- восприятие ситуации, благодаря чему администратор владеет доступной оперативной информацией о текущей ситуации и накапливает ее;
- оценка воздействия, позволяющая понимать характер и последствия влияния атаки;
- отслеживание ситуации, заключающееся в понимании ее дальнейшего развития;
- анализ тренда атаки и намерений нарушителей;
- анализ причинно-следственных связей;
- оценка достоверности данных о ситуации и ее развитии, заключающаяся в прогнозировании будущих возможных действий нарушителей, в понимании их намерений, возможностей и ресурсов, а также в понимании собственных уязвимостей, возможных контрмер.

Кроме того, при оценке состояния кибербезопасности применяется визуальный анализ данных с помощью стандартных или специально разработанных для этой цели нестандартных моделей визуализации.

Таким образом, в содержательном плане информационная технология аналитической обработки данных о кибербезопасности включает обнаружение в реальном времени компьютерных атак на основе аналитического и имитацион-

ного моделирования и аномальной активности и нарушений критериев и политик кибербезопасности, оперативную оценку защищенности информационных, телекоммуникационных и других критически важных ресурсов, оперативный анализ и управление рисками кибербезопасности, выработку и выбор критериев оценки состояния, поддержку принятия решений, расследование компьютерных инцидентов на основе аналитической обработки больших массивов гетерогенных данных о событиях кибербезопасности, оперативную визуализацию больших массивов данных о событиях кибербезопасности.

Основной целью информационной технологии аналитической обработки больших массивов данных о событиях кибербезопасности, основанной на применении суперкомпьютерных вычислений, является обеспечение надежного и устойчивого сбора, предварительной и итоговой интеллектуальной обработки больших информационных объектов – данных об атаках и иных инцидентах, их достоверном и масштабируемом оценивании в интересах оперативного анализа состояния, поддержки принятия решений и расследования.

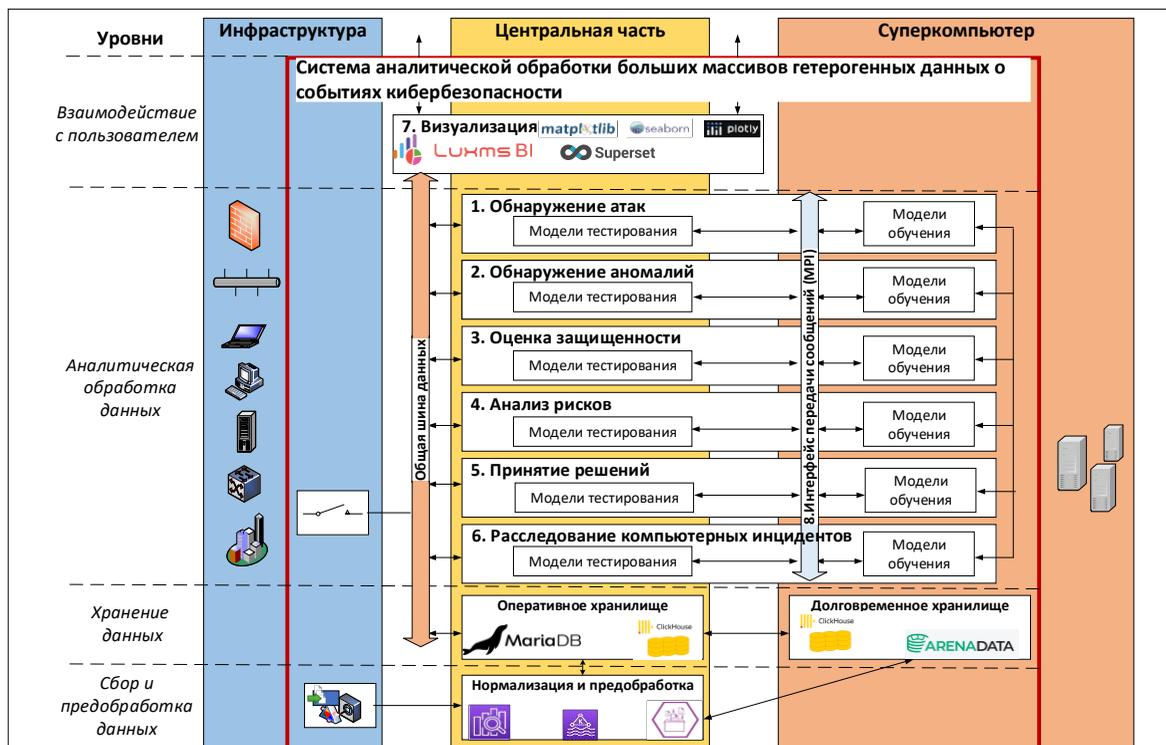
### Архитектура системы аналитической обработки, использующей суперкомпьютерные вычисления

Рассмотрим особенности архитектуры разработанной системы (см. рисунок).

– Система является распределенной, включает часть защищаемой инфраструктуры, часть оборудования суперкомпьютерного центра (СКЦ) и центральную часть (ядро), в которой находятся аналитики кибербезопасности.

– Компоненты 1–6 предназначены для решения конкретных задач по анализу событий безопасности и принятию решений по противодействию атакам. Компонент 7 предназначен для визуального анализа событий безопасности, а также для представления промежуточных и конечных результатов работы других компонентов.

– Каждый из компонентов 1–6 состоит из двух частей: центральной, где обученные модели применяются для тестирования (решаются задачи по выявлению атак, аномалий, по оценке защищенности и проч.), и удаленной, расположенной на стороне СКЦ, где осуществляется обучение моделей. Обмен между ними происходит в компоненте 8 по технологии MPI.



Архитектура системы аналитической обработки больших массивов данных о событиях кибербезопасности, использующей суперкомпьютерные вычисления

Architecture of a system for analytical processing of large data sets on cybersecurity events using supercomputing

– На уровнях «Сбор и предобработка данных» и «Хранение данных» выделяются два типа хранилищ: оперативное, которое находится в ядре, и долговременное, расположенное в СКЦ.

– Обмен между компонентами осуществляется через общую шину данных, которая отвечает за доведение данных до администратора и регуляторов.

Технологический стек предлагаемой системы менялся в ходе исследований в связи с обновлением карты доступных решений. Изначально рассматривались программные продукты с открытым исходным кодом или доступные к использованию по лицензии GNU. В конечном итоге был сделан выбор в пользу решений, которые не ограничивают территориальное использование своих компонентов. Так, для уровня 7 были протестированы платформы анализа и визуализации данных Superset BI (<https://superset.apache.org>), Datalens yandex (<https://cloud.yandex.com/services/datalens>) и Luxms BI (<https://luxmsbi.com>).

При выборе решения основными критериями стали требования соответствия следующим принципам: датацентричность (логика управления процессами кибербезопасности рядом с данными), открытая сервис-ориентированная архитектура (открытое API для интеграции со сторонними программными продуктами), платформенность (поддержка процесса настройки правил корреляции, обнаружения, выбора уровня риска силами пользователя).

Для реализации уровня «Хранение данных» была предложена двухзвенная клиент-серверная архитектура, при которой на сервере находятся балансировщик нагрузки и серверная часть. Сервер анализаторов располагается внутри БД, основная логика управления процессами кибербезопасности реализована на языке PL/pgSQL. Кроме того, на сервере выполняются приложения, реализующие функции хранения данных, управления очередями сообщений (<https://kafka.apache.org>), управления конфигурацией и мониторингом сервисов (<https://kubernetes.io>), выполнения задач извлечения, преобразования и загрузки, а также обмена данными с внешними системами для контроля эффективности и качества работы анализаторов.

Предложенная двухзвенная архитектура превосходит трехзвенную по показателям скорости обработки данных и времени отклика. За счет сокращения количества звеньев экономится время на выборку данных из БД в сервер приложений, что позволяет снизить объем се-

тевого трафика. Таким образом, за счет устранения лишних этапов передачи информации платформа способна эффективно обрабатывать практически неограниченные объемы данных. Это позволяет в полной мере использовать возможности внешних СУБД для аналитической обработки данных. Для хранения и управления метаданными в системе используется Maria DB, основная БД Clickhouse (<https://clickhouse.com>), а для долговременного хранения данных – Arenadata Hadoop (<https://www.arenadata.io/hadoop>).

### Экспериментальная оценка информационной технологии

Разработанная технология прошла экспериментальную оценку на кластере «РСК Торнадо» в СКЦ «Политехнический», который находится на 4-м месте в российском рейтинге и на 22-м – в мировом (<https://rscgroup.ru/project/spbstu-politechnic/>). Кластер содержит 612 узлов, каждый из которых имеет следующие характеристики: 2 процессора Intel Xeon CPU E5-2697 v3 @ 2.60 ГГц, 28 ядер и 56 потоков суммарно, 64 Гб оперативной памяти и 1 Пб общей для всех узлов памяти.

Оценка выполнялась с использованием набора данных HAI (<https://www.kaggle.com/datasets/icsdataset/hai-security-dataset>), который был собран на испытательном стенде промышленной системы управления паровыми турбинами, имитирующем выработку электрической и гидроаккумулирующей энергии. Длина временного ряда в наборе данных была равна 361 200, количество признаков – 86. Решалась задача прогнозирования будущих состояний на основании предыдущих, полученных путем кластеризации системных событий. Метод предсказания основан на рекуррентной нейронной сети, работающей в режиме классификации и состоящей из двух слоев – LSTM и Dense. Слой LSTM по умолчанию имел 512 входов, а слой Dense содержал количество выходов, равное количеству предсказуемых классов.

Реализованный метод прогнозирования показал, что его точность зависит от количества учитываемых предыдущих состояний (NPS) и дальности предсказания (PR). Так, для NPS = 1 получена точность 0,73 при PR = 1 и 0,61 при PR = 9. Для NPS = 4 получена точность 0,82 при PR = 1 и 0,68 при PR = 10.

Таблица представляет данные о времени, затраченном на построение матрицы состояний, при использовании СКЦ и обычного *per-*

сонального компьютера (ПК) при различном количестве потоков.

**Сравнительная оценка времени построения матрицы состояний в зависимости от количества потоков для СКЦ и ПК**

**Comparative assessment of state matrix construction time depending on the number of flows for SCC and PC**

Количество потоков	Время, сек.	
	СКЦ	ПК
1	724,5	538,0
3	306,3	283,8
10	147,8	259,4
15	132,0	220,1

Время построения матрицы состояний при работе на 15 потоках СКЦ уменьшается на 40 %, по сравнению с работой ПК на 15 потоках, и на 75 % при работе ПК на одном потоке.

Таким образом, эксперименты показывают, что разработанная информационная технология для обработки больших массивов данных о событиях кибербезопасности, использующая суперкомпьютерные вычисления, демонстрирует существенный выигрыш во времени решения задач аналитической обработки.

**Заключение**

В статье рассмотрены основные положения и обосновано содержание разработанной информационной технологии аналитической обработки больших массивов данных о событиях кибербезопасности, использующей суперкомпьютерные вычисления. Описана архитектура и представлены технологические аспекты реализующей ее системы. Приведены экспериментальные результаты оценки разработанной технологии на СКЦ «Политехнический». Дальнейшие исследования связаны с апробацией разработанной технологии на различных типах защищаемых инфраструктур.

**Список литературы**

1. Alani M.M. Big data in cybersecurity: A survey of applications and future trends. *J. of Reliable Intelligent Environments*, 2021, vol. 7, pp. 85–114. doi: 10.1007/s40860-020-00120-3.
2. Verma R., Bhatt R. Security issues and challenges of big data analytics. *Proc. Int. Conf. PDGC*, 2022, pp. 61–66. doi: 10.1109/PDGC56933.2022.10053205.
3. Arya A., Malhotra H., Dayanand, Jeberson W. Big data analytics in cyber security. *IJERT*, 2017, vol. 5, no. 10, pp. 1–3.
4. Andrade R.O., Ontaneda N., Silva A., Tello-Oquendo L. et al. Application of big data analytic in cybersecurity. *Proc. Int. Conf. ACC*, 2020, pp. 26–32.
5. Котенко И.В., Саенко И.Б., Браницкий А.А., Парашук И.Б., Гайфулина Д.А. Интеллектуальная система аналитической обработки цифрового сетевого контента для защиты от нежелательной информации // Информатика и Автоматизация. 2021. Т. 20. № 4. С. 755–788. doi: 10.15622/ia.20.4.1.
6. Parashchuk I., Doynikova E., Saenko I., Kotenko I. Selection of countermeasures against harmful information based on the assessment of semantic content of information objects in the conditions of uncertainty. *Proc. Int. Conf. INISTA*, 2020, pp. 1–7. doi: 10.1109/INISTA49547.2020.9194680.
7. Kotenko I.V., Saenko I.B., Parashchuk I.B., Doynikova E.V. An approach for selecting countermeasures against harmful information based on uncertainty management. *ComSIS*, 2022, vol. 19, no. 1, pp. 415–433. doi: 10.2298/CSIS210211057K.
8. Tf M.R., Singh Y. An exploration on big data analysis and data mining methods. *Proc. INCOFT*, 2022, pp. 1–6. doi: 10.1109/INCOFT55651.2022.10094454.
9. Kamara M.K. *Securing Critical Infrastructures*. Xlibris US, Bloomington, 2020, 224 p.
10. Samanis E., Gardiner J., Rashid A. Adaptive cyber security for critical infrastructure. *Proc. ICCPS*, 2022, pp. 304–305. doi: 10.1109/ICCPS54341.2022.00043.
11. Ekpo U. *Introduction to Cyber Security: Fundamentals*. Independently published, NY, 2018, 37 p.
12. Srivastava N., Jaiswal U.C. Big data analytics technique in cyber security: A review. *Proc. ICCMC*, 2019, pp. 579–585. doi: 10.1109/ICCMC.2019.8819634.
13. Bothos M.A., Thanos K.G., Kyriazanos D.M. et al. Correlation and dependence analysis on cyberthreat alerts. *ITU J.: ICT Discoveries*, 2018, vol. 1, no. 2, pp. 1–6.
14. Zhang K., Zhao F., Luo S., Xin Y., Zhu H. An intrusion action-based IDS alert correlation analysis and prediction framework. *IEEE Access*, 2019, vol. 7, pp. 150540–150551. doi: 10.1109/ACCESS.2019.2946261.
15. Zhu G., Zeng Y., Guo M. A security analysis method for supercomputing users' behavior. *Proc. Int. Conf. CSCloud*, 2017, pp. 287–293. doi: 10.1109/CSCloud.2017.19.
16. Баранов А.В., Корепанов П.М., Кузнецов Е.Е. Обеспечение информационной безопасности научного суперкомпьютерного центра // Программные продукты и системы. 2023. Т. 36. № 4. С. 615–631. doi: 10.15827/0236-235X.144.615.
17. Yang B., Yu Y., Wang Z., Li Sh. et al. Research on network security protection of application-oriented supercomputing center based on multi-level defense and moderate principle. *JPCS*, 2021, vol. 1828, art. 012114. doi: 10.1088/1742-6596/1828/1/012114.

18. Агеева А.Ф. Роль суперкомпьютеров в вопросах национальной безопасности // Вестн. академии. 2023. № 1. С. 49–62. doi: 10.51409/v.a.2023.03.01.005.

19. Yalcin H., Daim T., Moughari M.M., Mermoud A. Supercomputers and quantum computing on the axis of cyber security. *Tech. in Society*, 2024, vol. 77, art. 102556. doi: 10.1016/j.techsoc.2024.102556.

20. Несмиянова И.О. Информационные технологии: этапы развития, понятие и классификация // Изв. ТулГУ. Экономические и юридические науки. 2020. № 1. С. 149–155.

21. Котенко И.В., Саенко И.Б., Захарченко Р.И., Величко Д.В. Подсистема предупреждения компьютерных атак на объекты критической информационной инфраструктуры: анализ функционирования и реализации // Вопросы кибербезопасности. 2023. № 1. С. 13–27. doi: 10.21681/2311-3456-2023-1-13-27.

22. Ададуров С.Е., Глухов А.П., Котенко И.В., Саенко И.Б. Интеллектуальные сервисы обеспечения информационной безопасности // Автоматика, связь, информатика. 2022. № 3. С. 27–30.

Software &amp; Systems

doi: 10.15827/0236-235X.148.487-494

2024, 37(4), pp. 487–494

### Analytical processing of large data sets of cybersecurity events using supercomputing

Igor V. Kotenko <sup>1</sup>✉, Igor B. Saenko <sup>1</sup>, Igor B. Parashchuk <sup>1</sup>,  
Vasily A. Desnitsky <sup>1</sup>, Lydia A. Vitkova <sup>1</sup>

<sup>1</sup> St. Petersburg Federal Research Center of the Russian Academy of Sciences,  
St. Petersburg, 199178, Russian Federation

#### For citation

Kotenko, I.V., Saenko, I.B., Parashchuk, I.B., Desnitsky, V.A., Vitkova, L.A. (2024) 'Analytical processing of large data sets of cybersecurity events using supercomputing', *Software & Systems*, 37(4), pp. 487–494 (in Russ.). doi: 10.15827/0236-235X.148.487-494

#### Article info

Received: 20.06.2024

After revision: 27.08.2024

Accepted: 30.08.2024

**Abstract.** The issue of cybersecurity of critical infrastructures is complicated by the need to process large data sets of security events. This leads to the need to develop information technology that combines analytical processing with supercomputing. The authors proposed a general scheme of such technology and architecture of the system realizing it. The system contains components that realize real-time detection of computer attacks, abnormal activity and security policy violations. Furthermore, the system components allow promptly assessing the security of network resources, analyzing risks, making decisions on the protection of network resources, investigating computer incidents, visualizing large data sets of cybersecurity events and interacting with the supercomputer center. The authors used the principles of data-centricity, open service-oriented architecture and platform to select the solution. They presented a high-level and low-level description of the system architecture. The authors demonstrated experimental results obtained at the Polytechnic supercomputer center. They evaluated the developed technique using HAI dataset collected on a testbed of an industrial steam turbine control system. The authors solved the problem of predicting future states based on previous states obtained by clustering system events. The realized prediction method showed that the accuracy depends on the number of considered previous states and the prediction range. These results confirmed the effectiveness of the proposed information technology and demonstrated its high performance.

**Keywords:** information technology; cybersecurity; large data sets; cybersecurity event; supercomputing

**Acknowledgements.** The study was supported by the Russian Science Foundation grant no. 21-71-20078 in St. Petersburg Federal Research Center of the Russian Academy of Sciences

#### References

1. Alani, M.M. (2021) 'Big data in cybersecurity: A survey of applications and future trends', *J. of Reliable Intelligent Environments*, 7, pp. 85–114. doi: 10.1007/s40860-020-00120-3.
2. Verma, R., Bhatt, R. (2022) 'Security issues and challenges of big data analytics', *Proc. Int. Conf. PDGC*, pp. 61–66. doi: 10.1109/PDGC56933.2022.10053205.
3. Arya, A., Malhotra, H., Dayanand, Jeberson, W. (2017) 'Big data analytics in cyber security', *IJERT*, 5(10), pp. 1–3.
4. Andrade, R.O., Ontaneda, N., Silva, A., Tello-Oquendo, L. et al. (2020) 'Application of big data analytic in cybersecurity', *Proc. Int. Conf. ACC*, pp. 26–32.
5. Kotenko, I.V., Saenko, I.B., Branitskiy, A.A., Parashchuk, I.B., Gaifulina, D.A. (2021) 'Intelligent system of analytical processing of digital network content for protection against inappropriate information', *Informatics and Automation*, 20(4), pp. 755–788 (in Russ.). doi: 10.15622/ia.20.4.1.
6. Parashchuk, I., Doynikova, E., Saenko, I., Kotenko, I. (2020) 'Selection of countermeasures against harmful information based on the assessment of semantic content of information objects in the conditions of uncertainty', *Proc. Int. Conf. INISTA*, pp. 1–7. doi: 10.1109/INISTA49547.2020.9194680.

7. Kotenko, I.V., Saenko, I.B., Parashchuk, I.B., Doynikova, E.V. (2022) 'An approach for selecting countermeasures against harmful information based on uncertainty management', *ComSIS*, 19(1), pp. 415–433. doi: 10.2298/CSIS210211057K.
8. Tf, M.R., Singh, Y. (2022) 'An exploration on big data analysis and data mining methods', *Proc. INCOFT*, pp. 1–6. doi: 10.1109/INCOFT55651.2022.10094454
9. Kamara, M.K. (2020) *Securing Critical Infrastructures*. Bloomington: Xlibris US, 224 p.
10. Samanis, E., Gardiner, J., Rashid, A. (2022) 'Adaptive cyber security for critical infrastructure', *Proc. ICCPS*, pp. 304–305. doi: 10.1109/ICCPS54341.2022.00043.
11. Екпо, У. (2018) *Introduction to Cyber Security: Fundamentals*. NY: Independently published, 37 p.
12. Srivastava, N., Jaiswal, U.C. (2019) 'Big data analytics technique in cyber security: A review', *Proc. ICCMC*, pp. 579–585. doi: 10.1109/ICCMC.2019.8819634.
13. Bothos, M.A., Thanos, K.G., Kyriazanos, D.M. et al. (2018) 'Correlation and dependence analysis on cyberthreat alerts', *ITU J.: ICT Discoveries*, 1(2), pp. 1–6.
14. Zhang, K., Zhao, F., Luo, S., Xin, Y., Zhu, H. (2019) 'An intrusion action-based IDS alert correlation analysis and prediction framework', *IEEE Access*, 7, pp. 150540–150551. doi: 10.1109/ACCESS.2019.2946261.
15. Zhu, G., Zeng, Y., Guo, M. (2017) 'A security analysis method for supercomputing users' behavior', *Proc. Int. Conf. CSCloud*, pp. 287–293. doi: 10.1109/CSCloud.2017.19.
16. Baranov, A.V., Korepanov, P.M., Kuznetsov, E.E. (2023) 'Information security of a supercomputer center', *Software & Systems*, 36(4), pp. 615–631 (in Russ.). doi: 10.15827/0236-235X.144.615.
17. Yang, B., Yu, Y., Wang, Z., Li, Sh. et al. (2021) 'Research on network security protection of application-oriented supercomputing center based on multi-level defense and moderate principle', *JPCS*, 1828, art. 012114. doi: 10.1088/1742-6596/1828/1/012114.
18. Ageeva, A.F. (2023) 'The role of supercomputers in matters of national security', *Bull. of the Academy*, (1), pp. 49–62 (in Russ.). doi: 10.51409/v.a.2023.03.01.005.
19. Yalcin, H., Daim, T., Moughari, M.M., Mermoud, A. (2024) 'Supercomputers and quantum computing on the axis of cyber security', *Tech. in Society*, (77), art. 102556. doi: 10.1016/j.techsoc.2024.102556.
20. Nesmiyanova, I.O. (2020) 'Information technologies: Stages of development, concept and classification,' *Izvestiya TulGU. Economic and Legal Sci.*, (1), pp. 149–155 (in Russ.).
21. Kotenko, I.V., Saenko, I.B., Zakharchenko, R.I., Velichko, D.V. (2023) 'Subsystem for prevention of computer attacks against objects of critical information infrastructure: analysis of functioning and implementation', *Cybersecurity Issues*, (1), pp. 13–27 (in Russ.). doi: 10.21681/2311-3456-2023-1-13-27.
22. Adadurov, S.E., Glukhov, A.P., Kotenko, I.V., Saenko, I.B. (2022) 'Intelligent information security services', *Automation, communications, informatics*, (3), pp. 27–30 (in Russ.).

**Авторы**

**Котенко Игорь Витальевич**<sup>1</sup>, д.т.н.,  
профессор, главный научный сотрудник,  
руководитель лаборатории,  
ivkote@comsec.spb.ru

**Саенко Игорь Борисович**<sup>1</sup>, д.т.н.,  
профессор, главный научный сотрудник,  
ibsaen@comsec.spb.ru

**Парашук Игорь Борисович**<sup>1</sup>, д.т.н.,  
профессор, ведущий научный сотрудник,  
parashchuk@comsec.spb.ru

**Десницкий Василий Алексеевич**<sup>1</sup>, к.т.н.,  
доцент, старший научный сотрудник,  
desnitsky@comsec.spb.ru

**Виткова Лидия Андреевна**<sup>1</sup>, к.т.н.,  
старший научный сотрудник,  
vitkova@comsec.spb.ru

**Authors**

**Igor V. Kotenko**<sup>1</sup>, Dr.Sci. (Engineering),  
Professor, Chief Researcher,  
Head of Laboratory  
ivkote@comsec.spb.ru

**Igor B. Saenko**<sup>1</sup>, Dr.Sci. (Engineering),  
Professor, Chief Researcher,  
ibsaen@comsec.spb.ru

**Igor B. Parashchuk**<sup>1</sup>, Dr.Sci. (Engineering),  
Professor, Leading Researcher,  
parashchuk@comsec.spb.ru

**Vasily A. Desnitsky**<sup>1</sup>, Cand. of Sci. (Engineering),  
Senior Researcher,  
desnitsky@comsec.spb.ru

**Lydia A. Vitkova**<sup>1</sup>, Cand. of Sci. (Engineering),  
Senior Researcher,  
vitkova@comsec.spb.ru

<sup>1</sup> Санкт-Петербургский федеральный  
исследовательский центр РАН,  
г. Санкт-Петербург, 199178, Россия

<sup>1</sup> St. Petersburg Federal Research Center  
of the Russian Academy of Sciences,  
St. Petersburg, 199178, Russian Federation

## Высокопроизводительный сервис сбора и анализа файлов журналов сетевого и серверного оборудования в национальной исследовательской компьютерной сети

А.Г. Абрамов<sup>1,2</sup>✉<sup>1</sup> Межведомственный суперкомпьютерный центр РАН, Санкт-Петербургское отделение, г. Санкт-Петербург, 199034, Россия<sup>2</sup> Национальный исследовательский центр «Курчатовский институт», г. Москва, 123182, Россия

### Ссылка для цитирования

Абрамов А.Г. Высокопроизводительный сервис сбора и анализа файлов журналов сетевого и серверного оборудования в национальной исследовательской компьютерной сети // Программные продукты и системы. 2024. Т. 37. № 4. С. 495–503. doi: 10.15827/0236-235X.148.495-503

### Информация о статье

Группа специальностей ВАК: 2.3.5

Поступила в редакцию: 30.07.2024

После доработки: 10.09.2024

Принята к публикации: 16.09.2024

**Аннотация.** Построение бесперебойного и производительного решения для сбора, интеллектуальной обработки и анализа данных системных и сервисных журналов представляет собой содержательную и многоаспектную исследовательскую и прикладную задачу. Ее решение позволит обеспечить надежное функционирование научных телекоммуникационных сетей и сервисов на их основе. В настоящей статье приведен обзор разработанных и эксплуатируемых методик, технологий и инструментов работы с журналами с акцентом на программное обеспечение с открытым исходным кодом. Рассмотрены некоторые аспекты работы служб журналирования в операционных системах семейства Unix, основанных на протоколе syslog. Обозначены особенности построения и примеры типовых современных программных конвейеров обработки журналов и выполняемые ключевые функции, в том числе при задействовании методов и технологий машинного обучения. Приведено схематическое и детальное текстовое описание разработанного и внедренного в национальной исследовательской компьютерной сети России специализированного сервиса. Представленный сервис основан на наборе открытого программного обеспечения в составе широко применяемого в практике системного администрирования пакета сбора и обработки данных журналов Rsyslog, на высокопроизводительной колоночной СУБД ClickHouse и системе визуализации, статистики и аналитики данных Grafana.

**Ключевые слова:** сетевой мониторинг, системные и сервисные журналы, централизованный сбор и анализ журналов, национальная исследовательская компьютерная сеть, НИКС, свободно распространяемое программное обеспечение, Rsyslog, ClickHouse, Grafana

**Благодарности.** Публикация подготовлена в рамках госзадания СПбО МСЦ РАН и НИЦ «Курчатовский институт» по теме № FNEF-2024-0014 с использованием ресурсов Центра коллективного пользования научным оборудованием НИКС

**Введение.** В условиях массового распространения технологий и устройств Интернета вещей (*Internet of Things*, IoT), технологий виртуализации и облачных вычислений, искусственного интеллекта, киберфизических систем и иных интеллектуальных технологий концепции «Индустрия 4.0», мобильных сетей новых поколений информационно-коммуникационные системы стали включать в себя огромное количество распределенных компонентов, предназначенных для непрерывного предоставления потребителям разнообразных цифровых услуг и сервисов.

Постоянно повышается сложность, возрастает степень критичности и ответственности при мониторинге и управлении крупными цифровыми инфраструктурами. На решение этих задач нацелены наукоемкие методы, технологии и реализующие их коммерческие или базирующиеся на разработках с открытым исходным кодом программные инструменты [1, 2].

Развитые платформы могут реализовывать различные стратегии мониторинга и управления, которые способны органично дополнять друг друга и повышать общую эффективность и результативность процессов – методы на основе сетевых протоколов SNMP и ICMP, анализа системных и сервисных журналов, сетевой телеметрии (NetFlow, IPFIX и др.), а также на основе специальных подходов, предполагающих дублирование трафика, захват сетевых пакетов и выполнение глубокой инспекции трафика [3, 4].

В настоящей работе акцент сделан на методиках и программных решениях для централизованного сбора, обработки и анализа системных и сервисных журналов (логов), поступающих в платформы мониторинга от гетерогенного мультивендорного оборудования в составе наблюдаемых инфраструктур. Специалистам хорошо известно, что журналы обычно представляют собой текстовые файлы, которые со-

держат множество однострочных (в некоторых случаях – многострочных) записей, фиксирующих происходящие на оборудовании события аппаратного и программного уровней [5, 6]. Помимо файлов журналов, системы мониторинга могут агрегировать и представлять в своих пользовательских интерфейсах служебную информацию в виде SNMP-ловушек (трапов), MQTT-сообщений и т.п.

Крупные цифровые инфраструктуры способны генерировать десятки тысяч журнальных событий в минуту, причем далеко не все из них представляют ценность для целей мониторинга. Постоянно растущие объемы, слабая структурированность (или ее отсутствие), низкая плотность представления информации, широкий спектр форматов файлов журналов приводят к существенному повышению стоимости ресурсов долгосрочного хранения и к объективным трудностям при обработке и анализе данных.

Решению соответствующих задач посвящено достаточно большое число исследований и разработок, вместе с тем построение высокопроизводительного и надежного аппаратно-программного конвейера для автоматической работы с журналами в режиме реального времени по-прежнему является актуальной и важной для практики задачей.

Разработке методик, технологий и программных инструментов работы с большими данными, генерируемыми в научных телекоммуникационных сетях, включая информативные для целей аналитики данные о сетевых потоках и в большей степени служебное содержимое файлов журналов, уделяется существенное внимание при реализации проекта развития Национальной исследовательской компьютерной сети России (НИКС) [7, 8].

### **Обзор методов и инструментов сбора и анализа журналов**

**Основные методики сбора, обработки и анализа данных журналов.** Следует отметить, что какой-либо единый стандарт для формирования полей, составляющих запись о событиях, отсутствует, хотя попытки унификации формата журналов периодически предпринимаются, в том числе в отношении конкретного оборудования и сервисов. Несогласованность форматов записей создает сложности на пути выполнения типичных операций мониторинга на основе журнальных данных, например, классификация некоторого события как критичного

для инфраструктуры инцидента с оповещением ответственных служб, с генерацией проблемных билетов и при возможности с автоматическим принятием необходимых мер [9, 10].

Большинство доступных алгоритмов обработки и аналитики данных предполагают взаимодействие со структурированной информацией, кроме того, они неприменимы без определенных доработок и адаптаций к особенностям журналов. Наряду с большими объемами информации это обуславливает необходимость разработки специализированных алгоритмов, методик и программных пакетов, в том числе базирующихся на параллельной многопоточной обработке данных и машинном обучении.

К числу технических задач мониторинга, при решении которых могут оказаться полезными данные журналов, отнесем в частности

- выявление инцидентов на основе журнальных событий;
- обнаружение поведенческих аномалий, проблем с аппаратным и программным обеспечением и их производительностью;
- инициирование отправки оповещений о критических инцидентах и открытия проблемных билетов;
- анализ первопричин;
- диагностика и проактивное прогнозирование комплексных сбоев с целью предупреждения перерывов в предоставлении сервисов;
- анализ статистики использования сервисов [11–13].

Важным направлением является использование информации из журналов для обеспечения информационной безопасности, включая детектирование и предупреждение несанкционированных вторжений и действий пользователей, сетевых атак разных типов и целей и т.п. [14–16].

Поэтапный процесс обработки журнальных данных в рамках автоматизированного подхода обычно включает в себя централизованный сбор и хранение, предварительную обработку, анализ и представление (отчеты, визуализация) [17–19].

Сбор «сырых» журналов осуществляется на специально настроенных серверах-коллекторах, получающих данные от разнородных конечных устройств (сетевые маршрутизаторы и коммутаторы, серверы, системы IP-телефонии и видеонаблюдения и др.), на которых формируются события системного и прикладного уровней (операционные системы, среды виртуализации и контейнеризации, БД, сетевые сервисы и др.), а также из внешних систем сбора

журналов. Доставка данных в коллекторы может производиться в агентном и безагентном режимах с последующим сохранением в архивированном виде (в том числе для целей исполнения законодательства и технического учета), а также с обработкой и анализом в режиме реального времени.

Обработка преимущественно неструктурированной по своей природе первичной журнальной информации, ее автоматическое преобразование в структурированный унифицированный формат с компактным представлением и генерацией стандартных событий производятся в системах мониторинга с помощью сформированных правил и методов, которые выполняют операции по разбору (парсингу) журналов.

Основная идея парсинга заключается в классификации входных записей журнала на основе определенных процедур и синтаксиса известных событий, на их преобразовании в некий шаблон, что позволяет произвести семантическую интерпретацию содержимого и применить к данным доступные методы обнаружения. В условиях постоянного роста объемов журналов и применения большого числа разных форматов разработка парсеров вручную неэффективна и представляет собой задачу трудновыполнимую и требующую глубоких знаний в предметной области. В связи с этим передовые методы и алгоритмы генерации парсеров анализируют образцы данных журнала и автоматически создают шаблоны событий с выделением статических и динамических (переменных) частей. Наряду с рутинными регулярными выражениями и подготовленными вручную правилами при обработке и анализе журналов в последние годы стали широко применяться вычислительно эффективные алгоритмы, модели и библиотеки машинного обучения [20–22].

В процессе разбора и анализа журналов может производиться целый набор действий, таких как нормализация, фильтрация, сжатие, дубликация, классификация, извлечение признаков и проч.

К полезным функциям работы с журналами также можно отнести автоматическую отправку оповещений о детектированных инцидентах и эскалации, шаблоны для быстрой настройки интеграций с внешними системами, генерацию отчетов на основании задаваемых критериев, импорт/экспорт данных журналов, архивирование данных для длительного хранения и возможности восстановления, управление глубиной хранения.

### ***Доступные протоколы и инструменты для работы с журналами.***

Особенности структуры записей в журналах обуславливают использование протокола syslog как стандарта регистрации и отправки сообщений о происходящих событиях, используемого в Unix-подобных операционных системах и в большинстве сетевых устройств [23]. В соответствии со стандартом источники формируют текстовые сообщения о событиях и передают их на обработку локальному серверу syslog по протоколам UDP/TCP. Сообщения также могут передаваться на внешние syslog-серверы или в системы мониторинга, например, для целей централизованного сбора и обработки.

Типичная запись журнала может содержать отделяемые друг от друга символами-разделителями наборы полей в составе:

- временная метка, указывающая на момент наступления события;
- идентификатор и/или имя (IP-адрес) источника сообщения;
- идентификатор и/или имя сервиса/приложения (sshd, nginx, postfix, named, mysql и др.);
- уровень важности/критичности сообщения (alert, critical, error, warning, notice, info, debug и др.);
- идентификатор и/или имя категории источника записи (auth, cron, daemon, kern, mail, ntp, security, user и др.)
- идентификаторы и/или имена пользователей (при наличии);
- полезное содержание сообщения.

Место локального хранения журналов для разных типов устройств и платформ может отличаться, а в Unix-подобных системах для этих целей используется каталог файловой системы /var/log. Форматы хранения (текстовые или бинарные файлы, хранение в СУБД) и конкретные каталоги могут назначаться в каждом случае индивидуально в конфигурационных файлах службы журналирования. Специальная системная утилита logrotate ответственна за осуществление ротации и сжатия файлов журналов в соответствии с настроенными правилами.

Созданы десятки систем для работы с журнальной информацией, которые имеют разное качество и функциональные возможности, количество их растет, в первую очередь, в контексте решения задач информационной безопасности. Программные инструменты можно классифицировать по способу лицензирования, варианту доставки и по функциональным возможностям [24, 25]. Ключевыми критери-

ями качества систем являются аккуратность обработки журналов и производительность по количеству обрабатываемых событий в секунду.

Среди набора свободно распространяемых решений с открытым исходным кодом отметим пакеты Logstash, Fluentd, Graylog, Syslog-ng, NXlog. На машинном обучении базируются открытые методы и инструменты LogPAI, Logsy, LogAnomaly, DeepLog [26], DeepSyslog [27], DLLog [28].

Примерами коммерческих систем, традиционно отличающихся широким функционалом и развитыми пользовательскими интерфейсами, являются SolarWinds Loggly, ManageEngine EventLog Analyzer, Splunk Log Observer, Nagios Log Server; DataDog Log Management, Logic Monitor, Logz.io, Sematext Logs, Sumo Logic, Fluent Bit, FileBit.

Следует отметить открытые наборы датасетов с информацией из журналов, собранной из разных систем и программных продуктов, такие как Loghub, CFDR, OpenStack Fault Injection Dataset, SecRepo. Они могут служить полезным источником для машинного обучения моделей и их внедрения в практику мониторинга.

### Описание разработанного сервиса и его внедрение в НИКС

**Общие сведения о разработанном решении.** До недавнего времени в сети НИКС файлы журналов собирались на оборудовании и хранились разрозненно, форматы далеко не всегда были согласованы, что создавало серьезные трудности для администраторов сервисов и сетевых инженеров. В целях преодоления этой проблемы был проработан проект создания специализированного сервиса для централизованного сбора и анализа файлов журналов и реализован освещаемый в настоящей статье первый этап проекта.

На этом этапе разработаны методики и первая версия инструментария на основе обоснованно выбранного и протестированного стека ПО с открытым исходным кодом. В качестве целей создания сервиса были определены:

- реализация возможностей расширенного многоаспектного статистического и аналитического учета функционирования компонентов инфраструктуры НИКС и предоставляемых на их основе сервисов;
- централизованная агрегация и представление соответствующей оперативной и исторической информации с учетом категорий пользователей;

- предоставление возможностей качественной визуализации информации;

- повышение эффективности и результативности функционирования службы технической поддержки.

В сервисе реализуются следующие основные функции:

- авторизация и управление доступом пользователей к функциональным возможностям в соответствии с ролевым профилем;

- получение от сетевого и серверного оборудования информации о функционировании и использовании компонентов инфраструктуры и сервисов НИКС и ее централизованная статистическая обработка;

- отображение информации в различных видах с широкими возможностями настройки, формирования наглядных информационных панелей;

- интеграционное взаимодействие и обмен данными с внешними системами;

Сервис предоставил возможности для содержательного анализа статистики использования инфраструктуры и сервисов НИКС на базе накапливаемой информации журналов, оперативного мониторинга функционирования сетевого оборудования и совершаемых операций, а также для комплексной оценки востребованности отдельных организаций-пользователей и уровня вовлеченности в использование сервисных решений.

Аппаратно-программный комплекс сервиса развернут на нескольких выделенных виртуальных вычислительных серверах в собственной высокопроизводительной инфраструктуре МСЦ РАН (г. Москва). Инфраструктура обеспечения облачной платформы реализована на основе популярного решения виртуализации Proxmox VE, в качестве системного ПО используется операционная система семейства GNU/Linux.

Прикладное ПО сервисов (рис. 1) базируется на взаимодействующих свободно распространяемых решениях с открытым исходным кодом – системе сбора и обработки данных журналов rsyslog (<https://www.rsyslog.com>) и платформе визуализации, статистики и аналитики данных Grafana (<https://grafana.com>). Кроме того, задействуются аппаратно-программные ресурсы высокопроизводительной колоночной СУБД ClickHouse (<https://clickhouse.com>), которая используется в работе других высоконагруженных сервисов НИКС, в частности, сервиса статистики и аналитики данных о сетевых потоках [29, 30].

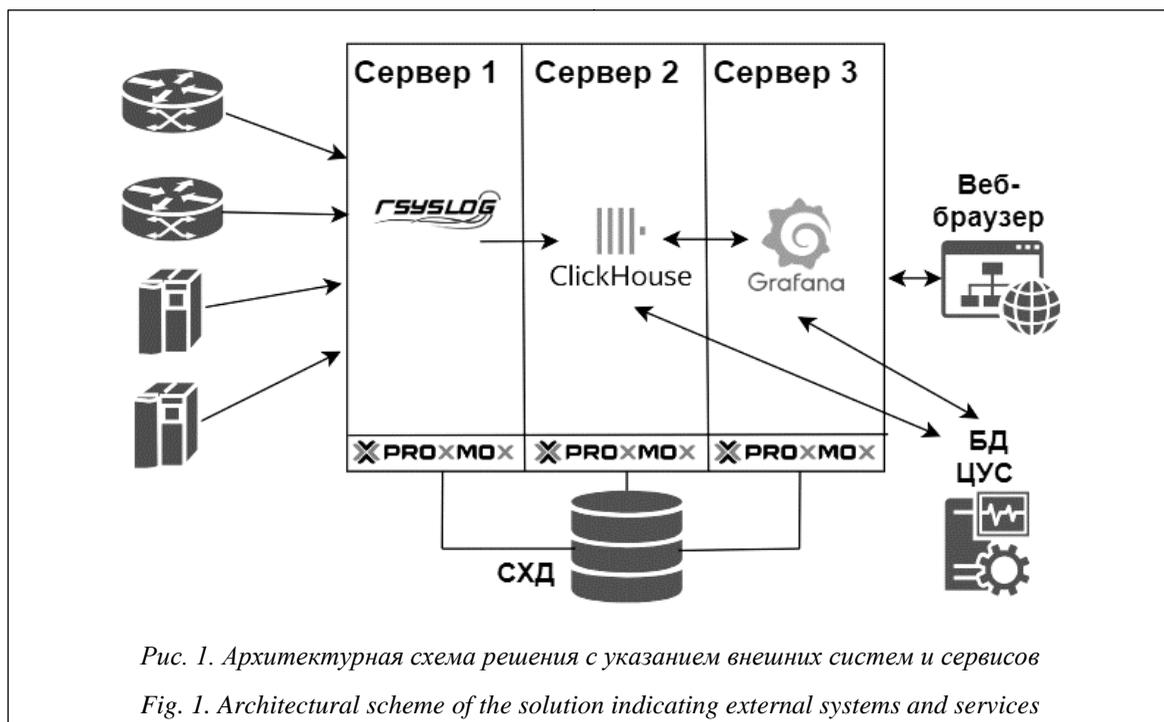


Рис. 1. Архитектурная схема решения с указанием внешних систем и сервисов

Fig. 1. Architectural scheme of the solution indicating external systems and services

Журналы пересылаются с сетевого и серверного оборудования по протоколу UDP на два виртуальных сервера с развернутым и настроенным решением rsyslog. Этот пакет имеет модульную архитектуру, расширение функционала и интеграция с внешними системами осуществляются путем установки дополнительных модулей. В данном случае на серверах используются модули препроцессинга rsyslog-omclickhouse (передача журналов в СУБД ClickHouse) и rsyslog-mmmnormalize (парсинг логов для целей статистики). Типичная интенсивность поступления журналов составляет несколько сотен записей в секунду.

На сервере с платформой Grafana установлен и настроен специальный плагин источника данных, обеспечивающий поддержку СУБД ClickHouse. Сервисы обеспечивают управление содержимым с помощью встроенных элементов веб-интерфейса Grafana и позволяют выполнять в нем SQL-запросы к СУБД ClickHouse для отбора требующих отображения данных и уточнения сформированных ранее и сохраненных запросов.

К преимуществам развернутого решения и его компонентов, помимо общей высокой производительности конвейера обработки, можно отнести доступный для использования совместно с rsyslog хорошо подходящий для обработки событий и настройки процесса обработки скриптовый язык RainerScript (<https://www.rsyslog.com/doc/rainerscript/>). Кроме того,

достоинствами являются инвертированные индексы СУБД ClickHouse, позволяющие эффективно производить полнотекстовый поиск по журналам, настраиваемые визуализации в пакете Grafana с разными типами графиков и диаграмм, а также потенциальная возможность формирования и отправки оповещений о выявленных в процессе анализа журналов и классифицированных инцидентах в наблюдаемой инфраструктуре.

### Визуализация журналов на информационных табло

Пример настроенного информационного табло (дашборда) с представлением журнальной информации, статистических данных и с элементами управления приведен на рисунке 2.

Два однотипных табло содержат сведения о событиях, сгруппированных по их источникам как серверные и сетевые. На табло размещено табличное представление записей с их сортировкой по временной метке. Каждую запись можно раскрыть для просмотра статических и динамических полей. В результате предварительной обработки и парсинга среди статических полей выделяются

- host (имя источника записи);
- facility (имя категории источника записи);
- priority (уровень важности, приоритет);
- service (имя сервиса/приложения, записанное в файл журнала).



Рис. 2. Общий вид информационной панели с отображением и визуализацией информации из журналов

Fig. 2. Dashboard with log data display

Помимо таблицы, на табло доступны для просмотра и аналитики статистические данные в форме временного ряда с агрегацией событий по уровню важности за назначенный период времени, кольцевая диаграмма с соответствующими интегрированными данными, а также топ-списки с обобщением в разрезе источников (хостов) и сервисов.

Элементы управления позволяют выполнять полнотекстовый поиск информации с возможностью уточнений вида включить/исключить из результатов, а также независимый множественный выбор из выпадающих списков по перечисленным выше статическим полям. Доступны возможности назначения диапазона времени для отображения информации, уточнения настроенных SQL-запросов, экспорта данных в формате CSV и ряд других, стандартно предоставляемых в интерфейсе Grafana.

Дополнительно сформировано специальное табло для работы с событиями информационной безопасности, в которое выводятся отображенные из общих журналов записи, рассматриваемые в качестве потенциальных кандидатов на инциденты безопасности, а также реализована интеграция с системой класса SIEM (*Security Information and Event Management*) в части экспорта журнальных данных.

В отношении некоторых сервисов НИКС организован альтернативный сбор и обработка данных журналов и созданы отдельные информационные панели, отражающие специфику их функционирования и использования [31], в том числе обслуживающие разные сервисы веб-сер-

веры, службы доменных имен (DNS) и точного времени (NTP), сервис тестирования пропускной способности сети iPerf, сервис роуминга в Wi-Fi-сетях для научно-образовательного сообщества Eduroam, сервис вебинаров BigBlueButton.

### Заключение

Результатом первого этапа выполнения представленного научно-технического проекта стали разработка и внедрение в НИКС высокопроизводительного сервиса, успешно решающего задачи полного цикла работы с журнальной информацией, которая накапливается на эксплуатируемом сетевом и серверном оборудовании. Сервис, построенный на компонентах свободно распространяемого ПО с открытым исходным кодом, реализует централизованный сбор, многопараметрическую обработку перенаправляемых от устройств журналов, а также элементы аналитики, визуализации и отчетности.

Реализация проекта предоставила сетевым инженерам и администраторам сервисов современные возможности по работе с журналами в едином интерфейсе с удобным для мониторинга и контроля отображением информации о функционировании и возникающих проблемах с оборудованием, системным и прикладным ПО. Доступная статистика обращений к сервисам является ценным источником сведений об их использовании, дает возможность обоснованно принимать решения о развитии сервисного направления НИКС и его отдельных компонентов.

На следующих этапах проекта предполагается интегрировать сервис с сайтом центра управления сетью НИКС (*Network Operations Center, NOC*) для отображения выявленных инцидентов на панели службы технической поддержки, а также реализовать автоматические оповещения об инцидентах средствами электронной почты, иных доступных механизмов и расширить возможности аналитики и визуализации.

В качестве одного из перспективных направлений развития рассматривается привлечение для автоматического анализа журналов методов машинного обучения. Накапливаемые наборы данных позволяют адаптировать и использовать их для обучения моделей машинного обучения с помощью разных подходов и методик в целях интеллектуализации анализа и проактивного предсказания аномалий в сетевой инфраструктуре.

### Список литературы

1. Julian M. *Practical Monitoring: Effective Strategies for the Real World*. California, Sebastopol, O'Reilly Media Publ., 2017, 229 p.
2. Blokdyk G. *Network Performance Monitoring and Diagnostics Tools*. Toronto, 5STARCOoks Publ., 2022, 324 p.
3. D'Alconzo A., Drago I., Morichetta A., Melli M., Casas P. A survey on big data for network traffic monitoring and analysis. *IEEE TNSM*, 2019, vol. 16, no. 3, pp. 800–813. doi: 10.1109/TNSM.2019.2933358.
4. Wilkins P. *Logging in Action: with Fluentd, Kubernetes and more*. Manning Publ., 2022, 392 p.
5. Абрамов А.Г., Гончар А.А., Евсеев А.В., Шабанов Б.М. Национальная исследовательская компьютерная сеть нового поколения: текущее состояние и концепция развития // *Информационные технологии*. 2021. Т. 27. № 3. С. 115–124. doi: 10.17587/it.27.115-124.
6. Абрамов А.Г., Гончар А.А., Евсеев А.В., Шабанов Б.М. Основные результаты первых этапов проекта развития национальной исследовательской компьютерной сети // *Информационные технологии и вычислительные системы*. 2024. № 1. С. 3–10. doi: 10.14357/20718632240101.
7. Kubacki M., Sosnowski J. Holistic processing and exploring event logs. In: *LNPSE. Proc. SERENE*, 2017, vol. 10479, pp. 184–200. doi: 10.1007/978-3-319-65948-0\_12.
8. He S., He P., Chen Z. et al. A survey on automated log analysis for reliability engineering. *ACM CSUR*, 2021, vol. 54, no. 6, art. 130. doi: 10.1145/3460345.
9. Lin H., Yan Z., Chen Y., Zhang L. A survey on network security-related data collection technologies. *IEEE Access*, 2018, vol. 6, pp. 18345–18365. doi: 10.1109/ACCESS.2018.2817921.
10. Драчев Г.А. Разработка алгоритма выделения и кодирования данных из журнальных сообщений вычислительной системы для систем обнаружения аномалий // *Информационные технологии*. 2023. Т. 29. № 7. С. 351–359. doi: 10.17587/it.29.351-359.
11. Zhang T., Qiu H., Castellano G. et al. System log parsing: a survey. *IEEE Transactions on Knowledge & Data Engineering*, 2023, vol. 35, no. 8, pp. 8596–8614. doi: 10.1109/TKDE.2022.3222417.
12. Ma J., Liu Y., Wan H., Sun G. Automatic parsing and utilization of system log features in log analysis: a survey. *Appl. Sci.*, 2023, vol. 13, no. 8, art. 4930. doi: 10.3390/app13084930.
13. Skopik F., Landauer M., Wurzenberger M. Online log data analysis with efficient machine learning: a review. *IEEE Security & Privacy*, 2022, vol. 20, no. 3, pp. 80–90. doi: 10.1109/MSEC.2021.3113275.
14. Худяков Д.А. Разработка системы выявления аномалий на основе распределенной трассировки логов // *Вестн. НГУ. Сер.: Информационные технологии*. 2023. Т. 21. № 1. С. 62–72. doi: 10.25205/1818-7900-2023-21-1-62-72.
15. Chen B., Jiang Z.M. A survey of software log instrumentation. *ACM CSUR*, 2021, vol. 54, no. 4, art. 90. doi: 10.1145/3448976.
16. Du M., Li F., Zheng G., Srikumar V. DeepLog: anomaly detection and diagnosis from system logs through deep learning. *Proc. ACM SIGSAC Conf. CCS*, 2017, pp. 1285–1298. doi: 10.1145/3133956.3134015.
17. Zhou J., Qian Y., Zou Q., Liu P., Xiang J. DeepSyslog: deep anomaly detection on syslog using sentence embedding and metadata. *IEEE Transactions on Information Forensics and Security*, 2022, vol. 17, pp. 3051–3061. doi: 10.1109/TIFS.2022.3201379.
18. Cheng H., Ying Sh., Duan X., Yuan W. DLLog: an online log parsing approach for large-scale system. *Int. J. of Intelligent Systems*, 2024, vol. 2024, art. 5961993. doi: 10.1155/2024/5961993.
19. Abramov A.G. Collection, analysis and interactive visualization of NetFlow data: Experience with big data on the base of the National Research Computer Network of Russia. *Lobachevskii J. Math.*, 2020, vol. 41, pp. 2525–2534. doi: 10.1134/S1995080220120021.
20. Abramov A.G., Porkhachev V.A., Yastrebov Yu.V. Methods and high-performance tools for collecting, analysis and visualization of data exchange with a focus on research and education networks. *Lobachevskii J. Math.*, 2023, vol. 44, pp. 4930–4938. doi: 10.1134/S1995080223110021.
21. Abramov A.G. Service portfolios of leading National Research and Education Networks and implementation on the basis of the National Research Computer Network of Russia. *Lobachevskii J. Math.*, 2021, vol. 42, pp. 2481–2492. doi: 10.1134/S1995080221110032.

## High-performance service for collecting and analyzing network and server hardware log files on a National Research Computer Network

Aleksy G. Abramov <sup>1,2</sup>✉<sup>1</sup> Joint Supercomputer Center of RAS – St. Petersburg Branch, St. Petersburg, 199034, Russian Federation<sup>2</sup> National Research Centre “Kurchatov Institute”, Moscow, 123182, Russian Federation

### For citation

Abramov, A.G. (2024) ‘High-performance service for collecting and analyzing network and server hardware log files on a National Research Computer Network’, *Software & Systems*, 37(4), pp. 495–503 (in Russ.). doi: 10.15827/0236-235X.148.495-503

### Article info

Received: 30.07.2024

After revision: 10.09.2024

Accepted: 16.09.2024

**Abstract.** Building a seamless and productive solution for collecting, intelligent processing and analyzing system and service log data is a meaningful and multidimensional research and application problem. This solution will ensure reliable functioning of scientific telecommunication networks and services based on them. This paper provides an overview of developed and operated journaling techniques, technologies and tools with an emphasis on open source software. It also considers some aspects of logging services in Unix operating systems based on the Rsyslog protocol. There are also outlined construction features and examples of typical modern software log processing pipelines and their key functions, including those using machine learning methods and technologies. The author gives a schematic and detailed textual description of a special-purpose service developed and implemented in the national research computer network of Russia. This service is based on a set of open source software comprising the widely used in system administration practice of Rsyslog log data collection and processing package, high-performance columnar ClickHouse DBMS and Grafana data visualization, statistics and analytics system.

**Keywords:** network monitoring, system and service log data, centralized log collection and analysis, National Research Computer Network, NIKS, free software, Rsyslog, ClickHouse, Grafana

**Acknowledgements.** The publication has been prepared within the framework of the state assignment of the Joint Supercomputer Center of RAS and the National Research Center “Kurchatov Institute” on topic no. FNEF-2024-0014 using the resources of the Center for collective use of scientific equipment “National Research Computer Network”

### References

1. Julian, M. (2017) *Practical Monitoring: Effective Strategies for the Real World*. California, Sebastopol: O'Reilly Media Publ., 229 p.
2. Blokdyk, G. (2022) *Network Performance Monitoring and Diagnostics Tools*. Toronto: 5STARCOOKS Publ., 324 p.
3. D'Alconzo, A., Drago, I., Morichetta, A., Melli, M., Casas, P. (2019) ‘A survey on big data for network traffic monitoring and analysis’, *IEEE TNSM*, 16(3), pp. 800–813. doi: 10.1109/TNSM.2019.2933358.
4. Wilkins, P. (2022) *Logging in Action: with Fluentd, Kubernetes and more*. Manning Publ., 392 p.
5. Abramov, A.G., Gonchar, A.A., Evseev, A.V., Shabanov, B.M. (2021) ‘The new generation national research computer network: Current status and concept for the development’, *Inform. Tech.*, 27(3), pp. 115–124 (in Russ.). doi: 10.17587/it.27.115-124.
6. Abramov, A.G., Gonchar, A.A., Evseev, A.V., Shabanov, B.M. (2024) ‘Main results of the first stages of the project for the development of national research computer network’, *J. of Inform. Tech. and Computing Systems*, (1), pp. 3–10 (in Russ.). doi: 10.14357/20718632240101.
7. Kubacki, M., Sosnowski, J. (2017) ‘Holistic processing and exploring event logs’, in *LNPSE. Proc. SERENE*, 10479, pp. 184–200. doi: 10.1007/978-3-319-65948-0\_12.
8. He, S., He, P., Chen, Z. et al. (2021) ‘A survey on automated log analysis for reliability engineering’, *ACM CSUR*, 54(6), art. 130. doi: 10.1145/3460345.
9. Lin, H., Yan, Z., Chen, Y., Zhang, L. (2018) ‘A survey on network security-related data collection technologies’, *IEEE Access*, 6, pp. 18345–18365. doi: 10.1109/ACCESS.2018.2817921.
10. Drachev, G.A. (2023) ‘Development of an algorithm for extracting and encoding data from log messages of a computing system for anomaly detection systems’, *Inform. Tech.*, 29(7), pp. 351–359 (in Russ.). doi: 10.17587/it.29.351-359.
11. Zhang, T., Qiu, H., Castellano, G. et al. (2023) ‘System log parsing: a survey’, *IEEE Transactions on Knowledge & Data Engineering*, 35(8), pp. 8596–8614. doi: 10.1109/TKDE.2022.3222417.
12. Ma, J., Liu, Y., Wan, H., Sun, G. (2023) ‘Automatic parsing and utilization of system log features in log analysis: a survey’, *Appl. Sci.*, 13(8), art. 4930. doi: 10.3390/app13084930.
13. Skopik, F., Landauer, M., Wurzenberger, M. (2022) ‘Online log data analysis with efficient machine learning: a review’, *IEEE Security & Privacy*, 20(3), pp. 80–90. doi: 10.1109/MSEC.2021.3113275.
14. Khudyakov, D.A. (2023) ‘Development of anomaly detection system based on distributed log tracing’, *Vestn. NSU. Ser.: Inform. Tech.*, 21(1), pp. 62–72 (in Russ.). doi: 10.25205/1818-7900-2023-21-1-62-72.
15. Chen, B., Jiang, Z.M. (2021) ‘A survey of software log instrumentation’, *ACM CSUR*, 54(4), art. 90. doi: 10.1145/3448976.

16. Du, M., Li, F., Zheng, G., Srikumar, V. (2017) 'DeepLog: anomaly detection and diagnosis from system logs through deep learning', *Proc. ACM SIGSAC Conf. CCS*, pp. 1285–1298. doi: 10.1145/3133956.3134015.
17. Zhou, J., Qian, Y., Zou, Q., Liu, P., Xiang, J. (2022) 'DeepSyslog: deep anomaly detection on syslog using sentence embedding and metadata', *IEEE Transactions on Information Forensics and Security*, 17, pp. 3051–3061. doi: 10.1109/TIFS.2022.3201379.
18. Cheng, H., Ying, Sh., Duan, X., Yuan, W. (2024) 'DLLog: an online log parsing approach for large-scale system', *Int. J. of Intelligent Systems*, 2024, art. 5961993. doi: 10.1155/2024/5961993.
19. Abramov, A.G. (2020) 'Collection, analysis and interactive visualization of NetFlow data: Experience with big data on the base of the National Research Computer Network of Russia', *Lobachevskii J. Math.*, 41, pp. 2525–2534. doi: 10.1134/S1995080220120021.
20. Abramov, A.G., Porkhachev, V.A., Yastrebov, Yu.V. (2023) 'Methods and high-performance tools for collecting, analysis and visualization of data exchange with a focus on research and education networks', *Lobachevskii J. Math.*, 44, pp. 4930–4938. doi: 10.1134/S1995080223110021.
21. Abramov, A.G. (2021) 'Service portfolios of leading National Research and Education Networks and implementation on the basis of the National Research Computer Network of Russia', *Lobachevskii J. Math.*, 42, pp. 2481–2492. doi: 10.1134/S1995080221110032.

**Авторы**

**Абрамов Алексей Геннадьевич**<sup>1,2</sup>,  
к.ф.-м.н., доцент,  
ведущий научный сотрудник,  
abramov@niks.su

<sup>1</sup> Межведомственный суперкомпьютерный  
центр РАН, Санкт-Петербургское отделение,  
г. Санкт-Петербург, 199034, Россия

<sup>2</sup> Национальный исследовательский центр  
«Курчатовский институт», г. Москва, 123182, Россия

**Authors**

**Aleksey G. Abramov**<sup>1,2</sup>,  
Cand. of Sci. (Physics and Mathematics),  
Associate Professor, Leading Researcher,  
abramov@niks.su

<sup>1</sup> Joint Supercomputer Center of RAS –  
St. Petersburg Branch,  
St. Petersburg, 199034, Russian Federation  
<sup>2</sup> National Research Centre “Kurchatov Institute”,  
Moscow, 123182, Russian Federation

## Повышение вычислительной мощности персонального компьютера за счет интеграции с распределенной системой из смартфонов

С.А. Балабаев<sup>1</sup>, С.А. Лупин<sup>1</sup>, П.Н. Телегин<sup>2,3</sup>✉, Б.М. Шабанов<sup>2,3</sup>

<sup>1</sup> Национальный исследовательский университет «МИЭТ», г. Москва, 124498, Россия

<sup>2</sup> Межведомственный суперкомпьютерный центр РАН, г. Москва, 119334, Россия

<sup>3</sup> Национальный исследовательский центр «Курчатовский институт», г. Москва, 123182, Россия

### Ссылка для цитирования

Балабаев С.А., Лупин С.А., Телегин П.Н., Шабанов Б.М. Повышение вычислительной мощности персонального компьютера за счет интеграции с распределенной системой из смартфонов // Программные продукты и системы. 2024. Т. 37. № 4. С. 504–513. doi: 10.15827/0236-235X.148.504-513

### Информация о статье

Группа специальностей ВАК: 2.3.1

Поступила в редакцию: 25.07.2024

После доработки: 27.08.2024

Принята к публикации: 30.08.2024

**Аннотация.** В статье рассмотрена возможность повышения производительности персонального компьютера за счет интеграции с ним распределенной системы из смартфонов. Приведен обзор основных аппаратных и программных особенностей мобильных устройств, проведена оценка влияния подобных узлов на производительность распределенной вычислительной системы. Было выявлено, что их основными отличиями от персональных компьютеров являются низкое качество охлаждения устройств и архитектура big.LITTLE. Эти особенности мобильных устройств не позволяют задействовать все ядра смартфона на полную мощность и должны быть учтены при их интеграции в единую вычислительную среду. Программное обеспечение для интеграции мобильных устройств с персональными компьютерами состоит из двух приложений – клиента и сервера, разработанных на языке программирования Java. Приложения позволяют загружать вычисляемую задачу на узлы, запускать ее, аккумулировать и отображать полученные с узлов результаты вычислений. Взаимодействие между устройствами происходит по сети. Вычисляемая задача представляет собой арк-приложение, содержащее Java-класс, методы которого могут быть вызваны из приложения, запущенного на клиенте. Используемый алгоритм балансировки нагрузки узлов позволяет интегрировать в единую среду смартфоны разных поколений, значительно отличающиеся по производительности. Показано, что для эффективного распределения нагрузки между узлами системы необходимо использовать значение их реальной производительности. В работе приводятся результаты решения задачи минимизации функции в распределенной среде, организованной при помощи разработанного программного обеспечения. Они подтверждают достижение поставленной цели. Полученные результаты могут быть полезны широкому кругу специалистов.

**Ключевые слова:** вычислительная мощность, распределенные вычисления, смартфон, операционная система Android, грид-системы, балансировка нагрузки узлов

**Благодарности.** Работа выполнена в МСЦ РАН и НИЦ «Курчатовский институт» по теме FNEF-2024-0016

**Введение.** Во второй половине XX века началось бурное развитие микроэлектронной промышленности. Размеры электронных устройств непрерывно уменьшаются, при этом их функциональность столь же неуклонно возрастает. Современные *персональные компьютеры* (ПК) по своей вычислительной мощности многократно превосходят огромные мейнфреймы недалекого прошлого. Благодаря прогрессу в электронике были получены первые фотографические снимки черной дыры, разработаны эффективные лекарства за доступную цену, сделаны многие открытия.

Однако с ростом мощности вычислительных устройств растет и потребность людей в высокопроизводительных вычислениях. Доминирующим способом ускорения работы компьютеров стало использование нескольких вычислительных устройств для решения одной задачи. Такой подход реализован и в рабочих станциях, и в суперкомпьютерах.

Наряду с мощными многоядерными процессорами в вычислениях могут участвовать и более слабые узлы, например, планшеты, смартфоны, телевизионные приставки. Объединив их в сеть, можно сформировать распределенную среду, способную решать ресурсоемкие задачи.

В работе описано специализированное ПО, позволяющее интегрировать в единую среду смартфоны. По характеристикам современные мобильные устройства лишь немногим уступают ПК и позволяют выполнять сложные вычисления [1]. На их основе можно построить, например, распределенную систему HGRID (*Home GRID*) [2].

Существуют похожие решения групп исследователей из разных стран. В работе [3] предлагается объединять мобильные устройства с помощью программной платформы BOINC. В качестве примера применения такой системы рассмотрено ее использование для поиска ла-

тинских квадратов. На особенности разработанной системы влияют ограничения, накладываемые платформой BOINC, – необходимость разработки программного кода на языках C, C++, использование достаточно грубого алгоритма балансировки нагрузки.

Нидерландские ученые разработали систему IBIS, позволяющую объединять различные устройства для распределенных вычислений [4]. Проект написан на языке Java и предоставляет возможность работы с устройствами Android. Авторы предлагают эффективный способ коммуникации устройств между собой с помощью разработанной библиотеки и поддержку MPI. Однако проект был заморожен и доступа к нему нет.

В большинстве случаев современные исследователи предлагают использовать распределенную систему из Android-смартфонов для решения частной задачи – федеративного обучения [5, 6]. Хотя подобные кластеры и показывают свою эффективность, решать на них другие задачи без изменения программного кода невозможно.

### Особенности разработки приложений для мобильных устройств

Разработка приложений для мобильных устройств является одним из ведущих IT-направлений. Число владельцев мобильных устройств, формирующих устойчивый спрос на новое ПО, увеличивается с каждым годом. Высокая производительность и широкая распространенность смартфонов позволяют без дополнительных затрат создать на их основе распределенную вычислительную систему, сопоставимую по мощности с ПК. Для оценки вычислительных возможностей смартфонов опишем особенности этих устройств.

#### Аппаратные характеристики

Современные мобильные телефоны, например, Нопог 8X (JSN-L22), обладают достаточными ресурсами для проведения высокопроизводительных вычислений:

- аппаратная платформа – HiSilicon Kirin710;
- число ядер – 8;
- наличие архитектуры big.LITTLE;
- архитектура процессора – 4x ARM Cortex-A73 @ 2.19 GHz;
- объем оперативной памяти – 3 680 Мб.

Представленные параметры показывают, что смартфон можно рассматривать в качестве полноценной вычислительной платформы. Наличие восьми ядер позволяет производить парал-

лельные вычисления, а высокая тактовая частота дает возможность выполнять их на сопоставимых с ПК скоростях.

#### Программные характеристики мобильных устройств

Семейство операционных систем, используемых в смартфонах, достаточно развито, среди них Android, iOS, Flume OS, Аврора. Наиболее популярной является система Android. Число мобильных устройств с Android превысило 3 млрд., в то время как устройств, выпущенных их основным конкурентом – компанией Apple, немногим более 1 млрд. штук.

В основе операционной системы Android лежит ядро Linux (рис. 1) [7]. Ядро взаимодействует с уровнем HAL (*Hardware Abstraction Layer* – аппаратный уровень абстракции), обеспечивающим связь между драйверами устройств и библиотеками. В пространство HAL входят несколько библиотечных модулей, каждый из которых реализует интерфейс для определенного аппаратного компонента.

Разработка ПО под архитектуру Android ведется на языках программирования Kotlin и Java. Для запуска программ, написанных на этих языках, требуется виртуальная машина. В Android используется виртуальная машина *Android Runtime* (ART). В версиях Android ниже 5.0 Lollipop использовалась Dalvik.

Отдельный уровень программной архитектуры – *Native C/C++ Libraries* – включает в себя набор библиотек, написанных на языке C или C++ и используемых различными компонентами ОС. При разработке ресурсоемких приложений со сложными вычислениями можно использовать нативный код для достижения большей производительности.

Для доступа к основным функциям операционной системы Android используется уровень *Java API Framework*, содержащий в себе набор приложений и библиотек для разработчиков [8].

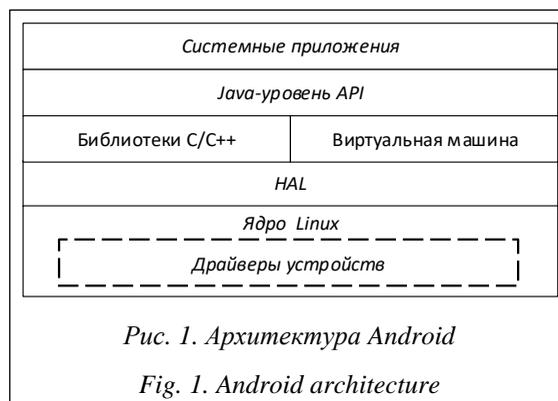


Рис. 1. Архитектура Android

Fig. 1. Android architecture

iOS – мобильная операционная система для устройств компании Apple. Для установки приложения на устройство с iOS требуются сертификаты, подписанные Apple, что делает разработку для этой платформы затруднительной.

Flume OS – операционная система на основе Android, поддерживаемая устройствами компании Meizu. Разработка приложений под нее аналогична разработке под Android.

Аврора – отечественная операционная система, разработанная компанией «Открытая мобильная платформа». В настоящее время она доступна только для ограниченного ряда устройств. Сегодня система менее перспективна для проведения исследований, но с учетом современных реалий ситуация может измениться [9].

Таким образом, можно сделать вывод, что наибольший интерес для использования в качестве узлов распределенной системы представляют мобильные устройства на основе операционной системы Android.

### **Использование мобильных устройств как вычислительных узлов**

Рассмотрим особенности мобильных платформ, которые необходимо учитывать при реализации на них вычислительных задач.

*Зависимость производительности процессора от температуры*

В отличие от ПК конструкция смартфонов не предусматривает принудительного охлаждения процессора. При проведении интенсивных вычислений, особенно в тех случаях, когда задействованы все ядра, температура процессора мобильного телефона сильно повышается. Например, в случае проведения расчетов на смартфоне ASUS температура повышается до 45 градусов [1]. При достижении критических для устройства значений температуры операционная система снижает нагрузку на процессор, уменьшая тактовую частоту ядер устройства, вследствие чего производительность системы падает. Использование внешнего охлаждения корпуса смартфона с помощью кулеров дает лишь незначительный эффект. Как следствие, при оценке вычислительной мощности мобильного узла невозможно ориентироваться на максимальное значение рабочей частоты процессора.

*Зависимость производительности процессора от заряда аккумулятора*

При низком уровне заряда аккумулятора, как и при высокой температуре, смартфоны снижают производительность процессора. Для

устранения этого эффекта при проведении вычислений в распределенной среде все его мобильные узлы должны быть подключены к сетевому питанию.

*Использование архитектуры big.LITTLE*

Большинство современных моделей смартфонов используют big.LITTLE. В основе архитектуры лежит совмещение относительно медленных энергосберегающих ядер с более мощными энергоемкими. В большинстве инсталляций число ядер процессора поровну делится между этими двумя типами. Фоновые задачи распределяются по LITTLE-ядрам, а ресурсоемкие выполняются на big-ядрах. Таким образом, при запуске интенсивных вычислений при высокой нагрузке могут быть задействованы и нагружены только big-ядра, на LITTLE-ядрах вычисления производиться не будут. Это снижает максимальную производительность мобильного узла почти в два раза по отношению к пиковому значению.

*Пропускная способность сети (модуль Wi-Fi)*

При реализации распределенной вычислительной системы необходимо обеспечить взаимодействие ее узлов между собой. Для этого можно использовать беспроводную локальную сеть. В современных смартфонах установлен Wi-Fi-модуль, работающий на частотах 2,4 или 5 МГц. Однако скорость передачи сообщений между устройствами достаточно низкая. Проведенный эксперимент по измерению скорости передачи данных между ПК и мобильным устройством Honor 8X с использованием точки доступа показал, что средняя скорость составляет 13.3 Мбит/с. Этого достаточно для распределенных вычислений, но не позволяет реализовывать параллельные программы с интенсивными межузловыми коммуникациями [10].

Если узлы расположены удаленно друг от друга, то единственным способом связи между устройствами будет Интернет. Для проверки качества связи были измерены входящий и исходящий трафики смартфона и ПК, находящихся в одной сети. Измерения проводились на расстоянии 1 метра от точки доступа при работе контроллера на различных частотах – 2,4 ГГц и 5 ГГц. На частоте 2.4 ГГц скорость оказалась значительно ниже, что связано с зашумленностью эфира (табл. 1).

Из таблицы видно, что скорость связи с сетью Интернет смартфона и ПК находится на одном уровне.

*Промежуточные выводы*

Приведенные особенности показывают, что смартфоны, как вычислительные узлы, значи-

тельно отличаются от ПК. Однако широкая распространенность и небольшие габариты позволяют использовать их в качестве узлов распределенной системы. Мощность нескольких смартфонов будет уже сопоставима с мощностью одного компьютера. При объединении мобильных устройств в распределенную систему важно учитывать разницу в характеристиках каждого из узлов и с помощью балансировки оптимально распределять вычислительную нагрузку между ними.

Таблица 1

**Результаты эксперимента по проверке скорости Интернета**

Table 1

**Experimental results of the Internet speed test**

Частота, ГГц	Трафик, Мбит/сек.			
	исходящий		входящий	
	смарт-фон	ПК	смарт-фон	ПК
2,4	117.05	55.83	35.10	37.86
5	699.52	300.67	198.68	199.92

Программные средства для организации распределенной системы из мобильных устройств

ПО, необходимое для интеграции смартфонов с ПК, должно обеспечивать

- объединение всех узлов в локальную сеть;
- передачу вычислительной задачи с иницилирующего узла исполнителям;
- сбор полученных узлами результатов;
- балансировку нагрузки узлов в соответствии с их характеристиками.

Исследуемая распределенная система представлена на рисунке 2. Разработанное ПО состоит из двух приложений: для смартфонов и для ПК. Оба приложения разработаны на языке программирования Java.

Иницилирующим узлом является ПК. С него вычислительная задача рассылается на исполняющие узлы и по команде запускается на каждом из них. В качестве среды для передачи данных используется беспроводная локальная сеть. Разработанное ПО может поддерживать вычисления и на узлах глобальной сети при условии соответствующей настройки маршрутизатора. После выполнения вычислений результаты передаются на иницилирующий узел.

Приложение для смартфона (Client) обеспечивает ввод IP-адреса и порта для подключения к хосту (рис. 3). В программе предусмотрена возможность сбора логов основных характеристик смартфона: температура CPU, тактовые

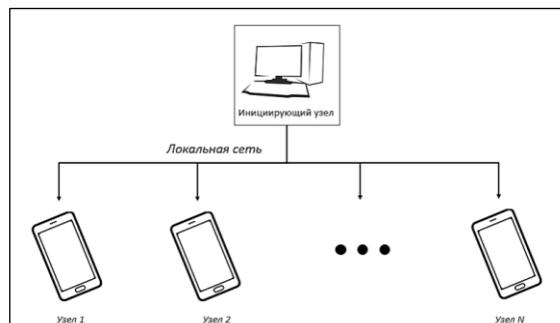


Рис. 2. Распределенная система с мобильными узлами

Fig. 2. Distributed system with mobile nodes



Рис. 3. Интерфейс ПО для смартфона (Client)

Fig. 3. Smartphone software interface (Client)

частоты каждого из ядер во время работы программы, время работы ядер на каждой из частот, температура, напряжение и уровень заряда батареи.

Интерфейс приложения для ПК (Server) представлен на рисунке 4. На экране в таблице отображаются все подключенные к нему узлы. Основные функции для работы запускаются нажатием кнопок, расположенных на панели справа.

Представим основные функции ПО иницилирующего узла (функции приложения Server):

Open File – открытие файла с вычислительной задачей;

Send File – пересылка приложения на выбранные узлы;

Check File – проверка наличия приложения на выбранных узлах;

Calculate – запуск вычислений;

Balance – запуск балансировщика нагрузки узлов;

Stress – нагрузочное тестирование; функция предназначена для определения пиковых характеристик узла.

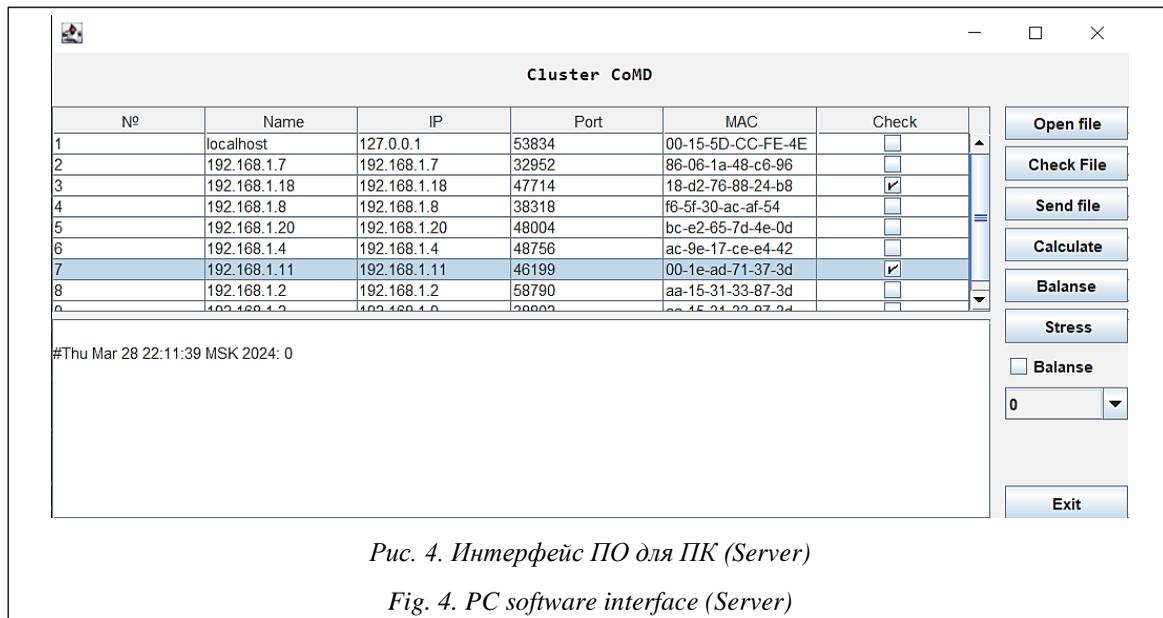


Рис. 4. Интерфейс ПО для ПК (Server)

Fig. 4. PC software interface (Server)

Разработанное ПО позволяет запускать распределенные приложения, не требующие интенсивного обмена сообщениями между узлами сети. Программный код и параметры рассылаются вычислительным узлам (рис. 4) перед началом работы. На каждом узле запускается приложение с полученными параметрами, а после завершения вычислений узел записывает полученный результат в файл. К этому классу задач относится, например, численное интегрирование.

#### Программные средства для организации распределенных вычислений

ПО для серверной части и клиентских узлов было разработано на языке Java, обеспечивающем необходимую производительность. Основной особенностью разработанного ПО является возможность включения в программу участков кода на языке C, которые и будут выполнять ресурсоемкие вычисления. Рассмотрим специфику реализации некоторых функций.

#### Механизм взаимодействия сервера и клиентов

После запуска приложения открывается порт (рис. 4), выбранный на прослушивание TCP-соединения, по которому к серверу подключаются клиенты-смартфоны. Взаимодействие между клиентами и сервером реализуется с помощью механизма сообщений, каждое из которых начинается и заканчивается ключевыми словами. Процесс взаимодействия происходит следующим образом: с иницилирующего устройства отправляется сообщение, которое содержит ключевое слово, определяющее тип дальнейшей операции.

Синтаксис ключевых слов:

BALANCE – расчет коэффициента мощности узла;

SNDF – отправка приложения на узел;

CHS – проверка наличия приложения на узле;

STRESS – проведение нагрузочного тестирования;

CALC – проведение вычислений.

Рассмотрим программную реализацию основных операций.

#### Расчет коэффициента мощности узла

Критерием эффективности работы распределенной системы является минимизация времени выполнения задачи. Гетерогенность среды не позволяет добиться этого простейшим способом: разбивая задачу на равные части, требуется учитывать характеристики каждого узла, то есть выполнять балансировку нагрузки узлов. Для настройки балансировки на сервере хранится таблица, содержащая в себе MAC-адрес устройства и значение, равное времени выполнения тестовой задачи на узле. Эта информация позволяет вычислить относительную мощность всех интегрированных в единую среду устройств.

Коэффициент относительной мощности используется для разбиения задачи на части, пропорциональные возможностям устройства. Клиентам с высокой производительностью необходимо выделить больший объем вычислений, а для смартфона с низкими вычислительными возможностями выделить маленький участок.

В качестве характеристики мощности узла можно использовать и значение пиковой про-

изводительности, основанное на его архитектуре и частоте работы ядер процессора. При этом нет необходимости тестировать смартфон перед выполнением задачи, но в то же время результат балансировки будет значительно уступать нагрузочному тестированию [11].

Полученные результаты подтверждают это утверждение (рис. 5). При балансировке с использованием тестовой задачи время работы узлов сопоставимо друг с другом, тогда как при балансировке, основанной на пиковой производительности, оно сильно отличается.

*Отправка приложения на узел*

Для разработки более гибкого приложения, позволяющего запускать разнообразные задачи на стороне клиента, исполняемый код задачи передается узлу в виде арк-приложения. Пересылаемая программа должна содержать java-класс, которому необходимо дать название DynamicCalculator. Класс должен содержать в себе метод Task, который будет вызван на клиенте. Выгрузка программного кода будет происходить с помощью класса DexClassLoader, разработанного для Android-приложений, и позволяющего загружать классы из файлов .jar и .apk.

Для приложения, отправляемого клиентам, на сервере вычисляется контрольная сумма файла с помощью алгоритма md5. После передачи файла контрольная сумма рассчитывается и на клиентском узле, а полученные значения сравниваются между собой. При совпадении в лог программы выводится сообщение об успешном завершении операции пересылки, в противном случае – о неудаче и предложении повторить пересылку. На рисунке 6 представлена соответствующая этой операции диаграмма обмена сообщениями.

В случае отправки вычислительной задачи приложению клиента обмен сообщениями будет происходить следующим образом. Первым сообщением отправляется ключевое слово SNDF (сокращение от *Sendfile*), запускающее на клиенте метод класса для приема файла. В ответ клиент отправляет сообщение о готовности к приему данных, после чего последовательно получает имя файла и его содержимое. По завершении процесса передачи на клиенте рассчитывается контрольная сумма полученного файла и отправляется на сервер, где сравнивается с исходной. В случае совпадения процесс передачи завершается.

*Проверка наличия задачи на узле*

Пересылка задачи с сервера на клиентский узел занимает продолжительное время. Чтобы

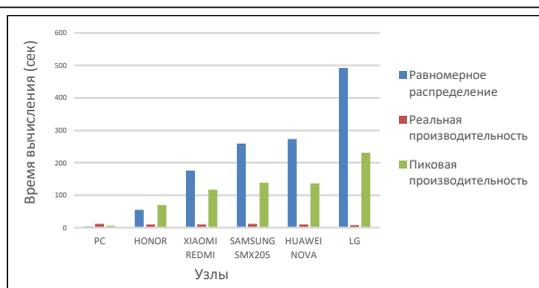


Рис. 5. Эффективность методов балансировки

Fig. 5. Effectiveness of balancing methods



Рис. 6. Диаграмма обмена сообщениями при отправке вычислительной задачи

Fig. 6. Diagram of message exchange when sending a computational task

избежать повторной передачи данных, можно проверить наличие на узлах выбранной на исполнение задачи. Для этого используется метод сравнения контрольной суммы файла с задачей на сервере и на клиентском узле. Если значения совпадают, то соответствующий файл уже находится на узле и дополнительная пересылка не требуется.

*Проведение нагрузочного тестирования*

Для оценки параметров устройства во время вычислений используется нагрузочное тестирование. Оно реализовано с помощью выполнения в бесконечном цикле нагрузочной задачи до получения сигнала прерывания со стороны сервера. Во время ее исполнения можно снимать с узла с определенным промежутком времени следующие характеристики: время работы каждого из ядер процессора, значение тактовой частоты ядер, температура процессора, напряжение на батарее, уровень заряда батареи, температура батареи.

*Проведение вычислений в распределенной системе*

После загрузки вычислительной задачи на каждый из узлов запускается процесс вычисления. Для этого на каждый из выбранных узлов (рис. 4) отправляется ключевое слово CALC и после получения подтверждения от принимающей стороны происходит отправка параметров для запуска задачи. По окончании вычисления его результат возвращается на сервер. Если алгоритм предполагает многократный запуск задачи на узле, то возможна повторная отправка запроса на вычисления с новыми параметрами вычислительной задачи.

*Эксперименты*

Для проверки функциональности разработанного ПО была организована распределенная система, состоящая из шести узлов (<http://www.swsys.ru/uploaded/image/2024-4/6.jpg>),

характеристики которых представлены в таблице 2.

Для узлов распределенной системы была рассчитана пиковая производительность по формуле

$$Perf = N_{cores} F_{cores} OpC,$$

где  $N_{cores}$  – число ядер процессора;  $F_{cores}$  – рабочая частота ядра (ГГц);  $OpC$  – число операций, выполняемых за такт.

Коэффициент относительной мощности узла определяется по формуле

$$Coef_i = \frac{\sum_1^6 Perf_i}{Perf_i}.$$

В таблице 3 приведены значения коэффициентов относительной мощности узлов, рассчитанные на основе их пиковой и реальной производительности.

Пиковая производительность распределенной системы составляет 210 Гигафлопс. Для

Таблица 2

**Характеристики узлов распределенной системы**

Table 2

**Characteristics of distributed system nodes**

Узел	Тип устройства	Марка	Модель	SoC	Процессор	RAM (Гб)	Число логических ядер	Тактовая частота (ГГц)
1	ПК	-	-	-	Intel core i5 11400H CPU @ 2.70 GHz	32	4	4*2.20
2	Смартфон	Huawei (Honor)	8X (JSN-L22)	Hisilicon Kirin 710	ARM 4×Cortex-A73 ARM 4×Cortex-A53	4	8	4*2,2 4*1,7
3	Смартфон	Xiaomi	Readme	Mt6877	ARM 2× Cortex-A78 ARM 6× Cortex-A55	8	8	2*2,6 6*2.0
4	Планшетный компьютер	Samsung	SM-X205	T618	ARM 6×Cortex-A55 ARM 2×Cortex-A75	4	8	8*2.02
5	Смартфон	Huawei	Huawei-Nova	Qualcomm Snapdragon 625	ARM 8×Cortex-A53	3	8	8*2.02
6	Смартфон	LG	Spirit LG-H422	My70ds	ARM 4×Cortex-A7	1	4	4*1.3

Таблица 3

**Коэффициенты относительной мощности узлов**

Table 3

**Coefficients of node relative power**

Узел	Тип устройства	Марка	Модель	Perf <sub>i</sub> (Гигафлопс)	Coef <sub>i</sub>		
					Пиковая производительность	Реальная производительность	
						Однопоточное выполнение	Многopotочное выполнение
1	ПК	-	-	64,8	0,31	0,633	0,455
2	Смартфон	Huawei (Honor)	8X (JSN-L22)	31,2	0,15	0,181	0,161
3	Смартфон	Xiaomi	Readme	44,8	0,21	0,115	0,242
4	Планшетный компьютер	Samsung	SM-X205	32,0	0,15	0,011	0,050
5	Смартфон			32,0	0,15	0,049	0,088
6	Смартфон			5,2	0,02	0,011	0,002

оценки реальной производительности распределенной системы были проведены вычислительные эксперименты, в ходе которых запускалось приложение, выполняющее поиск экстремума функции с помощью метода дихотомии. Подобные методы используются и при обучении нейронных сетей.

Эксперименты проводились как в однопоточном, так и в многопоточном режимах, причем задействовались все доступные ядра узлов. Для балансировки нагрузки узлов были использованы коэффициенты из таблицы 3 для реальной производительности. Полученные результаты представлены на рисунке 7 и в таблице 4. На оси абсцисс рисунка отражается состав распределенной среды, то есть ПК, ПК+HONOR, ПК+HONOR+XIAOMI, REDMI и т.д.

Таблица 4

**Эффективность HGRID**

Table 4

**HGRID effectiveness**

Режим	Время решения задачи (сек.)		Ускорение
	ПК	HGRID	
Однопоточный	27.3	16.7	1.6
Многопоточный	6.6	3.5	1.8

Время решения задачи в распределенной среде снизилось в 1,6 раза в однопоточном режиме и в 1,8 в многопоточном по сравнению с вычислением на ПК.

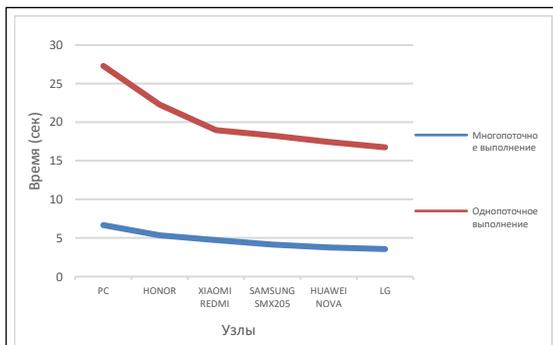


Рис. 7. Время решения задачи

Fig. 7. Problem solving time

**Заключение**

Проведенные исследования подтверждают возможность повышения вычислительной мощности ПК за счет интеграции с ним распределенной системы из смартфонов. Приложение, разработанное для проведения вычислений в HGRID, поддерживает узлы с операционными системами Android и Windows. Хотя использование старых смартфонов с низкими вычислительными возможностями и не оказывает существенного влияния на производительность HGRID, механизм балансировки позволяет использовать и их вклад в решение задачи. Конечно, HGRID-системы ориентированы прежде всего на реализацию распределенных приложений, но могут быть использованы, например, для отладки параллельных приложений.

Дальнейшие исследования будут направлены на оценку применимости HGRID для решения задачи обучения нейронной сети.

**Список литературы**

1. Балабаев С.А, Балабаев А.А. Применение CoMD-систем для обучения нейронных сетей // Микроэлектроника и информатика: матер. науч.-технич. конф. 2024. С. 24–28.
2. Балабаев С.А., Лупин С.А., Шакиров Р.Н. Вычислительный кластер на основе смартфонов Android и микрокомпьютеров Raspberry Pi // INJOIT. 2022. Т. 10. № 7. С. 86–93.
3. Kurochkin I., Dolgov A., Manzyuk M., Vatutin E. Using mobile devices in a voluntary distributed computing project to solve combinatorial problems. In: CCIS. Proc. RuSCDays, 2021, vol. 1510, pp. 525–537. doi: 10.1007/978-3-030-92864-3\_40.
4. Gurusamy V., Nandhini K. IBIS: The new era for distributed computing, IJESTR, 2018, 7(1), pp. 61–65.
5. Li K.H., de Gusmão P.P.B., Beutel D.J., Lane N.D. Secure aggregation for federated learning in flower. Proc. DistributedML, 2021, pp. 8–14. doi: 10.1145/3488659.3493776.
6. Hasumi M. Azumi T. Federated learning platform on embedded many-core processor with flower. Proc. RAGE Workshop, 2024, pp. 1–6. doi: 10.1109/RAGE62451.2024.00015.
7. Таненбаум Э.С., Херберт Б. Современные операционные системы. СПб: Питер, 2018. 1120 с.
8. Еранссон А. Эффективное использование потоков в операционной системе Android. М.: ДМК Пресс, 2015. 304 с.
9. Балабаев С.А., Лупин С.А. Оценка вычислительных возможностей мобильных устройств на платформе ОС Аврора // Микроэлектроника и информатика: матер. науч.-технич. конф. 2023. С. 51–56.
10. Таненбаум Э., Фимстер Н., Уэзеролл Д. Компьютерные сети; [пер. с англ.]. СПб: Питер, 2023. 992 с.

11. Мин Тху Кхайнг, Лупин С.А., Аунг Тху. Оценка эффективности методов балансировки нагрузки в распределенных вычислительных системах // INJOIT. 2021. Т. 9. № 11. С. 30–36.

Software &amp; Systems

doi: 10.15827/0236-235X.148.504-513

2024, 37(4), pp. 504–513

## Increasing the PC computing power: Integration with a distributed smartphone system

Sergey A. Balabaev<sup>1</sup>, Sergey A. Lupin<sup>1</sup>, Pavel N. Telegin<sup>2,3</sup>✉, Boris M. Shabanov<sup>2,3</sup>

<sup>1</sup> National Research University of Electronic Technology, MIET,  
Moscow, 124498, Russian Federation

<sup>2</sup> Joint Supercomputer Center of RAS, Moscow, 119334, Russian Federation

<sup>3</sup> National Research Centre “Kurchatov Institute”,  
Moscow, 123182, Russian Federation

### For citation

Balabaev, S.A., Lupin, S.A., Telegin, P.N., Shabanov, B.M. (2024) ‘Increasing the PC computing power: Integration with a distributed smartphone system’, *Software & Systems*, 37(4), pp. 504–513 (in Russ.). doi: 10.15827/0236-235X.148.504-513

### Article info

Received: 25.07.2024

After revision: 27.08.2024

Accepted: 30.08.2024

**Abstract.** The paper considers the possibility of improving PC performance by integrating a distributed smartphone system with it. There is an overview of the main hardware and software features of mobile devices. The authors evaluated the effect of such nodes on the performance of a distributed computing system. It was found that the main differences from PCs are low quality of device cooling and big.LITTLE architecture. These mobile device features do not allow using all smartphone cores at full capacity and should be taken into account when integrating them into a unified computing environment. Software for integrating mobile devices with personal computers consists of two applications: a client and a server developed in Java programming language. The applications allow uploading a computable task to nodes, running it, accumulating and displaying the computational results obtained from the nodes. The devices interact over the network. The computable task is an apk application containing a Java class with methods that can be called from the application running on the client. The utilized node load-balancing algorithm allows integrating smartphones of different generations, which differ significantly in performance, into a single environment. The authors show that in order to distribute the load efficiently between the system nodes, it is necessary to use their real performance value. The paper presents the results of solving the problem of function minimization in a distributed environment organized using the developed software. They confirm the achievement of the set goal. The obtained results are useful for a wide range of specialists.

**Keywords:** distributed computing, smartphone, Android OS, grid systems, node load balancing

**Acknowledgements.** The paper was carried out under the government assignment, project no. FNEF-2024-0016

### References

- Balabaev S., Balabaev A. (2024) ‘Application of CoMD systems for training neural networks’, *Proc. Sci. and Tech. Conf. Microelectronics and Informatics*, pp. 24–28 (in Russ.).
- Balabaev, S.A., Lupin, S.A., Shakirov, R.N. (2022) ‘Computing cluster based on Android smartphones and Raspberry Pi microcomputers’, *INJOIT*, 10(7), pp. 86–93 (in Russ.).
- Kurochkin, I., Dolgov, A., Manzyuk, M., Vatutin, E. (2021) ‘Using mobile devices in a voluntary distributed computing project to solve combinatorial problems’, in *CCIS. Proc. RuSCDays*, 1510, pp. 525–537. doi: 10.1007/978-3-030-92864-3\_40.
- Gurusamy V., Nandhini K. (2018) ‘IBIS: The new era for distributed computing’, *IJESTR*, 7(1), pp. 61–65.
- Li, K.H., de Gusmão, P.P.B., Beutel, D.J., Lane, N.D. (2021) ‘Secure aggregation for federated learning in flower’, *Proc. DistributedML*, pp. 8–14. doi: 10.1145/3488659.3493776.
- Hasumi, M., Azumi, T. (2024) ‘Federated learning platform on embedded many-core processor with flower’, *Proc. RAGE Workshop*, pp. 1–6. doi: 10.1109/RAGE62451.2024.00015.
- Tanenbaum, E.S., Herbert, B. (2014) *Modern Operating Systems*. NJ, Hoboken: Pearson Education Publ., 1185 p. (Russ.ed.: (2018) St. Petersburg, 1020 p.).
- Göransson, A. (2014) *Efficient Android Threading*. O’Reilly Publ., 277 p. (Russ. ed.: (2015) Moscow, 304 p.).
- Balabaev, S.A., Lupin, S.A. (2023) ‘Assessment of computing capabilities of mobile devices on the Aurora OS platform’, *Proc. Sci. and Tech. Conf. Microelectronics and Informatics*, pp. 51–56 (in Russ.).
- Tanenbaum, E., Feamster, N., Wetherall, D. (2021) *Computer Networks*. NJ, Hoboken: Pearson Education Publ., 944 p. (Russ. ed.: (2023) St. Petersburg, 992 p.).

11. Min Thu Khaing, Lupin, S.A., Aung Thu (2021) 'Evaluating the effectiveness of load balancing methods in distributed computing systems', *INJOIT*, 9(11), pp. 30–36 (in Russ.).

**Авторы**

**Балабаев Сергей Андреевич**,

аспирант, sergei.balabaev@mail.ru

**Лупин Сергей Андреевич**<sup>1</sup>, к.т.н.,

профессор, lupin@miec.ru

**Телегин Павел Николаевич**<sup>2,3</sup>, к.т.н.,

ведущий научный сотрудник, ptelegin@jscs.ru

**Шабанов Борис Михайлович**<sup>2,3</sup>,

д.т.н., чл.-корр. РАН, директор,

зам. директора, shabanov@jscs.ru

<sup>1</sup> Национальный исследовательский университет «МИЭТ»,

г. Москва, 124498, Россия

<sup>2</sup> Межведомственный суперкомпьютерный центр РАН, г. Москва, 119334, Россия

<sup>3</sup> Национальный исследовательский центр «Курчатовский институт», г. Москва, 123182, Россия

**Authors**

**Sergey A. Balabaev**<sup>1</sup>, Postgraduate Student,  
sergei.balabaev@mail.ru

**Sergey A. Lupin**<sup>1</sup>, Cand. of Sci. (Engineering),  
Professor, lupin@miec.ru

**Pavel N. Telegin**<sup>2,3</sup>, Cand. of Sci. (Engineering),  
Leading Researcher, ptelegin@jscs.ru

**Boris M. Shabanov**<sup>2,3</sup>, Dr.Sci. (Engineering),  
Corresponding Member of the RAS,  
Director, Deputy Director, shabanov@jscs.ru

<sup>1</sup> National Research University of Electronic Technology, MIET,  
Moscow, 124498, Russian Federation

<sup>2</sup> Joint Supercomputer Center of RAS,  
Moscow, 119334, Russian Federation

<sup>3</sup> National Research Centre “Kurchatov Institute”,  
Moscow, 123182, Russian Federation

## Диалоговый программный комплекс для формирования онтологии Единого цифрового пространства научных знаний

С.А. Власова<sup>1,2</sup>✉, Н.Е. Каленов<sup>1,2</sup>

<sup>1</sup> Межведомственный суперкомпьютерный центр РАН, г. Москва, 119334, Россия

<sup>2</sup> Национальный исследовательский центр «Курчатовский институт», г. Москва, 123182, Россия

### Ссылка для цитирования

Власова С.А., Каленов Н.Е. Диалоговый программный комплекс для формирования онтологии Единого цифрового пространства научных знаний // Программные продукты и системы. 2024. Т. 37. № 4. С. 514–523. doi: 10.15827/0236-235X.148.514-523

### Информация о статье

Группа специальностей ВАК: 2.3.5

Поступила в редакцию: 25.07.2024

После доработки: 23.08.2024

Принята к публикации: 27.08.2024

**Аннотация.** В статье рассматривается программный комплекс, формирующий онтологию Единого цифрового пространства научных знаний (ЕЦПНЗ). Онтология включает совокупность справочников, описывающих подпространства, классы объектов, атрибуты объектов и связи различных типов и видов. Каждый справочник атрибутов объектов содержит наименование атрибута, информацию о его формате и статусе, а также имя словаря значений и его тип. Справочники связей содержат необходимую информацию о связях пар объектов или значений атрибутов. Онтология ЕЦПНЗ включает именованные связи трех типов – универсальные, квазиуниверсальные и специфические. Конкретная связь любого типа может быть простой или составной. Представленный в статье программный комплекс реализует сервисы, связанные с формированием справочников подпространств, классов, атрибутов объектов и связей всех типов ЕЦПНЗ. Комплекс ориентирован на администраторов подпространств ЕЦПНЗ и позволяет в режиме диалога создавать структурные элементы ЕЦПНЗ. Эти элементы необходимы для формирования контента пространства с автоматическим определением списка атрибутов объектов, форматов их ввода и необходимых средств формально-логического контроля. При работе комплекса обеспечивается контроль вводимой информации, исключая дублирование элементов или некорректный ввод данных, противоречащий декларированной классом «форматы» структуре. Программный комплекс состоит из двух модулей – формирования новых элементов и редактирования уже имеющихся. Модуль редактирования позволяет удалять элементы онтологии без нарушения ее целостности. В результате работы комплекса строится структура, являющаяся основой для унифицированных программ формирования контента ЕЦПНЗ.

**Ключевые слова:** программный комплекс, цифровое пространство научных знаний, онтология, классы объектов, атрибуты, структуризация, связанные данные

**Благодарности.** Работа выполнена в МСЦ РАН и НИЦ «Курчатовский институт» в рамках госзадания № FNEF-2024-0014

**Введение.** Структурированная информационная среда *Единое цифровое пространство научных знаний* (ЕЦПНЗ) объединяет сведения о различных объектах, связанных с наукой, в качестве которых могут выступать цифровая копия физической сущности (например, книги, музейного предмета, архивного документа и т.п.), база данных, информация об ученом, научном мероприятии, научном факте и т.п. [1, 2].

Учитывая основные тренды современного развития информатики, связанные с автоматизацией обработки научной информации и применением искусственного интеллекта во многих областях науки, ЕЦПНЗ необходимо строить на принципах Semantic Web [3, 4], а при его разработке использовать онтологический подход [5].

Для описания онтологий и проектирования на их основе информационных систем существует много языков, основой которых является принятый в качестве стандарта OWL,

базирующийся на представлении данных в структурах RDF (<http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/>) и RDFS (<https://www.w3.org/TR/rdf-schema/>). На базе этих стандартов консорциумом W3 разработана модель организации знаний SKOS (*Simple Knowledge Organization System*), унифицирующая структуру представления тезаурусов, относящихся к различным областям знаний (<https://www.w3.org/TR/skos-reference/>).

В Сети представлено значительное количество публикаций, посвященных разработкам на основе OWL-онтологий, описывающих отдельные направления науки и техники как за рубежом (например, [6–8]), так и в России (например, [9–11]).

Однако, как показывает анализ, использование OWL затруднительно при описании объектов, обладающих многоуровневыми атрибутами (свойствами – в терминологии OWL), такими как численность населения страны или города

по годам, расход воды в водопадах по месяцам, работа персоны в той или иной организации в разных должностях в разное время и т.п. Эта проблема и сложности ее преодоления подчеркнуты, в частности, в [12]. Отражая динамику развития различных областей науки, в том числе в историческом плане, онтология ЕЦПНЗ должна учитывать такие аспекты, как переименование населенных пунктов, изменение принадлежности территорий тем или иным государствам, различные обозначения единиц измерения и т.п.

Необходимо отметить, что в информатике существует ряд трактовок термина «онтология» применительно к различным задачам. В данной работе под онтологией будем понимать структурированные данные, отражающие информацию об элементах ЕЦПНЗ и их связях.

### Элементы онтологии ЕЦПНЗ

ЕЦПНЗ представляет собой совокупность тематических *подпространств* (ПП), связанных единой онтологией – унифицированными правилами отражения информации об объектах и их связях, а также собственно связями между объектами различных ПП.

Структура онтологии ЕЦПНЗ описана в [13–15] и представлена в виде элементов справочника CDSSK на сайте ЕЦПНЗ (<http://dirsmc.ru/cdssk/structure/>). Чтобы облегчить понимание работы программного комплекса, приведем основные определения, относящиеся к структуре онтологии ЕЦПНЗ.

Объект ЕЦПНЗ характеризуется фиксированным набором атрибутов (свойств), каждый из которых принимает определенное значение (или ряд значений) для конкретного объекта. Перечень атрибутов объектов определяется задачами ЕЦПНЗ, к решению которых имеет отношение данный объект.

Объекты ПП с одинаковыми наборами атрибутов объединяются в классы.

Между парами любых объектов и значений атрибутов могут быть установлены связи, каждая из которых характеризуется видом, типом и (факультативно) значением.

Структуры классов, атрибутов объектов и связей определяются унифицированными справочниками и соответствующими объектами класса «форматы» универсального ПП.

Каждый элемент ЕЦПНЗ имеет свое уникальное имя (URN). Упростить обработку данных для формирования URN позволяют следующие введенные мнемонические правила.

Каждому формируемому ПП присваивается префикс, который в общем случае может состоять из фиксированного числа символов, определяемого в справочнике ЕЦПНЗ. В данной версии префикс ПП состоит из двух буквенно-цифровых символов, включающих стандартную латиницу и кириллицу.

URN класса объектов состоит из четырех буквенно-цифровых символов: первые два являются префиксом ПП, к которому принадлежит данный класс, два других – префиксом класса. Классы с одним и тем же префиксом, обозначающим группу описываемых сущностей, могут принадлежать различным ПП и иметь разный, но пересекающийся набор атрибутов. Например, персоны, включаемые в универсальное ПП, имеют более широкий набор атрибутов, чем персоны, включаемые в тематическое ПП; объекты класса «местоположение» универсального ПП имеют существенно меньше атрибутов, чем те же объекты, описываемые в ПП, относящихся к географии, и т.п. Эквивалентность отдельных объектов из разных ПП устанавливается с помощью соответствующей связи. В качестве URN класса не используются буквосочетания REUN, REQU, RESP, зарезервированные для обозначения универсальных, квазиуниверсальных и специфических связей.

Справочники ЕЦПНЗ представляют собой кортежи, их размерность определяется видом элемента, который они описывают (ПП, класс, атрибут, связи).

Значения атрибутов объектов и связей хранятся в словарях одного из двух видов – статических или динамических. Статические словари (в справочниках они обозначены буквой S) содержат стандартизированные значения атрибутов (перечень ученых степеней и званий персон, классификационные индексы, виды отношений между объектами различных классов и т.п.). Их элементы формируются либо при первоначальной загрузке системы в пакетном режиме, либо в процессе формирования справочников разработанным программным комплексом. Словари значений второго типа формируются в процессе наполнения контента ЕЦПНЗ (в справочниках они обозначены буквой D).

Предлагаемый программный комплекс предназначен для формирования справочников элементов ЕЦПНЗ всех видов и для наполнения статических словарей значений атрибутов объектов и связей. Он ориентирован на администратора, формирующего и редактирующего

структуру определенного ПП с его классами и связями объектов.

Результатом работы комплекса является система связанных справочников и статических словарей, URN которых построены по установленным правилам. Элементы справочников и статических словарей динамически формируются в виде реляционной базы данных, а также текстовых строк заданной структуры [14].

Одной из основных составляющих справочников ЕЦПНЗ является указание на формат представления элемента, к которому относится справочник. Каждый элемент формата представляет собой объект вспомогательного класса универсального ПП «форматы», имеющий 6 атрибутов [16] – «тип представления данных», «вид формата», «признак обязательности», «признак уникальности», «ограничения по структуре», «описание». Значения атрибутов формируются данным программным комплексом и хранятся в соответствующих статических словарях, а их комбинации, характеризующие правила представления конкретных атрибутов и связей, – в словаре форматов. URN соответствующих элементов словаря форматов являются обязательными составляющими всех справочников атрибутов объектов и связей.

### Классификация связей ЕЦПНЗ

Для упрощения алгоритмов формирования справочников связей и сокращения диалога с администратором введена классификация связей. Каждая связь определяется типом и видом.

Выделены пять типов связей:

- универсальная связь между объектами;
- универсальная связь между значениями атрибутов;
- универсальная связь между значениями конкретного объекта;
- квазиуниверсальная связь между объектами;
- специфическая связь между объектами.

Связь определенного типа может относиться к одному из следующих видов:

- простая связь, представляющая собой аналог триплетов RDF и содержащая указания на связи «субъект» и «объект» (здесь и далее приводимые в кавычках понятия «субъект» и «объект» понимаются в терминах триплета RDF, на рисунках они обозначены как  $O_1$  и  $O_2$  соответственно) (рис. 1);

- составная связь первого рода, содержащая (дополнительно к указаниям «субъект» и «объ-

ект») URN значений атрибутов связи, приведенных в соответствующих словарях, определяемых справочником атрибутов связи (рис. 2);

- составная связь второго рода, содержащая неразветвленную иерархическую цепочку значений атрибутов, когда все значения атрибута данного уровня имеют одни и те же подчиненные атрибуты (рис. 3);

- составная связь третьего рода – разветвленная иерархическая цепочка, в которой каждый атрибут верхнего уровня имеет свои атрибуты следующего уровня (рис. 4).

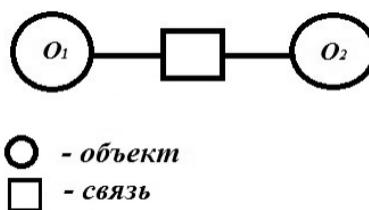


Рис. 1. Простая связь

Fig. 1. Simple relation

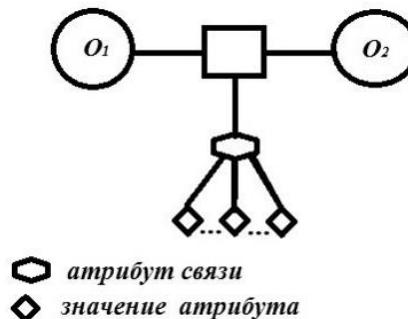


Рис. 2. Составная связь первого рода

Fig. 2. Composite relation of the first kind

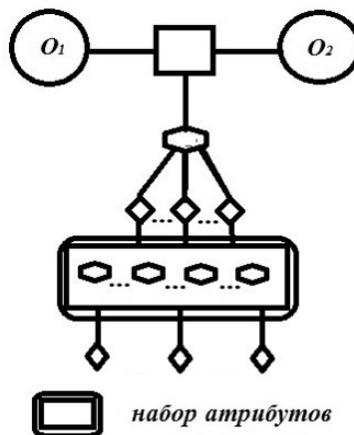
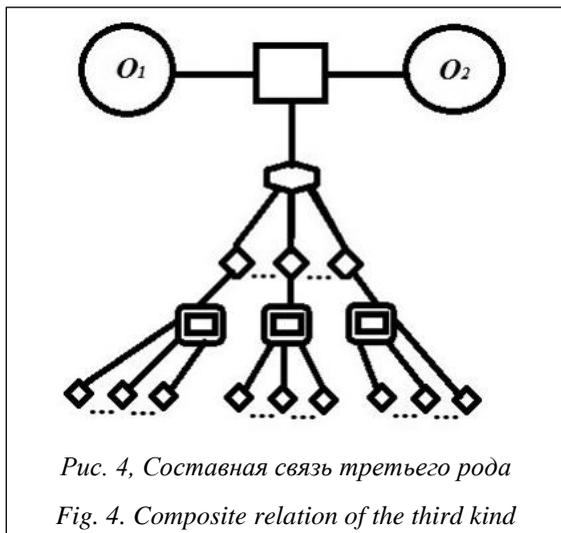


Рис. 3. Составная связь второго рода

Fig. 3. Composite relation of the second kind



### Типы атрибутов объектов ЕЦПНЗ

Атрибуты могут иметь простые или составные значения, а также принимать значения связей.

Простое значение атрибута определяется элементом словаря значений атрибута, заданного в его справочнике.

Составное значение атрибута может состоять из нескольких значений одного уровня (одноуровневый составной атрибут) или из нескольких иерархических значений (многоуровневый составной атрибут). К одноуровневым составным атрибутам относятся, например, количественные и временные характеристики объекта – высота, длина, скорость и т.п. Значение такого атрибута состоит из единицы измерения и ее числового выражения. Примером многоуровневого составного атрибута может служить расход воды в водопаде или реке (количественная характеристика в зависимости от времени года), численность населения административной единицы с указанием года, скорость звука в различных средах и т.п.

Значение атрибута как связи может вводиться для удобства пользовательского интерфейса при диалоговом вводе данных. Так, при формировании метаданных доклада на конференции удобнее запросить ввод атрибута «автор доклада», чем запрашивать связь доклада с персоной.

Вид атрибута определяется его справочником, содержащим указание на соответствующий элемент формата, имя словаря значений атрибута и имя справочника подчиненного атрибута (при его наличии). Для простого атрибута в справочнике будет указано имя словаря значений атрибута, для составного атрибута –

имя словаря соответствующего объекта, для атрибута как связи – имя словаря соответствующей связи.

Для автоматического формирования системы связанных справочников и статических словарей ЕЦПНЗ разработан специальный диалоговый программный комплекс ФОНТ ЕЦПНЗ. Он создан на основе технологии Microsoft ASP.NET на платформе Microsoft .NET Framework в среде разработки Microsoft Visual Studio.

### Структура программного комплекса ФОНТ ЕЦПНЗ

Комплекс включает модули создания элементов ЕЦПНЗ (ПП, классов объектов, атрибутов объектов и связей) и редактирования их справочников и статических словарей значений атрибутов.

В модуле создания ПП пользователю предоставляется список наименований имеющихся ПП (рис. 5). Наименования являются активными ссылками, обеспечивающими получение данных о ПП с возможностью их редактирования. Для регистрации нового ПП необходимо ввести наименование ПП, префикс (2 символа), описание ПП (рис. 5).

Модуль формирования классов предлагает список наименований имеющихся ПП, из него нужно выбрать то, в котором будет создан класс. Для регистрации нового класса нужно ввести его наименование и префикс (2 символа). Автоматически будет сформирован URN справочника атрибутов для данного класса, который при необходимости можно изменить.

Для создания атрибута из списка наименований имеющихся ПП выбирается нужное, затем из списка классов данного ПП выбирается класс, в котором будет создан атрибут. На рисунке 6 показана страница создания справочника нового атрибута в классе «административные единицы» ПП «Социально-экономическая география». В правой части экрана выдается список наименований имеющихся атрибутов данного класса. Наименования являются активными ссылками, обеспечивающими получение данных справочника атрибута и словаря его значений.

Просмотр данных атрибута «вид объекта»:

Наименование атрибута

**вид объекта**

URN атрибута **A\_39AU.2**

Формат представления значений атрибута

**UNFT . 42**

URN словаря значений атрибута

**N\_A\_39AU.2**

Тип словаря значений **S**

Словарь значений атрибута

**часть света**

**страна**

**регион / область / штат / край**

**город / поселок / деревня**

Создание нового атрибута начинается с ввода его названия в левой части экрана (рис. 6).

Далее определяется формат атрибута. Из списка имеющихся типов формата выбирается нужный тип, например, выбор типа формата для атрибута «численность населения»:

Подпространство: **Социально-экономическая география**

Класс: **административные единицы**

Наименование атрибута: **численность населения**

**Тип формата**

- текст
- изображение
- видео
- аудио
- любое число
- ссылка на внешний ресурс
- дата в формате гггг[.мм[.дд]]
- **количественная характеристика**
- временная характеристика

Затем из списка форматов указанного типа выбирается формат, например, для атрибута «численность населения»:

**Единое цифровое пространство научных знаний**

**Существующие подпространства**

- [Универсальное](#)
- [Научные мероприятия](#)

**Ввод нового подпространства**

\* Наименование подпространства

Социально-экономическая география

\* Префикс (2 символа)

\* Описание

система географических наук, изучающая территориальную организацию общества, её законы и закономерности

**Ввод**

Рис. 5. Регистрация нового ПП

Fig. 5. Registration of a new subspace

**Единое цифровое пространство научных знаний**

**Существующие атрибуты**

- [наименование объекта](#)
- [вид объекта](#)
- [площадь](#)
- [примечания](#)

**Создание справочника атрибутов**

Подпространство: **Социально-экономическая география**    Класс: **административные единицы**

\* Наименование атрибута

численность населения

**Ввод**

Рис. 6. Создание справочника атрибута класса «административные единицы»

Fig. 6. Creating a directory of the class attribute “administrative units”

**Выберите формат**

- Количественная характеристика атрибут обязательный, значение повторяется
- Количественная характеристика атрибут факультативный, значение повторяющееся
- Факультативная уникальная количественная характеристика
- Другой вид формата

Если нужного формата в списке нет, то система предоставляет возможность ввода нового формата.

После выбора формата автоматически формируется URN словаря значений атрибута. В данном случае выбран URN словаря значений атрибута «количественная характеристика» (рис. 7). При вводе простого формата по умолчанию формируется имя словаря его значений в форме N\_<URN справочника атрибута>, которое можно при необходимости изменить. В случае формирования атрибута «как связь» или составного атрибута (тип формата атрибута «количественная характеристика», «временная характеристика» и т.п.) URN словаря значений атрибута изменить нельзя. Для завершения формирования справочника атрибута нужно указать тип словаря значений (S – статический или D – динамический) и при необходимости ввести дополнительную информацию (рис. 7). Для словарей типа S предоставляется возможность ввода значений для данного атрибута.

После формирования справочника нового атрибута можно перейти к формированию атрибута второго уровня. Так, например, для атрибута «численность населения» можно сфор-

**Создание справочника атрибутов**

Подпространство: Социально-экономическая география административные единицы    Класс:

Наименование атрибута: численность населения  
Формат представления значений атрибута: UNFT.46

URN словаря значений атрибута: UNQC  
Тип словаря значений (S - статический, D - динамический)

S  
 D

Дополнительная информация

*Рис. 7. Создание справочника для атрибута «численность населения»*

*Fig. 7. Creating a directory for "population size" attribute*

мировать подчиненный атрибут «дата указания количества населения» с форматом «дата в формате гггг[.мм[.дд]], атрибут необязательный, значение уникальное».

В программном комплексе ФОНТ ЕЦПНЗ модули создания ПП, классов и атрибутов связаны между собой удобной навигацией. После регистрации нового ПП можно сразу перейти к созданию классов для него. После регистрации класса перейти к созданию справочников атрибутов данного класса. При создании нового атрибута следует сразу ввести его значения в соответствующий статический словарь.

Программный комплекс предусматривает возможность редактирования наименований и описаний ПП и классов. Специальный модуль редактирования атрибутов позволяет изменять наименование и описание атрибута, URN и тип словаря значений (если атрибут еще нигде не использовался), добавлять и редактировать значения словаря атрибута (<http://www.swsys.ru/uploaded/image/2024-4/4.jpg>).

**Модуль создания связей**

В данном модуле процедура формирования новой связи определяется ее форматом. При создании любой связи необходимо вводить ее наименование и описание. Формирование квазиуниверсальной связи требует определения объекта связи, а специфической – и субъекта, и объекта. При создании связи с форматами «универсальная составная связь первого рода между объектами», «универсальная составная связь первого рода между значениями атрибутов», «квазиуниверсальная составная связь первого рода», «специфическая составная связь первого рода» формируются справочник атрибута связи и словарь значений атрибута.

В связях с форматами «универсальная составная связь второго рода между объектами», «универсальная составная связь второго рода между значениями атрибутов конкретного объекта», «специфическая составная связь второго рода» формируется справочник атрибута связи. Если тип словаря значений атрибута статический, то вводятся значения атрибута. Также формируются справочники атрибутов второго уровня (относятся ко всем значениям атрибута первого уровня) и вводятся их значения (в случае статического типа словаря значений).

При создании связи с форматом «специфическая составная связь третьего рода» формируются справочник атрибута связи и словарь значений атрибута. Запрашивается количество

подчиненных связей. Выбирается значение атрибута, для него формируются справочники подчиненных атрибутов и вводятся значения для этих атрибутов (в случае статического типа словаря значений). Последняя процедура повторяется для всех значений атрибута первого уровня, у которых должны быть подчиненные атрибуты.

Работу модуля формирования связей рассмотрим на примере создания связи персоны с публикацией. После выбора типа связи «Специфическая связь» предоставляется список имеющихся специфических связей и предлагается ввести наименование новой связи и ее описание (<http://www.swsys.ru/uploaded/image/2024-4/5.jpg>). Затем формируются префикс субъекта связи (выбирается «Универсальное пространство», класс «Персоны»), префикс объекта связи (выбирается «Универсальное пространство», класс «Публикации») и выбирается формат связи «специфическая составная связь первого рода»:

Наименование связи: **связь персоны с публикацией**

Префикс класса субъекта: **UNPS**

Префикс класса объекта: **UNPB**

**Формат представления связи**

- o специфическая простая связь
- o **специфическая составная связь первого рода**
- o специфическая составная связь второго рода
- o специфическая составная связь третьего рода

Далее для формирования справочника атрибута вводится его наименование (в данном

примере – «роль персоны в создании публикации»), выбирается формат атрибута «текст, только буквы, атрибут обязательный, значение повторяющееся», указывается тип словаря значений атрибута «S – статический» (рис. 8). После ввода значений словаря данного атрибута (например, «автор», «редактор», «составитель») формирование связи будет полностью завершено. Представим все данные сформированной связи:

Наименование связи **связь персоны с публикацией**

URN связи **RESP.5**

Префикс класса субъекта **UNPS**

Префикс класса объекта **UNPB**

формат представления связи **UNFT.53 специфическая составная связь первого рода**

URN справочника атрибута связи **A\_RESP.5**

Наименование атрибута **роль персоны в создании публикации**

Формат значений атрибута **UNFT.9**

**Текст, только буквы, атрибут обязательный, значение повторяющееся**

URN справочника атрибута **A\_RESP.5.1**

URN справочника значений атрибута **N\_A\_RESP.5.1**

Тип словаря значений **S**

Словарь значений атрибута

автор

редактор

составитель

автор перевода

художник

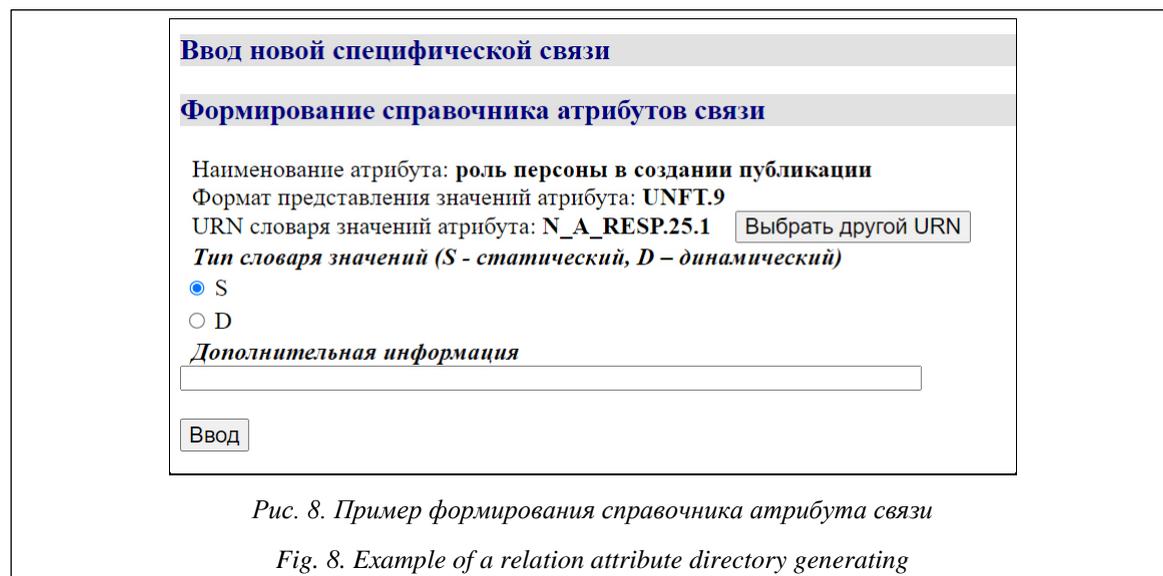


Рис. 8. Пример формирования справочника атрибута связи

Fig. 8. Example of a relation attribute directory generating

Специальный модуль редактирования связей позволяет вносить изменения в атрибуты связей и в словари их значений.

### Заключение

Комплекс ФОНТ ЕЦПНЗ находится в стадии опытной эксплуатации и используется для моделирования процессов формирования различных элементов ЕЦПНЗ и связей между разнородными объектами, отражающими научные знания. В частности, сформирована онтология и загружены атрибуты и связи объектов, им-

портированных из БД публикаций сотрудников МСЦ РАН, в стадии загрузки находится одна из коллекций музейных объектов. В дальнейшем предполагается реализовать многоплатформенную версию комплекса. Сформированные с помощью комплекса варианты онтологии отдельных ПП являются базой для реализации унифицированных программных средств ЕЦПНЗ, обеспечивающих пакетную и диалоговую технологии наполнения пространства разнородными реальными научными данными, что составляет одну из очередных задач в области формирования ЕЦПНЗ.

### Список литературы

1. Антопольский А.Б., Каленов Н.Е., Серебряков В.А., Сотников А.Н. О едином цифровом пространстве научных знаний // Вестн. РАН. 2019. Т. 89. № 7. С. 728–735. doi: 10.31857/S0869-5873897728-735.
2. Савин Г.И. Единое цифровое пространство научных знаний: цели и задачи // Информационные ресурсы России. 2020. № 5. С. 3–5. doi: 10.51218/0204-3653-2020-5-3-5.
3. Хорошевский В.Ф. Пространства знаний в сети Интернет и Semantic WEB (часть 1) // ИИИПР. 2008. № 1. С. 80–97.
4. Хорошевский В.Ф. Пространства знаний в сети Интернет и Semantic WEB (часть 2) // ИИИПР. 2009. № 4. С. 15–36.
5. Каленов Н.Е., Серебряков В.А. Об онтологии Единого цифрового пространства научных знаний // Информационные ресурсы России. 2020. № 5. С. 10–12. doi: 10.51218/0204-3653-2020-5-10-12.
6. Alromema W., Alahmadi A.H. Ontology building for patient bioinformatics of the smart card domain using OWL. *IJCC*, 2022, vol. 11, no. 4, pp. 316–329. doi: 10.1504/IJCC.2022.124796.
7. Luo Yu., Xiong X., Jin Sh., Liu Z. VIM: Research on OWL-based vocabulary ontology construction method for units of measurement. *Electronics*, 2023, vol. 12, no. 18, art. 3783. doi: 10.3390/electronics12183783.
8. Kosińska J., Brotoń G., Tobiasz M. Knowledge representation of the state of a cloud-native application. *STTT*, 2024, vol. 26, no. 1, pp. 21–32. doi: 10.1007/s10009-023-00705-2.
9. Загорюлько Ю.А., Боровикова О.И. Использование системы разнородных паттернов онтологического проектирования для разработки онтологий научных предметных областей // Программирование. 2020. № 4. С. 27–35. doi: 10.31857/S0132347420040068.
10. Загорюлько Ю.А., Сидорова Е.А., Загорюлько Г.Б., Ахмадеева И.Р., Серый А.С. Автоматизация разработки онтологий научных предметных областей на основе паттернов онтологического проектирования // Онтология проектирования. 2021. Т. 11. № 4. С. 500–520. doi: 10.18287/2223-9537-2021-11-4-500-520.
11. Митрофанова О.А., Митрофанов Е.П., Буре Н.А. Применение онтологического подхода к проектированию геопространственной базы опытных данных для информационного обеспечения исследований в точном земледелии // Вестн. СПбГУ. Прикладная математика. Информатика. Процессы управления. 2022. Т. 18. № 2. С. 253–262. doi: 10.21638/11701/spbu10.2022.206.
12. Марчук А.Г., Марчук П.А. Базовая онтология неспецифических сущностей BONE и ее использование для построения информационных систем // Вестн. СибГУТИ. 2014. № 4. С. 118–128.
13. Каленов Н.Е., Сотников А.Н. Структура онтологии единого цифрового пространства научных знаний // НТИ. Сер. 2. 2023. № 7. С. 20–26. doi: 10.36535/0548-0027-2023-07-3.
14. Kalenov N.E., Sobolevskaya I.N., Sotnikov A.N. Common digital space of scientific knowledge ontology structuring. *Lobachevskii J. of Math.*, 2023, vol. 44, pp. 2733–2743. doi: 10.1134/S1995080223070235.
15. Kalenov N.E., Sotnikov A.N. Common digital space of scientific knowledge as an integrator of polythematic information resources. *Doklady Math.*, 2024, vol. 109, pp. 93–99. doi: 10.1134/S106456242470176X.
16. Власова С.А., Каленов Н.Е., Сотников А.Н. Онтология вспомогательных и политематических предметных классов единого цифрового пространства научных знаний // Электронные библиотеки. 2024. Т. 27. № 1. С. 22–42. doi: 10.26907/1562-5419-2024-27-1-22-42.

**Dialogue software complex for forming Common Digital Space of Scientific Knowledge ontology**Svetlana A. Vlasova <sup>1,2</sup>✉, Nikolay E. Kalenov <sup>1,2</sup><sup>1</sup> Joint Supercomputer Center of RAS, Moscow, 119991, Russian Federation<sup>2</sup> National Research Centre “Kurchatov Institute”,  
Moscow, 123182, Russian Federation**For citation**Vlasova, S.A., Kalenov, N.E. (2024) ‘Dialogue software complex for forming Common Digital Space of Scientific Knowledge ontology’, *Software & Systems*, 37(4), pp. 514–523 (in Russ.). doi: 10.15827/0236-235X.148.514-523**Article info**

Received: 25.07.2024

After revision: 23.08.2024

Accepted: 27.08.2024

**Abstract.** The paper considers a software package that forms the ontology of the Common Digital Space of Scientific Knowledge (CDSSK). The ontology includes a set of directories describing subspaces, classes of objects, object attributes and relations of various types and kinds. Each object attribute directory contains an attribute name, information about its format and status (obligatory or optional, unique or multiple), as well as the name and type of the value dictionary. Relationship directories contain necessary information about the connections between object pairs or attribute values. The CDSSK ontology includes named connections of three types: universal, quasi-universal, and specific. A specific connection of any type can be simple or composite. The paper presents the software package that implements services related to the formation of reference books of subspaces, classes, attributes of objects and connections of all CDSSK types. The software package is for the administrators of CDSSK subspaces; it allows creating CDSSK structural elements in the dialog mode. These elements are necessary for forming the space content with automatic definition of the list of object attributes, their input formats and the necessary means of formal and logical control. The complex operation enables the input information control that excludes elements duplication or incorrect data input, which contradicts the structure declared by the class “formats”. The software package consists of two modules: forming new elements and editing existing ones. The editing module also allows deleting ontology elements without violating its integrity. The complex work enables building a structure that is a basis for unified programs of CDSSK content formation.

**Keywords:** software package, digital space of scientific knowledge, ontology, object classes, attributes, structuring, linked data

**Acknowledgements.** The paper was carried out under the government assignment, project No. FNEF-2024-0014

**References**

1. Antopolskii, A.B., Kalenov, N.E., Serebryakov, V.A., Sotnikov, A.N. (2019) ‘Common digital space of scientific knowledge’, *Bull. of RAS*, 89(7), pp. 728–735 (in Russ.). doi: 10.31857/S0869-5873897728-735.
2. Savin, G.I. (2020) ‘Common digital space of scientific knowledge: Goals and tasks’, *Inform. Resources of Russia*, (5), pp. 3–5 (in Russ.). doi: 10.51218/0204-3653-2020-5-3-5.
3. Khoroshevskiy, V.F. (2008) ‘Knowledge spaces on the Internet and Semantic WEB (part 1)’, *Artificial Intelligence and Decision Making*, (1), pp. 80–97 (in Russ.).
4. Khoroshevskiy, V.F. (2009) ‘Knowledge spaces on the Internet and Semantic WEB (part 2)’, *Artificial Intelligence and Decision Making*, (4), pp. 15–36 (in Russ.).
5. Kalenov, N.E., Serebryakov, V.A. (2020) ‘Ontology of the common digital space of scientific knowledge’, *Inform. Resources of Russia*, (5), pp. 10–12 (in Russ.). doi: 10.51218/0204-3653-2020-5-10-12.
6. Alromema, W., Alahmadi, A.H. (2022) ‘Ontology building for patient bioinformatics of the smart card domain using OWL’, *IJCC*, 11(4), pp. 316–329. doi: 10.1504/IJCC.2022.124796.
7. Luo, Yu., Xiong, X., Jin, Sh., Liu, Z. (2023) ‘VIM: Research on OWL-based vocabulary ontology construction method for units of measurement’, *Electronics*, 12(18), art. 3783. doi: 10.3390/electronics12183783.
8. Kosińska, J., Brotoń, G., Tobiasz, M. (2024) ‘Knowledge representation of the state of a cloud-native application’, *STTT*, 26(1), pp. 21–32. doi: 10.1007/s10009-023-00705-2.
9. Zagorulko, Yu.A., Borovikova, O.I. (2020) ‘Using a system of heterogeneous ontology design patterns to develop ontologies of scientific subject domains’, *Programming and Computer Software*, 46(4), pp. 273–280.
10. Zagorulko, Yu.A., Sidorova, E.A., Zagorulko, G.B., Akhmadeeva, I.R., Sery, A.S. (2021) ‘Automation of the development of ontologies of scientific subject domains based on ontology design patterns’, *Ontology of Designing*, 11(4), pp. 500–520 (in Russ.). doi: 10.18287/2223-9537-2021-11-4-500-520.
11. Mitrofanova, O.A., Mitrofanov, E.P., Bure, N.A. (2022) ‘Ontological approach application to the design of a geo-spatial experimental database for information support of research in precision agriculture’, *Vestn. of SPbSU. Applied Math. Comput. Sci. Control Processes*, 18(2), pp. 253–262 (in Russ.). doi: 10.21638/11701/spbu10.2022.206.
12. Marchuk, A.G., Marchuk, P.A. (2014) ‘Basic ontology of non-specific entities BONE and its use for building information systems’, *The Herald of the SibSUTIS*, (4), pp. 118–128 (in Russ.).
13. Kalenov, N.E., Sotnikov, A.N. (2023) ‘Structure of the common digital space of scientific knowledge ontology’, *STI. Ser. 2*, (7), pp. 20–26 (in Russ.). doi: 10.36535/0548-0027-2023-07-3.

14. Kalenov, N.E., Sobolevskaya, I.N., Sotnikov, A.N. (2023) 'Common digital space of scientific knowledge ontology structurization', *Lobachevskii J. of Math.*, 44, pp. 2733–2743. doi: 10.1134/S1995080223070235.

15. Kalenov, N.E., Sotnikov, A.N. (2024) 'Common digital space of scientific knowledge as an integrator of polythematic information resources', *Doklady Math.*, 109, pp. 93–99. doi: 10.1134/S106456242470176X.

16. Vlasova, S.A., Kalenov, N.E., Sotnikov, A.N. (2024) 'Ontology of the universal subspace of common digital space of scientific knowledge', *Russ. Digital Libraries J.*, 27(1), pp. 22–42 (in Russ.).

#### Авторы

**Власова Светлана Александровна**<sup>1,2</sup>, к.т.н.,  
вед. науч. сотр., vlas.svetlana2013@yandex.ru  
**Каленов Николай Евгеньевич**<sup>1,2</sup>, д.т.н.,  
профессор, гл. науч. сотр., nekalenov@mail.ru

<sup>1</sup> Межведомственный суперкомпьютерный  
центр РАН, г. Москва, 119334, Россия

<sup>2</sup> Национальный исследовательский центр  
«Курчатовский институт», г. Москва, 123182, Россия

#### Authors

**Svetlana A. Vlasova**<sup>1,2</sup>, Cand. of Sci. (Engineering),  
Leading Researcher, vlas.svetlana2013@yandex.ru  
**Nikolay E. Kalenov**<sup>1,2</sup>, Dr.Sci. (Engineering)  
Professor, Chief Researcher, nekalenov@mail.ru

<sup>1</sup> Joint Supercomputer Center of RAS,  
Moscow, 119334, Russian Federation

<sup>2</sup> National Research Centre “Kurchatov Institute”,  
Moscow, 123182, Russian Federation

## Организация поиска в базе данных со связанными сущностями

К.П. Погорелко<sup>1, 2</sup>✉, Г.И. Савин<sup>1, 2</sup>

<sup>1</sup> Межведомственный суперкомпьютерный центр РАН, г. Москва, 119334, Россия

<sup>2</sup> Национальный исследовательский центр «Курчатовский институт», г. Москва, 123182, Россия

### Ссылка для цитирования

Погорелко К.П., Савин Г.И. Организация поиска в базе данных со связанными сущностями // Программные продукты и системы. 2024. Т. 37. № 4. С. 524–531. doi: 10.15827/0236-235X.148.524-531

### Информация о статье

Группа специальностей ВАК: 2.3.5

Поступила в редакцию: 24.07.2024

После доработки: 22.08.2024

Принята к публикации: 30.08.2024

**Аннотация.** В базе данных со связанными сущностями возникает проблема организации поиска, при котором можно не только формулировать условия к атрибутам определенной сущности, но и задавать условия на связанные с ней сущности. В статье описан реализованный поисковый механизм, на базе которого организован поиск в информационной системе электронной библиотеки «Научное наследие России». Поисковый механизм обеспечивает работу на русском и английском языках. Настройка поискового механизма на структуру конкретной базы данных заключается в заполнении ряда таблиц. Поисковый механизм позволяет работать с такими типами атрибутов, как текстовое поле, числовое значение, булево значение, полнотекстовый индекс текстового поля, кодификаторы. Предусмотрен различный функционал для администраторов и обычных пользователей. Для формулировки запроса к базе данных разработан интуитивно понятный интерактивный построитель поисковых запросов, который позволяет в режиме диалога задавать условия к первоначально выбранной сущности, к связанным с ней сущностям, к их количеству и типам связи. Для работы с кодификаторами большого объема предусмотрен расширенный функционал. Описана организация html-разметки формы ввода и структуры данных, формируемых при построении запроса. Предложено алгоритмическое решение, которое позволяет построить SQL-предписание к базе данных по сформулированному в построителе запросу. Приведены пример запроса к связанным сущностям и построенное для этого запроса SQL-предписание. Описана организация системы визуализации результатов поиска. Предложенные алгоритмы построения поискового механизма могут служить моделью для организации поиска в Едином цифровом пространстве научных знаний, включающем разнородные именованные связи между объектами.

**Ключевые слова:** информационный поиск, документный поиск, база данных, автоматизированная система, связанные сущности, графовые системы

**Благодарности.** Работа выполнена в МСЦ РАН и НИЦ «Курчатовский институт» в рамках госзадания № FNEF-2024-0014

**Введение.** Начало проекту российской академической научной *электронной библиотеки «Научное наследие России»* (ЭБ ННР) было положено в 2009 г. [1]. Он создавался как интегратор многоаспектной цифровой информации о выдающихся российских ученых, внесших вклад в развитие фундаментальных естественных и гуманитарных наук [2]. По сравнению с традиционными цифровыми библиотеками БД этой системы имеет расширенный набор сущностей, в нее добавлены такие, как «музейный объект» и «коллекция». Кроме того, традиционные сущности имеют расширенный набор атрибутов, например, сущность «персона» имеет атрибуты, отражающие биографические сведения, направления научной деятельности и т.п. [3]. Первоначально поиск в ЭБ ННР был организован средствами программной оболочки Единого научного информационного пространства РАН [4] с достаточно ограниченными возможностями. Поскольку предполагается, что структура БД ЭБ ННР будет расширяться как модель Единого цифрового пространства научных знаний [5], требуется реализация по-

искового механизма, предоставляющего больше возможностей при формулировке поискового запроса.

В информационных системах, БД которых содержат связанные сущности, возникает потребность организации такого поиска, при котором обеспечивается возможность формулировать условия к атрибутам определенной сущности, а также задавать условия на количество связанных с ней сущностей, отвечающих своему набору заданных условий. Эти условия, в свою очередь, могут содержать требования на связи с какими-то другими сущностями. Кроме того, связи между сущностями могут иметь спецификации. Например, спецификация связи между сущностями «публикация» и «персона» может принимать значения «автор», «редактор», «переводчик», «о нем» и т.п. Таким образом, необходим поисковый механизм, который позволял бы формулировать подобный следующему запросу: «получить список авторов, публикации которых в области физики входят в коллекцию «Материалы по истории Российской академии наук».

В настоящее время для организации информационного поиска используется ряд моделей. Лидирующее положение среди них занимают векторная модель, вероятностные модели, такие как LSI и LDA, и предварительно обученные модели на основе искусственного интеллекта, например, BERT [6]. Все эти модели предназначены для организации поиска в слабо или совсем не структурированной текстовой информации большого объема. Также отмечается их основной недостаток – невозможность четко сформулировать требования к получаемым документам, что является актуальным, особенно в академической среде [7]. В силу этого для организации поиска для ЭБ ННР была выбрана булева модель. Сформулированный ранее запрос будет выглядеть в ней так: «найти экземпляры сущности «персона», которые имеют связь, специфицированную как «автор», хотя бы с одним экземпляром сущности «публикация», которая, в свою очередь, во-первых, отнесена к проблематике «физика» и, во-вторых, имеет связь хотя бы с одним экземпляром сущности «коллекция», название которого соответствует требуемому».

Известно, что БД, обеспечивающие работу с многосвязными данными, относят к категории графовых [8]. Поскольку использование стандартного языка SQL при формировании запросов к связанным сущностям приводит к достаточно громоздким конструкциям, для упрощения манипулирования данными в графовых базах разработан ряд специализированных языков. Прежде всего необходимо отметить CQL (*Cypher Query Language*), созданный в рамках проекта Neo4j (<https://technology.amis.nl/database/querying-connected-data-in-graph-databases-with-neo4j>), и GQL (*Graph Query Language*) (<https://github.com/OlofMorra/GQL-parser/blob/main/src/main/resources/report/A%20Semantics%20of%20GQL;%20a%20New%20Query%20Language%20forProperty%20Graphs%20Formalized.pdf>), который в апреле 2024 г. был опубликован ISO в качестве международного стандарта. Однако использование этих языков требует их изучения, знания специфики и структуры данных конкретной базы и предоставления соответствующих полномочий пользователю. Задачей описываемой разработки было создание поискового механизма с интуитивно понятным интерфейсом, который позволял бы формулировать запросы неподготовленному пользователю.

Организация поиска потребовала решения трех задач. Во-первых, необходим собственно

построитель поисковых запросов, который, предоставляя на каждом этапе выбор из возможных для этого места опций, позволяет пользователю формулировать требуемый запрос. Во-вторых, требуется алгоритмическое решение для преобразования сформулированного в построителе запроса в SQL-предписание к БД. В-третьих, необходима система визуализации найденных результатов.

Вариант поисковой системы с такими возможностями был реализован в ЭБ ННР [9]. С учетом опыта эксплуатации библиотеки в настоящее время появился новый вариант поисковой системы, отличающийся расширенным функционалом и возможностями его быстрой перенастройки на изменения структуры БД. Построитель поисковых запросов и система выдачи результатов реализованы на стороне клиента по технологии одностраничного приложения (SPA) на языке JavaScript. Сервисы для них на стороне сервера и функционал построения поискового предписания для SQL-сервера реализованы на языке C# в среде .NET 8.0. БД располагается на сервере PostgreSQL. Поисковая система для пользователей доступна по адресу <http://www.e-heritage.ru/search/>.

## Структура данных

Настройка поискового механизма на структуру конкретной БД сводится к заполнению ряда таблиц.

Прежде всего это перечень сущностей БД (классов объектов в терминологии онтологии Единого цифрового пространства научных знаний [10]). Для каждой сущности, которая может участвовать в поиске, задается следующая информация.

- Название таблицы БД, содержащей экземпляры сущности. Это название является идентификатором сущности, на который ссылаются другие таблицы настройки поискового механизма.

- Название поля таблицы, являющегося главным индексом.

- Название сущности на двух языках. Это связано с тем, что ЭБ ННР является двуязычной и построитель запросов работает в зависимости от затребованного языка.

- Признак, указывающий на то, что данная сущность доступна только администраторам системы. БД содержит данные, необходимые для обеспечения технологического процесса по вводу/редактированию информации, поэтому

в ней есть, например, такие сущности, как «ошибка», которые обычным пользователям не должны быть доступны.

– Условие на атрибуты, что данный экземпляр сущности имеет статус «опубликовано» и может быть выдан обычному пользователю. Администратору выдаются все найденные записи.

– Информация, необходимая для представления пользователю найденных результатов. Она включает перечень полей, из которых выбираются сведения, строку форматирования этих полей и форматы http-ссылок на конкретную сущность для администраторов и обычных пользователей.

– Перечень возможных сортировок сущностей с указанием названия сортировки на двух языках и перечня полей, по которым производится сортировка полученных результатов поиска.

– Перечень возможных элементов поискового условия для данной сущности.

Определяется перечень кодификаторов системы, которые могут использоваться при формулировке поискового запроса. Для каждого кодификатора задаются названия

– таблицы, содержащей кодификатор; это название является идентификатором кодификатора;

– кодификатора на русском и английском языках;

– полей, содержащих русские и английские тексты кодификатора;

– индексного поля.

Задается перечень связей между сущностями. Для каждой связи приводится следующая информация:

– название таблицы, связывающей сущности (оно является идентификатором связи);

– идентификаторы связанных сущностей;

– названия индексных полей, по которым осуществляется связь;

– если связь может иметь спецификации, указывается идентификатор соответствующего кодификатора.

Для каждого возможного элемента поискового условия для сущности задается следующая информация.

– Идентификатор элемента.

– Название на русском и английском языках.

– Признак того, что данный элемент доступен только администраторам.

– Тип элемента. Элемент может быть условием на значения как какого-либо атрибута

экземпляра сущности, так и связанных с ней кодификаторов. Он также может быть условием на связанную сущность, а точнее, на количество связанных с этим экземпляром других экземпляров связанных сущностей, отвечающих своему набору условий. Поддерживаются следующие типы элементов: текстовый, числовой, полнотекстовый индекс текстового поля, дата, двоичный (возможные значения «Да»–«Нет»), простой кодификатор – кодификатор небольшого размера, который может быть показан целиком, иерархический кодификатор – кодификатор большого размера с иерархической структурой, как, например, кодификатор ГРНТИ, условие на связанную сущность.

– Дополнительная информация. Для элемента, относящегося к условию на атрибут сущности, задается перечень названий полей, в которых содержится соответствующая информация. В определенных случаях поиск целесообразно выполнять сразу по нескольким полям. Так, в ЭБ ННР название публикации отражено в полях «оригинальное название» и «переведенное название», название коллекции присутствует на двух языках и т.д. Таким образом, условие, заданное для элемента «название публикации», проверяется как для поля «оригинальное название», так и для поля «переведенное название». Для кодификатора задаются идентификатор кодификатора и информация о таблице и индексах, связывающих кодификатор с сущностью. Для условия на связанную сущность указывается идентификатор связи.

Вся работа функциональных блоков поисковой системы основана на использовании данной информации.

### Построитель поискового запроса

При запросе на поиск построитель поискового запроса запрашивает у соответствующего сервиса на сервере поисковую информацию, передавая в качестве параметра код языка, на котором общается пользователь. На основе этого параметра, а также аутентификации пользователя как администратора сервис возвращает выборку из структуры данных, соответствующую запрошенным параметрам. Далее с учетом этой информации в зависимости от действий пользователя происходит модификация html-кода страницы и формируются необходимые структуры данных. По умолчанию сущностью для поиска выбирается первая из списка сущностей, а условие соответствует первому элементу из списка возможных элементов

поискового условия. При выборе другой сущности для поиска автоматически изменяется и условие по умолчанию.

При нажатии кнопки «Добавить условие...» дополнительные условия располагаются построчно с тем же отступом и имеют возможность выбора связывающего оператора И/ИЛИ. Перед каждым условием возможен выбор оператора НЕ. Вычисление логического значения ведется сверху вниз без учета старшинства операторов. Для возможности редактирования запроса каждая поисковая строка имеет кнопку, позволяющую удалить эту строку. В зависимости от типа выбранного элемента поискового условия формируются соответствующие элементы html-разметки.

Набор элементов для условий на атрибуты сущности:

- для текстового атрибута – возможность выбора опции «содержит», «начинается с» или «равно» и поля для ввода необходимых фрагментов текста;

- для числового атрибута – возможность выбора оператора «=», «>» или «<» и поля для ввода числового значения; поскольку числовые поля не всегда могут быть определены (например, если для года введено «первая половина XVIII века»), для обеспечения полноты поиска вводится возможность добавления к результату записи с неустановленным значением;

- для полнотекстового индекса – поле, позволяющее ввести слова, словоформы которых будут обнаруживаться в индексе, и возможность выбора взаимного расположения словоформ «рядом» или «в любом порядке»;

- для атрибута типа дата – возможность выбора оператора «=», «>» или «<» и поля для задания даты;

- для двоичного типа – возможность выбора «да»–«нет».

Для выбора условий на значения кодификатора имеется кнопка «Выбрать». Дальнейшие операции при ее нажатии зависят от типа кодификатора.

- Для простого кодификатора он выдается целиком. После завершения выбора в поисковой строке отображаются выбранные тексты рубрик. Повторное нажатие кнопки позволяет редактировать сделанный выбор.

- Для иерархического кодификатора предусмотрена более сложная функциональность, позволяющая производить поиск как по иерархии, так и по текстам рубрик. При выборе рубрики, содержащей подчиненные рубрики, можно указать необходимость их включения в условие. Поскольку выбранных рубрик может быть много, по завершении выбора в поисковой строке выдается только их количество. Повторное нажатие кнопки позволяет редактировать сделанный выбор.

Пример задания поисковых условий разного типа приведен на рисунке.

На связанную сущность имеются две группы условий: на количество связанных с этим экземпляром сущностей, отвечающих определенным условиям, и собственно эти условия. При выборе в поисковой строке элементов формируются соответствующие поля.

Условие на количество может иметь значения «хотя бы одна», «все» или количественное

*Использование поисковых условий разного типа*  
*Usage of search terms of different types*

значение с указанием числа и возможности выбора оператора « $\Leftarrow$ », « $\Rightarrow$ » или « $\Leftarrow$ ». Если связь между сущностями может иметь спецификацию, следует нажать кнопку для ее выбора. В результате выдается список спецификаций данной связи из соответствующего кодификатора, после выбора нужных значений их названия выводятся в соответствующую поисковую строку.

Набор условий на связанную сущность содержит поисковое условие по умолчанию для данной сущности и кнопку добавления условий. Этот список форматируется с отступом по отношению к родительскому списку условий, что позволяет наглядно представлять структуру формируемого запроса.

Возможности системы проиллюстрированы дополнительными примерами запросов (<http://www.swsys.ru/uploaded/image/2024-4/11.jpg>, <http://www.swsys.ru/uploaded/image/2024-4/12.jpg>).

### Структура информации строителя поискового запроса

Информация, вводимая пользователем в ходе конструирования поискового запроса, формируется следующим образом. Каждая поисковая строка получает уникальный номер. Все поля и элементы, относящиеся к данной строке, получают при генерации идентификатор (атрибуты 'id' и 'name' соответствующего html-элемента), состоящий из префикса, который определяет тип поля или элемента, и из номера поисковой строки. Так, для элемента, определяющего выбранный поисковый элемент, используется префикс 'f'. Значения его атрибута 'value' содержит идентификатор элемента в перечне возможных поисковых элементов сущности. Для поля, содержащего введенные данные, используется префикс 'd', для элемента, определяющего отношение данных, – префикс 't', для элемента связи условий – 'a', для отрицания – 'n' и т.д. Структура формируемого запроса сохраняется в виде дерева, каждый элемент которого соответствует строке поискового условия и содержит информацию о собственном числовом идентификаторе, а если это условие на связанную сущность, то еще и список идентификаторов строк, задающих условия на эту сущность. При запросе на поиск на сервер передаются структура запроса, свернутая в формат JSON, и вся информация, введенная пользователем в параметрах запроса в виде словаря «ключ–значение», где ключом является идентификатор соответствующих html-элементов, а значением – значение их атрибута 'value'.

### Генератор поискового предписания

Генерация поискового предписания для SQL-сервера происходит следующим образом. Определяется сущность, экземпляры которой требуется найти. Далее вызывается рекурсивная процедура построения SQL-запроса, которой в качестве параметров передаются идентификатор сущности и соответствующий ей узел структуры запроса.

Представим алгоритм процедуры. По идентификатору сущности на основе информации из перечня сущностей строится оператор SQL: 'SELECT <главный индекс таблицы сущности> FROM <таблица сущности> WHERE'. Затем перебираются номера дочерних элементов узла структуры запроса. Для каждого номера n из присланного в запросе словаря выбирается элемент с ключом 'f<n>', на основе которого отбирается информация из списка поисковых элементов соответствующей сущности. Далее в зависимости от типа элемента формируется поисковое условие.

- Для элементов, имеющих тип «текст», анализируется заданный тип отношения (элемент 't<n>'). Если задано отношение «содержит», то введенные данные (элемент 'd<n>') разбиваются на подстроки и для каждой подстроки формируется поисковое условие, все они затем соединяются оператором AND. Для типов «равно» и «начинается с» данные используются непосредственно. Для поиска в базе используется SQL-оператор 'ilike', который обеспечивает сравнение без учета регистра с заданием усечения, соответствующего указанному типу отношения. Для административного варианта поиска проверяется дополнительное условие. Если задано отношение «равно», а поле данных пустое, то формируется запрос на документы, у которых это поле не заполнено.

- Для элементов, имеющих типы «число», «дата» или «булево», проверяется корректность информации, заданной в соответствующем поле, и формируется соответствующее поисковое условие. В случае типа «число», если задана опция добавления документов с неопределенным значением, добавляется соответствующее условие.

- Для кодификаторов на основе информации об отобранных в запросе индексах (элемент 'd<n>') и о таблице связи кодификатора и сущности формируется условие '<главный индекс таблицы сущности> IN (SELECT <главный индекс таблицы сущности > FROM

<таблица связи с кодификатором> WHERE <индекс кодификатора> IN (<перечень отобранных индексов>))’.

- Для связанной сущности выбирается ее идентификатор и осуществляется рекурсивный вызов процедуры построения SQL-запроса с этим идентификатором и соответствующим элементом поисковой структуры. Возвращенное значение (оператор SELECT для связанной сущности) дополняется необходимыми параметрами, соответствующими количественным условиям на связанные сущности и таблицы связи сущностей.

Если в поисковой строке задан оператор НЕ (элемент ‘n<n>’), то полученное поисковое условие предваряется оператором NOT и на основе элемента ‘a<n>’ добавляется к генерируемому WHERE с использованием оператора AND или OR.

Если поисковый запрос не от администратора, после обработки всех дочерних элементов добавляется условие, что данный экземпляр опубликован, и процедура возвращает сформированный SQL-оператор.

Возвращенная процедурой команда для главной сущности дополняется параметрами выбранной пользователем сортировки и передается SQL-серверу. Полученные в результате выполнения запроса индексы передаются системе выдачи результатов поиска.

Сгенерированная SQL-команда для запроса на связанные сущности (<http://www.swsys.ru/uploaded/image/2024-4/7.jpg>) выглядит следующим образом:

```
SELECT personid FROM person WHERE
(person.personid in (SELECT personid
FROM publicationperson WHERE (publicationperson.roleid in (1)) AND (publicationid in (SELECT publicationid FROM
publication WHERE (((publication.publicationid in (SELECT publicationid FROM
publicationcollection WHERE collectionid in (SELECT collectionid FROM collection WHERE ((collectionname ilike '%академ%') and (collectionname ilike '%наук%')) OR ((collectionnameeng ilike '%академ%') and (collectionnameeng ilike '%наук%')))) and (ispublished)))) AND (publication.publicationid in (select distinct publicationid from publicationgrnti where codgrnti like '29%')))) and (ispublished)))) and (ispublished)
```

## Система визуализации результатов поиска

Поскольку объем полученной после выполнения запроса информации может быть значительным, в системе организована страничная выдача. К ее организации обычно используется подход, когда, при каждом запросе новой страницы выполняется процедура поиска и из нее выделяется нужный диапазон записей. Поскольку БД в ЭБ ННР меняется динамически, данный подход приводит к неприятным последствиям, так как результаты одного и того же запроса, полученные в разные моменты времени, могут отличаться и одна и та же страница при повторной выдаче может содержать разные документы. Кроме того, запросы бывают достаточно сложными и их повторное вычисление при пролистывании может привести к неоправданным задержкам. В данной поисковой системе результат поиска в виде списка найденных идентификаторов передается клиенту. Клиент при запросе страницы пользователем выделяет нужный диапазон идентификаторов и запрашивает у сервера необходимые документы. Сервер, получив нужные идентификаторы, на основе информации из описания сущности о формате представления пользователю найденных результатов и вида запроса (администратор–пользователь) формирует ответ, который возвращает клиенту.

## Выводы

Анализ протоколов работы поискового механизма позволяет сделать вывод, что при наличии ссылок на прежние, более привычные формы он пользуется устойчивым спросом, хотя и скрыт за ссылкой «расширенный поиск». Основными потребителями являются пользователи, у которых есть необходимость в получении ответов на сложные запросы, – администраторы системы и специалисты, занимающиеся аналитикой.

Разработанные подходы к формированию поисковых механизмов планируется реализовать в Едином цифровом пространстве научных знаний.

## Список литературы

1. Каленов Н.Е., Савин Г.И., Сотников А.Н. Электронная библиотека «Научное наследие России» // Информационные ресурсы России. 2009. № 2. С. 19–20.
2. Каленов Н.Е., Савин Г.И., Сотников А.Н. Электронная библиотека «Научное наследие России» как интегратор научной информации // Информационные системы и процессы: сб. науч. тр. 2016. Т. 15. С. 21–29.

3. Кириллов С.А., Соболевская И.Н., Сотников А.Н. Принципы формирования и представления междисциплинарных коллекций в цифровом пространстве научных знаний // Электронные библиотеки. 2021. Т. 24. № 2. С. 294–314. doi: 10.26907/1562-5419-2021-24-2-293-313.
4. Бездушный А.А., Бездушный А.Н., Серебряков В.А., Филиппов В.И. Интеграция метаданных Единого научного информационного пространства РАН. М.: изд-во ВЦ РАН, 2005. 238 с.
5. Каленов Н.Е., Погорелко К.П., Сотников А.Н. О развитии электронной библиотеки «Научное наследие России» как составляющей Единого цифрового пространства научных знаний // Информационные процессы. 2022. Т. 22. № 3. С. 155–166.
6. Lin J., Nogueira R., Yates A. Pretrained Transformers for Text Ranking: Bert and Beyond. Springer Cham Publ., 2022, 307 p. doi: 10.1007/978-3-031-02181-7.
7. Хьерланд Б. Информационный поиск и организация знаний: взгляд с точки зрения философии науки // Междунар. форум по информации. 2022. Т. 47. № 4. С. 36–57. doi: 10.36535/0203-6460-2022-04-3.
8. Robinson I., Webber J., Eifrem E. Graph Databases. O'Reilly Publ., CA, 2015, 220 p.
9. Погорелко К.П. Новая версия программного обеспечения электронной библиотеки «Научное наследие России» // Информационные ресурсы России. 2020. № 5. С. 27–29.
10. Каленов Н.Е., Сотников А.Н. Структура онтологии единого цифрового пространства научных знаний // Научно-техническая информация. Сер. 2. 2023. № 7. С. 20–26. doi: 10.36535/0548-0027-2023-07-3.

Software &amp; Systems

doi: 10.15827/0236-235X.148.524-531

2024, 37(4), pp. 524–531

### Search organization in a database with related entities

**Konstantin P. Pogorelko**<sup>1,2</sup>✉, **Gennady I. Savin**<sup>1,2</sup><sup>1</sup> Joint Supercomputer Center of RAS, Moscow, 119334, Russian Federation<sup>2</sup> National Research Centre “Kurchatov Institute”,  
Moscow, 123182, Russian Federation**For citation**Pogorelko, K.P., Savin, G.I. (2024) ‘Search organization in a database with related entities’, *Software & Systems*, 37(4), pp. 524–531 (in Russ.). doi: 10.15827/0236-235X.148.524-531**Article info**

Received: 24.07.2024

After revision: 22.08.2024

Accepted: 30.08.2024

**Abstract.** There is the problem of search organization in a database with related entities, in which it is possible to formulate conditions not only to the attributes of a certain entity, but also to set conditions on related entities. The paper describes the realized search mechanism. On the basis of this mechanism the author organized the search in the information system of the digital library “Scientific Heritage of Russia”. The search mechanism provides the work in the Russian and English languages. The author customize the search engine to the structure of a particular database by filling in a number of tables. The search engine allows working with such types of attributes as text field, numeric value, boolean value, full-text index of text field, codifiers. There are various functionalities for administrators and regular users. To formulate a query to the database, the author designed an intuitive interactive search query builder. It allows you to set conditions for the initially selected entity, related entities, their number and types of connection in the dialog mode. There is an extended functionality for working with codifiers of large volume. The author described the html markup organization of the input form and the data structure formed during query building. He gave an algorithmic solution that allows building an SQL instruction to the database according to the query formulated in the builder. He also gave an example of a query to related entities and the SQL instruction built for this query. The author described the organization of the search results visualization system. The proposed algorithms for building a search mechanism can serve as a model for organizing a search in the Unified Digital Space of Scientific Knowledge, including heterogeneous named relationships between objects.

**Keywords:** information search, document search, database, automated system, related entities, graph systems**Acknowledgements.** The paper was carried out under the government assignment, project no. FNEF-2024-0014**References**

1. Kalenov, N.E., Savin, G.I., Sotnikov, A.N. (2009) “‘Russian scientific heritage’ digital library as an integrator of scientific information’, *Information Resources of Russia*, (2), pp. 19–20 (in Russ.).
2. Kalenov, N.E., Savin, G.I., Sotnikov, A.N. (2016) “‘Russian scientific heritage’ digital library as an integrator of scientific information’, *Proc. Information Systems and Processes*, 15, pp. 21–29 (in Russ.).

3. Kirillov, S.A., Sobolevskaya, I.N., Sotnikov, A.N. (2021) 'Some aspects of the formation and representation principle of interdisciplinary collection in the digital space of scientific knowledge', *Russian Digital Libraries J.*, 24(2), pp. 294–314 (in Russ.). doi: 10.26907/1562-5419-2021-24-2-293-313.
4. Bezdushnyy, A.A., Bezdushnyy, A.N., Serebryakov, V.A., Filippov, V.I. (2005) *Metadata Integration of the Unified Scientific Information Space of RAS*. Moscow, 238 p. (in Russ.).
5. Kalenov, N.E., Pogorelko, K.P., Sotnikov, A.N. (2022) 'On the evolution of the digital library "Scientific heritage of Russia" as a component of the common digital space of scientific knowledge', *Information Processes*, 22(3) pp 155–166 (in Russ.).
6. Lin, J., Nogueira, R., Yates, A. (2022) *Pretrained Transformers for Text Ranking: Bert and Beyond*. Springer Cham Publ., 307 p. doi: 10.1007/978-3-031-02181-7.
7. Hjørland, B. (2022) 'Information retrieval and knowledge organization: a view from the point of view of the philosophy of science', *International Information Forum*, 47(4), pp. 36–57 (in Russ.). doi: 10.36535/0203-6460-2022-04-3.
8. Robinson, I., Webber, J., Eifrem, E. (2015) *Graph Databases*. O'Reilly Publ., CA, 220 p.
9. Pogorelko, K.P. (2020) 'A new version of the software for the electronic library "scientific heritage of Russia"', *Information Resources of Russia*, (5), pp. 27–29 (in Russ.).
10. Kalenov, N.E., Sotnikov, A.N. (2023) 'On the structure of the Common Digital Space of scientific knowledge ontology', *Sci. and Tech. Information. Ser. 2*, (7), pp. 20–26 (in Russ.). doi: 10.36535/0548-0027-2023-07-3.

**Авторы****Погорелко Константин Павлович**<sup>1,2</sup>,

к.т.н., старший научный сотрудник,

konstpog@yandex.ru

**Савин Геннадий Иванович**<sup>1,2</sup>,

д.ф.-м.н., профессор, академик РАН,

руководитель научного направления,

savin@jscc.ru

**Authors****Konstantin P. Pogorelko**<sup>1,2</sup>,

Cand. of Sci. (Engineering),

Senior Researcher, konstpog@yandex.ru

**Gennady I. Savin**<sup>1,2</sup>,

Dr.Sci. (Physics and Mathematics),

Professor, Academician of RAS,

Head of the Research Area, savin@jscc.ru

<sup>1</sup> Межведомственный суперкомпьютерный центр РАН, г. Москва, 119334, Россия<sup>2</sup> Национальный исследовательский центр

«Курчатовский институт», г. Москва, 123182, Россия

<sup>1</sup> Joint Supercomputer Center of RAS,

Moscow, 119334, Russian Federation

<sup>2</sup> National Research Centre "Kurchatov Institute",

Moscow, 123182, Russian Federation

## Особенности функционального моделирования сложных иерархических систем при различных способах организации управления исследуемыми процессами с целью выбора рациональной структуры системы

Н.В. Долгов<sup>1</sup>, В.А. Ильин<sup>1</sup>✉

<sup>1</sup> Военный институт дополнительного профессионального образования ВУНЦ ВМФ «Военно-морская академия», г. Санкт-Петербург, 195112, Россия

### Ссылка для цитирования

Долгов Н.В., Ильин В.А. Особенности функционального моделирования сложных иерархических систем при различных способах организации управления исследуемыми процессами с целью выбора рациональной структуры системы // Программные продукты и системы. 2024. Т. 37. № 4. С. 532–538. doi: 10.15827/0236-235X.148.532-538

### Информация о статье

Группа специальностей ВАК: 2.3.7

Поступила в редакцию: 03.05.2024

После доработки: 27.06.2024

Принята к публикации: 05.08.2024

**Аннотация.** В представленном исследовании выполнен общий анализ функционального моделирования сложных иерархически организованных систем. В таких системах отдельные процессы разных уровней иерархии взаимодействуют друг с другом непосредственно, кроме функции управления соответствующей иерархии при различных способах организации управления. В статье рассмотрены сложные системы двухуровневой иерархии управления при централизованном, децентрализованном и сетевом способах управления системами нижнего уровня иерархии. Также проанализированы несколько вариантов организации построения абстрактных иерархических процессов, представленных в нотации технологии IDEF0, и управления этими процессами. Данные процессы являются функциональными моделями сложных систем при различных способах управления. В ходе интерпретации функциональных моделей сложных систем при различных способах управления сформированы матрицы связей функций моделируемых процессов. Предложены математические зависимости для определения эффективности процесса функциональных моделей сложных систем при различных способах управления. Выполнен анализ функциональных моделей сложных систем и выявлена наиболее эффективная из них. Также доказано, что возрастание числа процедурных функций функциональных моделей ведет к уменьшению эффективности процессов, а увеличение прямых и обратных функциональных связей функциональных моделей – к ее увеличению.

**Ключевые слова:** функциональная модель, сложные иерархические системы, способы управления, эффективность функциональной модели, анализ, иерархия управления, технологии IDEF0

**Введение.** Постоянное усложнение систем, искусственных организационно-технических систем в первую очередь, создаваемых человеком в разных областях деятельности, привело не только к осознанию необходимости предвидеть результаты длительных и дорогостоящих работ по созданию таких систем, но и к поиску инструментария для этого. Поиск шел в направлении наработки способности увидеть и оценить функционирование разрабатываемой системы раньше, чем она будет создана физически.

Целью создания любой системы является ее функционирование. Характер функционирования системы в пределах ее морфологических возможностей определяется управлением работой системы. Именно особенности функционирования системы определяют и ее морфологию, и ее материально-информационные потоки.

Известно, что функциональная модель есть графическое представление функционирования систем и объектов [1]. Функциональная модель может использоваться как для анализа процессов протекающих в существующих си-

стемах, так и при проектировании новых систем для синтеза их структур [2]. Следует отметить, что анализ функциональных моделей в настоящее время выполняется на качественном уровне путем анализа структуры связей между отдельными функциями моделируемого процесса, сравнения различных структур процесса и процессов между собой. Незначительное количество имеющихся источников и документов по методологии функционального моделирования не приводят методов аналитического количественного анализа функциональных моделей и процессов ими описываемых, а приведенные источники относятся к базовым, что, в свою очередь, подчеркивает актуальность статьи, ее новизну и практическую значимость.

Функциональную модель, представленную в нотации технологии IDEF0 [2], можно рассматривать как оргграф и, следовательно, анализировать с помощью матрицы смежностей функций процесса между собой [3].

Эффективность процесса ( $E_n$ ) определяется как среднее арифметическое четырех коэффициентов: масштабного коэффициента процес-

са ( $K_m$ ), оценивающего количество функций в процессе; коэффициента связанности функций процесса ( $K_c$ ), оценивающего коммуникативность функций в процессе; коэффициента равномерности распределения функций в процессе ( $K_p$ ); коэффициента изолированности функций, определяющего общее число несвязанных функций, функций, не имеющих входов или выходов ( $K_u$ ) [3].

Сложные иерархические системы имеют многоуровневое взаимодействие между ее элементами. Система более высокого уровня выдвигает цель системе более низкого уровня, которая, в свою очередь, выступает как средство для реализации этой цели. Для функционирования системы эти цели и средства их реализации должны быть согласованы между собой [4].

Рассмотрим способы управления сложными системами [5]:

*Централизованный* способ управления обеспечивает использование всех функций и связей между отдельными функциями моделируемого процесса, при использовании единой функции управления на всех уровнях иерархии.

*Децентрализованное* управление предусматривает автономные действия отдельных функций без единого управления.

*Сетецентрическое управление* представляет собой матричный информационно-управляющий процесс, в основе которого лежит глобальная информационная взаимосвязь всех функций, интегрированных в единое информационное пространство.

Для сетецентрических систем характерна не только вертикальная интеграция между ее процессами, но и широко развитая сеть горизонтальных связей на одном и том же уровне управления между разнородными процессами системы, которые являются источниками и потребителями циркулирующей в системе информации.

Такие системы характеризуются принципами открытости, самоорганизации, слабой иерархии в контуре принятия решений и способностью порождать цели внутри себя [1].

При анализе сложных систем управления промоделируем каждый из способов более подробно. Моделирование процессов сложных систем выполним с использованием метода функционального моделирования [6–8].

Процесс действий сложных систем при *централизованном способе управления* представлен на рисунке 1 в нотации технологии IDEF0. Процесс включает в себя шесть процедурных функций (A1...A6), выполняющих, например,

обработку информации, и одну специфическую функцию управления процессом (A0), связанные между собой прямыми и обратными функциональными связями. Процесс имеет два входа из внешней среды (может быть из другого процесса – Out1, Out2) и два выхода во внешнюю среду (во внешние процессы – In1, In2). Как видим, входы из внешней среды осуществляется через функцию A1, A4, обеспечивающую обработку входной информации и ее передачу другим функциям. Выходы во внешнюю среду осуществляется функциями A3, A6, которые, решают основные задачи процесса, остальные функции выполняют обеспечивающие задачи [9]. Следует заметить, что количество входов и выходов в общем случае может быть любым и определяется только структурой исследуемого процесса [3].

Интерпретируя функциональную модель как оргграф, сформируем матрицу связей функций моделируемого процесса (табл. 1). В матрице по строкам единицами и нулями записаны входы в функции, по столбцам соответственно получаем выходы из функций. Суммы входов по строкам и выходов по столбцам позволяют ранжировать функции по степени важности в исследуемом процессе. Общая сумма входов и выходов дает возможность сравнивать структуры одного процесса и разные процессы по их сложности [3].

Таблица 1

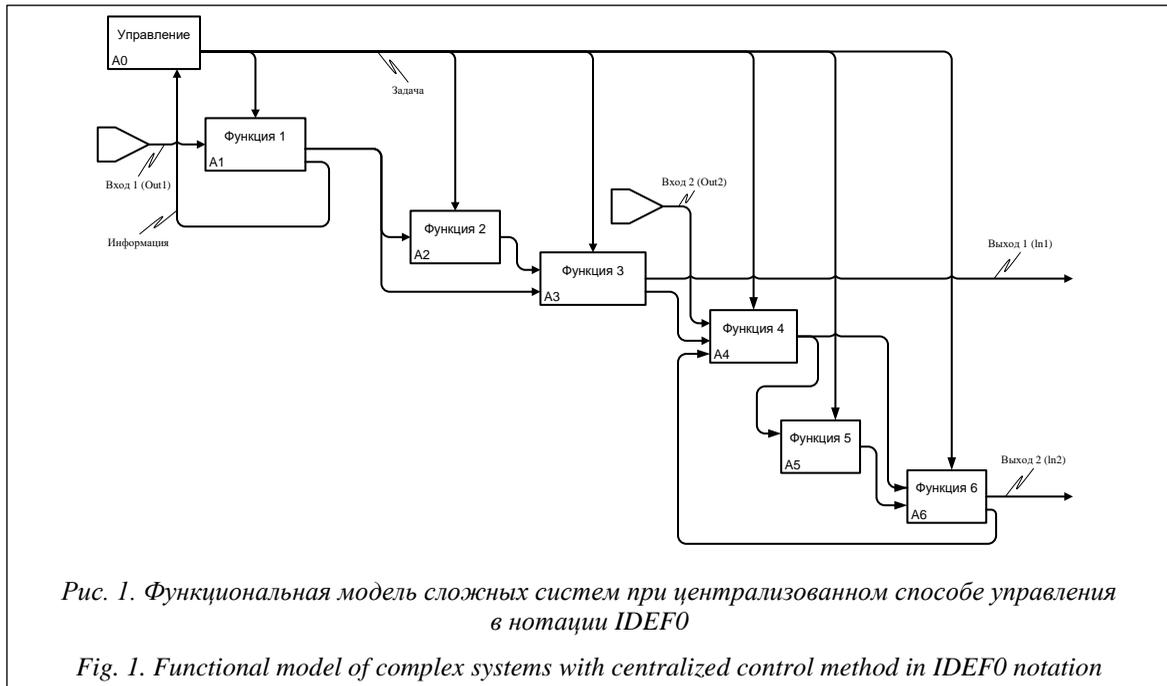
**Матрица связей функциональной модели действий сложных систем при централизованном способе управления**

Table 1

**Relationship matrix of functional model of complex systems actions under centralized control method**

	Out1-2	A0	A1	A2	A3	A4	A5	A6	In1-2	$\Sigma in$
Out1	0	0	1	0	0	0	0	0	0	1
Out2	0	0	0	0	0	1	0	0	0	1
A0	0	0	1	1	1	1	1	1	0	6
A1	0	1	0	1	1	0	0	0	0	3
A2	0	0	0	0	1	0	0	0	0	1
A3	0	0	0	0	0	1	0	0	0	1
A4	0	0	0	0	0	0	1	1	0	2
A5	0	0	0	0	0	0	0	1	0	1
A6	0	0	0	0	0	1	0	0	0	1
In1	0	0	0	0	1	0	0	0	0	1
In2	0	0	0	0	0	0	0	1	0	1
$\Sigma out$	0	1	2	2	4	4	2	4	0	19

По матрице связей выполним расчет коэффициентов и эффективности процесса.



Масштабный коэффициент процесса. Общее количество функций  $f_{\Sigma} = 6$ , тогда

$$K_m = 1 - e^{-\left(1 - \frac{2}{f_{\Sigma}}\right)} = 1 - e^{-\left(1 - \frac{2}{6}\right)} = 1 - e^{-0,6666} = 0,4865.$$

Коэффициент связанности функций процесса.

Общее число связей, то есть число единиц в матрице  $S_{\Sigma} = 19$ , тогда

$$K_c = e^{-\left[1 - \frac{S_{\Sigma}}{f_{\Sigma}(f_{\Sigma}-1)}\right]} = e^{-\left[1 - \frac{19}{6(6-1)}\right]} = e^{-0,3666} = 0,693.$$

Коэффициент равномерности распределения функций в процессе.

Среднее число входных связей  $\overline{in} = 8$ ; среднее число выходных связей  $\overline{out} = 12$ :

$$K_p = e^{-\frac{\sum |in-\overline{in}| + \sum |out-\overline{out}|}{2S_{\Sigma}}} = e^{-\frac{|19-8| + |19-12|}{2*19}} = e^{-0,1052} = 0,9.$$

Коэффициент изолированности функций.

Число пустых строк  $n_{стр} = 0$ , число пустых столбцов  $n_{стол} = 4$ , тогда

$$K_u = e^{-\frac{n_{стр} + n_{стол}}{2f_{\Sigma}}} = e^{-\frac{0+4}{2*6}} = e^{-0,3333} = 0,7165.$$

Эффективность процесса:

$$\mathcal{E}_n = \frac{K_m + K_c + K_p + K_u}{4} = \frac{0,4865 + 0,693 + 0,9 + 0,7165}{4} = 0,699.$$

Процесс действий сложных систем при децентрализованном способе управления пред-

ставлен на рисунке 2 в нотации технологии IDEF0. Он включает в себя шесть процедурных функций (A1...A6), выполняющих обработку информации, связанных между собой прямыми и обратными функциональными связями. Остальные действия выполняются так же, как и при централизованном способе управления.

Интерпретируя функциональную модель как оргграф, сформируем матрицу связей функций моделируемого процесса (табл. 2).

Таблица 2

Матрица связей функциональной модели действий сложных систем при децентрализованном способе управления

Table 2

Relationship matrix of the complex systems' functional model of actions under decentralized control method

	Out1-2	A1	A2	A3	A4	A5	A6	In1-2	$\Sigma in$
Out1	0	1	0	0	0	0	0	0	1
Out2	0	0	0	0	1	0	0	0	1
A1	0	0	1	1	0	0	0	0	2
A2	0	0	0	1	0	0	0	0	1
A3	0	0	0	0	1	0	0	0	1
A4	0	0	0	0	0	1	1	0	2
A5	0	0	0	0	0	0	1	0	1
A6	0	0	0	0	1	0	0	0	1
In1	0	0	0	1	0	0	0	0	1
In2	0	0	0	0	0	0	1	0	1
$\Sigma out$	0	1	1	3	3	1	3	0	12

По матрице связей (табл. 2) выполним расчет коэффициентов и эффективности процесса.

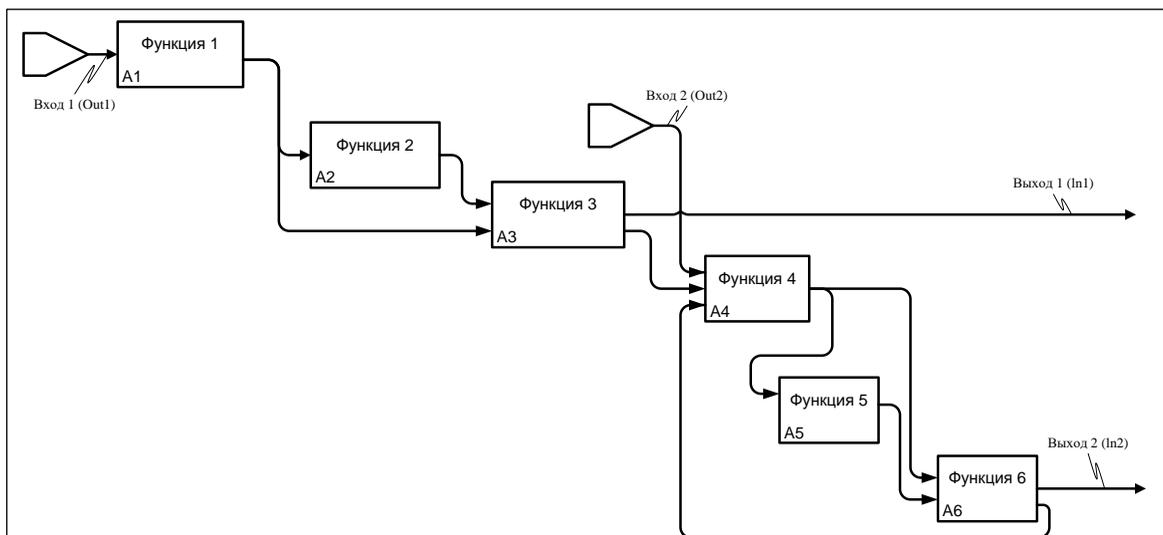


Рис. 2. Функциональная модель сложных систем при децентрализованном способе управления в нотации IDEF0

Fig. 2. Functional model of complex systems with decentralized control method in IDEF0 notation

Масштабный коэффициент процесса.

Общее количество функций  $f_{\Sigma} = 6$ , тогда

$$K_m = 1 - e^{-\left(1 - \frac{2}{f_{\Sigma}}\right)} = 1 - e^{-\left(1 - \frac{2}{6}\right)} = 1 - e^{-0.6666} = 0,4865.$$

Коэффициент связанности функций процесса.

Общее число связей, то есть число единиц в матрице  $S_{\Sigma} = 12$ , тогда

$$K_c = e^{-\left[1 - \frac{S_{\Sigma}}{f_{\Sigma}(f_{\Sigma}-1)}\right]} = e^{-\left[1 - \frac{12}{6(6-1)}\right]} = e^{-0.6} = 0,5488.$$

Коэффициент равномерности распределения функций в процессе.

Среднее число входных связей  $\bar{in} = 2$ ; среднее число выходных связей  $\bar{out} = 6$ :

$$K_p = e^{-\frac{\sum_f |\Sigma in - \bar{in}| - \sum_f |\Sigma out - \bar{out}|}{2 \cdot S_{\Sigma}}} = e^{-\frac{|12-2| - |12-6|}{2 \cdot 12}} = e^{-0.1666} = 0,8464.$$

Коэффициент изолированности функций.

Число пустых строк  $n_{стр} = 0$ , число пустых столбцов  $n_{стол} = 4$ , тогда

$$K_u = e^{-\frac{n_{стр} + n_{стол}}{2 \cdot f_{\Sigma}}} = e^{-\frac{0+4}{2 \cdot 6}} = e^{-0.3333} = 0,7165.$$

Эффективность процесса:

$$\mathcal{E}_n = \frac{K_m + K_c + K_p + K_u}{4} = \frac{0,4865 + 0,5488 + 0,8464 + 0,7165}{4} = 0,6496.$$

Процесс действий сложных систем при сетцентрическом способе управления представлен на рисунке 3 в нотации технологии IDEF0.

Процесс включает в себя шесть процедурных функций (A1...A6), выполняющих, например, обработку информации, связанных между собой прямыми и обратными функциональными связями. Остальные действия выполняются так же, как и при централизованном способе управления.

Интерпретируя функциональную модель как оргграф, сформируем матрицу связей функций моделируемого процесса, представленную в таблице 3.

Таблица 3

Матрица связей функциональной модели действий сложных систем при сетцентрическом способе управления  
Table 3  
Relationship matrix of the complex systems' functional model of actions under the network-centric management method

	Out1-2	A1	A2	A3	A4	A5	A6	In1-2	$\Sigma in$
Out1	0	1	0	0	0	0	0	0	1
Out2	0	0	0	0	1	0	0	0	1
A1	0	0	1	1	0	0	0	0	2
A2	0	1	0	1	0	0	0	0	2
A3	0	1	1	0	1	0	0	0	3
A4	0	1	0	0	0	1	1	0	3
A5	0	1	0	0	1	0	1	0	3
A6	0	1	0	0	1	1	0	0	3
In1	0	0	0	1	0	0	0	0	1
In2	0	0	0	0	0	0	1	0	1
$\Sigma out$	0	6	2	3	4	2	3	0	20

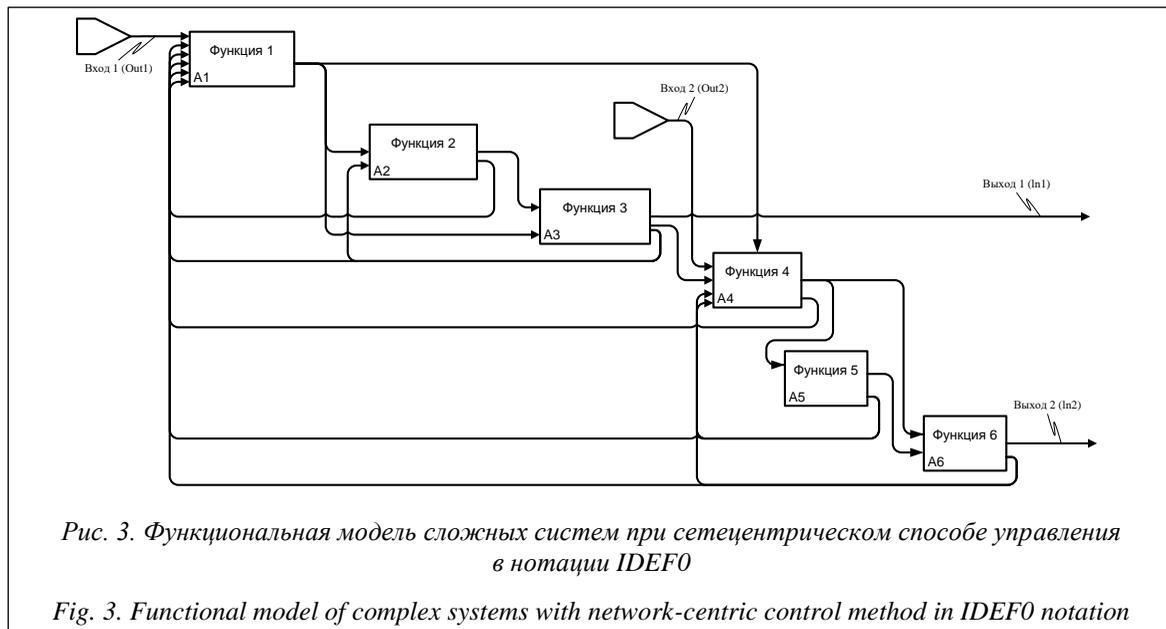


Рис. 3. Функциональная модель сложных систем при сетцецентрическом способе управления в нотации IDEF0

Fig. 3. Functional model of complex systems with network-centric control method in IDEF0 notation

По матрице связей (табл. 3) выполним расчет коэффициентов и эффективности процесса.

*Масштабный коэффициент процесса.*  
Общее количество функций  $f_{\Sigma} = 6$ , тогда

$$K_m = 1 - e^{-\left(1 - \frac{2}{f_{\Sigma}}\right)} = 1 - e^{-\left(1 - \frac{2}{6}\right)} = 1 - e^{-0,6666} = 0,4865.$$

*Коэффициент связанности функций процесса.*

Общее число связей, то есть число единиц в матрице  $S_{\Sigma} = 20$ , тогда

$$K_c = e^{-\left[1 - \frac{S_{\Sigma}}{f_{\Sigma}(f_{\Sigma}-1)}\right]} = e^{-\left[1 - \frac{20}{6(6-1)}\right]} = e^{-0,3333} = 0,7165.$$

*Коэффициент равномерности распределения функций в процессе.*

Среднее число входных связей  $\overline{in} = 10$ ; среднее число выходных связей  $\overline{out} = 14$ :

$$K_p = e^{-\frac{\sum_f |\sum in - \overline{in}| + \sum_f |\sum out - \overline{out}|}{2S_{\Sigma}}} = e^{-\frac{|20-10| + |20-14|}{2 \cdot 20}} = e^{-0,1} = 0,9048.$$

*Коэффициент изолированности функций.*

Число пустых строк  $n_{стр} = 0$ , число пустых столбцов  $n_{стол} = 4$ , тогда

$$K_u = e^{-\frac{n_{стр} + n_{стол}}{2f_{\Sigma}}} = e^{-\frac{0+4}{2 \cdot 6}} = e^{-0,3333} = 0,7165.$$

*Эффективность процесса:*

$$\mathcal{E}_n = \frac{K_m + K_c + K_p + K_u}{4} = \frac{0,4865 + 0,7165 + 0,9048 + 0,7165}{4} = 0,7061.$$

Расчет эффективности процессов функциональных моделей действий сложных систем в данном примере при различных способах управления при одинаковом числе функций позволяет сделать вывод, что наиболее эффективным является сетцецентрический способ управления.

### Заключение

В статье выполнен общий анализ функционального моделирования сложных систем при различных способах управления и выполнен расчет эффективности способов.

Предложенный метод позволяет количественно оценить эффективность систем и выбрать наиболее рациональную организацию процессов управления и функционирования сложных систем.

Также доказано, что возрастание числа процедурных функций функциональных моделей ведет к уменьшению эффективности процессов, а увеличение прямых и обратных функциональных связей функциональных моделей способствует ее увеличению.

### Список литературы

1. Черемных С.В., Семенов И.О., Ручкин В.С. Моделирование и анализ систем. IDEF-технологии. М.: Финансы и статистика, 2006. 192 с.
2. Дубейковский В.И. Практика функционального моделирования с AllFusion Process Modeler. М.: Диалог-МИФИ, 2004. 464 с.

3. Ильин В.А., Янча С.П. Методы анализа функциональных моделей // Программные продукты и системы. 2009. Т. 22. № 4. С. 32–34.
4. Шпак В.Ф., Директоров Н.Ф., Мирошников В.И. и др. Информационные технологии в системе управления силами ВМФ. СПб: Элмор, 2005. 827 с.
5. Ильин В.А., Козлов И.Л. К проблеме автоматизации управления силами и средствами, функциональное моделирование // Автоматизация процессов управления. 2009. № 4. С. 53–57.
6. Тебекин А. В. Методология функционального моделирования сложных технических систем модульного типа // Журнал технических исследований. 2021. Т. 7. № 2. С. 3–12.
7. Тебекин А.В., Тебекин П.А., Егорова А.А. Технологические трансформации XXI века как индуцирующий вектор перехода к новому качеству производства // Теоретическая экономика. 2021. № 1. С. 42–53.
8. Тебекин А.В., Петров В.С. Использование методологии моделирования IDEF при формировании структурно-параметрической модели реализации технологий обеспечения эффективного развития промышленных предприятий в условиях постиндустриальной экономики // Транспортное дело России. 2017. № 4. С. 43–49.
9. Харари Ф. Теория графов; [пер. с англ.]. М.: Книжный дом «Либроком», 2009. 296 с.

Software &amp; Systems

doi: 10.15827/0236-235X.148.532-538

2024, 37(4), pp. 532–538

### Features of functional modeling of complex hierarchical systems with different ways of organizing the control of the study processes for the purpose of selecting a rational system structure

Nikolay V. Dolgov<sup>1</sup>, Viktor A. Ilin<sup>1</sup> ✉

<sup>1</sup> Military Institute of Continuing Professional Education of the VUNTS Navy "Naval Academy", St. Petersburg, 195112, Russian Federation

#### For citation

Dolgov, N.V., Ilin, V.A. (2024) 'Features of functional modeling of complex hierarchical systems with different ways of organizing the control of the study processes for the purpose of selecting a rational system structure', *Software & Systems*, 37(4), pp. 532–538 (in Russ.). doi: 10.15827/0236-235X.148.532-538

#### Article info

Received: 03.05.2024

After revision: 27.06.2024

Accepted: 05.08.2024

**Abstract.** The paper presents a general analysis of functional modeling of complex hierarchically organized systems. In such systems, individual processes of hierarchy levels interact with each other directly, except for the control function of the corresponding hierarchy at different ways of control organization. The paper considers complex systems of two-level hierarchy of control at centralized, decentralized and network-centric ways of controlling the lower level systems of the hierarchy. The authors also analyzed several options for organizing the construction of abstract hierarchical processes represented in the notation of IDEF0 technology and management of these processes. These processes are functional models of complex systems at different methods of control. During the interpretation of complex systems functional models at different methods of control, the authors formed the function relationship matrices of the modeled processes. The authors proposed mathematical dependencies for determining the process efficiency of complex systems functional models under different methods of control. They analyzed the functional models of complex systems and identified the most efficient one. They also proved that increasing the number of procedural functions of functional models leads to a decrease in process efficiency. However, increasing the forward and backward functional relationships of functional models leads to an increase in the efficiency of processes.

**Keywords** functional model, complex hierarchically systems, ways of control, efficiency of a functional model, analysis, control hierarchy, IDEF0 technologies

#### References

1. Cheremnykh, S.V., Semenov, I.O., Ruchkin, V.S. (2006) *Modeling and Analysis of Systems. IDEF Technologies*. Moscow, 191 p. (in Russ.).
2. Dubeykovskiy, V.I. (2004) *Practice Functional Modeling with AllFusion Process Modeler*. Moscow, 464 p. (in Russ.).
3. Ilin, V.A., Yancha, S.P. (2009) 'Methods analysis of functional models', *Software & Systems*, 22(4), pp. 32–34 (in Russ.).
4. Shpak, V.F., Direktorov, N.F., Miroshnikov, V.I. (2005) *Information Technologies in the Navy Force Management System*. St. Petersburg, 827 p. (in Russ.).

5. Пыин V.A., Kozlov I.L. 2009. On the problem of automation of control of forces and means, functional modeling', *Automation of Control Processes*, 4(18), pp. 53–57.
6. Tebekin, A.V. (2021) 'Methodology for functional modeling of complex technical systems of modular type', *J. of Tech. Research*, 7(2), pp. 3–12 (in Russ.).
7. Tebekin, A.V., Tebekin, P.A., Egorova, A.A. (2021) 'Technological transformations of the 21st century as an inducing vector of transition to a new quality of production', *Theoretical Economics*, (1), pp. 42–53 (in Russ.).
8. Tebekin, A.V., Petrov, V.S. (2017) 'Using the modeling methodology IDEF in the formation of structural-parametric model for the implementation of technology ensure the efficient development of industrial enterprises in the condition of postindustrial economy', *Transport Business of Russia*, (4), pp. 43–49 (in Russ.).
9. Harari, F. (1972) *Graph Theory*. Basic Books Publ., 274 p. (Russ. ed.: (2009) Moscow, 296 p.).

**Авторы**

**Долгов Николай Валерьевич**<sup>1</sup>,

адъюнкт, d\_n\_v92@mail.ru

**Ильин Виктор Алексеевич**<sup>1</sup>, д.в.н.,

профессор, профессор кафедры,

vikil.45@yandex.ru

**Authors**

**Nikolay V. Dolgov**<sup>1</sup>, Adjunct,

d\_n\_v92@mail.ru

**Viktor A. Ilin**<sup>1</sup>, Dr.Sci. (Military Sciences),

Professor, Professor of the Department,

vikil.45@yandex.ru

<sup>1</sup> Военный институт дополнительного профессионального образования ВУНЦ ВМФ «Военно-морская академия», г. Санкт-Петербург, 195112, Россия

<sup>1</sup> Military Institute of Continuing Professional Education of the VUNTS Navy "Naval Academy", St. Petersburg, 195112, Russian Federation

## Реализация темпоральной базы данных для интеллектуальных систем поддержки принятия решений реального времени на примере образовательного процесса вуза

А.П. Еремеев <sup>1</sup>✉, Н.А. Панявин <sup>1</sup>, М.А. Маренков <sup>1</sup>

<sup>1</sup> НИУ «МЭИ», г. Москва, 111250, Россия

### Ссылка для цитирования

Еремеев А.П., Панявин Н.А., Маренков М.А. Реализация темпоральной базы данных для интеллектуальных систем поддержки принятия решений реального времени на примере образовательного процесса вуза // Программные продукты и системы. 2024. Т. 37. № 4. С. 539–546. doi: 10.15827/0236-235X.148.539-546

### Информация о статье

Группа специальностей ВАК: 2.3.5, 2.3.8

Поступила в редакцию: 13.05.2024

После доработки: 15.06.2024

Принята к публикации: 28.06.2024

**Аннотация.** Целью данной работы является исследование и разработка темпоральной базы данных и на ее основе темпоральной базы знаний для современных интеллектуальных систем реального времени. В качестве предметной области, требующей интеллектуального анализа и обработки динамических (темпоральных) данных, рассматривается образовательный процесс вуза. Прототип предназначен для контроля успеваемости и помощи студентам вуза в проблемных ситуациях в процессе обучения. Для реализации темпоральной модели представления информации и на ее основе темпоральных баз данных и знаний в работе предлагается использование нереляционной (NoSQL) организации базы данных. Эта база данных более всего подходит для оперирования пространственными динамическими данными. Для их обработки классические реляционные модели и реляционные (SQL) базы данных плохо применимы. Предложена общая схема прототипа интеллектуальной системы поддержки принятия решений реального времени с темпоральной базой данных и знаний. Темпоральная база данных организована на основе графовой модели и дополнена системой поиска решений на основе интервальной логики Аллена и метрической логики. В качестве реальных данных для организации темпоральной базы данных в прототипе были использованы данные 2 000 студентов НИУ «МЭИ» по ряду профильных дисциплин, полученные из специализированных информационных систем университета за последние несколько лет. Программная реализация выполнена с применением языка программирования Python v.3.10, библиотеки для распределенных и параллельных вычислений Dask и нереляционной графовой СУБД Neo4j.

**Ключевые слова:** темпоральная база данных, искусственный интеллект, система поддержки принятия решений, реальное время, классификация, база знаний, образовательный процесс

**Благодарности.** Работа выполнена при финансовой поддержке РФФИ, проект № 24-11-00285, <https://rscf.ru/project/24-11-00285/>

**Введение.** При интеллектуальном анализе динамических данных и разработке соответствующей темпоральной БД необходимо учитывать специфику предметной области, в данном случае специфику образовательного процесса. Так, например, в зарубежных вузах эта специфика представлена в работах [1–3], а в отечественных – в [4, 5]. Необходимо применение интеллектуальных методов анализа для выявления и извлечения паттернов в исходных данных обучающихся. Решение поставленной задачи путем формирования и применения традиционных select-запросов при использовании реляционных БД неэффективно из-за возможных скрытых данных. В образовательном процессе посредством анализа данных определяются ключевые характеристики и скрытые данные, что позволяет проводить более качественную обработку имеющейся информации и прогнозирование последствий принимаемых решений, в частности, прогноз успеваемости обучающихся в реальном времени и на заданный

временной интервал (например, до ближайшей контрольной недели или зачетной сессии). Необходимо также своевременное выявление проблемных ситуаций, связанных, например, со студентами, которые в настоящий момент по различным причинам находятся в группе риска и могут показать неудовлетворительные результаты при ближайшем контрольном мероприятии. Также в функции интеллектуальной системы (ИС) поддержки принятия решений реального времени (ИСППР РВ) входит оперативное оповещение соответствующего преподавателя или руководства университета (ЛПР) о проблемной ситуации и помощь в ее разрешении [6–8]. Для качественного выполнения перечисленных функций и поддержки образовательного процесса необходима реализация темпоральных моделей и на их основе темпоральных БД и БЗ для современных ИС/ИСППР РВ. Данной проблематике и посвящена статья. К сожалению, к настоящему времени авторам неизвестны коммерческие про-

граммные продукты, поддерживающие темпоральные БД, а в свободном доступе есть только темпоральный язык запросов TSQL2.

### Темпоральная модель данных

Классическая модель данных характеризуется такими показателями, как структура представления данных, целостность, операции с хранимыми данными. Темпоральное расширение модели предполагает внедрение средств обработки темпоральных зависимостей в виде темпоральной алгебры или логики (например, известной и активно используемой в ИС РВ интервальной логики Аллена [9]), которые могут поддерживаться на уровнях модели в целом, ограничений целостности, ключей и зависимостей, запросов.

В темпоральных моделях выделяют три основных типа времени:

- *модельное* – действительное время, представляющее собой метку или промежуток времени актуальности данных с точки зрения моделируемого мира;
- *транзакционное* – время регистрации события с точки зрения времени в системе;
- *пользовательское* – временная метка, представляющая момент или моменты времени, которые пользователь намерен хранить в атрибуте сущности.

Для поиска решения на основе темпоральной модели и соответствующей темпоральной БД предлагается применить метод анализа иерархий (*Analytical Hierarchy Process*, АНР), позволяющий представить сложные многофакторные или многокритериальные задачи в виде иерархических структур [10–12]. Иерархия в виде дерева определяется посредством декомпозиции исходной сложной проблемы посредством многоуровневой структуры, где первым уровнем является цель, вниз по иерархии следуют уровни подцелей и так далее до последнего уровня с рекомендуемыми решениями. Применение этого подхода к процессам высшего образования рассмотрено, например, в работах [12, 13].

К достоинствам метода АНР можно отнести:

- более удобную для анализа ЛПР при наличии многих критериев иерархическую структуру в виде дерева решений;
- наличие иерархии, что позволяет редуцировать сложные или неструктурированные задачи на более простые подзадачи, пока не будут получены так называемые элементарные задачи, решение которых известно или легко находится;

- допустимость в определенных пределах несоответствия различных альтернатив, выбранных ЛПР.

Основные этапы, входящие в процедуру АНР:

- 1) декомпозиция исходной задачи/проблемы и выявление критериев оценки;
- 2) установка приоритетов и попарная оценка результатов, когда процесс декомпозиции завершен и иерархия хорошо структурирована;
- 3) обобщение результатов и формирование общих приоритетов в иерархии;
- 4) проверка непротиворечивости полученных результатов;
- 5) принятие решения, заключающегося в выборе наилучшей альтернативы на основе заданных критериев.

При поиске решения на основе предложенной темпоральной графовой модели применена интеграция метода АНР и вывода средствами интервальной логики Аллена.

### Разработка темпоральной модели и БД для представления и обработки динамических данных

Рассмотрим построение темпоральной модели и на ее основе темпоральной БД для ИСПР РВ, призванных помочь в организации учебного процесса кафедры на примере подготовки специалистов по направлению «Прикладная математика и информатика». Образовательная программа «Математическое и программное обеспечение вычислительных машин и компьютерных сетей» включает ряд профильных дисциплин (математический анализ, теория вероятностей и математическая статистика, основы программирования, основы искусственного интеллекта и др.), направленных на формирование у специалиста набора компетенций [13, 14].

Каждая из дисциплин построена таким образом, чтобы студенты осваивали темы последовательно. Контрольные мероприятия, включая тестирование, организуются для проверки знаний и практических навыков, полученных при освоении разделов.

Процесс обучения и контроля может быть представлен и визуализирован в виде графических структур, например, диаграмм Хассе [15], сетей Петри, включая темпоральные сети Петри, ориентированные на моделирование динамических систем [16].

Для обеспечения высокой производительности при обработке многомерных графов, со-

держащих значительный объем динамических данных по обучающимся, была применена OLAP-структура и концепция нереляционной (NoSQL) БД. Это позволит оперировать пространственными данными, состоящими в разные моменты времени как из отдельных точек, линий, так и из целых областей – срезов, поверхностей или гиперкубов с более высокими размерностями, для обработки которых классические реляционные таблицы и реляционные (SQL) БД и СУБД плохо применимы.

В качестве реальных данных для организации темпоральной БД в прототипе ИСППР РВ были использованы обезличенные данные 2 000 студентов НИУ «МЭИ» по ряду профильных дисциплин, полученные из специализированных систем университета за последние несколько лет.

Рассмотрим следующую ситуацию. В соответствии с учебным планом студенту необходимо выполнить контрольные мероприятия к 4-й (KM1-2), 8-й (KM3), 12-й (KM4-6) и 16-й (KM7-8) учебным неделям. На рисунке 1 представлен пример фрагмента темпоральной модели, отображающей этап процесса обучения некоторого студента. Начальная вершина (START) соответствует началу процесса обучения по выбранной дисциплине, конечная вершина (END) – концу обучения по данной дисциплине. Зеленые вершины означают успешно пройденные студентом контрольные мероприятия учебного процесса, красные – несданные или еще недостигнутые мероприятия на текущий момент времени, оранжевые – сданные с опозданием. Связи между вершинами отражают темпоральные зависимости интервальной логики Аллена – «быть раньше» (Before) и «одновременно» (Equal).

Любое из контрольных мероприятий направлено на выработку определенных компетенций в рамках каждой из предметных областей.

Отслеживание состояния процесса обучения одного из студентов, а также группы студентов или потока в целом представляет собой достаточно сложную процедуру формирования огромных массивов динамических данных, содержащих информацию о контрольных мероприятиях, текущие и средние баллы обучаемых и проч. Темпоральная БД, организованная на основе нереляционной графовой СУБД Neo4J (<https://habr.com/ru/articles/650623/>) и дополненная системой темпорального вывода на базе интервальной логики Аллена и метрической логики [17, 18], позволяет существенно упростить процедуру. На рисунке 2 представлена схема прототипа ИСППР РВ со средствами обработки динамических данных.

### Реализация прототипа ИСППР РВ

На основе имеющихся данных о студентах и учебном процессе с применением языка программирования Python v.3.10, библиотеки для распределенных и параллельных вычислений Dask, и графовой СУБД Neo4j был разработан прототип ИСППР РВ с применением средств для обработки темпоральной информации. Выбор библиотеки Dask обусловлен возможностью разбиения данных (массивы, csv-таблицы и др.), которые не помещаются в оперативной памяти, на множество мелких частей и их параллельной обработкой. Рассмотрим основные компоненты прототипа (рис. 2.)

БД *общеуниверситетской системы* содержит информацию о статусе каждого студента (обучается, в академическом отпуске, отчис-

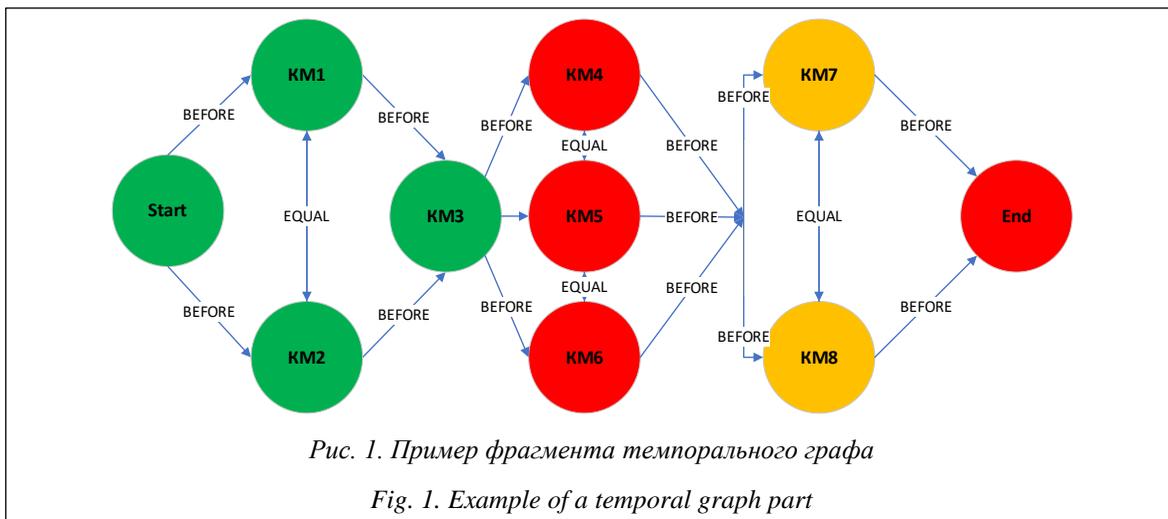
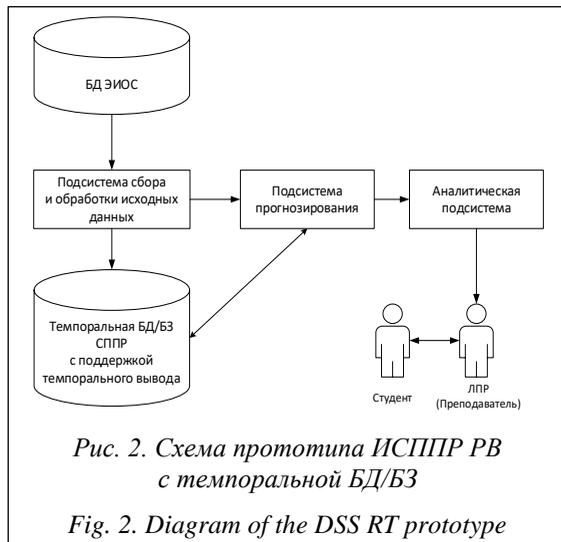


Рис. 1. Пример фрагмента темпорального графа

Fig. 1. Example of a temporal graph part



лен), его успеваемости по каждой дисциплине в соответствии с образовательной программой, а также его различные рейтинги, показывающие в реальном времени активность обучающегося в учебной и других областях.

Подсистема сбора и обработки исходных данных отвечает за автоматизацию фильтрации, сбора, предобработки (препроцессинга) исходных данных с целью представления их в формате, пригодном для дальнейшего применения с использованием средств темпоральной БД и темпорального вывода. Темпоральными данными являются планируемые сроки начала и окончания прохождения каждого контрольного мероприятия обучающимся, фактическое состояние обучаемого по каждой дисциплине на текущий момент и другие характеристики. Для организации темпорального вывода используются методы интервальной логики Аллена, согласования темпоральных ограничений, темпоральной модификации алгоритма К-ближайших соседей (*K-Nearest Neighbors*, KNN) и метод сравнения графовых структур Jaccard Graph Similarity (<https://habr.com/ru/articles/650623/>, [17]). Вывод осуществлялся посредством преобразования темпорального графа из текущего состояния обучаемого в планируемое.

База знаний содержит темпоральные правила. Для регистрации запросов, вызываемых при изменениях (при создании, обновлении, удалении данных) в темпоральной БД, созданной на основе СУБД Neo4j, использованы триггеры, которые могут выполняться до или после фиксации данных командой

```
Call apoc.trigger.add(<name>, <statement>, <selector>)
```

Для организации на основе темпоральной БД темпоральной БЗ, содержащей продукционные правила с заданной частотой их применения, используются команды запроса и отмены правил следующего вида:

пример запроса на правило:

```
Call apoc.periodic.repeat(<правило>, <частота повторений>),
```

пример отмены правила:

```
Call apoc.periodic.cancel(<правило>);
```

Чтобы отобразить все активные правила, необходимо ввести команду

```
Call apoc.periodic.list( );
```

Так как в день экзамена студент может сдавать только один экзамен, то в конце дня БД следует обновить и установить необходимые связи между узлами. Если студент сдает экзамен не с первого раза, то у обучаемого на темпоральном графе будет отмечено, что экзамен был сдан после определенной даты. Для этого в модель введено отношение `passed_exam_after`.

Пиведем пример темпорального правила «после», устанавливающего отсутствующие связи между узлами студентов и узлом экзамена 05 мая 2024 и считающего число восстановлений:

```
CALL apoc.periodic.commit ("
MATCH (s:Student)
WHERE s.passed_exam IS NOT NULL
AND NOT (s)-[passed_exam]->(e:Exam)
WITH s LIMIT {limit}
MATCH (e:Exam) WHERE s.passed_exam =
e.name
AND e.examdate < datetime('2024-05-
-05')
CREATE (s)-[:passed_exam_after]->(e)
RETURN count(*)
"", {limit:10000})"
```

Следует отметить, что расширенная версия языка запросов Cypher позволяет создавать темпоральные запросы на уровне целых узлов или связей. Так, темпоральный запрос на поиск всех студентов, получивших оценки после момента времени 05.05.2024, выглядит следующим образом:

```
MATCH (s:Student)
WHERE s.MarkDate > datetime('2024-
-05-05')
return s
```

Установлено, что средства Neo4j обеспечивают три базовых типа темпоральных запросов – текущие, последовательные и произвольные. В реализованной темпоральной модели был использован язык запросов TSQL2 [19], позволяющий задавать темпоральные характеристики

объектов на уровне отдельных кортежей; БЗ демонстрационного прототипа ИСППР РВ содержит правила, отражающие темпоральные зависимости между данными.

Представим алгоритм формирования темпорального графа  $G$  состояний объекта Граф  $G$  состоит из множества вершин, связанных темпоральными отношениями, где  $H_i = \{H_{i1}, H_{i2}, \dots, H_{in}\}$  –  $n$ -мерное описание состояния объекта в момент  $t$  регистрации события на  $i$ -м такте записи:

```
i=0
G←G∪Hi(t)
For k←1 to τ-1
    if Hk(t)≠Hi(t) then
        G←G∪Hi(t)
        k←i
    end if
end for
return G
```

Помимо классических продукционных правил «Если ...То ...» (в синтаксисе языка запросов Neo4j Cypher «Case When ... Then ...»), построенных с применением логических операций конъюнкции (And), дизъюнкции (Or) и отрицания (Not), средства Neo4j позволяют расширить базу правил с помощью введения правил логики предикатов первого порядка с применением кванторов существования (Exists) и общности (All), а также модальностей «для любого» (Any), «ни одного» (None). Это дает возможность реализовать темпоральный вывод на основе интервальной логики Аллена, а также вводить новые правила вывода. С помощью графовой модели также определяются темпоральные отношения на уровне модели, кроме того, она может быть расширена метрическими данными.

Средства СУБД Neo4j позволяют обрабатывать хранимую информацию в различных плоскостях-срезах. За счет распределенного хранения данных в отдельных файлах во время обработки данных в оперативную память подгружаются только необходимые файлы с нужной информацией. Пример темпорального запроса возвращает имена любых (ANY) студентов, имевших оценки «0» или «2» на момент 5 мая 2024 года:

```
MATCH (s:Student)
WHERE s.ExamDate = '2024-05-05' AND
ANY (x IN s.marks WHERE x = 0 OR x = 2)
RETURN s.name, s.marks
```

По итогам анализа и кластеризации графов состояний студентов, изучающих дисциплину

«Теория принятия решений», методами KNN, K-means, и Jaccard Graph Similarity были получены результаты, отраженные в таблице.

**Результаты кластеризации графов состояний**

**State graph clustering results**

Группа	Среднее значение балла выборки после классификации	
	Метод KNN	Метод K-means
Отличники	4,8341	4,7065
Хорошисты	4,4369	4,4137
Группа риска	2,0841	1,9836

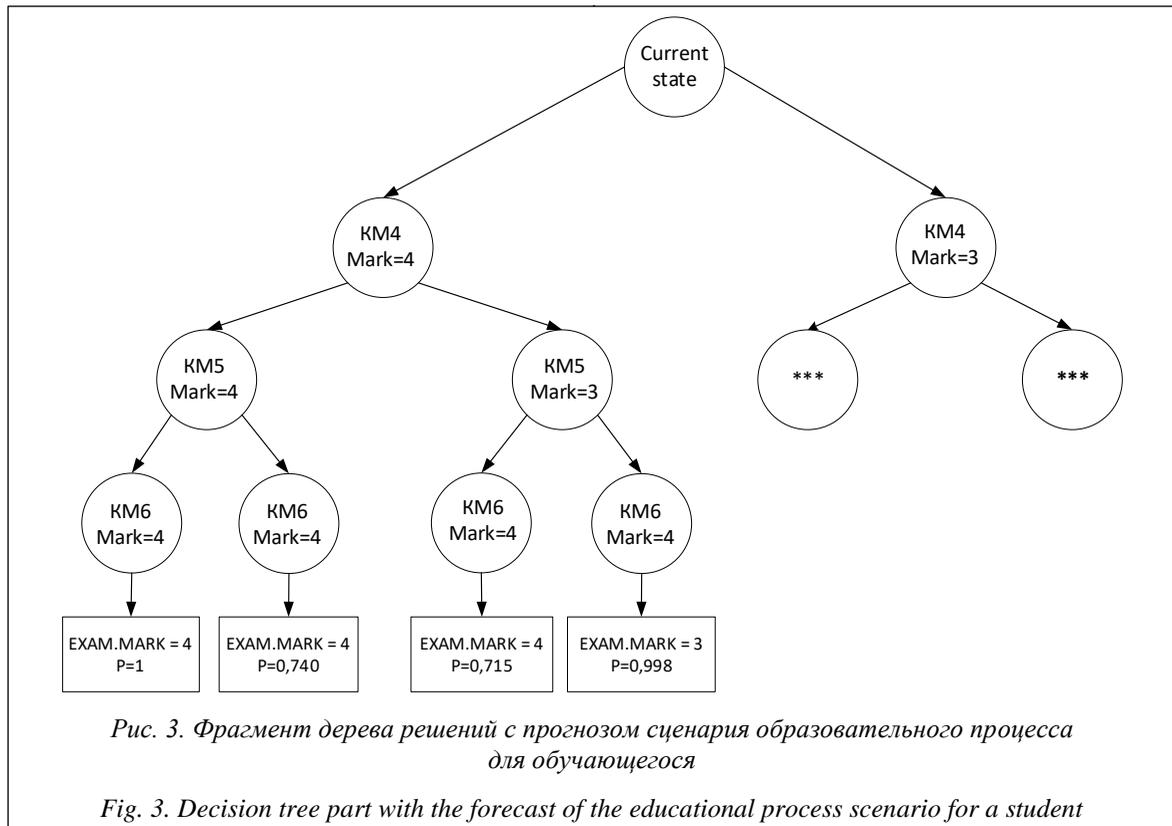
Для прогнозирования сроков сдачи задолженностей отстающими студентами был предложен алгоритм, строящий граф состояний, содержащий наборы контрольных мероприятий, необходимых для сдачи экзамена на удовлетворительную оценку или выше. При прогнозировании использовались материалы образовательной программы и данные студентов, успешно сдавших профильную дисциплину.

На рисунке 3 приведен фрагмент дерева решений относительно текущего состояния (Current\_state) с прогнозом сценария образовательного процесса, когда обучающийся со степенью уверенности 0,998 (практически 100 %) сдаст дисциплину с оценкой «удовлетворительно», со степенью уверенности 0,740 (74 %) с оценкой «хорошо» при условиях выполнения соответствующих контрольных мероприятий. При расчете прогноза учитываются такие параметры, как вес и темп выполнения контрольных мероприятий, темпоральные отношения типа Before, After относительно сроков промежуточных контрольных точек.

**Заключение**

В работе изложен подход к организации темпоральной модели и БД (и на ее основе темпоральной БЗ) с применением графовой нереляционной СУБД Neo4j для ИСППР РВ для помощи в организации учебного процесса и повышения его эффективности.

За счет реализации временных зависимостей и правил вывода темпоральной интервальной логики Аллена, дополненной средствами представления точечных событий, можно обрабатывать достаточно сложные запросы и проводить исследования учебного процесса в различных плоскостях-срезах как для отдельного обучающегося, так и для групп и потоков в целом.



Предложенная распределенная предобработка и интеллектуальный анализ данных с применением библиотеки Dask обеспечивают возможность кластеризации обучающихся на потоке по наиболее важным признакам, а также формирование предупреждений и рекомендаций студентам о возможном попадании их

в группу риска. В системе прогнозирования на этапе предобработки данных производится поиск закономерностей, зависимостей данных с построением темпоральных графов и алгоритмов темпорального вывода и построение моделей поведения студентов на основе выявленных паттернов.

**Список литературы**

1. Celik I. Towards Intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education. *Computers in Human Behavior*, 2023, vol. 138, art. 107468. doi: 10.1016/j.chb.2022.107468.
2. Hung H.C., Liu I.F., Liang C.T., Su Y.S. Applying educational data mining to explore students' learning patterns in the flipped learning approach for coding education. *Symmetry*, 2020, vol. 12, no. 2, art. 213. doi: 10.3390/sym12020213.
3. Tasrif E., Saputra H.K., Kurniadi D., Hidayat H., Mubai A. Designing website-based scholarship management application for teaching of Analytical Hierarchy Process (AHP) in Decision Support Systems (DSS) subjects. *IJIM*, 2021, vol. 15, no. 9, pp. 179–191. doi: 10.3991/ijim.v15i09.23513.
4. Баранова Е.В., Швецов Г.В. Методы и инструменты для анализа цифрового следа студента при освоении образовательного маршрута // *Перспективы науки и образования*. 2021. № 2. С. 415–430.
5. Гафаров Ф.М., Руднева Я.Б., Шарифов У.Ю. Прогностическое моделирование в высшем образовании: определение факторов академической успеваемости // *Высшее образование в России*. 2023. Т. 32. № 1. С. 51–70. doi: 10.31992/0869-3617-2023-32-1-51-70.
6. Khan A., Ghosh S.K. Student performance analysis and prediction in classroom learning: A review of educational data mining studies. *EAIT*, 2021, vol. 26, no. 1, pp. 205–240. doi: 10.1007/s10639-020-10230-3.
7. Aleem A., Gore M.M. Educational data mining methods: A survey. *Proc. IEEE CSNT*, 2020, pp. 182–188. doi: 10.1109/CSNT48778.2020.9115734.
8. Терентьев А.В. Методы и алгоритмы интеллектуального анализа данных в образовании // *Вестник науки*. 2024. Т. 4. № 5. С. 1545–1550.
9. Allen J.F., Ferguson G. Actions and events in interval temporal logic. *Journal of logic and computation*, 1994, vol. 4, no. 5, pp. 531–579.
10. Kukreja V., Aggarwal A. Multi-expert and multi-criteria evaluation of Online Education Factors: A fuzzy AHP approach. *JEET*, 2021, vol. 35, no. 2, pp. 140–148. doi: 10.16920/jeet/2021/v35i2/22081.

11. Barcelona A.B. An analytic hierarchy process for quality action researches in education. *IJERE*, 2020, vol. 9, no. 3, pp. 517–523. doi: 10.11591/ijere.v9i3.20626.
12. Yu D., Kou G., Xu Z., Shi S. Analysis of collaboration evolution in AHP research: 1982–2018. *IJTDM*, 2021, vol. 20, no. 01, pp. 7–36. doi: 10.1142/S0219622020500406.
13. Chernetsov A., Ereemeev A., Izhutkin V., Kutepov V., Varshavskii P., Fominykh I. Experience of development and application of intelligent IT and software in the educational process. *Proc. Int. Conf. Inforino*, 2024, pp. 1–5. doi: 10.1109/Inforino60363.2024.10551929.
14. Chernetsov A., Shevtsov D., Ivanov O. Development of a cross-platform interpreter for the muLISP language. *Proc. Int. Conf. Inforino*, 2024, pp. 1–6. doi: 10.1109/Inforino60363.2024.10551992.
15. Zhang Q., Cheng Y., Zhao F., Wang G., Xia S. Optimal scale combination selection integrating three-way decision with Hasse diagram. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, vol. 33, no. 8, pp. 3675–3689. doi: 10.1109/TNNLS.2021.3054063.
16. Еремеев А.П., Королев Ю.И. Анализ и верификация моделей процессов в сложных динамических системах // Искусственный интеллект и принятие решений. 2015. № 1. С. 45–56.
17. Еремеев А.П., Панявин Н.А. Унификация модели представления данных и преобразование форматов на основе нереляционной СУБД Neo4j // Программные продукты и системы. 2022. Т. 35. № 4. С. 549–556. doi: 10.15827/0236-235X.140.549-556.
18. Еремеев А.П., Куриленко И.В. Реализация вывода в темпоральных моделях ветвящегося времени // Известия РАН. Теория и системы управления. 2017. № 1. С. 107–127. doi: 10.7868/S0002338816060044.
19. Snodgrass R.T. The TSQL2 temporal query language. In: *SECS*, 1995, vol. 330, 674 p. doi: 10.1007/978-1-4615-2289-8.

Software &amp; Systems

doi: 10.15827/0236-235X.148.539-546

2024, 37(4), pp. 539–546

### Temporal database in intelligent real-time decision support systems: The case of a university educational process

Alexander P. Ereemeev <sup>1</sup>✉, Nikolay A. Panyavin <sup>1</sup>, Michail A. Marenkov <sup>1</sup><sup>1</sup> NRU “MPEI”, Moscow, 111250, Russian Federation

#### For citation

Ereemeev, A.P., Panyavin, N.A., Marenkov, M.A. (2024) ‘Temporal database in intelligent real-time decision support systems: The case of a university educational process’, *Software & Systems*, 37(4), pp. 539–546 (in Russ.). doi: 10.15827/0236-235X.148.539-546

#### Article info

Received: 13.05.2024

After revision: 15.06.2024

Accepted: 28.06.2024

**Abstract.** The purpose of the paper is to research and develop a temporal database, and then a temporal knowledge base for modern intelligent real-time systems by making such system prototype for an educational institution. The authors consider the educational process of a higher education institution (technical university) as a subject area that requires intellectual analysis and dynamic (temporal) data processing. The prototype aims at controlling academic progress and helping university students in problematic situations during the learning process. The paper proposes using a non-relational (NoSQL) database organization to implement the temporal model of information representation and temporal databases and knowledge bases on its basis. Such database organization is the most suitable for operating spatial dynamic data. Classical relational models and relational (SQL) databases are poorly applicable for their processing. The authors propose a prototype general diagram of a real-time intelligent decision support system with a temporal database and a knowledge base. The temporal database is based on a graph model and supplemented with a decision search system based on Allen’s interval logic and metric logic. The temporal database in the prototype used the data of 2,000 students of the National Research University “MPEI” on a number of profile disciplines. The data for the last several years was obtained from university special information systems. The software implementation involved Python v.3.10 programming language, Dask library for distributed and parallel computing and Neo4j non-relational graph DBMS.

**Keywords:** temporal database, artificial intelligence, decision support system, real time, classification, knowledge base, educational process

**Acknowledgements.** The work was financially supported by RSF, project no. 24-11-00285, <https://rscf.ru/project/24-11-00285/>

#### References

1. Celik, I. (2023) ‘Towards Intelligent-TPACK: An empirical study on teachers’ professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education’, *Computers in Human Behavior*, 138, art. 107468. doi: 10.1016/j.chb.2022.107468.
2. Hung, H.C., Liu, I.F., Liang, C.T., Su, Y.S. (2020) ‘Applying educational data mining to explore students’ learning patterns in the flipped learning approach for coding education’, *Symmetry*, 12(2), art. 213. doi: 10.3390/sym12020213.

3. Tasrif, E., Saputra, H.K., Kurniadi, D., Hidayat, H., Mubai, A. (2021) 'Designing website-based scholarship management application for teaching of Analytical Hierarchy Process (AHP) in Decision Support Systems (DSS) subjects', *IJIM*, 15(9), pp. 179–191. doi: 10.3991/ijim.v15i09.23513.
4. Baranova, E.V., Shvetsov, G.V. (2021) 'Methods and tools for analyzing a student's digital footprint when mastering an educational route', *Perspectives of Sci. and Education*, (2), pp. 415–430 (in Russ.).
5. Gafarov, F.M., Rudneva, Ya.B., Sharifov, U.Yu. (2023) 'Predictive modeling in higher education: determining factors of academic performance', *Higher Education in Russia*, 32(1), pp. 51–70 (in Russ.).
6. Khan, A., Ghosh, S.K. (2021) 'Student performance analysis and prediction in classroom learning: A review of educational data mining studies', *EAIT*, 26(1), pp. 205–240. doi: 10.1007/s10639-020-10230-3.
7. Aleem, A., Gore, M.M. (2020) 'Educational data mining methods: A survey', *Proc. IEEE CSNT*, pp. 182–188. doi: 10.1109/CSNT48778.2020.9115734.
8. Terentyev, A.V. (2024) 'Methods and algorithms of data mining in education', *Bull. of Sci.*, 4(5), pp. 1545–1550 (in Russ.).
9. Allen J. F., Ferguson G. Actions and events in interval temporal logic. *Journal of logic and computation*, 1994, vol. 4, no. 5, pp. 531–579.
10. Kukreja, V., Aggarwal, A. (2021) 'Multi-expert and multi-criteria evaluation of Online Education Factors: A fuzzy AHP approach', *JEET*, 35(2), pp. 140–148. doi: 10.16920/jeet/2021/v35i2/22081.
11. Barcelona, A.B. (2020) 'An analytic hierarchy process for quality action researches in education', *IJERE*, 9(3), pp. 517–523. doi: 10.11591/ijere.v9i3.20626.
12. Yu, D., Kou, G., Xu, Z., Shi, S. (2021) 'Analysis of collaboration evolution in AHP research: 1982–2018', *IJITDM*, 20(01), pp. 7–36. doi: 10.1142/S0219622020500406.
13. Chernetsov, A., Ereemeev, A., Izhutkin, V., Kutepov, V., Varshavskii, P., Fominykh, I. (2024) 'Experience of development and application of intelligent IT and software in the educational process', *Proc. Int. Conf. Inforino*, pp. 1–5. doi: 10.1109/Inforino60363.2024.10551929.
14. Chernetsov, A., Shevtsov, D., Ivanov, O. (2024) 'Development of a cross-platform interpreter for the muLISP language', *Proc. Int. Conf. Inforino*, pp. 1–6. doi: 10.1109/Inforino60363.2024.10551992.
15. Zhang, Q., Cheng, Y., Zhao, F., Wang, G., Xia S. (2021) 'Optimal scale combination selection integrating three-way decision with Hasse diagram', *IEEE Transactions on Neural Networks and Learning Systems*, 33(8), pp. 3675–3689. doi: 10.1109/TNNLS.2021.3054063.
16. Ereemeev, A.P., Korolev, Yu.I. (2015) 'Analysis and verification of process models in complex dynamic systems', *Artificial Intelligence and Decision Making*, (1), pp. 45–56 (in Russ.).
17. Ereemeev, A.P., Paniavin, N.A. (2022) 'Unification of a data presentation model and format conversion based on a non-relational Neo4j DBMS', *Software & Systems*, 35(4), pp. 549–556 (in Russ.). doi: 10.15827/0236-235X.140.549-556.
18. Ereemeev, A.P., Kurilenko, I.E. (2017) 'Realization of inference in temporal models of branching time', *News of RAS. Theory and Management Systems*, (1), pp. 107–127 (in Russ.). doi: 10.7868/S0002338816060044.
19. Snodgrass, R.T. (1995) 'The TSQL2 temporal query language', in *SECS*, 330, 674 p. doi: 10.1007/978-1-4615-2289-8.

**Авторы**

**Еремеев Александр Павлович**<sup>1</sup>, д.т.н.,  
профессор, профессор кафедры,  
eremeev@appmat.ru

**Паниявин Николай Александрович**<sup>1</sup>,  
аспирант, PaniavinNA@mpei.ru

**Маренков Михаил Андреевич**<sup>1</sup>,  
студент, MarenkovMA@mpei.ru

**Authors**

**Alexander P. Ereemeev**<sup>1</sup>, Dr.Sci. (Engineering),  
Professor, Professor of Department,  
eremeev@appmat.ru

**Nikolay A. Panyavin**<sup>1</sup>,  
Postgraduate Student, PaniavinNA@mpei.ru

**Michail A. Marenkov**<sup>1</sup>, Student,  
MarenkovMA@mpei.ru

<sup>1</sup> НИУ «МЭИ», г. Москва, 111250, Россия

<sup>1</sup> NRU "MPEI", Moscow, 111250, Russian Federation

## Совершенствование метода оценки стойкости пароля аутентификации пользователя компьютерных систем на основе использования известных уязвимостей

И.Г. Сидоркина<sup>1</sup>, С.В. Михалищев<sup>1</sup>✉

<sup>1</sup> Поволжский государственный технологический университет, г. Йошкар-Ола, 424000, Россия

### Ссылка для цитирования

Сидоркина И.Г., Михалищев С.В. Совершенствование метода оценки стойкости пароля аутентификации пользователя компьютерных систем на основе использования известных уязвимостей // Программные продукты и системы. 2024. Т. 37. № 4. С. 547–553. doi: 10.15827/0236-235X.148.547-553

### Информация о статье

Группа специальностей ВАК: 2.3.6

Поступила в редакцию: 26.07.2024

После доработки: 25.08.2024

Принята к публикации: 27.08.2024

**Аннотация.** Актуальность исследования обусловлена растущими угрозами безопасности пользовательских данных в современной цифровой среде, где выбор надежных паролей играет критическую роль в защите информации. Усовершенствованный метод основан на применении модели машинного обучения CatBoost, он улучшает существующие подходы к анализу структуры паролей и автоматическому формулированию требований к их сложности. При этом учитываются известные уязвимости, связанные с выбором пользователями слабых или скомпрометированных паролей. Особенностью усовершенствованного метода является глубокий анализ обучающего набора данных и использование словарей запрещенных и скомпрометированных паролей. Это позволяет повысить точность обученной модели и учесть динамически изменяющиеся требования безопасности. Основные результаты работы демонстрируют значительное повышение точности определения стойкости паролей аутентификации пользователей компьютерных систем. При анализе обучающего набора данных, используемого при обучении модели машинного обучения, были выявлены: неактуальное отнесение паролей к определенному классу надежности новым стандартом безопасности; наличие скомпрометированных паролей; отсутствие классифицированных паролей, у которых расстояние от левого символа кодовой таблицы unicode до правого минимально. В отличие от известных решений, проверка пароля по словарям проводилась на начальном этапе, до обучения модели. Это не создает дополнительной нагрузки на модель и не позволяет пользователям использовать запрещенные и скомпрометированные пароли. Практическая значимость работы заключается в интеграции предложенного усовершенствованного метода в системы аутентификации пользователей компьютерных систем. Это позволит исключить использование слабых и скомпрометированных паролей, повысит их эффективность и уровень защиты пользовательских данных, а также снизит риск успешных атак злоумышленников. Кроме того, предложенный усовершенствованный метод может быть адаптирован для различных систем безопасности при интеграции в существующие механизмы проверки паролей. Данное исследование вносит вклад в развитие методов цифровой безопасности и может быть полезно для специалистов в области информационной безопасности и при разработке программного обеспечения.

**Ключевые слова:** метод определения стойкости пароля, машинное обучение, аутентификация, безопасность информации, модель прогнозирования

**Введение.** Сохранность пользовательских данных является ключевым аспектом в современной цифровой среде, где злоумышленники постоянно совершают новые атаки и пытаются проникнуть в системы для получения конфиденциальной информации. В этом контексте одной из наиболее критических уязвимостей является выбор ненадежных паролей, которые могут быть легко взломаны с помощью различных методов атаки, таких как перебор, словарные атаки и использование радужных таблиц. Несмотря на широкое использование различных схем аутентификации, проблема слабых паролей остается актуальной и требует надежных методов защиты.

Ранее проведенные исследования в этой области посвящены различным аспектам про-

верки паролей и методам улучшения стойкости паролей аутентификации. Некоторые из них фокусировались на анализе сложности паролей с использованием статистических методов, в то время как другие исследования предлагали алгоритмы проверки паролей на основе сложных математических моделей.

Однако большинство разработок из-за недостаточной точности или сложности внедрения в реальные системы сталкиваются с определенными ограничениями. Кроме того, эти методы часто не учитывают возможное использование словарей паролей и динамически изменяющиеся требования безопасности.

Данная работа стремится заполнить этот пробел в исследованиях, предлагая усовершенствованный метод проверки стойкости паролей

аутентификации. Формируя базу запрещенных и скомпрометированных паролей, можно повысить точность обученной модели и исключить фактор использования злоумышленниками словарей паролей. Этот подход позволяет учитывать динамические изменения в требованиях к безопасности и обеспечивать более высокий уровень защиты данных.

Цель работы заключается в исключении возможности использования слабых и скомпрометированных паролей для обеспечения защиты пользовательских данных и в усовершенствовании метода определения стойкости пароля аутентификации.

### Анализ существующих методов

Согласно отчету [1], в 2020 году 89 % всех типов взломов связано со злоупотреблением учетными данными, включая атаки грубой силы и повторное использование учетных данных. Это обстоятельство подчеркивает важность использования машинного обучения для определения стойкости пароля, что может уменьшить риск большинства возможных атак. В таком контексте значительное количество современных исследований фокусируется на разработке эффективных методов определения стойкости пароля с помощью машинного обучения [2, 3].

Основные уязвимости стойкости паролей аутентификации пользователей компьютерных систем выявила компания Nord Security Inc. (разработчик приложений для хранения и генерации паролей). Ее ежегодные исследования показали, что большая часть пользователей

компьютеров и Интернета избегают использования надежных паролей, выбирая простые, легко запоминающиеся комбинации цифр и слов, которые уже давно известны даже рядовым пользователям [4]. Наиболее популярные пароли, применяемые пользователями во всем мире, представлены в таблице 1. Как следствие, такие системы взламываются за секунды, нанося урон персональным данным и конфиденциальной информации.

Для обеспечения безопасности пользовательских данных под запрет должны попадать пароли, включенные в словари запрещенных паролей, содержащих популярные имена, города, дни месяца, части логинов пользователей, номера компьютеров и т.д., и в словари скомпрометированных паролей.

Использование методов машинного обучения [5] для построения верификаторов позволяет алгоритмам формулировать требования к надежности паролей наиболее гибко. Алгоритмы машинного обучения, такие как классификация, регрессия и нейронные сети, анализируют структуру и характеристики вводимых паролей, определяя, какие из них являются потенциально слабыми или уязвимыми. Обучаясь на данных о различных типах паролей и их стойкости, модели машинного обучения могут более точно оценивать стойкость, выявляя скрытые паттерны и слабые места, неочевидные на первый взгляд.

Существующие методы к созданию верификаторов надежности паролей (<https://www.kaggle.com/datasets/morphlmax/password-security-sber-dataset/data>, [6]), как правило, используют популярные модели машинного обучения, такие

Таблица 1

Наиболее распространенные (скомпрометированные) пароли 2021–2023 гг.

Table 1

Most common (compromised) passwords in 2021–2023

2021		2022		2023	
Пароль	Количество пользователей	Пароль	Количество пользователей	Пароль	Количество пользователей
123456	103 170 552	password	4 929 113	123456	4 524 867
123456789	46 027 530	123456	1 523 537	admin	4 008 850
12345	32 955 431	123456789	413 056	12345678	1 371 152
qwerty	22 317 280	guest	376 417	123456789	1 213 047
password	20 958 297	qwerty	309 679	1234	969 811
12345678	14 745 771	12345678	284 946	12345	728 414
111111	13 354 149	111111	229 047	password	710 321
123123	10 244 398	12345	188 602	123	528 086
1234567890	9 646 621	col123456	140 505	Aa123456	319 725
1234567	9 396 813	123123	127 762	1234567890	302 709

как SGDClassifier, KNeighborsClassifier, CatBoost [7– 9].

Сравнительный результат моделей машинного обучения показал сопоставимые результаты всех проанализированных методов. В данной работе использовалась модель градиентного бустинга CatBoost» [10], создающая решающую модель прогнозирования в виде ансамбля слабых моделей прогнозирования, обычно деревьев решений.

Сохраним предложенную в существующих методах классификацию паролей по уникальным признакам:

- длина пароля (len);
- количество заглавных букв (count\_upper\_sym);
- количество прописных букв (count\_lower\_sym);
- количество спецсимволов (count\_special\_sym);
- количество цифр (count\_digit);
- расстояние от левого символа до правого на основе кодовой таблицы unicode (distance\_unicode);
- количество уникальных символов в пароле (count\_unique\_sym).

Итоги экспериментов подтвердили, что наибольшее влияние на стойкость оказывает длина пароля, за ней идут регистр, спецсимволы и наличие цифр (см. рисунок).

Рассмотрим результаты проведенных исследований.

Обучающий набор данных, используемый в существующих методах, содержал в себе 100 000 уникальных классифицированных паролей. Данные в наборе не сбалансированы: 74 % паролей с классом «1», 14 % слабых паролей с классом «0», 12 % сложных паролей с классом «2».

Поскольку набор паролей с классом сложности «1» считается несбалансированным, сравнивать его можно только по метрике  $F1_{macro}$ , которая учитывает производительность модели для каждого класса независимо от его размера в наборе [11]:

$$F1_{macro} = \frac{1}{N} \sum_{i=1}^N F1_i,$$

где  $N$  – количество классов;  $F1_i$  – F1-мера для каждого отдельного класса.

При работе с обучающим набором данных выполнялись случайное перемешивание и разбивка на обучающий (80 %) и тестовый (20 %) наборы. Для тестирования моделей использовалась перекрестная проверка с пятью делениями набора данных.

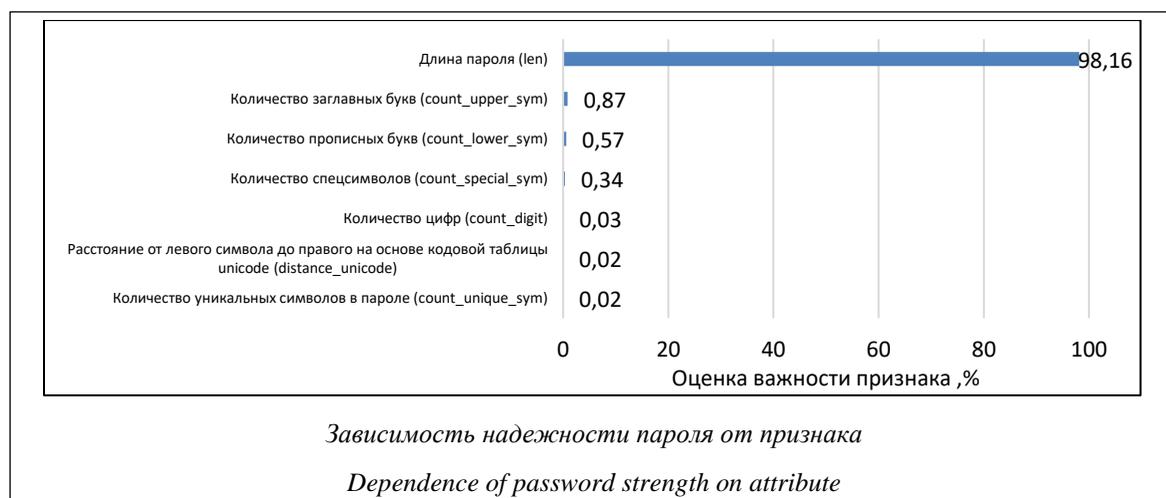
Предложенный метод показал, что применение модели машинного обучения на обучающем наборе методом кросс-валидации по метрике Precision дает почти 100-процентный результат. Однако тестирование модели на валидной выборке, содержащей 70 уникальных классифицированных паролей, показало невысокий результат:  $F1_{macro} \approx 0,7(7)$ .

Выявленные недостатки данного метода могут быть обусловлены

а) отсутствием фильтрации обучающего набора данных перед началом обучения на предмет использования скомпрометированных и запрещенных паролей;

б) использованием обучающего набора данных на основе кодовой таблицы unicode, в котором не представлены пароли с минимальным расстоянием от левого до правого символа;

в) доверительным отношением к классификации паролей в обучающем наборе данных, которая не подвергалась сомнению и не проходила дополнительных проверок.



Данные недостатки и достоинства учтены при разработке усовершенствованного метода оценки стойкости пароля аутентификации пользователя компьютерных систем.

### Определение стойкости пароля аутентификации

В соответствии с документом ФСТЭК (<https://fstec.ru/dokumenty/vse-dokumenty/prikazy/prikaz-fstek-rossii-ot-18-fevralya-2013-g-p-21>) для достижения требований четвертого уровня защищенности персональных данных пароль должен содержать не менее 8 символов при алфавите пароля не менее 70 символов. Максимальное количество неуспешных попыток аутентификации (ввода неправильного пароля) до блокировки – 4.

Проанализируем валидную выборку, используемую для тестирования рассмотренных моделей.

Наибольший процент ошибок пришелся на пароли, у которых расстояние от левого символа до правого на основе кодовой таблицы unicode минимально. Представим примеры паролей, стойкость которых отнесена обученной моделью к классу надежности «2»: «bbaacdaa aabdcbada», «dadaaabaacbbadbade», «31222a23 2331131312», «32a331132312121232» и аналогичные. В действительности их стойкость по валидному набору данных равна единице.

Для устранения ошибки, связанной с отсутствием в обучающем наборе данных паролей, характеризующихся минимальной дисперсией кодов unicode для крайних символов, в обучающий набор данных будет добавлено 20 паролей, подобных «deeddfdfeddddfdee», «4456554 6d455644664» и аналогичных, со значением класса надежности, равным единице.

Перед началом обучения модели применим базу скомпрометированных паролей (<https://www.kaggle.com/datasets/joebeachcapital/top-10-million-passwords?resource=download>), содержащую 999 997 записей, и исключим эти записи из обучающего набора данных. После применения обучающий набор данных, содержащий 100 020 уникальных значений, классифицированных по уровням сложности, сократился до 96 776 значений.

Протестируем усовершенствованную модель. Для обучения будем использовать модель градиентного бустинга CatBoost. Классификацию паролей выполним по признакам, приведенным на рисунке.

Основную результирующую оценку модели будем выполнять по метрике  $F1_{macro}$ , используя формулу, представленную выше.

Результат обучения и оценки модели по метрике  $F1_{macro}$  равен 1.

Повторное тестирование модели на валидной выборке, используемой в рассмотренных аналогах, показало:  $F1_{macro} \approx 1$ .

Сгенерируем валидный набор данных из 100 паролей. Для генерации использовался сервис «Рандомус» (<https://randomus.ru/password>). Для проверки надежности каждого пароля и отнесения его к определенному классу стойкости использовался сервис от лаборатории Касперского (<https://password.kaspersky.com/ru>). В результате получили валидный набор, содержащий несбалансированные данные: 60 паролей с классом «1», 15 слабых паролей с классом «0», 25 сложных паролей с классом «2».

Проверим обученную модель на валидной выборке по метрике  $F1_{macro}$ , используя следующий алгоритм.

- Преобразование массивов:
  - преобразуем массив предсказанных значений `pred_valid` в одномерный;
  - преобразуем столбец `strength DataFrame df_valid_features` в одномерный массив.
- Вычисление F1-меры:
  - используем функцию `f1_score` для вычисления с макроусреднением на основе одномерных массивов `pred_valid` и `df_valid_features.strength`.
- Вывод значения F1-меры с макроусреднением.
- Создание и обработка массива ошибок:
  - создаем массив ошибок со значением типа булево, сравнивая значения массивов `pred_valid` и `df_valid_features.strength`;
  - выбираем номера строк из `DataFrame df_valid_features`, где обнаружены ошибки.
- Вывод номеров строк с ошибками.

В результате выполнения алгоритма установлено, что значение F1-меры с макроусреднением составляет 0,8154121863799283. Ошибки обнаружены в валидном наборе данных с номерами строк: 5, 6, 7, 8, 9, 10, 11, 12, 13 и 14.

Анализ допущенных ошибок показал, что обученная модель отнесла 10 слабых паролей к классу надежности «1» и ни разу не ошиблась при классификации сильных паролей. Пароли, стойкость которых отнесена обученной моделью к классу надежности «1»: «e\_4wKT~1», «\_3>hN84O», «\_S\$8/I/x», «-JNkz{s6», «-3@Vy Kl.», «e\_4wKT~1», «\_3>hN84O», «\_S\$8/I/x», «-JNkz{s6» и «-3@VyKl.». В действительности

их стойкость по валидному набору данных равна нулю. Все представленные пароли, на которых обученная модель показала низкую точность при классификации, имеют одинаковую длину, равную 8 знакам.

Обучающий набор данных содержит 16 816 паролей длиной 8 знаков, все из них были отнесены к классу надежности «1». Однако валидный набор данных включал 10 паролей той же длины, которые были отнесены к классу надежности «0». Это расхождение в классификации обучающего и валидного наборов данных и привело к значительным ошибкам.

Для улучшения точности модели было принято решение о пересмотре и корректировке обучающего набора данных. Все 16 816 паролей длиной 8 знаков переклассифицированы из класса надежности «1» в класс надежности «0». Это изменение отражает актуальные требования к надежности паролей и необходимость их пересмотра в соответствии с новыми стандартами безопасности.

После внесения этих корректировок точность обученной модели значительно улучшилась. Результаты тестирования на валидной выборке показали точность по метрике  $F1_{macro}$ , равную 1, что свидетельствует о правильной классификации всех паролей в валидной выборке (табл. 2). Таким образом, реализация усовершенствований метода привела к увеличению точности обученной модели на 23,2 %.

Для получения результата, близкого к стопроцентной точности, были предприняты следующие шаги.

1. Исключение из обучающего набора паролей по словарю скомпрометированных и запрещенных паролей, что позволило снизить вероятность переобучения модели на неинформативных данных для обобщения результатов на новых, ранее не рассматриваемых паролях.

2. Добавление в обучающую выборку паролей, у которых расстояние от левого символа до правого на основе кодовой таблицы unicode минимально. Этот шаг был направлен на повышение энтропии генерируемых паролей, что затрудняет их взлом методами перебора и криптоанализа.

3. Переклассификация паролей длиной 8 знаков. Все 16 816 паролей длиной 8 знаков были переклассифицированы из класса надежности «1» в класс надежности «0». Это изменение отражает актуальные требования к надежности паролей и необходимости их пересмотра в соответствии с новыми стандартами безопасности.

Такую модель следует применять только после проверки пользовательского пароля на наличие его в словарях скомпрометированных и запрещенных паролей. Исключение слабых паролей из анализа не только повышает эффективность работы модели, но и снижает объем вычислений, необходимых для обработки данных. Такой подход обеспечивает более быструю и точную оценку стойкости паролей, что в конечном итоге способствует надежной защите пользовательских данных.

### Заключение

Таким образом, показано, что применение модели машинного обучения «градиентный бустинг с CatBoost», использование словаря скомпрометированных и запрещенных паролей при ее обучении, а также применение анализа и корректировка обучающего набора данных повышают эффективность метода точности определения стойкости пароля аутентификации до значений, близких к 100 %, что является существенным усилением защиты пользовательских данных от атак злоумышленников.

Результат тестирования обученной модели на валидной выборке

Таблица 2

Table 2

Result of testing the trained model on a valid sample

Выборка	Точность (precision)	Полнота (recall)	F1-мера (f1-score)	Поддержка (support)
Класс надежности 0	1,00	1,00	1,00	15
Класс надежности 1	1,00	1,00	1,00	60
Класс надежности 2	1,00	1,00	1,00	25
Точность (accuracy)			1,00	100
Среднее (macro avg)	1,00	1,00	1,00	100
Средневзвешенное (weighted avg)	1,00	1,00	1,00	100

При создании модели любой набор данных перед использованием должен быть проверен и приведен в соответствие действующим требованиям к защищенности персональных данных. Это осуществляется с учетом словаря скомпрометированных и запрещенных паролей и технических возможностей злоумышленников. Модель машинного обучения (обученная на специально подготовленном наборе данных с вероятностью, близкой к 100 %) позволит классифицировать надежность пароля аутентификации и проинформирует пользователя об ограничениях на использование слабых паролей.

В процессе разработки усовершенствованного метода оценки стойкости пароля были учтены достоинства существующих решений, включая использование алгоритмов машин-

ного обучения и статистического анализа. Проведенный анализ также выявил их недостатки, такие как отсутствие комплексной проверки данных и неэффективное использование известных уязвимостей. Внедрение БД скомпрометированных и запрещенных паролей позволило устранить указанные недостатки, что способствовало повышению надежности оценки стойкости паролей.

Применение усовершенствованного метода определения стойкости пароля аутентификации повысит эффективность защиты информации и обеспечит повышение безопасности в системах управления доступом, в мобильных/экранных приложениях, в веб-приложениях, таких как онлайн-банкинг и электронная почта.

#### Список литературы

1. Jartelius M. The 2020 data breach investigations report – a CSO's perspective. *Network Security*, 2020, vol. 2020, no. 7, pp. 9–12. doi: 10.1016/S1353-4858(20)30079-9.
2. Sarkar S., Nandan M. Password strength analysis and its classification by applying machine learning based techniques. *Proc. ICCSEA*, 2022, pp. 1–5. doi: 10.1109/ICCSEA54677.2022.9936117.
3. Sakya S.S., Mauparna M.N. Building a multi-class password strength generator and classifier model by augmenting supervised machine learning techniques. *Research Square*, 2022. doi: 10.21203/rs.3.rs-1820885/v1.
4. Zaidi T., Garai S., Biradar T.V. Exploring the landscape of password managers for individual users through innovative solution. In: *Information Technology Security. STEEE*, 2024, pp. 69–99. doi: 10.1007/978-981-97-0407-1\_4.
5. Назарова А.Д. Анализ надежности паролей для защиты данных // *Умная цифровая экономика*. 2022. Т. 2. № 4. С. 41–46.
6. Беликов В.В., Прокуронов И.А. Построение верификатора стойкости пароля с использованием классических методов машинного обучения и рекуррентной LSTM нейронной сети // *Russ. Tech. J.* 2023. Т. 11. № 4. С. 7–15. doi: 10.32362/2500-316X-2023-11-4-7-15.
7. Li X., Orabona F. On the convergence of stochastic gradient descent with adaptive stepsizes. *Proc. Int. Conf. on Artificial Intelligence and Statistics. PMLR*, 2019, vol. 89, pp. 983–992.
8. Lubis A.R., Lubis M., Khowarizmi A. Optimization of distance formula in K-Nearest Neighbor method. *BEEI*, 2020, vol. 9, no. 1, pp. 326–338. doi: 10.11591/eei.v9i1.1464.
9. Coronado-Blázquez J. Classification of Fermi-LAT unidentified gamma-ray sources using CatBoost gradient boosting decision trees. *MNRAS*, 2022, vol. 515, no. 2, pp. 1807–1814. doi: 10.1093/MNRAS/STAC1950.
10. Prokhorenkova L., Gusev G., Vorobev A., Dorogush A.V., Gulina A. CatBoost: Unbiased boosting with categorical features. *Proc. NeurIPS*, 2018, vol. 31, pp. 6638–6648.
11. Lever J., Krzywinski M., Altman N. Classification evaluation. *Nat. Methods*, 2016, vol. 13, pp. 603–604. doi: 10.1038/nmeth.3945.

#### Assessing the strength of a computer system user authentication password: Improving the method based on using known vulnerabilities

Irina G. Sidorkina <sup>1</sup>, Stanislav V. Mikhailishchev <sup>1</sup>✉

<sup>1</sup> Volga State University of Technology, Yoshkar-Ola, 424000, Russian Federation

#### For citation

Sidorkina, I.G., Mikhailishchev, S.V. (2024) 'Assessing the strength of a computer system user authentication password: Improving the method based on using known vulnerabilities', *Software & Systems*, 37(4), pp. 547–553 (in Russ.). doi: 10.15827/0236-235X.148.547-553

**Article info**

Received: 26.07.2024

After revision: 25.08.2024

Accepted: 27.08.2024

**Abstract.** The relevance of the research is due to the growing threats to user data security in the modern digital environment; the choice of strong passwords plays a critical role in protecting information. The improved method is based on applying the CatBoost machine-learning model. It improves existing approaches to analyzing password structure and automatic formulating requirements for password complexity. It takes into account well-known vulnerabilities associated with users choosing weak or compromised passwords. A special feature of the improved method is a deep analysis of the training data set and using dictionaries of forbidden and compromised passwords. This allows improving the accuracy of the trained model and take into account dynamically changing security requirements. The main results of the work demonstrate a significant improvement in the accuracy of password strength determination for authentication of computer system users. When analyzing the training dataset used to train the machine-learning model, the authors identified some issues. These were irrelevant assignment of passwords to a certain security class by a new security standard; compromised passwords; the absence of classified passwords that have a minimum distance from the left character of the Unicode code table to the right one. In contrast to known solutions, password verification by dictionaries took place at the initial stage, before training the model. It does not create additional load on the model and does not allow users to use forbidden and compromised passwords. The practical significance of this work is in implementing the improved method in user authentication systems. This will eliminate weak and compromised passwords, increase their efficiency and the level of user data protection, reduce the risk of successful attacks. Furthermore, the authors can adapt the proposed method for various security systems when integrated into existing password verification mechanisms. This research contributes to the development of digital security techniques. It can be useful for information security and software engineering professionals.

**Keywords:** password strength determination method, machine learning, authentication, information security, prediction model

**References**

1. Jartelius, M. (2020) 'The 2020 data breach investigations report – a CSO's perspective', *Network Security*, 2020(7), pp. 9–12. doi: 10.1016/S1353-4858(20)30079-9.
2. Sarkar, S., Nandan, M. (2022) 'Password strength analysis and its classification by applying machine learning based techniques', *Proc. ICCSEA*, pp. 1–5. doi: 10.1109/ICCSEA54677.2022.9936117.
3. Sakya, S.S., Mauparna, M.N. (2022) 'Building a multi-class password strength generator and classifier model by augmenting supervised machine learning techniques', *Research Square*. doi: 10.21203/rs.3.rs-1820885/v1.
4. Zaidi, T., Garai, S., Biradar, T.V. (2024) 'Exploring the landscape of password managers for individual users through innovative solution', in *Information Technology Security. STEEE*, pp. 69–99. doi: 10.1007/978-981-97-0407-1\_4.
5. Nazarova, A.D. (2022) 'Password strength analysis for data protection', *Smart Digital Economy*, 2(4), pp. 41–46 (in Russ.).
6. Belikov, V.V., Prokuronov, I.A. (2023) 'Password strength verification based on machine learning algorithms and LSTM recurrent neural networks', *Russ. Tech. J.*, 11(4), pp. 7–15 (in Russ.). doi: 10.32362/2500-316X-2023-11-4-7-15.
7. Li, X., Orabona, F. (2019) 'On the convergence of stochastic gradient descent with adaptive stepsizes', *Proc. Int. Conf. on Artificial Intelligence and Statistics. PMLR*, 89, pp. 983–992.
8. Lubis, A.R., Lubis, M., Khowarizmi, A. (2020) 'Optimization of distance formula in K-Nearest Neighbor method', *BEEI*, 9(1), pp. 326–338. doi: 10.11591/eei.v9i1.1464.
9. Coronado-Blázquez, J. (2022) 'Classification of Fermi-LAT unidentified gamma-ray sources using CatBoost gradient boosting decision trees', *MNRAS*, 515(2), pp. 1807–1814. doi: 10.1093/MNRAS/STAC1950.
10. Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulina, A. (2018) 'CatBoost: Unbiased boosting with categorical features', *Proc. NeurIPS*, 31, pp. 6638–6648.
11. Lever, J., Krzywinski, M., Altman, N. (2016) 'Classification evaluation', *Nat. Methods*, 13, pp. 603–604. doi: 10.1038/nmeth.3945.

**Авторы**

**Сидоркина Ирина Геннадьевна**<sup>1</sup>, д.т.н.,  
профессор, заведующий кафедрой,  
igs592000@mail.ru

**Михалищев Станислав Вячеславович**<sup>1</sup>,  
магистр, соискатель,  
mihalischevstas@gmail.com

**Authors**

**Irina G. Sidorkina**<sup>1</sup>, Dr.Sci. (Engineering),  
Professor, Head of Chair,  
igs592000@mail.ru

**Stanislav V. Mikhailishev**<sup>1</sup>,  
Master of Science, Candidate,  
mihalischevstas@gmail.com

<sup>1</sup> Поволжский государственный  
технологический университет,  
г. Йошкар-Ола, 424000, Россия

<sup>1</sup> Volga State University of Technology,  
Yoshkar-Ola, 424000,  
Russian Federation

## Рекуррентные нечеткие ситуационно-прецедентные модели для оперативного управления сложными техническими объектами

А.М. Соколов<sup>1</sup>, В.В. Борисов<sup>2</sup>✉

<sup>1</sup> Национальный исследовательский университет «Московский энергетический институт»,  
г. Москва, 111250, Россия

<sup>2</sup> Смоленский филиал Национального исследовательского университета МЭИ,  
г. Смоленск, 214013, Россия

### Ссылка для цитирования

Соколов А.М., Борисов В.В. Рекуррентные нечеткие ситуационно-прецедентные модели для оперативного управления сложными техническими объектами // Программные продукты и системы. 2024. Т. 37. № 4. С. 554–565. doi: 10.15827/0236-235X.148.554-565

### Информация о статье

Группа специальностей ВАК: 2.3.1

Поступила в редакцию: 14.06.2024

После доработки: 17.07.2024

Принята к публикации: 30.07.2024

**Аннотация.** В статье рассматриваются вопросы оперативного управления сложными техническими объектами (СТО), функционирующими в динамически изменяющихся, неопределенных условиях внешней среды, рисков возникновения опасных и аварийных ситуаций. Проведен анализ подходов к ситуационному управлению СТО, который показал, что перспективным решением для оперативного управления этими объектами является применение нечетких моделей типа «ситуация–действие». Основными ограничениями моделей данного типа являются необходимость дополнительной экспертной информации для актуализации нечеткой базы знаний, а также отсутствие учета возможности переходов в различные ситуации при применении сходных управляющих воздействий. Для устранения указанных недостатков предлагается применение прецедентного подхода, который предполагает накопление информации о функционировании СТО в базе нечетких прецедентов с целью их дальнейшего использования при выборе управленческих решений. Предложены новый тип рекуррентных нечетких ситуационно-прецедентных моделей, а также способ и алгоритмы оперативного управления СТО по прецедентам на основе этих моделей. Авторское решение отличается применением итерационной процедуры поиска последовательности прецедентных управляющих решений. Данная процедура обеспечивает повышение эффективности функционирования СТО при условии адаптации к изменениям внутренних и внешних факторов. Разработаны алгоритмические и программные средства, реализующие предложенные модели, способ и алгоритмы управления СТО. В ходе вычислительных экспериментов подтверждена адекватность предложенных моделей. Выполнена оценка эффективности оперативного управления СТО по прецедентам с использованием созданных моделей, способа и алгоритмов в условиях неопределенности по сравнению с экспертной нечеткой ситуационной моделью «ситуация–действие».

**Ключевые слова:** рекуррентная нечеткая ситуационно-прецедентная модель, нечеткая ситуационная модель «ситуация–действие», прецедент, оперативное управление

**Благодарности.** Работа выполнена в рамках госзадания, проект № FSWF-2023-0012

**Введение.** В различных отраслях промышленности *сложные технические объекты* (СТО) используются для решения широкого спектра задач. Эти объекты характеризуются многокомпонентной структурой со сложными взаимосвязями между компонентами, а также большим числом влияющих на их работу внутренних и внешних факторов в условиях неопределенности. Указанные особенности значительно затрудняют, а в ряде случаев делают невозможным построение аналитических моделей, что обуславливает целесообразность использования ситуационных моделей управления.

Среди СТО можно выделить отдельный класс специальных роботизированных объектов, предназначенных для решения задач в открытом мире и функционирующих в динамически изменяющихся и неопределенных усло-

виях внешней среды. К таким СТО могут быть отнесены оснащенные манипуляторами и системами технического зрения автономные робототехнические комплексы, обеспечивающие контроль и проведение технологических операций при строительстве и техническом обслуживании, например, в добывающей промышленности, в подводной инфраструктуре, на линиях коммуникаций [1, 2]. Примером таких СТО являются также мобильные роботы, выполняющие исследовательские и разведывательные работы в полевых условиях [3], и снабженные манипуляторами беспилотные аппараты, обслуживающие технику в космическом пространстве [4].

Отличительными особенностями СТО рассматриваемого класса являются сложность прогнозирования ситуаций в процессе эксплуатации объектов, а также большое число опас-

ных и аварийных ситуаций, требующих оперативного реагирования [5]. В таких случаях в качестве критерия управления целесообразно рассматривать достижение целевой ситуации за наименьшее число управляющих решений при обеспечении требуемого уровня безопасности. При этом из-за сложности оценки и неопределенности влияния совокупности факторов на результаты управления СТО возможны случаи, когда одинаковым управляющим решениям будут соответствовать различные итоговые ситуации.

Таким образом, развитие моделей, способов и алгоритмов для ситуационного управления СТО в условиях неопределенности, учитывающих возможность вариативности результатов применения управляющих решений, является актуальной и практически значимой задачей.

### Ситуационный подход к управлению СТО в условиях неопределенности

Создание адекватных аналитических моделей для управления СТО в условиях неопределенности является нетривиальной и ресурсоемкой задачей, что обосновывает целесообразность применения интеллектуальных моделей, базирующихся на положениях теории нечетких множеств и нечетких вычислений [6].

Одним из современных подходов к решению проблемы учета влияния динамически изменяющихся воздействий внешней среды при управлении СТО является создание интеллектуальных систем на основе комбинирования нечеткой логики и алгоритмов обучения с подкреплением [7, 8]. Данный подход предполагает непрерывную адаптацию в процессе управления СТО с учетом мониторинга их взаимодействия со средой функционирования и закрепления положительного опыта управления [9]. Эвристический характер формирования паттернов управления СТО снижает результативность обучения с подкреплением при наличии аварийных режимов работы и чрезвычайных ситуаций, поскольку при этом цена ошибки значительно возрастает.

Для управления СТО в условиях неопределенности распространение получили типы моделей состояний и управления, абстрагирующиеся от внутренней структуры объекта и ориентированные на представление различных ситуаций и/или состояний, а также моделей реализации различных способов переходов между ними, а именно: нечеткие когнитивные модели, нечеткие ситуационно-сетевые модели, нечеткие модели типа «ситуация–действие».

Нечеткие когнитивные модели, представляющие СТО в виде множества концептов и нечетких отношений между ними, позволяют при решении задач управления этими объектами осуществлять предварительный анализ СТО [10], вырабатывать рекомендации по их усовершенствованию [11], использовать такие модели СТО в составе более сложных композиционных моделей [12]. Несмотря на наглядность и возможность учета динамики поведения СТО и изменений внешней среды, данный тип моделей затруднительно использовать для ситуационного управления при наличии множества параметров (характеристик) возможных ситуаций.

Нечеткие ситуационно-сетевые модели представляются в виде графа, вершины которого соответствуют нечетким ситуациям, а дуги – управляющим решениям, переводящим СТО из одного состояния в другое. Они предназначены для определения результата применения последовательности управляющих решений, а также стратегии управления, приводящей к требуемой нечеткой ситуации.

Авторами работы [13] предложен способ нечеткого ситуационного управления СТО на основе комбинирования нечетких ситуационно-сетевых моделей и композиционного гибридного моделирования, применение которого проиллюстрировано на примере моделирования управления центробежными компрессорами. Основным преимуществом данного подхода является возможность формирования наилучших с точки зрения выбранных критериев стратегий управления СТО. К недостаткам метода управления относится то, что учет динамики изменения СТО осуществляется только на уровне композиционных моделей, но не отражается непосредственно на структуре нечетких ситуационно-сетевых моделей, а также отсутствие в этом методе гибкого механизма адаптации модели к изменяющимся условиям функционирования СТО. Это обуславливает необходимость привлечения дополнительной экспертной информации для построения и актуализации модели.

В работе [14] рассмотрены нечеткие ситуационные модели (НСМ) типа «ситуация–действие», которые в явном виде задают соответствия между ситуациями и управляющими решениями, тем самым формируя нечеткую базу знаний. Эти модели ориентированы на определение наиболее близкой эталонной ситуации и на выдачу соответствующего (для ее достижения) управляющего решения, обладают более высоким быстродействием, что определяет

целесообразность их применения для оперативного управления СТО в условиях неопределенности. НСМ типа «ситуация–действие» не учитывают различные стратегии управления СТО, что является их основным ограничением. Однако, как отмечалось ранее, при управлении в аварийных и опасных ситуациях, если в качестве критерия эффективности управления рассматривается минимизация количества управляющих решений для достижения целевой ситуации, указанное ограничение не настолько существенно. Отдельным вопросом при построении НСМ типа «ситуация–действие» являются формирование и адаптация нечеткой базы знаний, которая, как правило, задается экспертно, что не позволяет в полной мере учитывать влияние динамически изменяющихся факторов внешней среды в условиях неопределенности. Еще одним ограничением этих моделей следует назвать отсутствие учета возможности переходов в различные нечеткие ситуации при применении сходных управляющих воздействий.

Как известно, при моделировании функционирования СТО, для которых полное описание невозможно или затруднено из-за структурной и параметрической сложности, а также влияния на их работу стохастических и нестохастических внешних воздействий, применяется прецедентный подход [15–17]. Он предполагает накопление и использование ретроспективной информации о функционировании объекта.

В данной статье для повышения эффективности оперативного управления СТО в условиях неопределенности предлагаются новый тип *рекуррентных нечетких ситуационно-прецедентных моделей* (РНСПМ), а также способ и алгоритмы управления СТО по прецедентам на основе этих моделей.

### Способ оперативного управления СТО по прецедентам на основе РНСПМ

Использование прецедентного подхода для построения и адаптации РНСПМ предполагает создание *базы нечетких прецедентов* (БНП) для накопления актуальной информации о положительном опыте управления СТО и результатов моделирования.

Прецеденты в БНП представляются в виде  $\tilde{c}_n = \langle \tilde{S}_n^h, \tilde{S}_n^k, \tilde{R}_n \rangle$ , (1)

где  $n = 1, \dots, N$  – номер прецедента,  $N$  – число прецедентов;  $\tilde{S}_n^h$  – начальная нечеткая ситуа-

ция, в которой находится СТО;  $\tilde{S}_n^k$  – конечная нечеткая ситуация, в которую переходит СТО в результате воздействия управляющего решения  $\tilde{R}_n$ .

Каждая нечеткая ситуация  $\tilde{S}_i$  задается набором ситуационных признаков  $\tilde{P}$  в виде нечетких множеств 2-го уровня [18]:

$$\tilde{P} = \{ \tilde{p}_m^i \} = \begin{cases} \mu_h(p_1)/p_1, \\ \dots \\ \mu_h(p_m)/p_m, \\ \dots \\ \mu_h(p_M)/p_M, \end{cases}$$

$$\tilde{S}_i = \{ \{ \mu_h(p_m)/p_m \}, h = 1, \dots, H \}, m = 1, \dots, M \}, i = 1, \dots, I,$$

где  $p_m$  – значение  $m$ -го нечеткого ситуационного признака,  $m = 1, \dots, M$ ;  $\mu_h(p_m)/p_m$  – значение функции принадлежности признака  $\tilde{p}_m$  к  $h$ -му терм-множеству,  $h = 1, \dots, H$ .

Оценка степени близости нечетких ситуаций может осуществляться на основе вычисления псевдометрических расстояний, например, нечеткого расстояния Евклида:

$$d_E(\tilde{S}_i, \tilde{S}_j) = \frac{1}{M} \sum_{m=1}^M \sqrt{(\mu_i(\tilde{p}_m^i) + \mu_j(\tilde{p}_m^j))^2}, \quad (2)$$

при этом ситуации считаются нечетко схожими, если степень нечеткой близости между ними не превышает заданного порога:

$$\tilde{S}_i \cong \tilde{S}_j : d_E(\tilde{S}_i, \tilde{S}_j) \leq d_{кр}.$$

Если ситуации являются нечетко схожими, то может быть проведено их объединение:

$$\tilde{S}_{avg}(\tilde{S}_i, \tilde{S}_j) = \left\{ \frac{1}{2} \cdot \left( \left\{ \begin{matrix} \mu_i(\tilde{p}_m^i) / \tilde{p}_m^i + \\ \mu_j(\tilde{p}_m^j) / \tilde{p}_m^j \end{matrix} \right\} \right) \right\}_{m=1, \dots, M}. \quad (3)$$

При достаточном объеме статистических данных возможность перехода СТО из нечеткой ситуации  $\tilde{S}_i$  в нечеткую ситуацию  $\tilde{S}_j$  в результате воздействия управляющего решения  $\tilde{R}_{ij}$  можно оценить следующим образом:

$$Z(\tilde{S}_i, \tilde{R}_{ij}) = N_{\tilde{R}_{ij}}^* / N_{\tilde{R}_{ij}},$$

где  $N_{\tilde{R}_{ij}}^*$  – число раз, когда СТО перешло из  $\tilde{S}_i$  в  $\tilde{S}_j$  при общем числе  $N_{\tilde{R}_{ij}}$  воздействий управляющего решения  $\tilde{R}_{ij}$ .

Учитывая, что подобные сведения могут быть неопределенными, целесообразно перейти к нечеткой оценке возможности перехода из ситуации  $\tilde{S}_i$  в ситуацию  $\tilde{S}_j$ :

$$\tilde{Z}(\tilde{S}_i, \tilde{R}_{ij}) = \left\{ \mu_{\tilde{Z}(\tilde{S}_i, \tilde{R}_{ij})}(Z) / Z \right\}, \quad (4)$$

где  $\mu_{\tilde{Z}(\tilde{S}_i, \tilde{R}_j)}(Z)$  – степень принадлежности нечеткого множества  $\tilde{Z}(\tilde{S}_i, \tilde{R}_j)$  к значению базового множества  $Z \in [0, 1]$ .

Предлагается способ оперативного управления СТО на основе РНСПМ, заключающийся в рекуррентном поиске прецедентов и применении последовательности соответствующих им управляющих решений для достижения целевой ситуации с использованием наименьшего количества прецедентов:

$$w_{\tilde{S}_i, \tilde{S}_j}^* = \min_n \{w_{\tilde{S}_{тек}, \dots, \tilde{S}_n, \dots, \tilde{S}_{цел}}\},$$

где  $w_{\tilde{S}_i, \tilde{S}_j}^*$  – последовательность, состоящая из наименьшего числа прецедентов, применение управляющих решений которых позволяет перевести СТО из текущей ситуации  $\tilde{S}_{тек}$  в целевую  $\tilde{S}_{цел}$ ;  $w_{\tilde{S}_{тек}, \dots, \tilde{S}_n, \dots, \tilde{S}_{цел}}$  –  $n$ -я последовательность применения прецедентных управляющих решений для перевода СТО из ситуации  $\tilde{S}_{тек}$  в ситуацию  $\tilde{S}_{цел}$ .

Другим условием целесообразности применения способа является достаточно большое число возможных ситуаций и управляющих переходов.

Предлагаемый способ оперативного управления СТО по прецедентам на основе РНСПМ включает в себя следующие этапы.

Этап 1. Идентификация нечеткой текущей ситуации  $\tilde{S}_{тек}$ , в которой находится СТО, на основе информации от контрольно-измерительной аппаратуры и датчиков.

Этап 2. Задание целевой нечеткой ситуации  $\tilde{S}_{цел}$  предельного значения  $d_{кр}$  для оценки степени сходства между ситуациями.

Этап 3. Поиск в БНП прецедента, для которого выполняется следующее условие:

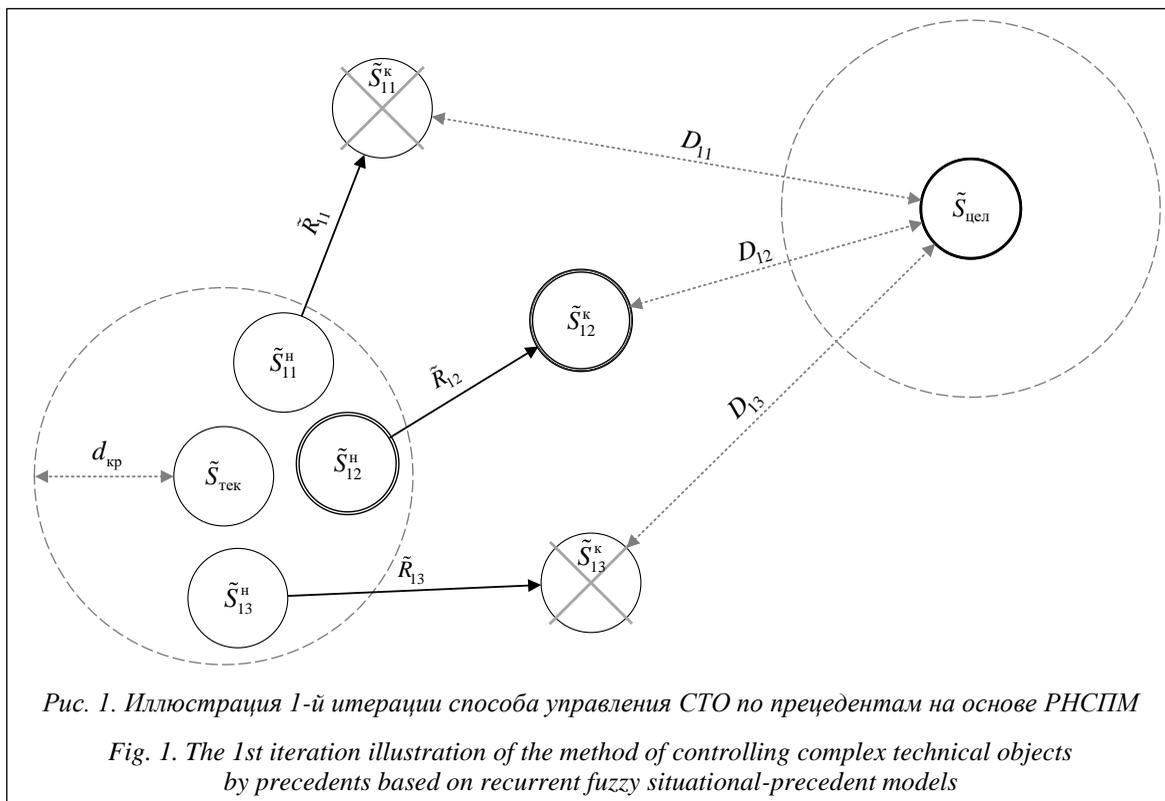
$$\tilde{c}_n : \min_{n=1, \dots, N} \{D_n\},$$

$$D_n = d_E(\tilde{S}_{тек}, \tilde{S}_n^h) + d_E(\tilde{S}_n^k, \tilde{S}_{цел}), \quad (5)$$

где  $n$  – номер прецедента в БНП,  $n = 1, \dots, N$ ,  $N$  – количество прецедентов в БНП;  $\tilde{S}_n^h, \tilde{S}_n^k$  – начальная и конечная ситуации  $n$ -го нечеткого прецедента соответственно;  $d_E$  – мера близости (например, нечеткое расстояние Евклида).

Этап 4. Итерационный поиск прецедентов и применение соответствующих управляющих решений до достижения  $\tilde{S}_{цел}$ . При этом на очередной итерации в качестве  $\tilde{S}_{тек}$  рассматривается конечная ситуация последнего из найденных прецедентов.

На рисунках 1 и 2 изображены итерации предложенного способа; для наглядности используется двойная индексация нечетких ситуаций  $\tilde{S}_{ij}$  и расстояний  $D_{ij}$ , где  $i$  – номер итерации при поиске прецедента;  $j$  – порядковый номер прецедента.



На рисунках 3 и 4 проиллюстрирован поиск последовательности прецедентов с наименьшим количеством прецедентных управляющих решений.

Предложенный способ может быть применен и для случаев, если критерием эффективности управления СТО является минимизация числа управляющих решений для перехода в

целевую ситуацию без учета сопутствующих затрат, например, при игнорировании переходов СТО через промежуточные нежелательные ситуации, соответствующие аварийным режимам работы.

Отдельно следует рассмотреть случай, когда имеются прецеденты, у которых начальные ситуации и управляющие решения близки,

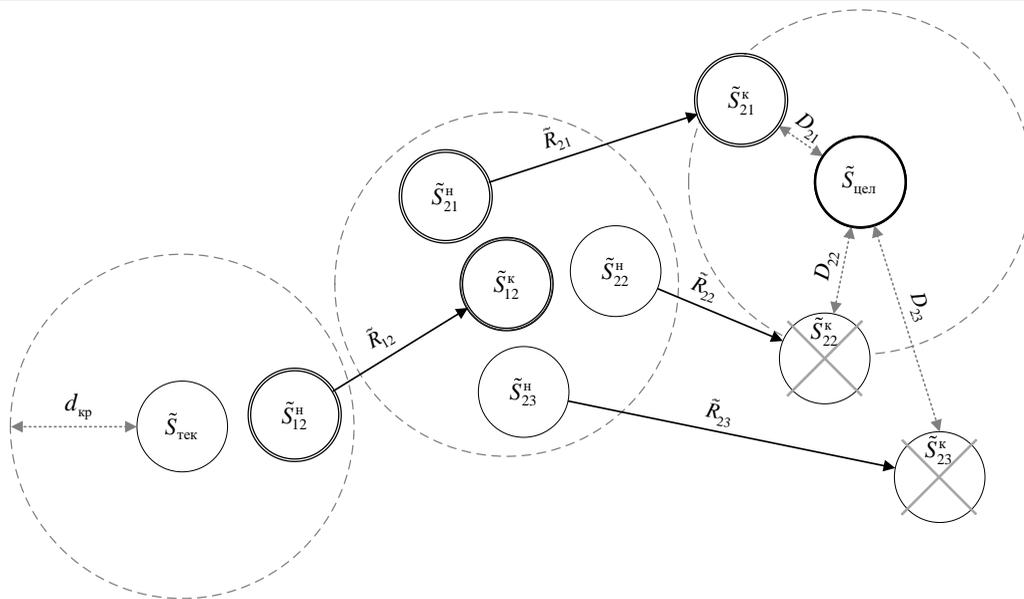


Рис. 2. Иллюстрация 2-й итерации способа управления СТО по прецедентам на основе РНСПМ

Fig. 2. The 2nd iteration illustration of the method of controlling complex technical objects by precedents based on recurrent fuzzy situational-precedent models

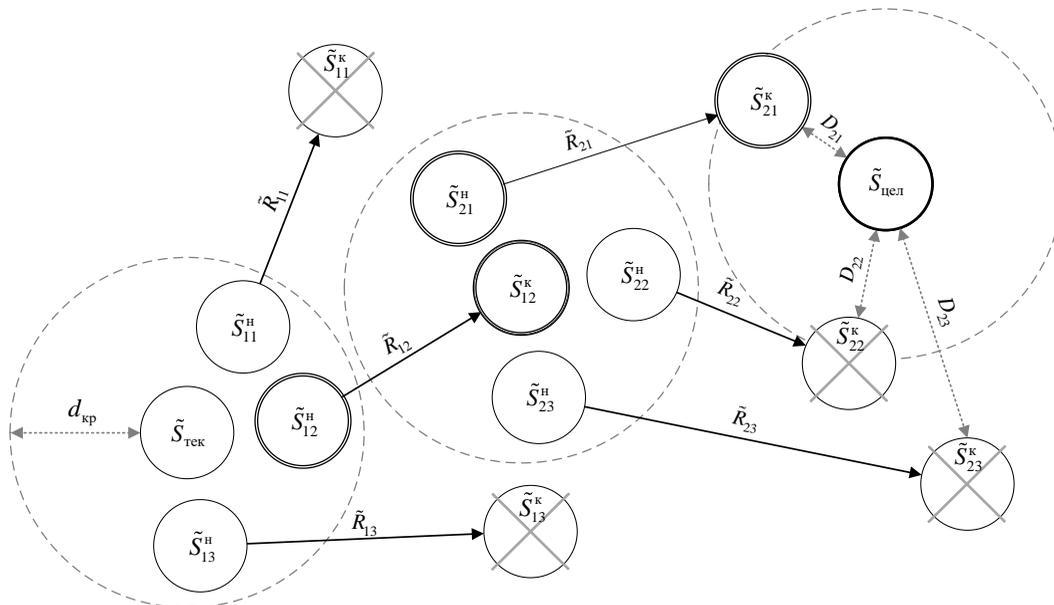


Рис. 3. Иллюстрация поиска последовательности прецедентов с наименьшим количеством прецедентных управляющих решений

Fig. 3. Example of searching for a sequence of precedents with the least number of precedent decisions

а конечные ситуации различны. Далее будем называть такие прецеденты связанными. Выбор и применение одного из связанных прецедентов не исключает возможности перехода СТО в незапланированные или даже нежелательные ситуации. Пример подобной ситуации приведен на рисунке 5.

Предположим, установлено, что при воздействии одного и того же управляющего решения  $\tilde{R}_n$  или группы схожих управляющих решений  $\{\tilde{R}_{n,h}\}$ , где  $\tilde{R}_{n,1} \cong \tilde{R}_{n,2} \cong \dots \cong \tilde{R}_{n,H}$ , возможны переходы из ситуации  $\tilde{S}_n$  сразу в несколько ситуаций  $\{\tilde{S}_h\}$ ,  $h = 1, \dots, H$ . Считаем, что переход в незапланированную ситуацию (если она не является нежелательной) будет допустимым, тогда расстояние для всей группы связанных прецедентов может быть определено с использованием выражения

$$D_n^{(H)} = (D_1 \tilde{\tau} \tilde{Z}(\tilde{S}_n, \tilde{R}_{n1})) \tilde{\tau} \dots \tilde{\tau} (D_h \tilde{\tau} \tilde{Z}(\tilde{S}_n, \tilde{R}_{nh})) \tilde{\tau} \dots \tilde{\tau} (D_H \tilde{\tau} \tilde{Z}(\tilde{S}_n, \tilde{R}_{nH})), \quad (6)$$

где  $H$  – общее число связанных прецедентов в группе;  $D_h$  – расстояние до  $h$ -го прецедента,

вычисляемое по (5);  $\tilde{Z}(\tilde{S}_n, \tilde{R}_{nh})$  – нечеткая возможность перехода в нечеткую ситуацию  $\tilde{S}_h$ ; « $\tilde{\tau}$ », « $\tilde{\rightarrow}$ » – операции нечеткого умножения и сложения соответственно.

Выражение (6) является обобщением выражения (5), которое используется для случая несвязанных прецедентов.

Если хотя бы один из связанных прецедентов предполагает возможность перехода в нежелательную ситуацию, все эти прецеденты исключаются из дальнейшего рассмотрения. Пример исключения связанных прецедентов с нежелательной ситуацией  $\tilde{S}_{14}$  приведен на рисунке 6.

Для повышения оперативности реагирования СТО на аварийные и опасные ситуации значения ситуационных признаков целевой ситуации  $\tilde{S}_{\text{цел}}$  и предельное значение  $d_{\text{кр}}$  могут быть заранее согласованы в соответствии с требованиями к обеспечению безопасности СТО (см. 2-й этап способа). Также может быть определено множество допустимых целевых ситуа-

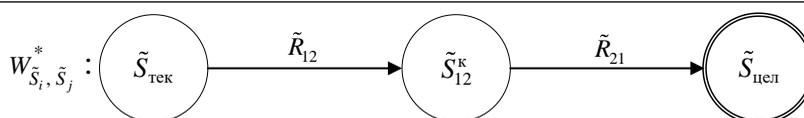


Рис. 4. Результат поиска последовательности прецедентов с наименьшим количеством прецедентных решений

Fig. 4. Search result for the sequence of precedents with the smallest number of precedent solutions

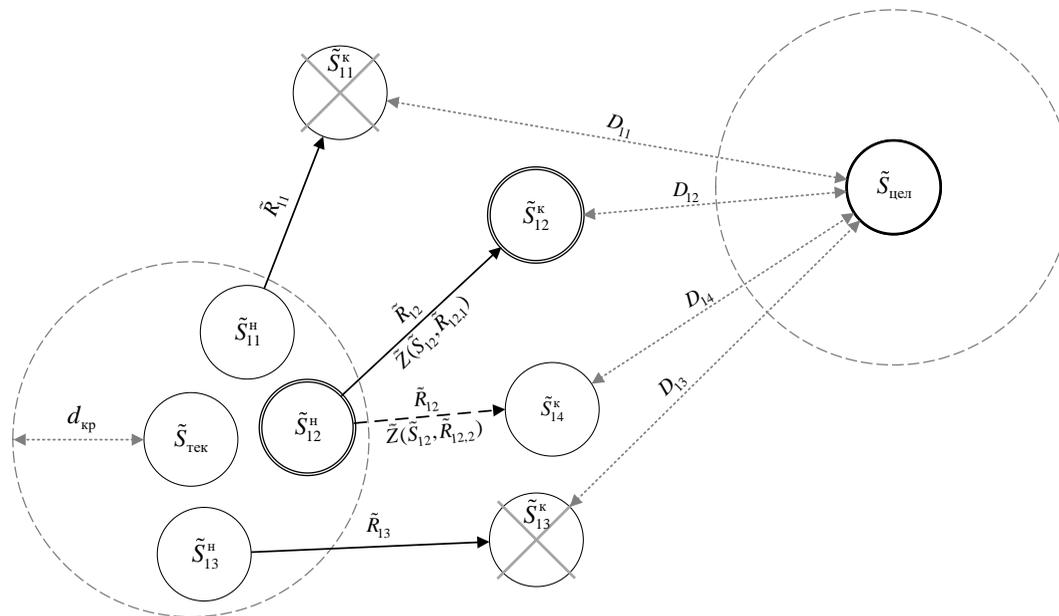


Рис. 5. Пример случая с возможностью перехода в незапланированную ситуацию

Fig. 5. Case example with the possibility of transition to an unplanned situation

ций, в которые возможен перевод СТО при возникновении аварийных и опасных ситуаций:

$$\{\tilde{S}_{\text{цел}}^1, \dots, \tilde{S}_{\text{цел}}^l, \dots, \tilde{S}_{\text{цел}}^L\},$$

где  $l = 1, \dots, L$ ,  $L$  – число допустимых целевых ситуаций.

Особенностью данного варианта является то, что выбор предпочтительной целевой ситуации будет выполняться не на основе выбранного критерия эффективности управления СТО, а на результатах оценки степени ее бли-

зости к текущей ситуации. С учетом этого на первом этапе способа дополнительно может осуществляться выбор допустимой целевой ситуации, наиболее близкой к текущей  $\tilde{S}_{\text{тек}}$ :

$$\tilde{S}_{\text{цел}} : \min_{l=1, \dots, L} (d(\tilde{S}_{\text{тек}}, \tilde{S}_{\text{цел}}^l)).$$

На рисунке 7 проиллюстрирован выбор из множества допустимых целевых ситуаций.

Применение предложенного способа управления СТО с определением набора допустимых целевых ситуаций целесообразно также, если

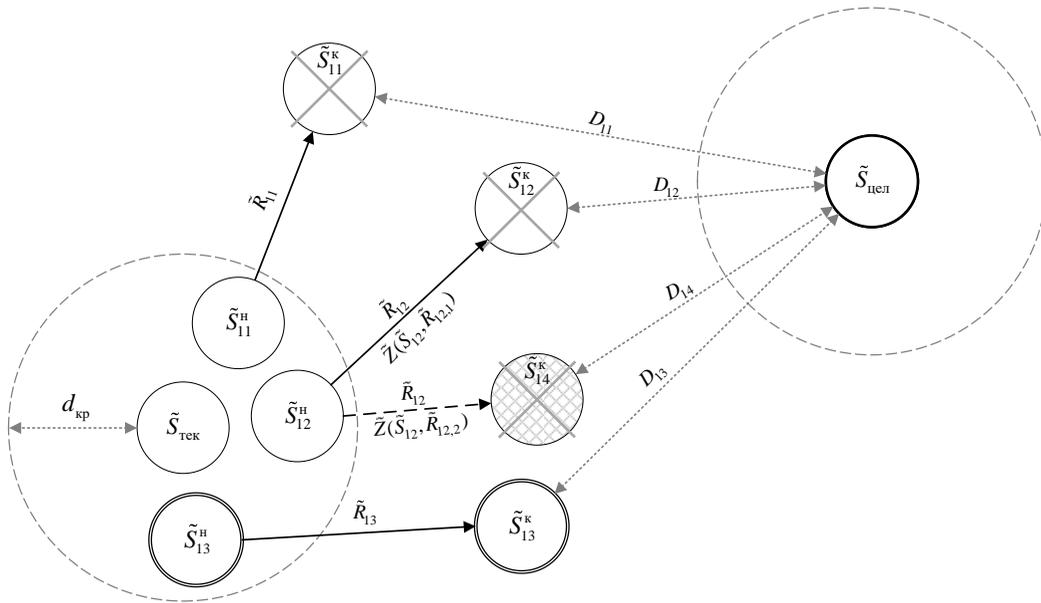


Рис. 6. Пример исключения связанных прецедентов с нежелательной ситуацией

Fig. 6. Exclusion example of related precedents with undesirable situation

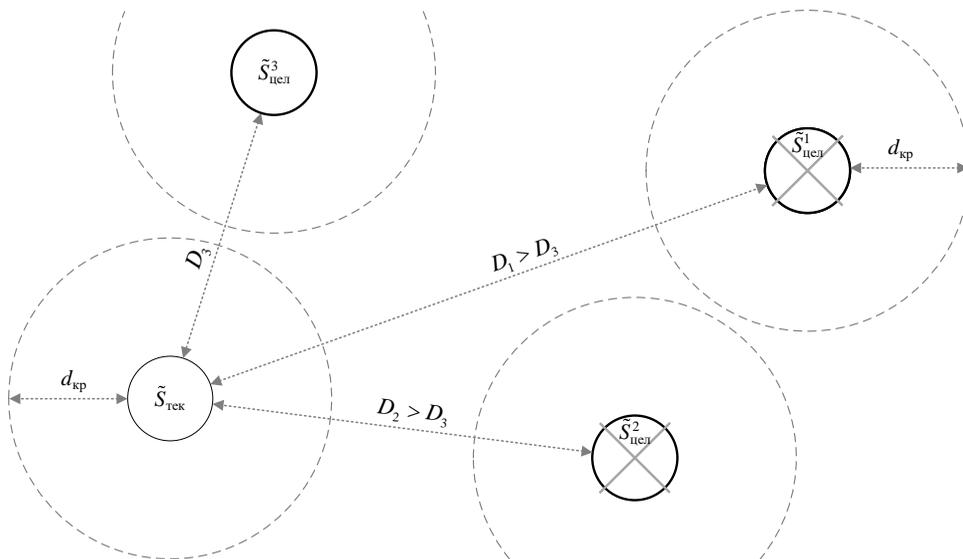


Рис. 7. Пример выбора из множества нечетких целевых ситуаций

Fig. 7. Choice example from a set of fuzzy target situations

в качестве нечетких ситуационных признаков, помимо показателей самих СТО, используются параметры, существенно зависящие от динамически изменяющихся факторов внешней среды.

### Алгоритмическая и программная реализация

На рисунке 8 приведена блок-схема алгоритма предложенного способа оперативного управления СТО по прецедентам на основе РНСПМ с критерием эффективности управления, заключающимся в минимизации числа управляющих решений для достижения целевой ситуации.

Исходными данными алгоритма является информация о нечеткой текущей ситуации  $\tilde{S}_{\text{тек}}$  и нечеткой целевой ситуации  $\tilde{S}_{\text{цел}}$ , в которую необходимо перевести СТО. Рассматривается пример алгоритма, в котором целевая ситуация задается в явном виде, при этом при необходимости может быть реализован дополнительный этап выбора ближайшей ситуации из набора  $\{\tilde{S}_{\text{цел}}^1, \dots, \tilde{S}_{\text{цел}}^l, \dots, \tilde{S}_{\text{цел}}^L\}$  допустимых целевых ситуаций. Также задается предельное значение  $d_{\text{кр}}$  степени близости ситуаций.

В блоке 4 алгоритма в качестве начальной устанавливается текущая нечеткая ситуация. Далее последовательно для каждого прецедента  $\tilde{c}_n, n = 1, \dots, N$ , из БНП определяются связанные прецеденты  $\{\tilde{c}_h\}, h = 1, \dots, H$ , то есть такие, у которых начальные ситуации  $\tilde{S}_h^n$  близки к начальной ситуации  $\tilde{S}_n^n$  рассматриваемого  $n$ -го прецедента, а схожие управляющие решения  $\tilde{R}_h^n \cong \tilde{R}_n^n$  приводят к различным конечным ситуациям  $\tilde{S}_h^k \not\cong \tilde{S}_n^k$ . Если связанные прецеденты не найдены, то для прецедента  $\tilde{c}_n$  определяется значение  $D_n$  (блок 11), вычисляемое в соответствии с выражением (5). Если для рассматриваемого  $n$ -го прецедента найдены связанные прецеденты, то для каждого из них, помимо вычисления значения  $D_h$  (блок 12), дополнительно оценивается возможность перехода  $\tilde{Z}_h$  (блок 10) в соответствии с выражением (4).

В блоке 13 рассчитывается скорректированное значение  $D$  с учетом риска перехода в незапланированные ситуации на основе обобщенной формулы (6), при этом, если связанные прецеденты не были найдены, возможность перехода устанавливается равной единице.

Если полученное значение  $D$  является минимальным, информация об управляющем решении и конечной ситуации данного прецедента сохраняется.

После рассмотрения всех прецедентов в БНП и определения наилучшего управляющего решения последнее включается в искомую последовательность. Затем осуществляется проверка достижения целевой ситуации (блок 19) в соответствии с выражением  $d(\tilde{S}_n^k, \tilde{S}_{\text{цел}}) \leq d_{\text{кр}}$ . Если целевая ситуация не была достигнута, то в качестве новой начальной ситуации устанавливается конечная ситуация  $\tilde{S}_n^k$  выбранного прецедента и алгоритм переходит к следующей итерации (блок 5). В противном случае осуществляется возврат сформированной последовательности управляющих решений и алгоритм завершает работу.

Для проверки адекватности и оценки эффективности предложенных РНСПМ и способа управления СТО по прецедентам разработаны программные средства, включающие в себя программные модули формирования БНП и прецедентного управления.

Реализация программных средств выполнена на языке Python 3.12.1, сильной стороной которого является гибкий механизм реализации объектно-ориентированного подхода к написанию кода, что позволяет значительно упростить процесс проектирования и поддержки сложных программных систем. Также важным преимуществом языка Python является наличие большого числа сторонних расширений и библиотек с открытым исходным кодом для решения широкого класса задач. Для осуществления математических операций с аппаратом нечеткой арифметики разработано собственное решение на основе библиотеки NumPy, предоставляющей поддержку векторных и матричных операций, а также возможность работы с многомерными массивами данных. Указанная библиотека реализована на языке C, что позволяет повысить скорость вычислений. Для формирования и администрирования БНП, взаимодействия с программными модулями, обмена данными применялась объектно-реляционная СУБД PostgreSQL.

### Результаты вычислительных экспериментов

С использованием разработанных программных средств проведены вычислительные эксперименты, целью которых явились про-

верка адекватности предложенных РНСПМ и оценка эффективности их применения. С помощью разработанной цифровой модели последовательного робота-манипулятора [19] была получена выборка из 1 000 опытов для формирования БНП. Эксперимент предполагал моделирование захвата деталей при трех различных начальных ситуациях:  $\tilde{S}_1$  – функциони-

рование в нормальных условиях;  $\tilde{S}_2$  – работа в условиях ограниченной видимости;  $\tilde{S}_3$  – наличие внешних воздействий на привод последовательного робота-манипулятора. Для сравнения использовалась экспертная НСМ типа «ситуация–действие», нечеткая база знаний которой сформирована на основе экспертной информации.

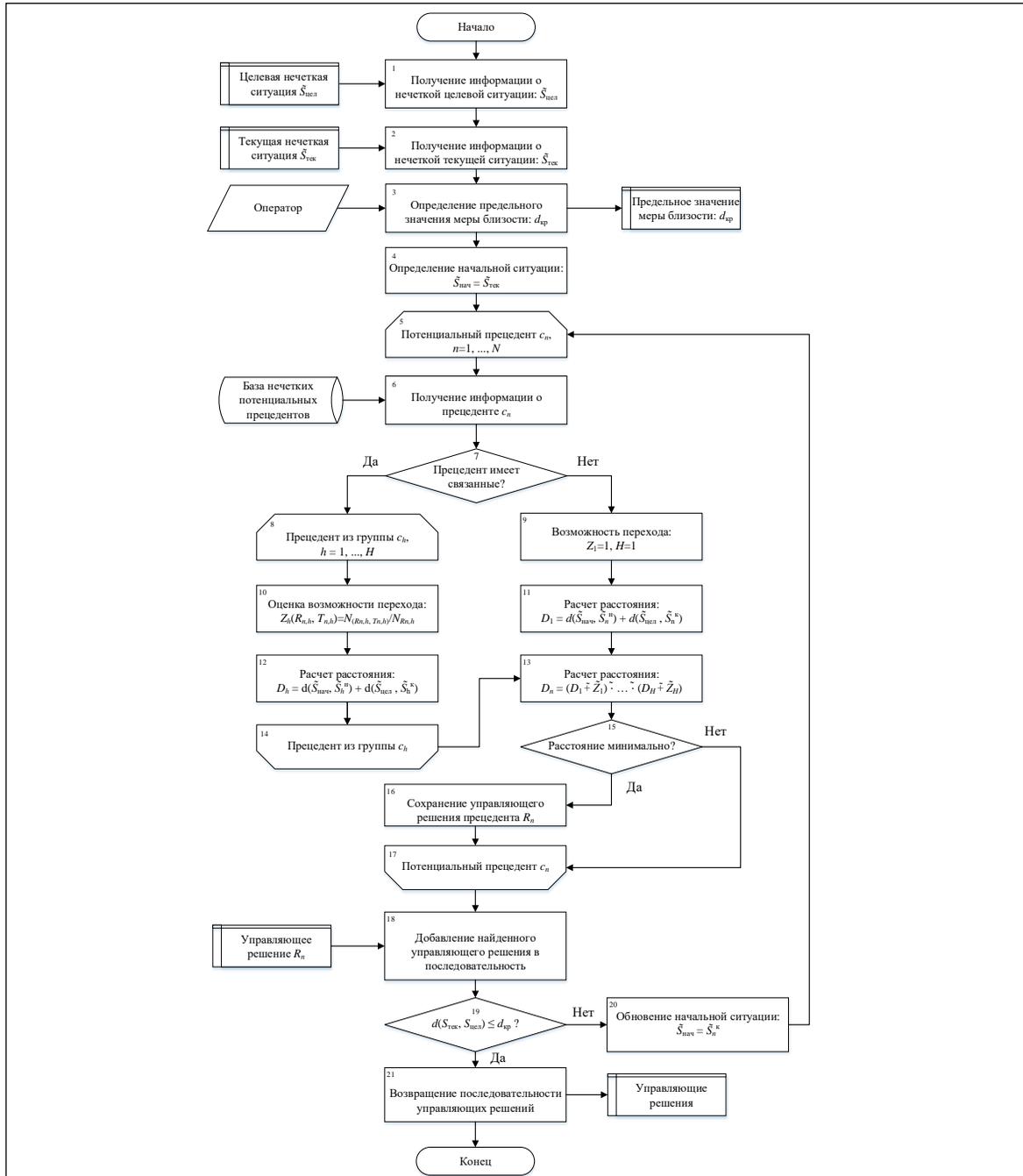


Рис. 8. Алгоритм управления СТО по прецедентам на основе РНСПМ

Fig. 8. Algorithm of controlling complex technical objects by precedents on the basis of recurrent fuzzy situational-precedent models

В таблице приведены результаты сравнительного анализа применения предложенной РНСПМ и экспертной НСМ типа «ситуация–действие» для различных вариантов начальной ситуации.

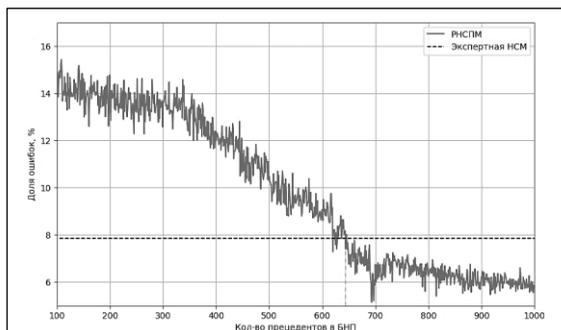
**Доля успешных захватов детали последовательного робота-манипулятора, %**

**Percentage of successful a detail by a robot-manipulator, %**

Начальная нечеткая ситуация	Экспертная НСМ «ситуация–действие»	РНСПМ
$\tilde{S}_1$	96	97
$\tilde{S}_2$	83	94
$\tilde{S}_3$	71	93

Из таблицы видно, что РНСПМ обеспечивает более высокую долю успешных захватов последовательного робота-манипулятора в условиях неопределенности и при возможности различной реакции СТО на одинаковые управляющие воздействия.

На рисунке 9 показано изменение эффективности применения РНСПМ при увеличении числа прецедентов в БНП (начальная ситуация  $\tilde{S}_2$ ). Для сравнения также приведены ре-



*Рис. 9. Доля ошибок при захвате детали роботом-манипулятором*

*Fig. 9. The percentage of errors when capturing a detail by a robot-manipulator*

зультаты применения экспертной НСМ типа «ситуация–действие» (которые не зависят от накопления прецедентов).

На основании проведенных экспериментов можно сделать вывод, что использование разработанных РНСПМ и способа управления СТО по прецедентам на базе этих моделей обеспечивает повышение эффективности управления последовательным роботом-манипулятором и при превышении определенного числа сформированных прецедентов (на рисунке 9 этому значению соответствуют 643 прецедента) обеспечивает лучшие результаты по сравнению с аналогом.

### Заключение

Для повышения эффективности оперативного управления СТО в условиях неопределенности предлагаются новый тип РНСПМ, а также способ и алгоритмы управления СТО по прецедентам на основе таких моделей.

Предложенный способ оперативного управления СТО по прецедентам отличается использованием итерационной процедуры поиска последовательности прецедентных управляющих решений. Он также обеспечивает повышение эффективности функционирования СТО при условии адаптации к изменению внутренних и внешних факторов.

Разработаны алгоритмические и программные средства, реализующие предложенные модели, способ и алгоритмы управления СТО.

В результате вычислительных экспериментов (на примере управления последовательным роботом-манипулятором) подтверждена адекватность предложенных моделей и выполнена оценка эффективности оперативного управления СТО по прецедентам с использованием созданных моделей, способа и алгоритмов в условиях неопределенности по сравнению с экспертной нечеткой ситуационной моделью типа «ситуация–действие».

### Список литературы

1. Филаретов В.Ф., Климчик А.С., Юхимец Д.А., Коноплин А.Ю., Зуев А.В. Интеллектуальные подводные робототехнические комплексы для выполнения осмотровых и технологических операций на морских газопроводах и добычных комплексах // Газовая промышленность. 2020. № 8. С. 30–38.
2. Балабанов А.Н., Безуглая А.Е., Шушляпин Е.А. Управление манипулятором подводного робота // Информатика и автоматизация. 2021. Т. 20. № 6. С. 1307–1332. doi: 10.15622/ia.20.6.5.
3. Тимофеев А.Н., Ракшин Е.А., Заруцкий Н.В. Робототехническая система для ремонта и мониторинга труднодоступного оборудования // Наука и бизнес: пути развития. 2022. № 4. С. 193–196.
4. Мещанов А.С., Губаева Э.И. Управление манипуляторами летательных аппаратов на скользящих режимах при инерционности приводов и неинвариантности к возмущениям // Вестн. Технологического университета. 2020. Т. 23. № 7. С. 100–108.

5. Косицын А.А., Богомолов А.С., Кушников В.А. Задача снижения вероятности развития аварийных комбинаций событий в беспилотных авиационных системах // Прикладная информатика. 2024. Т. 19. № 4. С. 18–34. doi: 10.37791/2687-0649-2024-19-4-18-34.
6. Игнатъев В.В. Компенсация разных типов неопределенностей при управлении техническим объектом с помощью интеллектуальных регуляторов // Программные продукты и системы. 2023. Т. 36. № 3. С. 423–431. doi: 10.15827/0236-235X.143.423-431.
7. Еремеев А.П., Сергеев М.Д., Петров В.С. Интеграция методов обучения с подкреплением и нечеткой логики для интеллектуальных систем реального времени // Программные продукты и системы. 2023. Т. 36. № 4. С. 600–606. doi: 10.15827/0236-235X.144.600-606.
8. Goharimanesh M., Mehrkish A., Janabi-Sharifi F. A fuzzy reinforcement learning approach for continuum robot control. *J. of Intelligent & Robotic Systems*, 2020, vol. 100, pp. 809–826. doi: 10.1007/s10846-020-01237-6.
9. Пучков А.Ю., Дли М.И., Тиндова М.Г. Метод решения обратной задачи кинематики на основе обучения с подкреплением при управлении роботами-манипуляторами // Прикладная информатика. 2023. Т. 18. № 6. С. 120–133.
10. Борисов В.В. Нечеткие когнитивные модели как основа для исследования сложных систем и процессов // Речевые технологии. 2020. № 1-2. С. 48–62. doi: 10.58633/2305-8129\_2020\_1-2\_48.
11. Chen C.-T., Chiu Y.-T. A study of dynamic fuzzy cognitive map model with group consensus based on linguistic variables. *Technological Forecasting and Social Change*, 2021, vol. 171, art. 120948. doi: 10.1016/j.techfore.2021.120948.
12. Nair A., Reckien D., van Maarseveen M.F.A.M. A generalised fuzzy cognitive mapping approach for modelling complex systems. *Applied Soft Computing*, 2019, vol. 84, art. 105754. doi: 10.1016/j.asoc.2019.105754.
13. Борисов В.В., Авраменко Д.Ю. Нечеткое ситуационное управление сложными системами на основе их композиционного гибридного моделирования // Системы управления, связи и безопасности. 2021. № 3. С. 207–237.
14. Starodubtsev M., Bilousov M., Kamardin S., Makarenko H., Shybanov S. Models of technological processes management under conditions of uncertainty. *J. of Natural Sci. and Tech.*, 2024, vol. 3, no. 1, pp. 254–258.
15. Добрынин А.С., Гудков М.Ю., Койнов Р.С. Прецедентный подход к управлению инцидентами в автоматизированных системах управления технологическими процессами // Программные системы и вычислительные методы. 2020. № 2. С. 45–52. doi: 10.7256/2454-0714.2020.2.31040.
16. Черновалова М.В. Нечеткие прецедентные модели для управления проектами с использованием мультионтологического подхода // Прикладная информатика. 2021. Т. 16. № 2. С. 4–16. doi: 10.37791/2687-0649-2021-16-2-4-16.
17. Дли М.И., Черновалова М.В., Соколов А.М., Моргунова Э.В. Нечеткая динамическая онтологическая модель для поддержки принятия решений по управлению энергоемкими системами на основе прецедентов // Прикладная информатика. 2023. Т. 18. № 5. С. 59–76. doi: 10.37791/2687-0649-2023-18-5-59-76.
18. Мелихов А.Н., Берштейн Л.С., Коровин С.Я. Ситуационные советующие системы с нечеткой логикой. М.: Наука, 1990. 272 с.
19. Борисов В.В., Соколов А.М., Жарков А.П., Култыгин О.П. Решение обратной задачи кинематики для последовательных роботов манипуляторов на основе нечетких численных методов // Прикладная информатика. 2022. Т. 17. № 4. С. 113–126. doi: 10.37791/2687-0649-2022-17-4-113-126.

Software &amp; Systems

doi: 10.15827/0236-235X.148.554-565

2024, 37(4), pp. 554–565

### Recurrent fuzzy situational-precedent models for operational control of complex technical objects

Andrey M. Sokolov<sup>1</sup>, Vadim V. Borisov<sup>2</sup>✉<sup>1</sup> National Research University “MPEI”, Moscow, 111250, Russian Federation<sup>2</sup> Smolensk Branch of the Moscow Power Engineering Institute, Smolensk, 214013, Russian Federation

#### For citation

Sokolov, A.M., Borisov, V.V. (2024) ‘Recurrent fuzzy situational-precedent models for operational control of complex technical objects’, *Software & Systems*, 37(4), pp. 554–565 (in Russ.). doi: 10.15827/0236-235X.148.554-565

#### Article info

Received: 14.06.2024

After revision: 17.07.2024

Accepted: 30.07.2024

**Abstract.** The paper deals with the issues of operational management of complex technical objects (CTO) functioning in dynamically changing, uncertain environmental conditions, risks of dangerous and emergency situations. The authors analyzed the approaches to situational management of CTO. The analysis showed that a promising solution for operational management of these objects is the use of fuzzy models of the “Situation-Action” type. The main limitation of these models is the need for additional expert information to update the fuzzy knowledge base, as well as the lack of consideration of transition to different situations when applying similar control actions. To eliminate these disadvantages, the authors propose the application of the precedent approach. It involves accumulation of information about the CTO functioning in the base of fuzzy precedents for the purpose of their further use in the choice of management decisions. The authors proposed a new type of recurrent fuzzy situational-precedent models, as well as a method and algorithms for operational management of CTO by precedents based on these models. The author's solution differs in the application of iterative procedure of search and application of a sequence of precedent control decisions. This procedure provides an increase in the efficiency of CTO functioning under the condition of adaptation to changes in internal and external factors. The authors designed algorithmic and software tools that implement the proposed models, method and algorithms of CTO control. In the course of computational experiments, they confirmed the adequacy of the proposed models. The authors evaluated the effectiveness of oper-

ational CTO management by precedents using the created models, method and algorithms in conditions of uncertainty in comparison with the expert fuzzy situational network of the “Situation-Action” model.

**Keywords:** recurrent fuzzy situational-precedent model, fuzzy situational network of the “Situation-Action” model, precedent, operational management

**Acknowledgements.** The paper was carried out under the government assignment, project no. FSWF-2023-0012

### References

1. Filaretov, V.F., Klimchik, A.S., Yukhimets, D.A., Konoplin, A.Yu., Zuev, A.V. (2020) ‘Intelligent underwater robotics for inspection and process operations on offshore gas pipelines and production systems’, *Gas Industry J.*, (8), pp. 30–38 (in Russ.).
2. Balabanov, A.N., Bezuglaya, A.E., Shushlyapin, E.A. (2021) ‘Underwater robot manipulator control’, *Informatics and Automation*, 20(6), pp. 1307–1332 (in Russ.). doi: 10.15622/ia.20.6.5.
3. Timofeev, A.N., Rakshin, E.A., Zarutskiy, N.V. (2022) ‘A robotic system for repair and monitoring of hard-to-reach equipment’, *Science and Business: Ways of Development*, (4), pp. 193–196 (in Russ.).
4. Meshchanov, A.S., Gubaeva, E.I. (2020) ‘Control of aircraft manipulators in sliding modes with drive inertia and non-invariance to perturbations’, *Herald of Technological University*, 23(7), pp. 100–108 (in Russ.).
5. Kositsyn, A.A., Bogomolov, A.S., Kushnikov, V.A. (2024) ‘The task of reducing the probability of developing combinations of emergency events in unmanned aerial vehicles’, *J. of Applied Informatics*, 19(4), pp. 18–34 (in Russ.). doi: 10.37791/2687-0649-2024-19-4-18-34.
6. Ignatyev, V.V. (2023) ‘Compensation of various types of uncertainties when controlling a technical object using intelligent controllers’, *Software & Systems*, 36(3), pp. 423–431 (in Russ.). doi: 10.15827/0236-235X.143.423-431.
7. Eremeev, A.P., Sergeev, M.D., Petrov, V.S. (2023) ‘Integration of reinforcement learning methods and fuzzy logic for intelligent real-time systems’, *Software & Systems*, 36(4), pp. 600–606 (in Russ.). doi: 10.15827/0236-235X.144.600-606.
8. Goharimanesh, M., Mehrkish, A., Janabi-Sharifi, F. (2020) ‘A fuzzy reinforcement learning approach for continuum robot control’, *J. of Intelligent & Robotic Systems*, 100, pp. 809–826. doi: 10.1007/s10846-020-01237-6.
9. Puchkov, A.Yu., Dli, M.I., Tindova, M.G. (2023) ‘A method for solving the inverse kinematics problem based on reinforcement learning for controlling robotic manipulators’, *J. of Applied Informatics*, 18(6), pp. 120–133 (in Russ.).
10. Borisov, V.V. (2020) ‘Fuzzy cognitive models for the research of complex systems and processes’, *Speech Technology*, (1-2), pp. 48–62 (in Russ.). doi: 10.58633/2305-8129\_2020\_1-2\_48.
11. Chen, C.-T., Chiu, Y.-T. (2021) ‘A study of dynamic fuzzy cognitive map model with group consensus based on linguistic variables’, *Technological Forecasting and Social Change*, 171, art. 120948. doi: 10.1016/j.techfore.2021.120948.
12. Nair, A., Reckien, D., van Maarseveen, M.F.A.M. (2019) ‘A generalised fuzzy cognitive mapping approach for modelling complex systems’, *Applied Soft Computing*, 84, art. 105754. doi: 10.1016/j.asoc.2019.105754.
13. Borisov, V.V., Avramenko, D.Yu. (2021) ‘Fuzzy situational control of complex systems based on their composite hybrid modeling’, *Systems of Control, Communication and Security*, (3), pp. 207–237 (in Russ.).
14. Starodubtsev, M., Bilousov, M., Kamardin, S., Makarenko, H., Shybanov, S. (2024) ‘Models of technological processes management under conditions of uncertainty’, *J. of Natural Sci. and Tech.*, 3(1), pp. 254–258.
15. Dobrynin, A.S., Gudkov, M.Yu., Koynov, R.S. (2020) ‘A precedent approach to incident management in automated process control systems’, *Software Systems and Computational Methods*, (2), pp. 45–52 (in Russ.). doi: 10.7256/2454-0714.2020.2.31040.
16. Chernovalova, M.V. (2021) ‘Fuzzy case models for project management using a multi-ontology approach’, *J. of Applied Informatics*, 16(2), pp. 4–16 (in Russ.). doi: 10.37791/2687-0649-2021-16-2-4-16.
17. Dli, M.I., Chernovalova, M.V., Sokolov, A.M., Morgunova, E.V. (2023) ‘Fuzzy dynamic ontological model for decision support of energy-intensive systems management based on precedents’, *J. of Applied Informatics*, 18(5), pp. 59–76 (in Russ.). doi: 10.37791/2687-0649-2023-18-5-59-76.
18. Melikhov, A.N., Bershteyn, L.S., Korovin, S.Ya. (1990) *Situational Advising Systems with Fuzzy Logic*. Moscow, 272 p. (in Russ.).
19. Borisov, V.V., Sokolov, A.M., Zharkov, A.P., Kulygin, O.P. (2022) ‘Solving the inverse kinematics problem for sequential robot manipulators based on fuzzy numerical methods’, *J. of Applied Informatics*, 17(4), pp. 113–126 (in Russ.). doi: 10.37791/2687-0649-2022-17-4-113-126.

### Авторы

**Соколов Андрей Максимович**<sup>1</sup>,  
аспирант, ansokol98@mail.ru

**Борисов Вадим Владимирович**<sup>2</sup>,  
д.т.н., профессор, vbor67@mail.ru

### Authors

**Andrey M. Sokolov**<sup>1</sup>, Postgraduate Student,  
ansokol98@mail.ru

**Vadim V. Borisov**<sup>2</sup>, Dr.Sci. (Engineering),  
Professor, vbor67@mail.ru

<sup>1</sup> Национальный исследовательский университет  
«Московский энергетический институт»,  
г. Москва, 111250, Россия

<sup>2</sup> Смоленский филиал Национального  
исследовательского университета МЭИ,  
г. Смоленск, 214013, Россия

<sup>1</sup> National Research University “MPEI”,  
Moscow, 111250,  
Russian Federation

<sup>2</sup> Smolensk Branch of the Moscow  
Power Engineering Institute,  
Smolensk, 214013, Russian Federation

## Нечеткая кластеризация как способ повышения эффективности управления в автоматических системах

А.С. Игнатьева <sup>1</sup>✉

<sup>1</sup> Южный федеральный университет, Институт радиотехнических систем и управления, г. Ростов-на-Дону, 344006, Россия

### Ссылка для цитирования

Игнатьева А.С. Нечеткая кластеризация как способ повышения эффективности управления в автоматических системах // Программные продукты и системы. 2024. Т. 37. № 4. С. 566–575. doi: 10.15827/0236-235X.148.566-575

### Информация о статье

Группа специальностей ВАК: 2.3.1

Поступила в редакцию: 18.06.2024

После доработки: 23.07.2024

Принята к публикации: 30.07.2024

**Аннотация.** Статья посвящена поиску новых способов повышения эффективности управления в автоматических системах. Предложен новый метод оптимизации процедуры синтеза интеллектуальных регуляторов в системах автоматического управления техническими объектами. С целью организации гибридного управления предлагается метод, в основе которого лежит применение алгоритмов нечеткой кластеризации, а также специально разработанные алгоритмы, реализованные в интеллектуальном регуляторе. Регулятор разработан на основе методов нечеткой логики, нейросетей и биоинспирированных методов. В рамках описания предложенного метода показаны существующие подходы к оптимизации интеллектуальных регуляторов. На концептуальном уровне предложен собственный подход для решения задачи оптимизации параметров интеллектуальных регуляторов, представлено его обобщение. В общем виде описаны модели, для которых применим разработанный метод. В частности, показаны результаты работы метода для ANFIS-системы в гибридных моделях управления. Представлены ANFIS-архитектуры до и после применения кластеризации. Для достижения полученных результатов разработана многоуровневая архитектура комбинированного управления в интеллектуальном регуляторе. Показаны структуры нечетких регуляторов при реализации ANFIS до и после применения нечеткой кластеризации. Выполнен анализ результатов проведенного исследования. В результате применения предложенного метода значительно сокращается количество функций принадлежности лингвистических переменных интеллектуального регулятора и количество управляющих правил. Это позволяет без потери качества управления и без увеличения вычислительных ресурсов снизить трудоемкость разработки интеллектуального регулятора и время его проектирования.

**Ключевые слова:** нечеткая кластеризация, система управления, технический объект, интеллектуальный регулятор, оптимизация

**Введение.** В источниках, посвященных проектированию интеллектуальных регуляторов для систем автоматического управления (САУ) особое внимание уделяется разрабатываемым подходам, методам, алгоритмам для оптимизации параметров данных регуляторов [1–3]. Известно, что ПИД-регуляторы широко применяются в САУ, но для современных систем управления, которые функционируют в условиях неопределенности, их ресурсов не всегда достаточно из-за отсутствия оптимальной настройки, наличия задержек управляющих воздействий, сложности настройки их параметров, невозможности использования качественной информации для управления объектами и т.п. [4].

Поэтому для управления объектом, который работает в недетерминированной среде, предпочтение все чаще отдается интеллектуальным регуляторам. Их основа, как правило, – нечеткая логика, нейросети и биоинспирированные методы (генетические алгоритмы). Следует отметить, что перечисленные способы построения

интеллектуальных регуляторов могут быть объединены в систему. Этим реализуется гибридное управление, способствующее наиболее эффективному функционированию объектов в условиях неопределенности.

Данная работа посвящена представлению возможностей нечеткой кластеризации для повышения эффективности управления в автоматических системах, реализованных на основе гибридного подхода.

### Обзор релевантных работ

Проблема сокращения вычислительных ресурсов при обработке больших объемов данных в современных системах автоматизации и в других отраслях обусловила разработку эффективных систем управления с применением интеллектуальных технологий, основанных на поиске методов оптимизации баз правил регуляторов. Решению этой задачи посвящено немало научных исследований.

В работе [5] авторы предлагают использовать перекрывающиеся кластеры и систему не-

четкого вывода Такаги-Сугено для обеспечения надежности работы интеллектуальных структур в системе управления. Решения направлены на применение алгоритмов кластеризации для адаптации функций принадлежности, так как эффективность работы интеллектуальных систем управления зависит от наличия неопределенности в физических параметрах, измерениях датчиков и во внешних возмущениях. Полученные результаты демонстрируют более высокую точность работы и способность к обобщению схемы управления на основе данных по сравнению с другими стандартными техническими стратегиями машинного обучения.

Изменения сигнала атмосферного электрического поля (*Atmospheric Electric Field Signal – AEFS*) тесно связаны с изменениями погоды, особенно с грозовой активностью. В работе [6] авторами впервые используется метод кластеризации нечетких *C-средних (Fuzzy C-means – FCM)* для разработки инновационного подхода к характеристике погодных атрибутов, переносимых AEFS. Обнаружение грозовой активности с помощью AEFS с точки зрения технологии нечетких множеств дает содержательное руководство для интерпретируемых гроз. Надежность предлагаемого метода проверяется путем объединения радиолокационных диаграмм и опыта экспертов.

Точное и надежное управление автоматическим регулятором напряжения (*Automatic Voltage Regulators – AVR*) необходимо для поддержания стабильного выходного напряжения синхронного генератора в изменяющихся условиях эксплуатации. Целью исследования [7] является улучшение переходных характеристик и характеристик надежности систем AVR путем разработки новой схемы управления ПИД-*LP*-регулятора на основе нечеткой логики. Для достижения этой цели авторы сначала разрабатывают новый ПИД-*LP*-регулятор, затем ПИД-*LP*-регулятор на основе нечеткой логики для адаптивной настройки параметров первого регулятора с учетом ошибки и ее изменений. Результаты подтверждают применимость и превосходство регулятора в поддержании стабильной работы при флуктуационной динамике энергосистемы с колебаниями до 50 %.

Управление процессом необходимо для обеспечения стабильной и эффективной работы систем управления температурой транскритического CO<sub>2</sub> электромобилей. *Пропорционально-интегральное (ПИ)* управление обычно вызывает недостаточную динамическую реакцию

для переменных и сложных условий эксплуатации, а затем вызывает неустойчивое управление процессом. В исследовании [8] была предложена адаптивная стратегия нечеткого ПИ-регулятора, обеспечивающая изменение коэффициентов усиления ПИ-регулятора в реальном времени для быстрого реагирования и подавления колебаний с учетом давления нагнетания компрессора, температуры охлаждающей жидкости и воздуха на выходе. Высокоточная транскритическая тепловая модель CO<sub>2</sub> в электромобилях построена на платформе Amesim для моделирования работы в реальном времени, разработан адаптивный нечеткий алгоритм ПИ-регулятора для оптимизации управления процессом. Результаты показали, что предложенный адаптивный нечеткий ПИ-регулятор может обеспечить подавление перерегулирования на 9,22 % для давления нагнетания, максимальное подавление перерегулирования на 22,29 % и сокращение времени установления на 86 с для температуры воздуха на выходе, а также сокращение времени установления на 25 % для регулирования температуры охлаждающей жидкости. Кроме того, нечеткий ПИ-регулятор улучшил характеристики запуска. Полученные результаты показали, что предлагаемый нечеткий алгоритм ПИ-регулятора может эффективно оптимизировать динамическую работу транскритической тепловой системы CO<sub>2</sub> в электромобилях.

В последнее время бесконтактные методы измерения, такие как лазерное сканирование, приобрели популярность при сборе разрозненных данных из-за их способности генерировать облака точек высокого разрешения, содержащие подробную информацию об объекте. В статье [9] авторами предложен алгоритм оптимизации, основанный на нечеткой кластеризации. Результаты показали, что оптимизированный метод справился с кластеризацией без априорного предположения о числе кластеров для удовлетворения различных требований.

Исследования по использованию различных индексов достоверности кластеров для проверки качества нечеткой кластеризации многомерных наборов данных проведены в работе [10]. По мнению авторов, валидация нечетких кластеров возможна только при применении надежного индекса достоверности кластера (*Cluster Validity Index – CVI*), который правильно распознает структуру данных, при этом его валидации должны быть независимы от любого параметра алгоритма кластеризации или свойства данных. Тем не менее, некоторые классические

нечеткие CVI, такие как коэффициент разделения, энтропия разделения и индекс Фукуяма-Сугено (*Fukuyama-Sugeno* – FS), имеют монотонную тенденцию в зависимости от количества кластеров.

Несмотря на то, что в литературе представлены обширные исследования подобной тенденции, они были проведены для низкоразмерных данных, в которых это свойство данных не влияет на поведение кластеризации. Чтобы исследовать, как такие аспекты влияют на результаты нечеткой кластеризации многомерных данных, авторы сгруппировали объекты из тринадцати реальных наборов данных, используя алгоритм С-средних. Нечеткие разделы были проверены с помощью коэффициента разделения, энтропии разделения индекса FS и с некоторыми предложенными улучшениями, чтобы привести к монотонной тенденции. Помимо анализа кластеров (в сумме восемь проанализированных нечетких CVI), был проведен тест Манна-Кендалла для статистической проверки монотонного тренда результатов CVI. Проведенные анализы показали, что индексы модифицированного коэффициента разделения и индексы энтропии с масштабированным разделением оказались успешными в улучшении по сравнению с индексами энтропии разделения и коэффициента разделения.

В настоящее время алгоритм нечеткой кластеризации широко применяют для сегментации медицинских изображений. Известен метод автоматической сегментации, который объединяет нечеткую кластеризацию с методом установки уровня через динамический ограниченный параметр [11]. Способ установки уровня широко используется в сегментации медицинских изображений, при которой производительность серьезно зависит от инициализации и настройки параметров. Предложенный авторами метод может использовать результаты нечеткой кластеризации непосредственно, что позволяет контролировать эволюцию набора уровней. Более того, добавленный ограниченный параметр постоянно меняется до получения окончательных результатов. Такой алгоритм значительно уменьшает ручную работу и приводит к более надежным результатам сегментации. Предложенный алгоритм был протестирован на некоторых медицинских изображениях, проведено сравнение с другими моделями наборов уровней и современным методом U-Net. Количественные и качественные экспериментальные результаты показывают его эффективность и очевидное улучшение для медицинской сегментации изображения.

В работе [12] авторами решается проблема разложения большой взаимосвязанной электрической сети на более мелкие слабосвязанные зоны для облегчения и гибкого управления системами передачи электроэнергии. Такое разделение обеспечивает контроль напряжения вторичного уровня на региональных уровнях и контролируемое выделение, что может быть использовано для предотвращения распространения отключений большой площади. Система передачи электроэнергии рассматривается как полностью связанный взвешенный ориентированный граф, где узлы и его граничные веса представляют соответственно шины и количественную оценку электрического подобия между любыми двумя шинами.

В отличие от мер подобия, основанных на импедансе или допуске, которые в значительной степени ограничивают и не учитывают топологию электрических сетей, электрическое сходство между любыми двумя шинами в этой работе рассматривается с точки зрения их влияния на оставшуюся часть сети. В частности, электрическое сходство между двумя шинами определяется количественно с точки зрения соответствующих колебаний напряжения на всех шинах в сети в результате возмущений реактивной мощности на этих шинах. Кроме того, количественная оценка электрического влияния не имеет существенного влияния на вычислительную сложность, поскольку она рассчитывается с использованием якобианов, полученных в качестве побочных продуктов решения уравнений потока мощности. Результирующий ориентированный граф затем группируется в заданное количество зон, которые электрически слабо связаны, с помощью теоретико-графического алгоритма кластеризации. Авторами предложена основанная на правилах стратегия децентрализованного управления для эффективного управления напряжениями шины в слабосвязанных зонах, которые получаются в результате процесса кластеризации. Предложенный подход затем тестируется на испытательных системах для таких приложений, как диспетчерское управление напряжением и изоляция, и дает хорошую идентификацию взаимно разделенных подсетей в большой электрической сети.

В работе [13] представлена новая методология прогнозирования SSOFC-Apriori-WRP, которая представляет прогнозирование энергии ветра и скорости на один день вперед. Задача вероятностного прогноза энергии ветра решается с помощью комбинированной интеллекту-

альной структуры и алгоритма нечеткой кластеризации.

Исследования характеристик эксплуатационного поведения ветроэнергетической системы с использованием метода нечеткой кластеризации для мониторинга состояния системы и для исключения ее неисправностей проведены в [14].

Новый метод проектирования систем с помощью входных данных для автоматического определения типов нечетких множеств на основе нечеткости рассмотрены в работе [15]. В нечеткой системе гибридного типа тип нечеткого множества определяется нечеткостью для повышения производительности системы, а интерпретируемость данной системы определяется целостностью, различимостью и избыточностью нечетких множеств. Во-первых, нечеткая кластеризация используется для инициализации базы правил и типов нечетких множеств. Во-вторых, метод роя частиц используется для оптимизации параметров нечеткой системы гибридного типа, а типы нечетких множеств определяются нечеткостью, что означает, что типы нечетких множеств будут динамически меняться в процессе оптимизации. Таким образом, в статье предлагается метод проектирования нечетких систем управления данными, которые могут автоматически определять типы нечетких множеств за счет введения нечеткости.

Многоуровневая нечеткая модель, основанная на кластеризации нечетких правил для задач прогнозирования, предложена в [16]. Такой подход может быть полезен к применению в системах мониторинга эксплуатации и прогноза технического состояния оборудования. В [17] для диагностики и предсказания сбоя ветряной турбины применяется нечеткая кластеризация.

В работе [18] для повышения точности стандартного алгоритма FCM применен лесной алгоритм оптимизации. Нечеткая кластеризация в данном случае представляет собой комбинацию лесного алгоритма оптимизации с градиентным методом, что позволяет получить оптимизированные кластерные центры.

Задача декомпозиции больших электросетей на малые и слабосвязанные с целью упрощения процесса управления системами передачи электроэнергии рассматривается в [12]. В экспертной системе управления на основе правил применена нечеткая кластеризация для проектирования локальных управляющих действий во время перегрузки, недогрузки и разделения.

Представленный обзор показывает широкое распространение в различных областях методов управления, основанных на применении кластеризации, что говорит об актуальности проведенных в настоящей работе исследований.

### **Разработка метода оптимизации процедуры синтеза интеллектуальных регуляторов в САУ техническими объектами**

Анализ существующих подходов к оптимизации управления в САУ позволил сделать вывод о том, что в настоящее время распространение получили комбинированные подходы, которые в своей работе объединяют классическую теорию управления, нечеткую логику, нейросетевые технологии и генетические алгоритмы [19]. Данный комбинированный метод при своей эффективности требует анализа и обработки больших объемов объективной и субъективной информации, что, в свою очередь, влечет за собой необходимость использования повышенных вычислительных мощностей для сокращения времени производимых вычислений.

В результате эффективность работы САУ значительно снижается. Устранить данную проблему предлагается с помощью нечеткой кластеризации с разработанными алгоритмами ее реализации, которая позволяет оптимизировать параметры интеллектуальных регуляторов на основе нечеткого кластерного анализа. В настоящее время для повышения качества предлагается оптимизация интеллектуальных регуляторов (<http://www.swsys.ru/uploaded/image/2024-4/9.jpg>).

В данной работе предлагается метод для решения задачи оптимизации параметров интеллектуальных регуляторов, графически представленный на рисунке 1.

С учетом того, что генетические алгоритмы часто могут выполняться на стадии обучения нейросетей и с учетом модели, на которой выполняется апробация предложенного метода, решение задачи оптимизации параметров интеллектуальных регуляторов приобретает вид, представленный на рисунке 2.

В используемой для апробации модели управления техническими объектами заложен определенный принцип взаимодействия. Реализуется гибридное управление техническим объектом за счет применения интеллектуального регулятора на основе классического регулятора, нечеткого регулятора и нейросети. Его отличительной особенностью является то,

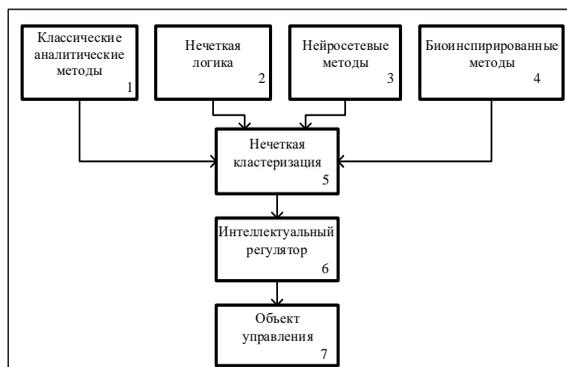


Рис. 1. Решение задачи оптимизации параметров интеллектуальных регуляторов

Fig. 1. Solving the problem of optimization of intelligent controller parameters



Рис. 2. Решение в обобщенном виде

Fig. 2. General solution

что классический регулятор – это источник информации для нечеткого регулятора, который обучает нейросеть, используя информацию с классического.

Подробное описание гибридной модели управления приведено в [20]. Применяв нечеткую кластеризацию (получаемую путем обучения нечеткого регулятора с помощью нейросети данными с классического регулятора) со специально разработанными алгоритмами для подобных моделей, удастся достичь существенной оптимизации синтезируемой адаптивной системы нейро-нечеткого вывода (*Adaptive Neuro-fuzzy Inference System – ANFIS*).

В общем виде ANFIS-архитектура выглядит несколько громоздкой (<http://www.swsys.ru/uploaded/image/2024-4/10.jpg>). После применения разработанного автором метода ANFIS-архитектура упрощается, отличаясь существенным сокращением количества функций принадлежности для каждой из входных линг-

вистических переменных системы нечеткого вывода и значительным сокращением количества продукционных правил (рис. 3).

Достижение представленных результатов оказалось возможным благодаря разработанной автором многоуровневой архитектуре комбинированного управления в интеллектуальном регуляторе (рис. 4).

Рассмотрим подробнее представленную архитектуру. На первом уровне управление реализуется с помощью классического регулятора с рассчитанными параметрами и объектом управления с известной (рассчитанной) моделью [21]. Для управления объектом в условиях неопределенности на втором этапе синтезируется нечеткий регулятор на основе данных с классического регулятора (используются значения сигналов отклонения, интеграла или дифференциала отклонения для разработки входных лингвистических переменных и сигналы управляющего воздействия для разработки выходной лингвистической переменной) [22]. Формирование базы правил нечеткого регулятора выполняется автоматически (на основе модифицированного алгоритма Мамдани) [20].

Рассмотрим подробнее представленную архитектуру:

На первом уровне управление объектом осуществляется с помощью классического регулятора (модель заранее известна и рассчитана) [21].

На втором этапе на основе данных, полученных с классического регулятора, синтезируется нечеткий регулятор для управления объектом в условии неопределенности. Для разработки выходной лингвистической переменной используются полученные с классического регулятора значения сигналов отклонения, интеграла или дифференциала отклонения и сигналы управляющего воздействия [19]. Формирование базы правил нечеткого регуля-

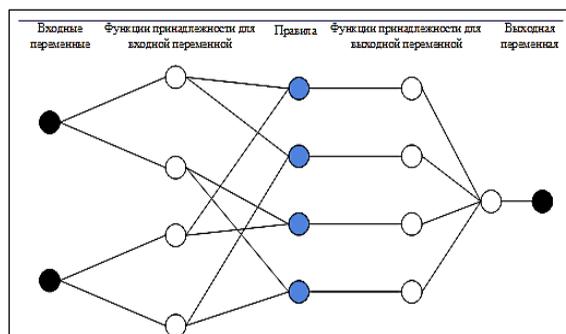
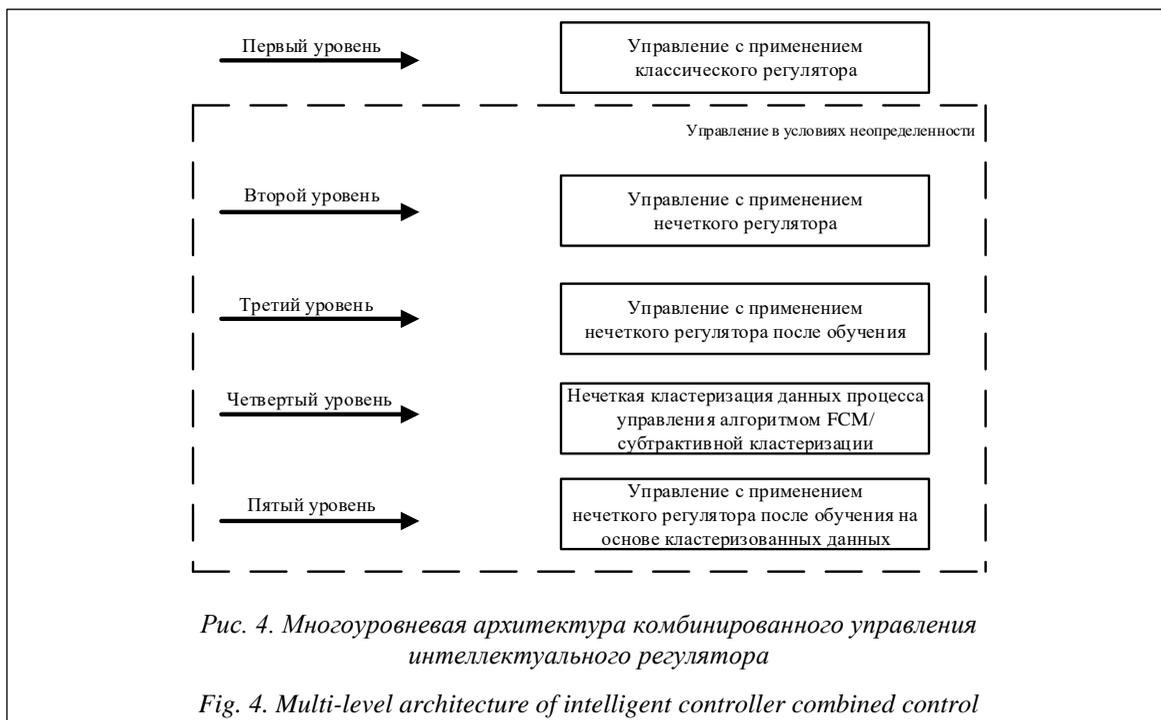


Рис. 3. ANFIS-архитектура после применения кластеризации

Fig. 3. ANFIS architecture after clustering



тора выполняется автоматически (на основе модифицированного алгоритма Мамдани) [20].

На третьем этапе выполняется разработка системы ANFIS для обучения нечеткого регулятора. В качестве обучающих данных используются значения сигналов отклонения, интеграла или дифференциала отклонения и управляющего воздействия, полученные после работы нечеткого регулятора. В качестве проверочных данных используются значения сигналов отклонения, интеграла или дифференциала отклонения и управляющего воздействия, полученные после работы классического регулятора [20, 23].

На четвертом уровне реализуется предложенный автором метод оптимизации, который можно разделить на два отдельных. Первый метод оптимизации предусматривает выполнение нечеткой кластеризации алгоритмом FCM, основу которого составляет метод нечетких множителей Лагранжа, всех значений сигналов ошибки, интеграла ошибки или дифференциала ошибки и управляющего воздействия, полученных при функционировании классического регулятора. Второй метод оптимизации предусматривает выполнение нечеткой кластеризации всех значений этих сигналов алгоритмом субтрактивной кластеризации, которая основана на оценке плотности точек данных вокруг соответствующего центра кластера.

На пятом уровне выполняется разработка системы ANFIS (повторяются работы третьего уровня) для обучения нечеткого регулятора, но

уже используются кластеризованные значения сигналов.

В результате реализации описанной многоуровневой архитектуры комбинированного управления синтезируется новый интеллектуальный регулятор без потери качества управления, позволяющий управлять объектом в условиях неопределенности. Желаемое управление с применением такого регулятора достигается при значительном сокращении количества производимых вычислений.

Предложенный метод позволяет в несколько раз сократить следующие параметры:

- количество требуемых для разработки ANFIS значений для каждого регулятора по показателям отклонения, интеграла или дифференциала отклонения и управляющего воздействия;
- количество функций принадлежности лингвистических переменных интеллектуального регулятора;
- количество правил в базе правил интеллектуального регулятора.

Благодаря этому значительно сократится время выдачи управляющего воздействия в САУ с применением интеллектуального регулятора.

Представим практическую реализацию предложенного метода. На рисунке 5 продемонстрирована структура полученного нечеткого регулятора ANFIS, обученного с помощью нейросети [19] для управления техническим объектом без применения нечеткой класте-

ризации. Из рисунка видно, что количество управляющих правил равно двадцати пяти.

После применения разработанного метода с использованием нечеткой кластеризации алгоритмом FCM и субтрактивного алгоритма кластеризации в этой же модели для этого же объекта управления структура нечеткого регулятора ANFIS принимает вид, представленный на рисунке 6. Количество управляющих правил сократилось до четырех, то есть более чем в шесть раз без потери качества управления [19]. Соответственно, сократилось и количество функций принадлежности с пяти до двух у каждой входной лингвистической переменной.

### Анализ полученных результатов

Результаты исследования разработанного метода оптимизации процедуры синтеза интеллектуальных регуляторов позволяют судить о возможности его применения для различных моделей управления техническими объектами в системах автоматизации. Применение метода значительно сокращает количество функций

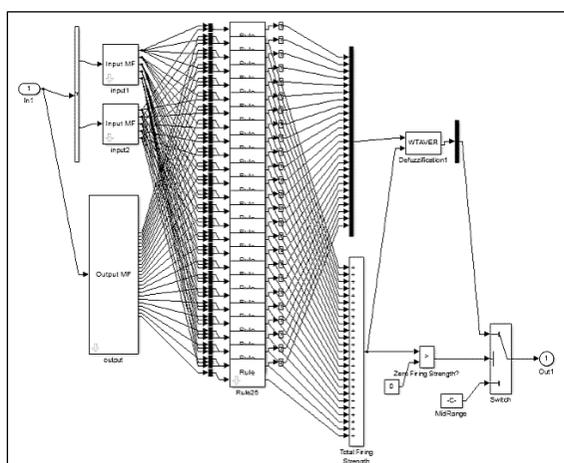


Рис. 5. Структура нечеткого регулятора до применения кластеризации

Fig. 5. Fuzzy controller structure before clustering

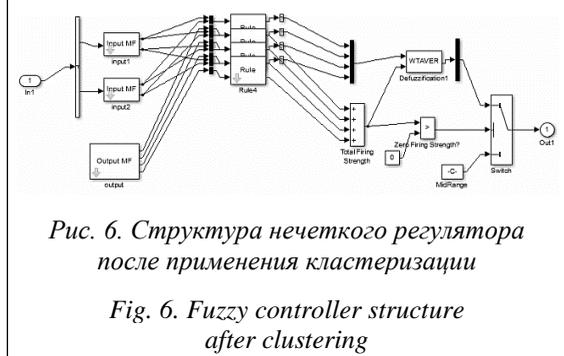


Рис. 6. Структура нечеткого регулятора после применения кластеризации

Fig. 6. Fuzzy controller structure after clustering

принадлежности лингвистических переменных интеллектуального регулятора, количество управляющих правил. Следовательно, можно значительно сократить время выдачи управляющих воздействий без потери качества управления и без увеличения вычислительных ресурсов. Также метод позволяет снизить трудоемкость разработки интеллектуального регулятора и время его проектирования.

### Обсуждение результатов

Рассматривая нечеткую кластеризацию в целом как способ решения той или иной научно-практической задачи и анализируя работы исследователей по ее конкретному применению, можно сделать вывод о том, что она является эффективным инструментом для сокращения вычислений в различных областях знаний. Например, подход из работы [15] может привести к ошибкам при разработке базы правил из-за отсутствия точных данных для начальной разработки системы нечеткого вывода. В предлагаемом в данной статье методе кластерный анализ выполняется уже на этапе синтеза нечеткого регулятора, исключая возможные сложности и неточности с проектированием его базы правил.

Анализ работ в области проектирования интеллектуальных САУ показывает, что нечеткий кластерный анализ в данной области или не применяется вовсе, или применяется для отдельных подзадач, без стремления сократить трудоемкость при разработке интеллектуальных регуляторов в САУ. В основном усилия разработчиков направлены на оптимизацию параметров регуляторов с помощью одного или нескольких объединенных (гибридных) методов эффективного управления. Как показали результаты исследований, представленные в данной работе, нечеткая кластеризация может являться полноценным элементом гибридных методов управления, реализация которой позволяет получать требуемое управление без потери качества. При этом фактическим результатом от применения нечеткой кластеризации является значительное улучшение количественных показателей при синтезе интеллектуальных регуляторов.

### Заключение

В данном научном исследовании представлен новый метод оптимизации процедуры синтеза интеллектуальных регуляторов. Алгоритмы нечеткого кластерного анализа используются в системах управления, спроектированных на

основе гибридного подхода, поэтому трудоемкость разработки таких систем значительно ниже. Предложенный метод может найти свое применение в интеллектуальных системах в развитии известных подходов к организации эффективного управления без потери качества и с сокращением вычислительных ресурсов с помощью нечеткой кластеризации.

### Список литературы

1. Jalali N., Razmi H., Doagou-Mojarrad H. Optimized fuzzy self-tuning PID controller design based on Tribe-DE optimization algorithm and rule weight adjustment method for load frequency control of interconnected multi-area power systems. *Applied Soft Computing*, 2020, vol. 93, art. 106424. doi: 10.1016/j.asoc.2020.106424.
2. Solihin M.I., Chuan Ch.Yo., Astuti W. Optimization of fuzzy logic controller parameters using modern meta-heuristic algorithm for gantry crane system (GCS). *Materials Today: Proc.*, 2020, vol. 29, pt. 1, pp. 168–172. doi: 10.1016/j.matpr.2020.05.641.
3. Tripathi S., Shrivastava A., Jana K.C. Self-tuning fuzzy controller for sun-tracker system using Gray Wolf Optimization (GWO) technique. *ISA Transactions*, 2020, vol. 101, pp. 50–59. doi: 10.1016/j.isatra.2020.01.012.
4. Карабцов, Р.Д., Денисова Л.А. Проектирование нечеткой системы регулирования с использованием генетического алгоритма оптимизации // Омский научн. вестн. 2017. № 6. С. 137–142.
5. Zand J.P., Katebi J., Yaghmaei-Sabegh S. A hybrid clustering-based type-2 adaptive neuro-fuzzy forecasting model for smart control systems. *Expert Systems with Applications*, 2024, vol. 239, art. 122445. doi: 10.1016/j.eswa.2023.122445.
6. Yang X., Xing H., Ji X., Xu W., Pedrycz W. Fuzzy clustering for electric field characterization and its application to thunderstorm interpretability. *Digital Communications and Networks*, 2024. doi: 10.1016/j.dcan.2024.03.010.
7. Ahmadnia M., Hajipour A., Tavakoli H. Robust variable-order fractional PID-LP fuzzy controller for Automatic Voltage Regulator systems. *Applied Soft Computing*, 2024, vol. 167, pt. A, art. 112268. doi: 10.1016/j.asoc.2024.112268.
8. Li J., Wang W., Song Yu., Yang X., Cao F. Adaptive fuzzy PI control for the dynamic operation of the transcritical CO<sub>2</sub> thermal system in electric vehicles. *Applied Thermal Eng.*, 2024, art. 124737. doi: 10.1016/j.applthermaleng.2024.124737.
9. Zhou J., Chen J., Li H. An optimized fuzzy K-means clustering method for automated rock discontinuities extraction from point clouds. *Int. J. of Rock Mechanics and Mining Sciences*, 2024, vol. 173, art. 105627. doi: 10.1016/j.ijrmm.2023.105627.
10. Eustaquio F., Nogueira T. Evaluating the numerical instability in fuzzy clustering validation of high-dimensional data. *Theoretical Computer Sci.*, 2020, vol. 805, pp. 19–36. doi: 10.1016/j.tcs.2019.10.039.
11. Yang Yu., Wang R., Feng Ch. Level set formulation for automatic medical image segmentation based on fuzzy clustering. *Signal Processing: Image Communication*, 2020, vol. 87, art. 115907. doi: 10.1016/j.image.2020.115907.
12. Baranwal M., Salapaka S. Clustering and supervisory voltage control in power systems. *Int. J. of Electrical Power & Energy Systems*, 2019, vol. 109, pp. 641–651. doi: 10.1016/j.ijepes.2019.02.025.
13. Li L., Yin X.-L., Jia X.-Ch., Sobhani B. Day ahead powerful probabilistic wind power forecast using combined intelligent structure and fuzzy clustering algorithm. *Energy*, 2020, vol. 192, art. 116498. doi: 10.1016/j.energy.2019.116498.
14. Yang M., Shi Ch., Liu H. Day-ahead wind power forecasting based on the clustering of equivalent power curves. *Energy*, 2021, vol. 218, art. 119515. doi: 10.1016/j.energy.2020.119515.
15. Tan T., Zhao T. A data-driven fuzzy system for the automatic determination of fuzzy set type based on fuzziness. *Inform. Sci.*, 2023, vol. 642, art. 119173. doi: 10.1016/j.ins.2023.119173.
16. Fan Z., Chiong R., Hu Z., Lin Y. A multi-layer fuzzy model based on fuzzy-rule clustering for prediction tasks. *Neurocomputing*, 2020, vol. 410, pp. 114–124. doi: 10.1016/j.neucom.2020.04.031.
17. Ruiz de la Hermosa Gonzalez-Carrato R. Wind farm monitoring using Mahalanobis distance and fuzzy clustering. *Renewable Energy*, 2018, vol. 123, no. C, pp. 526–540. doi: 10.1016/j.renene.2018.02.097.
18. Chaghari A., Feizi-Derakhshi M.-R., Balafar M.-A. Fuzzy clustering based on Forest optimization algorithm. *J. of King Saud University – Computer and Inform. Sci.*, 2018, vol. 30, pp. 25–32. doi: 10.1016/j.jksuci.2016.09.005.
19. Игнатъев В.В. Методы управления техническими объектами с помощью интеллектуальных регуляторов на основе самоорганизации баз знаний. Ростов н/Д; Таганрог: изд-во ЮФУ, 2020. 142 с. doi: 10.18522/801273622.
20. Ignatyeva A., Kureychik V., Ignatyev V., Soloviev V., Beloglazov D., Kovalev A. Optimization of the intelligent controller rule base based on fuzzy clustering for controlling an object operating in changing conditions. In: *CCIS. Proc. CIT&DS*, 2021, vol. 1448, pp. 149–162. doi: 10.1007/978-3-030-87034-8\_12.
21. Игнатъев В.В., Соловьев В.В. Оптимизация параметров интеллектуального регулятора с помощью генетического алгоритма для управления неустойчивым нелинейным техническим объектом // *Актуальные научные исследования в современном мире*. 2021. № 12-11. С. 76–83.
22. Игнатъев В.В. Компенсация разных типов неопределенностей при управлении техническим объектом с помощью интеллектуальных регуляторов // *Программные продукты и системы*. 2023. Т. 36. № 3. С. 423–431. doi: 10.15827/0236-235X.143.423-431.
23. Игнатъев В.В., Спиридонов О.Б., Курейчик В.М., Ковалев А.В., Игнатъева А.С. Метод гибридного управления в интеллектуальных системах на основе ПИД и ПИД-FUZZY-регуляторов // *Вестн. РГРТУ*. 2017. № 62. С. 110–118. doi: 10.21667/1995-4565-2017-62-4-110-118.

**Fuzzy clustering for improving control efficiency in automatic systems**Alexandra S. Ignatyeva <sup>1</sup>✉<sup>1</sup> Southern Federal University, Institute of Radio Engineering Systems and Control, Rostov-on-Don, 344006, Russian Federation**For citation**Ignatyeva, A.S. (2024) 'Fuzzy clustering for improving control efficiency in automatic systems', *Software & Systems*, 37(4), pp. 566–575 (in Russ.). doi: 10.15827/0236-235X.148.566-575**Article info**

Received: 18.06.2024

After revision: 23.07.2024

Accepted: 30.07.2024

**Abstract.** The paper focuses on finding new ways to improve control efficiency in automatic systems. It proposes a new method for optimizing the procedure of synthesizing intelligent controllers in automatic control systems for technical objects based on the analysis of advanced approaches to solving this scientific problem. Hybrid control as the most effective when operating objects under conditions of uncertainty. In order to organize hybrid control, the author proposes a method that is based on fuzzy clustering algorithms, as well as specially developed algorithms implemented in an intelligent controller. The controller is based on fuzzy logic methods, neural networks and bioinspired methods. The paper also shows the existing approaches to optimizing intelligent controllers in terms of the proposed method. At the conceptual level, the author proposes an approach to solving the problem of optimizing intelligent controller parameters, presents its generalization. In general, the paper describes the models for applying the developed method. In particular, there are the results of the ANFIS system method in hybrid control models. There is a presentation of ANFIS architectures before and after clustering. The obtained results are due to developing a multi-level architecture of combined control in an intelligent controller. The paper shows fuzzy controller structures when implementing ANFIS before and after fuzzy clustering. The author analyzes the study results. After applying the proposed method, the number of membership functions of linguistic variables of the intelligent controller and the number of control rules are significantly reduced. This allows reducing labor intensity of developing an intelligent controller and the time of its design without losing control quality and increasing computing resources.

**Keywords:** fuzzy clustering, control system, technical object, intelligent controller, optimization

**References**

1. Jalali, N., Razmi, H., Doagou-Mojarrad, H. (2020) 'Optimized fuzzy self-tuning PID controller design based on Tribe-DE optimization algorithm and rule weight adjustment method for load frequency control of interconnected multi-area power systems', *Applied Soft Computing*, 93, art. 106424. doi: 10.1016/j.asoc.2020.106424.
2. Solihin, M.I., Chuan, Ch.Yo., Astuti, W. (2020) 'Optimization of fuzzy logic controller parameters using modern meta-heuristic algorithm for gantry crane system (GCS)', *Materials Today: Proc.*, 29(1), pp. 168–172. doi: 10.1016/j.matpr.2020.05.641.
3. Tripathi, S., Shrivastava, A., Jana, K.C. (2020) 'Self-tuning fuzzy controller for sun-tracker system using Gray Wolf Optimization (GWO) technique', *ISA Transactions*, 101, pp. 50–59. doi: 10.1016/j.isatra.2020.01.012.
4. Karabtsov, R.D., Denisova, L.A. (2017) 'Designing a fuzzy control system using a genetic optimization algorithm', *Omsk Sci. Bull.*, (6), pp. 137–142 (in Russ.).
5. Zand, J.P., Katebi, J., Yaghmaei-Sabegh, S. (2024) 'A hybrid clustering-based type-2 adaptive neuro-fuzzy forecasting model for smart control systems', *Expert Systems with Applications*, 239, art. 122445. doi: 10.1016/j.eswa.2023.122445.
6. Yang, X., Xing, H., Ji, X., Xu, W., Pedrycz, W. (2024) 'Fuzzy clustering for electric field characterization and its application to thunderstorm interpretability', *Digital Communications and Networks*. doi: 10.1016/j.dcan.2024.03.010.
7. Ahmadnia, M., Hajipour, A., Tavakoli, H. (2024) 'Robust variable-order fractional PID-LP fuzzy controller for Automatic Voltage Regulator systems', *Applied Soft Computing*, 167(A), art. 112268. doi: 10.1016/j.asoc.2024.112268.
8. Li, J., Wang, W., Song, Yu., Yang, X., Cao, F. (2024) 'Adaptive fuzzy PI control for the dynamic operation of the transcritical CO<sub>2</sub> thermal system in electric vehicles', *Applied Thermal Eng.*, art. 124737. doi: 10.1016/j.applthermaleng.2024.124737.
9. Zhou, J., Chen, J., Li, H. (2024) 'An optimized fuzzy K-means clustering method for automated rock discontinuities extraction from point clouds', *Int. J. of Rock Mech. and Mining Sci.*, 173, art. 105627. doi: 10.1016/j.ijrmms.2023.105627.
10. Eustaquio, F., Nogueira, T. (2020) 'Evaluating the numerical instability in fuzzy clustering validation of high-dimensional data', *Theoretical Computer Sci.*, 805, pp. 19–36. doi: 10.1016/j.tcs.2019.10.039.
11. Yang, Yu., Wang, R., Feng, Ch. (2020) 'Level set formulation for automatic medical image segmentation based on fuzzy clustering', *Signal Processing: Image Communication*, 87, art. 115907. doi: 10.1016/j.image.2020.115907.
12. Baranwal, M., Salapaka, S. (2019) 'Clustering and supervisory voltage control in power systems', *Int. J. of Electrical Power & Energy Systems*, 109, pp. 641–651. doi: 10.1016/j.ijepes.2019.02.025.
13. Li, L., Yin, X.-L., Jia, X.-Ch., Sobhani, B. (2020) 'Day ahead powerful probabilistic wind power forecast using combined intelligent structure and fuzzy clustering algorithm', *Energy*, 192, art. 116498. doi: 10.1016/j.energy.2019.116498.
14. Yang, M., Shi, Ch., Liu, H. (2021) 'Day-ahead wind power forecasting based on the clustering of equivalent power curves', *Energy*, 218, art. 119515. doi: 10.1016/j.energy.2020.119515.

15. Tan, T., Zhao, T. (2023) 'A data-driven fuzzy system for the automatic determination of fuzzy set type based on fuzziness', *Inform. Sci.*, 642, art. 119173. doi: 10.1016/j.ins.2023.119173.
16. Fan, Z., Chiong, R., Hu, Z., Lin, Y. (2020) 'A multi-layer fuzzy model based on fuzzy-rule clustering for prediction tasks', *Neurocomputing*, 410, pp. 114–124. doi: 10.1016/j.neucom.2020.04.031.
17. Ruiz de la Hermosa Gonzalez-Carrato, R. (2018) 'Wind farm monitoring using Mahalanobis distance and fuzzy clustering', *Renewable Energy*, 123(C), pp. 526–540. doi: 10.1016/j.renene.2018.02.097.
18. Chaghari, A., Feizi-Derakhshi, M.-R., Balafar, M.-A. (2018) 'Fuzzy clustering based on Forest optimization algorithm', *J. of King Saud University – Computer and Inform. Sci.*, 30, pp. 25–32. doi: 10.1016/j.jksuci.2016.09.005.
19. Ignatyev, V.V. (2020) *Methods for Controlling Technical Objects Using Intelligent Controllers Based on Self-Organization of Knowledge Bases*. Rostov-on-Don; Taganrog, 142 p. (in Russ.). doi: 10.18522/801273622.
20. Ignatyeva, A., Kureychik, V., Ignatyev, V., Soloviev, V., Beloglazov, D., Kovalev, A. (2021) 'Optimization of the intelligent controller rule base based on fuzzy clustering for controlling an object operating in changing conditions', in *CCIS. Proc. CIT&DS*, 1448, pp. 149–162. doi: 10.1007/978-3-030-87034-8\_12.
21. Ignatyev, V.V., Soloviev, V.V. (2021) 'Optimization of the parameters of an intelligent controller using a genetic algorithm to control an unstable nonlinear technical object', *Actual Sci. Research in the Modern World*, (12-11), pp. 76–83 (in Russ.).
22. Ignatyev, V.V. (2023) 'Compensation of various types of uncertainties when controlling a technical object using intelligent controllers', *Software & Systems*, 36(3), pp. 423–431 (in Russ.). doi: 10.15827/0236-235X.143.423-431.
23. Ignatyev, V.V., Spiridonov, O.B., Kureychik, V.M., Kovalev, A.V., Ignatyeva, A.S. (2017) 'Method of hybrid control in intelligent systems based on pid and pid-fuzzy-controllers', *Vestn. RGRU*, (62), pp. 110–118 (in Russ.). doi: 10.21667/1995-4565-2017-62-4-110-118.

**Авторы**

**Игнатъева Александра Сергеевна**<sup>1</sup>, инженер,  
alexandra\_25@mail.ru

<sup>1</sup> Южный федеральный университет,  
Институт радиотехнических систем и управления,  
г. Ростов-на-Дону, 344006, Россия

**Authors**

**Alexandra S. Ignatyeva**<sup>1</sup>, Engineer,  
alexandra\_25@mail.ru

<sup>1</sup> Southern Federal University,  
Institute of Radio Engineering Systems and Control,  
Rostov-on-Don, 344006, Russian Federation

## Разработка программного комплекса многоканального распознавания и коррекции речевых сообщений на основе алгоритмов машинного обучения в структуре импортозамещения

Ф.Н. Абу-Абед <sup>1</sup>✉, С.А. Жиронкин <sup>1</sup>

<sup>1</sup> Тверской государственный технический университет, г. Тверь, 170026, Россия

<sup>2</sup> Национальный исследовательский Томский политехнический университет, г. Томск, 634050, Россия

### Ссылка для цитирования

Абу-Абед Ф.Н., Жиронкин С.А. Разработка программного комплекса многоканального распознавания и коррекции речевых сообщений на основе алгоритмов машинного обучения в структуре импортозамещения // Программные продукты и системы. 2024. Т. 37. № 4. С. 576–584. doi: 10.15827/0236-235X.148.576-584

### Информация о статье

Группа специальностей ВАК: 2.3.1

Поступила в редакцию: 09.07.2024

После доработки: 14.08.2024

Принята к публикации: 19.08.2024

**Аннотация.** Предметом данного исследования является разработка программного комплекса многоканального распознавания и коррекции речевых сообщений на основе машинного обучения. Комплекс призван решать задачу повышения эффективности факторов производства как составляющую моделирования структуры импортозамещения в российской экономике. Актуальность работы заключается в отсутствии программных аналогов и разработок на отечественном рынке программных продуктов в условиях усиления технологических санкционных ограничений. Целью исследования является повышение эффективности внутренних коммуникаций российских компаний. В качестве методов исследования применяются системный анализ, машинное обучение, информационно-телекоммуникационное системное проектирование и объектно-ориентированное программирование. В настоящей статье приведены архитектура и основные компоненты программного комплекса, такие как бот для взаимодействия с пользователями, кэширование данных, долговременное хранение информации и сервис для распознавания и коррекции речевых сообщений с использованием методов машинного обучения. В разработанном авторами приложении решаются задачи развертывания и контейнеризации речевых сообщений с реализацией сервиса распознавания и расшифровки речевых сообщений. Предложенная система основана на кэшировании данных, распределяет нагрузку между независимыми сервисными компонентами с поддержкой контейнеризации, она адаптирована к масштабированию и работе на различных платформах и облачных средах. Интерфейс приложения дает пользователю возможность произвести необходимые настройки с целью автоматического распознавания, диаризации, коррекции и обобщения речевых сообщений. Научная новизна заключается в получении результатов, способствующих оптимизации внутренних коммуникаций с помощью алгоритмов машинного обучения для повышения точности и адаптивности корпоративных систем связи. Данные системы позволяют эффективно решать важную задачу моделирования структуры импортозамещения в условиях усиления внешних шоков и технологических ограничений.

**Ключевые слова:** структура, распознавание речевых сообщений, машинное обучение, программное средство, пользовательский интерфейс, команды и процедуры, база данных, импортозамещение

**Благодарности.** Исследование выполнено за счет гранта РФФИ № 23-28-01423, <https://rscf.ru/project/23-28-01423/>

**Введение.** Решение задачи моделирования структуры импортозамещения в российской экономике связано с определением потенциала рынка отечественных программных продуктов, в наибольшей степени пострадавшего от санкционных ограничений и технологических шоков. В целом острота проблемы импортозамещения технологий и компонентов систем искусственного интеллекта, машинного зрения и обучения, бортовых систем [1] связана с их диффузией в структуре российской экономики как сквозных технологий, глубоко модернизирующих целые отрасли обрабатывающего и добывающего комплекса. Моделирование отраслевой и секторальной структуры импортозамещения должно учитывать перспективы по-

вышения эффективности факторов производства (производительность труда, фондоотдача, рентабельность), которые используются в целевой функции структуры импортозамещения как основные переменные [2]. Возможности их радикального повышения за счет разработки и внедрения отечественных технологий Индустрии 4.0, таких как машинное обучение, распознавание и коррекция речи, отличаются особой актуальностью [3].

Существующие подходы к учету технологического компонента в повышении эффективности факторов производства включают в себя:

– общее положительное влияние технологий Индустрии 4.0 на производительность труда в обрабатывающем секторе, в частности,

за счет сокращения времени вспомогательных операций [4, 5];

- рост производительности труда от внедрения в промышленности широкого ряда АСУ, основанных на искусственном интеллекте [6];

- прирост производительности и рентабельности за счет внедрения систем машинного видения и обучения [7];

- падение производительности и рентабельности предприятий ряда обрабатывающих отраслей из-за санкционных ограничений и разрыва связей с ведущими мировыми производителями программных продуктов.

Учитывая потери российской экономики от санкций и ограничений в сфере передовых производственных технологий Индустрии 4.0 [8], поиск подходов к опережающей разработке программных компонентов, соответствующих уровню сквозных (отраслевых) технологий, играет важную роль в моделировании структуры импортозамещения. В частности, разработка отечественного программного комплекса для распознавания речевых сообщений внутри цифровых коммуникационных платформ, в том числе промышленных, имеет стратегическое значение, поскольку способствует долгосрочному снижению зависимости от зарубежных поставщиков принципиально новых передовых производственных технологий, в которых доля отечественных составляет порядка 1 % от используемых сегодня в России [9].

Для анализа используемых в России программных комплексов распознавания речевых сообщений были рассмотрены популярные зарубежные системы Microsoft Azure Speech, Amazon Transcribe, Google Cloud Speech-to-Text, а также отечественные Сбер Салют и Яндекс Speech API. Выявлено, что при высокой точности распознавания, которая является преимуществом зарубежных систем, отмечены такие недостатки, как высокие затраты, внешний сервис, невозможность улучшать модели, существенные затруднения доступа к сервисам в условиях санкций, непредсказуемость введения новых ограничений функционала при ужесточении санкционного противостояния. К преимуществам отечественных систем следует отнести сам сервис и высокую точность на общих тестах, а к недостаткам – низкое качество распознавания в домене, невозможность радикального улучшения основного функционала, затруднения отраслевой кастомизации, неудовлетворительный внешний сервис.

Для разработки комплекса многоканального распознавания и коррекции речевых сооб-

щений на основе алгоритмов машинного обучения был выбран язык Python как один из самых популярных в мире, а значит, есть большое сообщество разработчиков, которое поддерживает инструменты разработки, создает значительную базу специализированных знаний, в которой можно найти решение для любой возникающей проблемы [10]. Практическим следствием этого является наличие значительного количества библиотек и фреймворков, упрощающих разработку новых продуктов. Большинство библиотек языка написано на C и C++, что способствует достаточно высокой производительности итогового продукта [11]. Наряду с этим язык Python поддерживает достаточное число инструментов для улучшения качества разработки программных продуктов, таких как Jupyter Notebook с интерактивной средой для выполнения и визуализации исследований, а также Hydra с возможностью логирования всех экспериментов и их гибкого конфигурирования. Это имеет важное значение для диффузии технологий машинного обучения в отраслевой структуре промышленности, при создании коллаборативных роботов, использование которых вместе с людьми, а не вместо них, является основой будущей Индустрии 5.0 (ее экспансия ожидается во второй половине 21 в. [12]), направленной на радикальное повышение производительности труда за счет синергии человеко-интеллектуально-машинных систем.

### Основа архитектуры системы

Архитектура сервиса многоканального распознавания и коррекции речевых сообщений состоит из следующих ключевых компонентов:

- модуль распознавания – основан на моделях машинного обучения и включает в себя компоненты для обнаружения голоса, диаризации (разделения речи разных говорящих), устранения шумов, а также средства для распознавания речи;

- Discord-бот – основной модуль, реализованный на языке Python с использованием библиотеки discord.py; предназначен для прямого взаимодействия с пользователями Discord и для управления процессом записи речевых сообщений;

- система Redis – используется для временного кэширования данных о голосовых сообщениях, таких как их идентификатор, идентификатор канала Discord и идентификатор сообщения с информацией о сообщении;

– PostgreSQL – реляционная БД для долговременного хранения аудиофайлов и информации о записанных голосовых сообщениях;

– сервер gRPC – обеспечивает API для взаимодействия с записанными голосовыми сообщениями, предоставляя возможность получения метаданных и скачивания аудиофайлов.

Рассматриваемая в данной статье система распознавания и коррекции речевых сообщений спроектирована с учетом требований масштабируемости и высокой производительности при работе с большим количеством голосовых сообщений и пользователей. Этого возможно добиться при реализации следующих решений.

1. Использование СУБД Redis для кэширования данных о голосовых сообщениях, что обеспечивает быстрый доступ к информации и снижает нагрузку на БД PostgreSQL.

2. Разделение ответственности между компонентами системы (Discord-бот, gRPC-сервер, сервис распознавания речи), что позволяет масштабировать их независимо друг от друга пропорционально нагрузке.

3. Контейнеризация с помощью инструмента развертывания приложений Docker Compose, что упрощает масштабирование системы на различных платформах и облачных средах. Это особо ценно для диффузии разработки в разных отраслях экономики.

Общий вид архитектуры предлагаемой системы распознавания и коррекции речевых сообщений представлен на рисунке 1.

Взаимодействие пользователя с системой для записи речевых сообщений происходит следующим образом.

Шаг 1. Пользователь переходит в один из голосовых каналов Discord и ожидает присоединения остальных участников.

Шаг 2. После этого пользователь вводит в любом текстовом канале команду /record.

Шаг 3. Discord-бот подключается к голосовому каналу, в котором находится пользователь, и начинает запись речевых сообщений всех присутствующих участников.

Записанные аудиофайлы сохраняются в Redis и PostgreSQL. API, реализованный на основе gRPC, обеспечивает доступ к аудиофайлам и позволяет интегрироваться с другими системами.

Представим листинг процесса взаимодействия пользователя при записи сообщений:

```
@bot.command()
async def record(ctx): logging.info
    ("Attempting_to_record_call") voice
    = ctx.author.voice
    if not voice:
```



Рис. 1. Архитектура системы распознавания и коррекции речевых сообщений

Fig. 1. Architecture of the speech message recognition and correction system

```
await ctx.
respond(JOIN_TO_CHANNEL_TO_
RECORD)
return
vc = await voice.channel.connect() #
    Connect to the voice channel the
    author is in.
call_id = uuid.uuid4().hex
start_recording(vc, ctx.channel,
call_id) embed_data = create_new_
record_embed_msg(vc, voice, call_id)
```

Далее Discord-ботом отправляется сообщение в gRPC-сервер с основной информацией о голосовом сообщении (идентификатор, название канала) и кнопкой для завершения записи. После завершения записи, аудиофайлы с голосовыми сообщениями сохраняются в БД PostgreSQL, а информация о сообщении кэшируется в Redis.

**Реализация сервиса распознавания и расшифровки речевых сообщений**

Для распознавания речи были использованы следующие технологии:

- модель для обнаружения голосовой активности в аудиофайле Silero VAD (Россия);
- программный инструмент с открытым кодом для диаризации речи Ryannote (Франция);
- программный продукт для устранения шумов из аудиозаписей Denoiser (Россия);
- модель для распознавания речи с открытым кодом GigaAM CTC (Россия);
- модель для разделения речевых сообщений Sepformer (США);
- модель для исправления базовых ошибок Yandex Speller (Россия);
- авторская модель для расстановки пунктуации.

В представленном списке большая часть используемых моделей российского производства, а иностранные имеют открытый код.

Что касается валидации, для каждого компонента были выбраны наиболее подходящие технологии и алгоритмы. Для диаризации использовались Audacity и Ryannote, так как они показывали высокие результаты в разделении речи разных говорящих в условиях многоканальных записей (рис. 2).

Для проверки точности моделей были использованы следующие валидационные данные (<http://www.swsys.ru/uploaded/image/2024-4/1.jpg>):

- порядка 5 часов вручную размеченных аудиоданных;
- оптимизация аудиоданных под задачи обнаружения и для диаризации;
- 200 000 сообщений с мессенджера Discord;
- 30 000 задач из приложения Jira.

Для обнаружения голоса изначально рассматривались две модели машинного обучения: Silero VAD и MarbleNet VAD; по данным авторских тестов, Silero VAD оказался предпочтительнее по основным метрикам – по охвату и по точности. Точность сообщает, насколько положительные предсказания совпадают с фактическими данными. Охват показывает, насколько хорошо модель находит все положительные предсказания. Авторами была проведена оптимизация порогового значения при помощи

использованных валидационных данных, что значительно повысило эффективность предлагаемой системы распознавания и коррекции речевых сообщений.

Использование Silero VAD позволило сократить вычисления на 22,26 %, при этом уровень ошибок составил всего 0,6 %. Данная модель машинного обучения, используя метод скользящего окна, анализирует аудиофайлы и с помощью задачи бинарной классификации [13] определяет, присутствует ли человеческий голос в текущем сегменте (около 5-30 мс). Хотя использование модели Silero VAD не является обязательным в предложенной авторами системе распознавания речи, она может быть очень полезна в определенных сценариях. Например, когда аудио уже разделено на отдельные файлы (как это часто бывает с записью, полученной с цифровых платформ), поскольку логично предположить, что люди перебивают друг друга редко (лишь в 2,1 % случаев в рассматриваемом неделимом наборе данных (датасете)). Таким образом, для обработки аудиозаписи длиной 60 минут, содержащей речь пяти участников, необходимо обработать 300 минут аудио. Это значительно превышает объем полезного сигнала, который составляет максимум 60 минут, что позволяет не превышать максимальную длительность обработки объекта полезного сигнала.

Результат работы модели Silero VAD представлен на рисунке 3.

Для обнаружения человеческого голоса были рассмотрены только модели MarbleNet и Silero VAD, другие известные методы Audiotok и WebRTC уже подробно исследованы (<https://github.com/wiseman/py-webrtcvad/issues/68>).

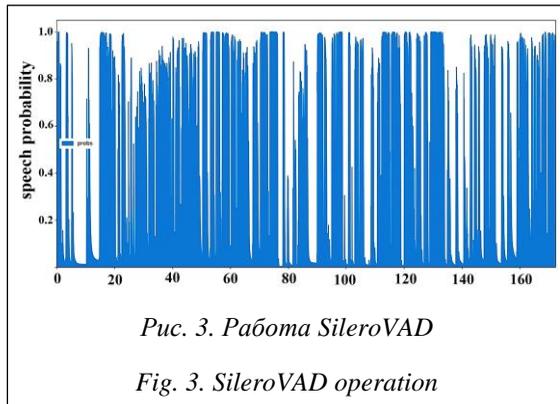
Для распознавания речи выбор предпочтительной модели производился из трех перспективных: Whisper v3, GigaAM CTC и RNN-T.

Удобство модели Whisper в том, что нет необходимости распознавать пунктуацию. Кроме того, можно распознавать сторонние звуки (например, звуковой фон), однако из-за этого возникают неоднозначные ситуации, когда мо-



Рис. 2. Пример разметки в Audacity

Fig. 2. Audacity markup example



дель распознает не существующие в реальности музыку, шуршание и тому подобное во многом из-за обучения на субтитрах видео. К тому же фактическое качество распознавания русского языка при помощи Whisper оказалось низким. Поэтому выбор происходил между моделями CTC и RNN-T, и на основе полученных данных авторы отдали предпочтение CTC.

После этого была обучена N-граммная модель для рескоринга (процесса валидации гипотез CTC-модели с помощью внешней модели). То есть после предсказаний CTC-модели ее обрабатывают языковой моделью, которая определяет, насколько данный текст правдоподобен. Основной метрикой здесь является WER (*Word Error Rate*) – сумма перестановок, удалений и вставок слов по сравнению с эталоном. С помощью языковой модели WER удалось сократить примерно до 10 %.

Для оценки качества классификации моделей обнаружения голоса зачастую используют метрики Precision (точность) и Recall (полнота). Метрика Precision показывает, насколько точно модель предсказывает положительные случаи. Она измеряет долю истинно положительных предсказаний (TruePositive) среди всех предсказанных положительных случаев (TruePositive и FalsePositive).

Метрика Recall оценивает, насколько полно модель обнаружения голоса охватывает все положительные случаи. Она измеряет долю истинно положительных предсказаний (TruePositive) среди всех фактических положительных случаев (TruePositive и FalseNegative).

После выбора модели необходимо определить оптимальный порог для классификации. Так как данная модель важна в контексте оптимизации прогноза, важно найти компромисс между прогнозами FalseNegative (не определяем фразы, когда они есть) и TrueNegative (места в аудио, которые можно не обрабатывать).

Итоги определения оптимального прогноза отражены в таблице.

**Результаты при разном пороговом значении классификации**

**Results at different classification thresholds**

Порог	TruePositive, %	FalsePositive, %	TrueNegative, %	FalseNegative, %
0,05	55,77	40,22	3,67	0,35
0,1	55,77	40,22	3,67	0,35
0,15	55,77	37,05	6,84	0,35
0,2	55,77	36,85	7,03	0,35
0,25	55,77	34,74	9,15	0,35
0,3	55,50	34,74	9,15	0,61
0,35	55,50	31,50	12,38	0,61
0,4	55,50	29,35	14,53	0,61
0,45	55,50	27,21	16,68	0,61
<b>0,5</b>	<b>55,50</b>	<b>21,62</b>	<b>22,26</b>	<b>0,61</b>
0,55	55,24	18,19	25,70	0,87
0,6	54,03	15,67	28,22	2,09
0,65	51,12	10,68	33,21	4,99
0,7	47,68	8,56	35,32	8,43
0,75	47,42	7,24	36,64	8,69
0,8	46,04	3,60	40,29	10,07
0,85	44,98	3,34	40,55	11,13
0,9	40,40	3,07	40,81	15,71
0,95	35,68	2,81	41,08	20,44

Из табличных данных видно, что наиболее подходящим значением является порог 0,5, который позволяет терять лишь 0,61 % фреймов при сокращении вычислений на 22,26 %.

**Диаризация и устранение шумов**

Для обработки аудиофайлов, не разделенных по спикерам, важно выявить моменты, когда говорят несколько участников, и разделить их аудиодорожки [14]. Это позволяет обрабатывать их отдельно и параллельно, значительно ускоряя процесс. Для выполнения задачи была использована библиотека ruannotate, которая специализируется на диаризации.

Для валидации модели был адаптирован датасет, использованный ранее, содержащий метки моментов речи нескольких человек. Это позволило разделить аудиофайл на фрагменты, каждый из которых содержит речь одного человека, и затем объединить их, имитируя одновременную речь нескольких участников. Процесс синтеза данных для данной задачи представлен на рисунке 4.

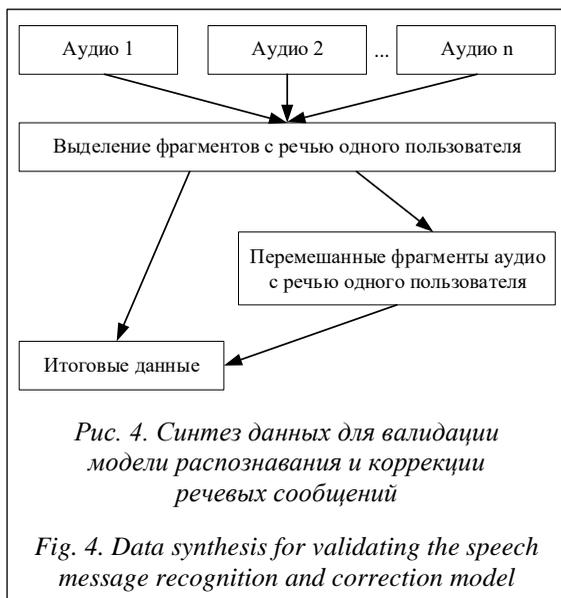


Fig. 4. Data synthesis for validating the speech message recognition and correction model

### Расстановка пунктуации

Ввиду отсутствия подходящих решений для расстановки пунктуации была обучена авторская модель. Контекст обработки был расширен в два раза по сравнению с базовой моделью, что позволило обрабатывать до 1 500 слов за один прогон (модель была адаптирована к работе с полилогами). Модель обучалась с помощью разбиения Discord-сообщений на небольшие диалоги и удаления после этого всей пунктуации (рис. 5). Затем пользователей нумеровали, и это добавлялось в контекст;

потом производилось обучение на три эпохи в течение 6 часов на одной видеокарте с 24 Гб видеопамати (NVidia GeForce RTX 3090 Ti).

Обучающая выборка состояла из 200 000 внутренних сообщений компании и предварительной очистки (удаление эмодзи, ссылок, фрагментов кода) с помощью регулярных выражений.

Фрагмент кода для загрузки и сохранения данных представлен в листинге:

```
load_dotenv()
async def load_messages(guild, excluded_channels: list[str]):
    messages = []
    for channel in guild.text_channels:
        if channel.id not in excluded_channels:
            logging.info("Reading channel: %s", channel)
            try:
                async for message in channel.history(limit=None):
                    msg_data = DiscordMessage(channel.id, message.content, message.created_at.timestamp(), message.author.id)
                    messages.append(msg_data)
            except discord.Forbidden:
                logging.error(f"Do not have permissions to read {channel.name}")
            except discord.HTTPException:
                logging.error(f"Failed to read history for {channel.name}, skipping...")
        else:
            logging.info("Skipping channel: %s", channel)
    return messages
```

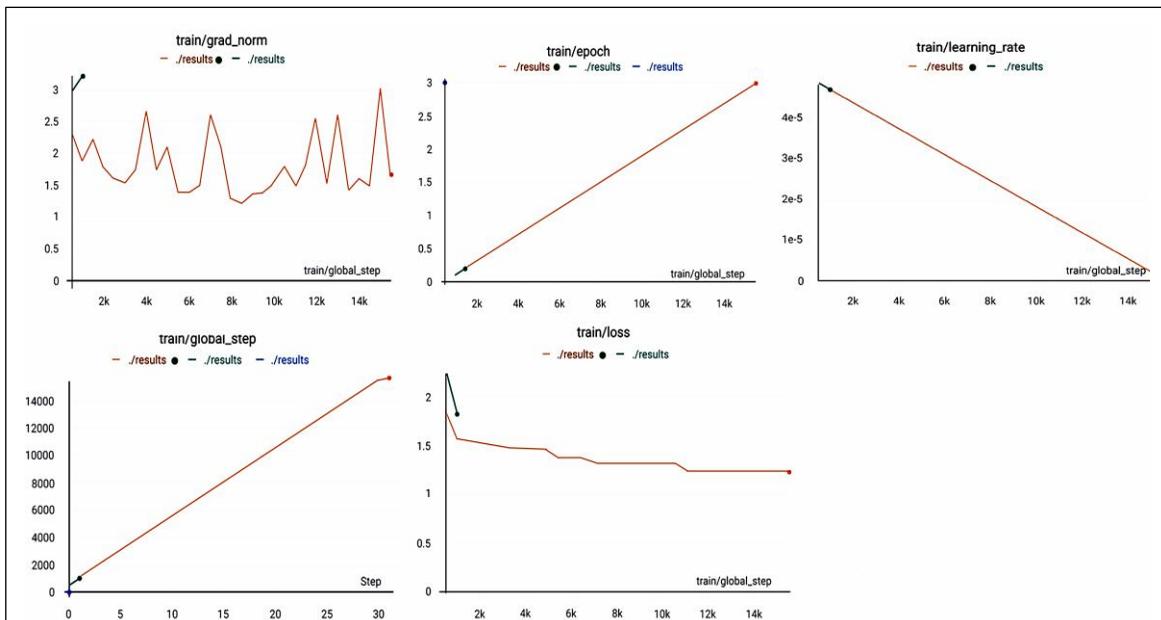


Рис. 5. Динамика обучения пунктуационной модели

Fig. 5. Punctuation model training dynamics

## Коррекция речевых сообщений и разделение их на беседы

В результате обоснования моделей для коррекции речевых сообщений было установлено, что для большинства и охват, и точность были ниже 50 % (см. таблицу) даже на общих данных, внутри домена данные отличались еще меньшей точностью. Поэтому было принято решение использовать Yandex Speller, обогащенный данными из диалогов пользователей внутри беседы, что позволило уменьшить количество ложноположительных результатов практически до нуля. Наиболее эффективные решения получены при комбинации Yandex Speller и словарей из Discord и Jira. Иначе говоря, при отсутствии в словаре искомого слова (гипотеза отвергалась), данная модель предлагала заменить его на слово, которое присутствует в словаре

В целях разделения сообщений на беседы данные после загрузки и очистки были преобразованы для выполнения задач с диалоговыми данными. Добавлены токены для обозначения пользователей < \_ >, для определения конца сообщения < // >, а также для разделения контекстных и генерируемых данных < / >.

### Выводы

Основной целью работы явилось создание программного комплекса, который повышает эффективность внутрикорпоративных коммуникаций и поддерживает политику импортозамещения. Для достижения этой цели были поставлены задачи, включающие исследование существующих методов распознавания и коррекции речи, разработку ПО, интеграцию с цифровыми коммуникационными платформами и создание средств для обобщения речевых сообщений. В ходе анализа существующих программных продуктов были выделены функциональные и нефункциональные требования к программному комплексу. Функциональные требования включают точное распознавание

речи, коррекцию орфографических и грамматических ошибок, автоматическую расстановку пунктуации, обнаружение голоса и удаление шума. В состав нефункциональных требований входят высокая производительность, масштабируемость, надежность, совместимость с различными платформами и удобство эксплуатации.

Разработанная система записи речевых сообщений включает ключевые компоненты, такие как Discord-бот для взаимодействия с пользователями, система Redis для кэширования данных, БД PostgreSQL для долговременного хранения информации, а также сервис для распознавания и коррекции речевых сообщений, реализованный с использованием современных методов машинного обучения. В ходе экспериментов и тестирования использованы различные модели для обнаружения голоса, диаризации и удаления шума, что позволило достичь высокой точности и производительности системы.

### Заключение

В данной работе продемонстрировано, что современные системы распознавания речи являются результатом многолетних исследований и развития технологий. Использование методов глубокого обучения и облачных технологий значительно расширило возможности применения распознавания речи в реальных приложениях, делая эту технологию неотъемлемой частью современных коммуникационных систем и умных устройств. Разработанная система успешно интегрирована в цифровые коммуникационные платформы, обеспечивает высокую точность и надежность распознавания и коррекции речевых сообщений. Во время работы были выполнены все поставленные задачи. Дальнейшее развитие систем распознавания речи будет связано с улучшением моделей машинного обучения, с повышением их адаптивности к различным условиям и с интеграцией их с другими технологиями искусственного интеллекта.

### Список литературы

1. Горелиц Н.К., Гукова А.С., Краснощеков Д.В. Анализ российского программного обеспечения для поддержки жизненного цикла разработки бортовых систем в условиях политики импортозамещения // Тр. ИСП РАН. 2020. Т. 32. № 2. С. 175–190.
2. Абу-Абед Ф.Н., Жиронкин С.А. Моделирование структуры импортозамещения на базе модели системы оптимального распределения // Программные продукты и системы. 2023. Т. 36. № 4. С. 644–653. doi: 10.15827/0236-235X.144.644-653.
3. Абу-Абед Ф.Н. Киберфизические системы и человек в контексте интеллектуального производства Индустрии 4.0 // Экономика и управление инновациями. 2022. № 3. С. 78–87. doi: 10.26730/2587-5574-2022-3-78-87.

4. Тарасов И.В. Технологии Индустрии 4.0: влияние на повышение производительности промышленных компаний // Стратегические решения и риск-менеджмент. 2018. № 2. С. 62–69. doi: 10.17747/2078-8886-2018-2-62-69.
5. Гасанов М.А., Гасанов Э.А., Ашванян С.К., Жаворонок А.В., Жиронкин С.А. Цифровой структурный сдвиг: подход к анализу в современной экономике // Экономика и управление инновациями. 2024. № 2. С. 23–34. doi: 10.26730/2587-5574-2024-2-23-34.
6. Макаров М.Ю. Влияние искусственного интеллекта на производительность труда // Экономика и управление. 2020. Т. 26. № 5. С. 479–486.
7. Miranda S.A.D., Aguilar R.R. Machine learning models in health prevention and promotion and labor productivity: A co-word analysis. *Iberoamerican J. of Sci. Measurement and Communication*, 2024, vol. 4, no. 1, pp. 1–16. doi: 10.47909/ijsmc.85.
8. Таран Е.А., Слесаренко Е.В., Жиронкин В.С. К вопросу о формировании модели структуры импортозамещения в российской экономике и ее ограничениях в условиях внешних шоков // Экономика и управление инновациями. 2024. № 2. С. 12–22. doi: 10.26730/2587-5574-2024-2-12-22.
9. Таран Е.А., Жиронкин С.А. Структура импортозамещения в российской экономике в условиях внешних шоков. Томск: СТТ, 2023. 144 с.
10. Магеррамов И.М., Акперов Г.И. Решение задач интернет-маркетинга средствами Python // Вестн. кибернетики. 2022. № 1. С. 29–37.
11. Subero A. The C programming language. In: *Programming PIC Microcontrollers with XC8*. Berkeley, CA, APress, 2024, pp. 15–33.
12. Абу-Абед Ф.Н. Применение технологий интеллектуального управления и бизнес-проектирования Индустрии 5.0 в Майнинге 5.0 // Экономика и управление инновациями. 2022. № 3. С. 50–59. doi: 10.26730/2587-5574-2022-3-50-59.
13. Lu L., Guo M., Renals S. Knowledge distillation for small-footprint highway networks. *Proc. ICASSP*, 2017, pp. 4820–4824. doi: 10.1109/ICASSP.2017.7953072.
14. Watanabe S., Barker J.P., Vincent E., Mandel M. CHiME-6 challenge: Tackling multi-speaker speech recognition for unsegmented recordings. *Proc. Int. Workshop CHiME*, 2020, pp. 1–7. doi: 10.21437/CHiME.2020-1.

Software &amp; Systems

doi: 10.15827/0236-235X.148.576-584

2024, 37(4), pp. 576–584

### Software complex for multi-channel recognition and correction of speech messages based on machine learning algorithms in terms of import substitution

Fares N. Abu-Abed <sup>1</sup>✉, Sergey A. Zhironkin <sup>2</sup><sup>1</sup> Tver State Technical University, Tver, 170026, Russian Federation<sup>2</sup> National Research Tomsk Polytechnic University, Tomsk, 634050, Russian Federation

#### For citation

Abu-Abed, F.N., Zhironkin, S.A. (2024) 'Software complex for multi-channel recognition and correction of speech messages based on machine learning algorithms in terms of import substitution', *Software & Systems*, 37(4), pp. 576–584 (in Russ.). doi: 10.15827/0236-235X.148.576-584

#### Article info

Received: 09.07.2024

After revision: 14.08.2024

Accepted: 19.08.2024

**Abstract.** The research focuses on developing a software complex for multichannel recognition and correction of speech messages based on machine learning. The complex solves the problem of increasing the efficiency of production factors as a part of modeling an import substitution structure in the Russian economy. This work is relevant due to the lack of software analogs among domestic software products under conditions of increasing technological sanction restrictions. The study aims to improve the efficiency of internal communications of Russian companies. The authors use system analysis, machine learning, information and telecommunication system design and object-oriented programming as research methods. This paper presents the architecture and the main components of the software complex. For example, a bot for interaction with users, data caching, long-term storage of information and a service for recognizing and correcting speech messages using machine learning methods. The application implements a speech message recognition and decryption service to solve the problems of speech message deployment and containerization. The proposed system is based on data caching. It distributes the load among independent service components with containerization support. It is adapted to scale and work on different platforms and cloud environments. The application interface enables the user to make necessary adjustments in order to automatically recognize, diarize, correct and summarize speech messages. The scientific novelty consists in obtaining results that contribute to the optimization of internal communications using machine-learning algorithms to improve the

accuracy and adaptability of corporate communication systems. These systems allow solving the important problem of modeling the structure for import substitution under the conditions of increasing external shocks and technological constraints.

**Keywords:** structure, speech message recognition, machine learning, software tool, user interface, commands and procedures, database, import substitution

**Acknowledgements.** The study was supported by the Russian Science Foundation grant no. 23-28-01423, <https://rscf.ru/en/project/23-28-01423/>

### References

1. Gorelits, N.K., Gukova, A.S., Krasnoschekov, D.V. (2020) 'Analysis of Russian software supporting onboard systems development lifecycle in context of import substitution policy', *Proc. of ISP RAS*, 32(2), pp. 175–190 (in Russ.).
2. Abu-Abed, F.N., Zhironkin, S.A. (2023) 'Russian import-substitution structure based on an optimal distribution system model', *Software & Systems*, 36(4), pp. 644–653 (in Russ.). doi: 10.15827/0236-235X.144.644-653.
3. Abu-Abed, F.N. (2022) 'Cyber-physical systems and human in the context of intelligent production of Industry 4.0', *Economics and Innovation Management*, (3), pp. 78–87 (in Russ.). doi: 10.26730/2587-5574-2022-3-78-87.
4. Tarasov, I.V. (2018) 'Industry 4.0: Technologies and their impact on productivity of industrial companies', *Strategic Decisions and Risk Management*, (2), pp. 62–69 (in Russ.). doi: 10.17747/2078-8886-2018-2-62-69.
5. Gasanov, M.A., Gasanov, E.A., Ashvanyan, S.K., Zhavoronok, A.V., Zhironkin, S.A. (2024) 'Digital structural shift: An approach to analysis in modern economy', *Economics and Innovation Management*, (2), pp. 23–34 (in Russ.). doi: 10.26730/2587-5574-2024-2-23-34.
6. Makarov, M.Yu. (2020) 'The influence of artificial intelligence on labor productivity', *Economics and Management*, 26(5), pp. 479–486 (in Russ.).
7. Miranda, S.A.D., Aguilar, R.R. (2024) 'Machine learning models in health prevention and promotion and labor productivity: A co-word analysis', *Iberoamerican J. of Sci. Measurement and Communication*, 4(1), pp. 1–16. doi: 10.47909/ijsmc.85.
8. Taran, E.A., Slesarenko, E.V., Zhironkin, V.S. (2024) 'On the issue of the formation of a model of the structure of import substitution in the Russian economy and its limitations under conditions of external shocks', *Economics and Innovation Management*, (2), pp. 12–22 (in Russ.). doi: 10.26730/2587-5574-2024-2-12-22.
9. Taran, E.A., Zhironkin, S.A. (2023) *The Structure of Import Substitution in the Russian Economy in Conditions of External Shocks*. Tomsk, 144 p. (in Russ.).
10. Magerramov, I.M., Akperov, G.I. (2022) 'Solving internet marketing problems via Python', *Bull. of Cybernetics*, (1), pp. 29–37 (in Russ.).
11. Subero, A. (2024) 'The C programming language', in: *Programming PIC Microcontrollers with XC8*, Berkeley, CA: Apress, pp. 36–45.
12. Abu-Abed, F.N. (2022) 'Application of intelligent management technologies and business design of Industry 5.0 in Mining 5.0', *Economics and Innovation Management*, (3), pp. 50–59 (in Russ.). doi: 10.26730/2587-5574-2022-3-50-59.
13. Lu, L., Guo, M., Renals, S. (2017) 'Knowledge distillation for small-footprint highway networks', *Proc. ICASSP*, pp. 4820–4824. doi: 10.1109/ICASSP.2017.7953072.
14. Watanabe, S., Barker, J.P., Vincent, E., Mandel, M. (2020) 'CHiME-6 challenge: Tackling multi-speaker speech recognition for unsegmented recordings', *Proc. Int. Workshop CHiME*, pp. 1–7. doi: 10.21437/CHiME.2020-1.

### Авторы

**Абу-Абед Фарес Надимович**<sup>1</sup>, к.т.н.,

доцент, декан, aafares@mail.ru

**Жиронкин Сергей Александрович**<sup>2</sup>,

д.э.н., профессор, zhironkin@tpu.ru

### Authors

**Fares N. Abu-Abed**<sup>1</sup>, Cand. of Sci. (Engineering),

Associate Professor, Dean, aafares@mail.ru

**Sergey A. Zhironkin**<sup>2</sup>, Dr.Sci. (Economics),

Professor, zhironkin@tpu.ru

<sup>1</sup> Тверской государственный технический университет, г. Тверь, 170026, Россия

<sup>2</sup> Национальный исследовательский Томский политехнический университет, г. Томск, 634050, Россия

<sup>1</sup> Tver State Technical University, Tver, 170026, Russian Federation

<sup>2</sup> National Research Tomsk Polytechnic University, Tomsk, 634050, Russian Federation

## Совершенствование информационной технологии автоматизации мониторинга технической готовности кораблей Военно-Морского Флота

М.И. Елецкий <sup>1</sup>✉, А.Ф. Зальмарсон <sup>1</sup>, А.А. Захарьев <sup>2</sup>

<sup>1</sup> НИИ ОСИС ВМФ ВУНЦ ВМФ «Военно-морская академия», г. Санкт-Петербург, 197045, Россия

<sup>2</sup> НПО «Марс», г. Ульяновск, 432022, Россия

### Ссылка для цитирования

Елецкий М.И., Зальмарсон А.Ф., Захарьев А.А. Совершенствование информационной технологии автоматизации мониторинга технической готовности кораблей Военно-Морского Флота // Программные продукты и системы. 2024. Т. 37. № 4. С. 585–591. doi: 10.15827/0236-235X.148.585-591

### Информация о статье

Группа специальностей ВАК: 2.3.1, 2.3.7

Поступила в редакцию: 08.04.2024

После доработки: 03.06.2024

Принята к публикации: 10.06.2024

**Аннотация.** Статья посвящена актуальной проблеме повышения уровня автоматизации, оперативности и качества решения функциональных задач мониторинга технической готовности кораблей за счет разработки компонентов информационной технологии. Для их реализации необходима математическая постановка задачи, основанная на методах логико-лингвистического моделирования. Потребность в совершенствовании информационной технологии автоматизации мониторинга технической готовности кораблей тесно связана с решением ряда теоретических и практических вопросов. Анализ существующей информационной технологии показал, что выделенные авторами этапы обработки информации по технической готовности кораблей являются весьма трудоемкими, а их качественное выполнение требует сохранения и анализа больших массивов информации. Однако реализованная в настоящее время в АСУ информационная технология представлена фрагментарно и несистемно. В то же время неавтоматизированная технология не обеспечивает полную и качественную реализацию всех выделенных в ходе исследований этапов обработки информации. Для решения данных проблем требуется разработка компонентов информационной технологии, обеспечивающих комплексную автоматизацию мониторинга технической готовности кораблей. Реализация указанных требований предусматривает, прежде всего, разработку формальной математической модели, адекватно и полно отображающей компоненты информационной технологии, семантические и прагматические свойства объектов мониторинга. Предложенный авторами подход на практике должен обеспечить создание из информационного ресурса качественного информационного продукта, удовлетворяющего требованиям оператора (пользователя).

**Ключевые слова:** автоматизация, информация, информационная технология, мониторинг технической готовности, логико-лингвистическая модель, обработка информации

**Введение.** Эффективное управление *технической готовностью* (ТГ) кораблей требует постоянного мониторинга различных систем и оборудования, своевременного обнаружения и предотвращения неисправностей, а также оперативного устранения возникших проблем [1].

Мониторинг – важный элемент управленческой деятельности, который позволяет контролировать процессы и результаты выполнения задач [2].

Вопросы совершенствования автоматизации управления силами флота, в частности *мониторинга технической готовности* (МТГ) кораблей, находятся в центре внимания руководства ВМФ. Данный факт подтверждается актуальными задачами органов военного управления ВМФ и основными приоритетами научно-технического и информационного обеспечения АСУ ВМФ [3].

Недостаточная автоматизация процесса мониторинга может привести к задержкам в об-

служивании кораблей, к увеличению расходов на ремонт и техническое обслуживание.

Для повышения эффективности МТГ кораблей необходимо разработать компоненты *информационной технологии* (ИТ) [4], позволяющие комплексно автоматизировать процессы сбора, обработки, анализа и визуализации данных о состоянии кораблей как сложных информационных объектов.

Развитие современных ИТ в области автоматизации МТГ кораблей позволит улучшить оперативность принятия решений, снизить риск возникновения критических (аварийных) ситуаций.

Анализ публикаций (например, [5–7]) позволил сделать вывод о том, что вопросы комплексной автоматизации МТГ кораблей в основном рассматриваются только с точки зрения автоматизации отдельных его процессов, в частности, вопросов учета информации по ТГ. В то время как содержательной семантической

обработке информации по ТГ кораблей значительного внимания не уделяется. Проведенные работы системно проработаны, но не в полной мере учитывают современные возможности и тенденции развития ИТ.

Одним из основных источников мониторинга являются донесения от кораблей. Постоянный МТГ кораблей преследует решение взаимосвязанных задач:

- проверка целостности и актуальности информации по ТГ;
- контроль соблюдения нормативных требований к параметрам ТГ;
- определение состава мероприятий по приведению параметров ТГ к нормативному значению.

Решение данных задач представляет собой фрагментарное использование возможностей существующей ИТ, позволяющей автоматизировать лишь отдельные процессы, что не обеспечивает возможность комплексной автоматизации МТГ кораблей [3].

Причиной такого положения является нехватка новых методов обработки информации, обеспечивающих комплексное решение задач, и выполнение последовательных и взаимосвязанных технологических процедур МТГ кораблей. Потребность в этом вызвана противоречием между необходимостью принятия решений в возможно более короткие сроки и увеличением сложности, объема и динамичности информационных потоков, требующих большого времени на их обработку.

Состояние информатизации в ВМФ рассмотрено авторами на основе анализа и сравнения с позициями государств-лидеров в контексте текущих достигнутых результатов в области применения ИТ [8, 9]. Так, США и Китай активно внедряют новейшие технологии в военную сферу (искусственный интеллект, кибербезопасность, кибервоенное дело, квантовые вычисления и др.). Кроме того, они активно разрабатывают системы управления информационными потоками, а также военной связью, сетевой инфраструктурой, цифровыми платформами для принятия оперативных решений и проч.

В целом информатизация в военной сфере играет ключевую роль в обеспечении безопасности и защите интересов государства. Поэтому развитие и внедрение новейших технологий и информационных систем военного назначения является приоритетной задачей для всех стран, стремящихся к укреплению своей обороноспособности [10].

Информационное обследование объектов автоматизации ОВУ ВМФ показало следующее.

1. Средства автоматизации отображают большой объем исходных данных по ТГ кораблей без их анализа.

2. Существующие технологии не обеспечивают учет и анализ большого количества нормативных требований к параметрам ТГ кораблей, изложенных в многочисленных руководящих документах.

3. Разработано достаточно много автономных задач, связанных с формированием и контролем выполнения различных планов, в то же время единая технология автоматизации МТГ кораблей не создана.

Результаты анализа существующей технологии мониторинга обусловили решение задачи разработки компонентов ИТ, обеспечивающих комплексную автоматизацию единой технологической цепочки обработки информации по ТГ кораблей. Реализация требований предусматривает прежде всего разработку формальной математической модели, адекватно и полно отображающей компоненты ИТ, семантические и прагматические свойства объектов МТГ.

Представим основные свойства объектов мониторинга.

1. Связи между отдельными параметрами ТГ, а также между параметрами и сведениями о нормативах, обобщенных и статистических показателях выражаются не количественно, а задаются с помощью лингвистических и теоретико-множественных отношений.

2. Все сведения по ТГ могут быть сообщены в виде некоторой совокупности фраз на естественном языке.

3. Сведения о методах обработки информации могут быть сообщены в виде некоторой совокупности правил, описанных фразами на естественном языке.

4. Количество параметров ТГ и их возможных значений велико, в то же время число условий активизации правил обработки информации и количество самих правил невелико и конечно.

5. Невелико и конечно количество результатов обработки информации.

6. Большинство правил обработки информации носит ситуационный характер, то есть условия их активизации определяются текущим состоянием объектов анализа (текущими значениями параметров ТГ).

Для формализации подобных объектов используются *логико-лингвистические модели* (ЛЛМ), которые активно применяются в тео-

рии и практике системного анализа, они способны качественно описывать и изучать процессы, системы и явления [11]. Данные модели характеризуются использованием логических средств обработки информации для анализа данных, представленных в лингвистическом виде. Они позволяют формализовать содержательные знания об объектах управления и связанных с ними методах обработки информации.

Главным достоинством ЛЛМ следует принять большие возможности используемых языковых средств расчетов по формализации реальных ситуаций по сравнению с традиционными математическими языками.

Отметим, что ИТ, применяющую ЛЛМ, характеризуют следующие принципиальные особенности:

- представление как декларативных, так и процедурных знаний с помощью специальных формализмов (ЛЛМ) в электронном виде для более эффективного решения задач с помощью средств автоматизации;

- понятийный характер используемой информации;

- прямой доступ конечных пользователей к средствам автоматизации в диалоговом режиме за счет образования интеллектуального интерфейса, при этом взаимодействие происходит на профессиональном языке пользователя;

- адаптация и гибкость по отношению к задачам проблемной области.

Четкая технологическая схема разработки ЛЛМ отсутствует, так как она всегда ориентирована на конкретную предметную область.

Однако комплексный анализ методологических и прикладных аспектов решения проблемы разработки ЛЛМ предполагает исследование и разработку двух взаимосвязанных типовых вопросов, вместе с тем имеющих самостоятельное научное и практическое значение [11], а именно: системы знаний об объектах и методах обработки информации, а также способов, технологии и организации программной реализации, полученной ЛЛМ.

Опираясь на определения ЛЛМ, математическую постановку задачи совершенствования ИТ автоматизации МТГ кораблей можно свести к разработке формальной модели ( $M$ ) следующего вида:

$$M = \langle T, I, P \rangle, \quad (1)$$

где  $T$  – система понятий и терминов процесса МТГ кораблей, используемых при конструировании правил и моделировании предметной области;  $I$  – информационная модель процесса

мониторинга, фиксирующая закономерности и семантический связи объектов, понятий и терминов предметной области;  $P$  – система правил-продукций обработки информации, непосредственно реализующая необходимые компоненты ИТ автоматизации МТГ кораблей.

Важнейшая проблема построения ЛЛМ заключается в разработке множества базовых понятий, или лексики ЛЛМ, позволяющей создать информационную модель и модель знаний (в виде совокупности правил обработки информации) с необходимой для практики детализацией и допускающей общение с пользователем путем однозначного понимания всех используемых лексических единиц.

### Информационная модель МТГ кораблей

Результатом изучения ЛЛМ стала разработка информационной модели МТГ кораблей. Такая информационная модель (обозначим ее  $I$ ) – это формализованная по определенным правилам совокупность всех сведений о составе, структуре и взаимосвязях следующих элементов модели:

$$I = \langle K, T, P, N, O \rangle, \quad (2)$$

где  $K$  – классификаторы (словари) процесса мониторинга – систематизированный перечень наименований объектов, каждому из которых в соответствие дан уникальный код, предназначенный для упрощения работы по сбору информации и систематизации данных, а также для указания кодов в различных унифицированных документах;  $T$  – параметры текущего состояния ТГ (техническое состояние корабля; сроки проведения мероприятий по поддержанию ТГ корабля, его систем, оружия и вооружения; состояние и сроки замеров физических полей корабля; запасы материально-технических средств; допуск корабля к выполнению отдельных мероприятий);  $P$  – формальные данные из текущих и годовых планов по установленным формам;  $N$  – нормативно-справочная информация;  $O$  – обобщенные показатели ТГ кораблей.

Такая модель должна отображать исходное представление об объектах и результатах обработки информации в ходе мониторинга. Для информационного моделирования используются различные средства [12, 13], определяющие правила структурирования данных и позволяющие реализовать их интерпретацию в соответствии с определенными требованиями, например, для данного случая:

– обеспечение возможности достаточно легкого отображения разработанной информационной модели в логическую структуру БД с помощью современных, перспективных систем управления БД как реляционного, так и нереляционного типа;

– максимальное отображение семантики проблемной области.

Информационная модель в каждый момент времени принимает вполне определенный вид, обусловленный следующими показателями.

1. Действующими множествами значений объектов, понятий и терминов предметной области –  $S1(t)$ , то есть  $K \Rightarrow S1(t)$ .

2. Текущими значениями параметров ТГ –  $S2(t)$ , то есть  $T \Rightarrow S2(t)$ .

3. Формальными данными из текущих и годовых планов –  $S3(t)$ , то есть  $P \Rightarrow S3(t)$ .

4. Нормативно-справочной информацией –  $S4(t)$ , то есть  $N \Rightarrow S4(t)$ .

5. Рассчитанными текущими значениями обобщенных показателей ТГ –  $S5(t)$ , то есть  $O \Rightarrow S5(t)$ .

Будем называть такую конкретизированную модель текущей ситуацией на объекте мониторинга –  $\{S_i(t)\}$  ( $i$  есть отличный номер ситуации, связанной с конкретным кораблем).

Каждой ситуации из  $\{S_i(t)\}$  может быть предписано одно или несколько правил обработки информации процесса МТГ кораблей. В самом общем виде система таких правил может быть задана системой продукций вида

$$\{P(k)\}: \{S_i(t) \Rightarrow Q_j; F\}, \quad (3)$$

где  $P(k)$  – имя продукции, с помощью которой данная продукция выделяется из всего множества продукций (в качестве имени может выступать некоторая лексема, отражающая суть данной продукции или порядковый номер продукции в их множестве);  $S_i(t)$  – текущая ситуация на объекте мониторинга в момент времени  $t$ ;  $Q_j$  – конкретное правило обработки информации, которое должно быть активизировано при возникновении ситуации  $S_i(t)$ ;  $F$  – постусловия продукции (актуализируются после реализации действия).

Постусловия продукции описывают действия и процедуры (продукции), которые необходимо выполнить после реализации  $Q_j$ .

Непосредственно активизируемые в правилах действия по обработке информации  $\{Q_j\}$  чаще всего представляются посредством элементарных арифметических, логических и теоретико-множественных выражений над отдельными элементами информационной модели.

Основное предназначение информационной модели – формирование и сохранение в БД АСУ ситуаций, каждой из которой соответствует одно или несколько правил обработки информации, а также сведений о результатах обработки информации по технической готовности в результате выполнения  $\{Q_j\}$ .

Реализация информационной модели МТГ кораблей обеспечивается за счет системы правил-продукций.

### Система правил-продукций обработки информации по ТГ кораблей

Данная система правил ( $P$ ) должна включать правила трех видов:

$$P = \langle P1, P2, P3 \rangle, \quad (4)$$

где  $P1$  – правила проверки целостности и актуальности информации по ТГ кораблей:  $P1: \{S1_i(t) \Rightarrow Q1_j\}$ ;  $P2$  – правила контроля выполнения нормативных требований к параметрам ТГ кораблей, обеспечивающие формализацию нормативных требований, интерпретацию текущего состояния:  $P2: \{S2_i(t) \Rightarrow Q2_j\}$ ;  $P3$  – правила формирования суммарных показателей ТГ кораблей:  $P3: \{(S3_i(t)) \Rightarrow Q3_j\}$ .

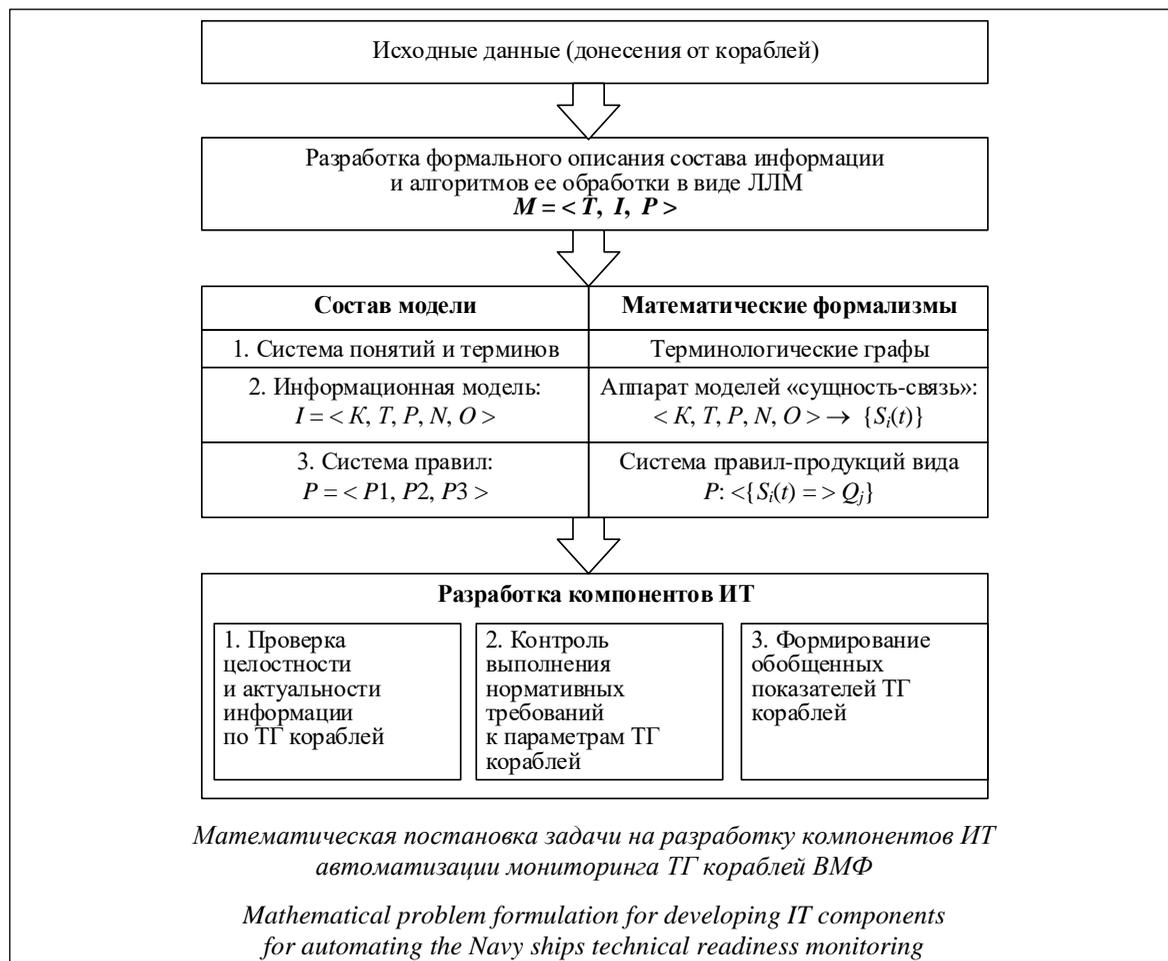
При этом действия по обработке информации, специфицированные в правилах, должны обеспечивать формирование текущей ситуации на объекте мониторинга в момент времени  $t$ .

Такой метод, основанный на введении понятий ситуации, классификации ситуации и их преобразовании, называется методом ситуационного управления [14]. Формализация правил в виде системы продукций обладает рядом свойств, делающих их весьма удобным средством описания методов обработки информации и их программной реализации.

В результате математическую постановку задачи на разработку компонентов ИТ автоматизации МТГ кораблей можно представить в графическом виде (см. рисунок).

Формирование правил в виде продукций является наиболее популярным средством представления знаний в информационных системах [12]. Продукционная модель, с одной стороны, близка к логической, фреймовой, сетевой моделям, что позволяет организовать на них эффективные процедуры вывода, а с другой, лучше сформировать и структурировать знания, делая их более доступными и понятными для пользователей.

Важно отметить, что эффективность таких продукций зависит от умения правильно структурировать информацию.



**Заключение**

В статье приведена разработанная авторами математическая постановка задачи совершенствования ИТ автоматизации МТГ кораблей ВМФ России, которая сводится к разработке

формальной модели. Реализация модели позволит обеспечить на качественно новом уровне информационную поддержку деятельности операторов органов военного управления в части управления и поддержания ТГ сил флота на заданном уровне.

**Список литературы**

1. Мухитов Э.И., Бабошин В.А., Колесников А.В., Кузнецов А.В. Автоматизация управления техническим обеспечением флота на основе использования трансдисциплинарной интеграции в логических информационно-интеллектуальных системах поддержки принятия решений // Автоматизация процессов управления. 2023. № 1. С. 4–14. doi: 10.35752/1991-2927\_2023\_1\_71\_4.
2. Цельковских А.А., Мосентз Т.А., Дубовский В.А. Концептуальная модель подсистемы мониторинга технического состояния в структуре управления полным жизненным циклом вооружения, военной и специальной техники // Вооружение и экономика. 2019. № 2. С. 36–42.
3. Елецкий М.И., Зальмарсон А.Ф. Возможности цифровых технологий проведения мониторинга технической готовности сил флота // Технологии. Инновации. Связь: матер. науч.-практич. конф. 2023. С. 267–274.
4. Елецкий М.И., Зальмарсон А.Ф. Решение задач мониторинга технической готовности кораблей Военно-Морского Флота с использованием современных компьютерных технологий // Проблемы технического обеспечения войск в современных условиях: сб. тр. по матер. науч.-практич. конф. 2022. С. 294–299.
5. Алексеев А.В., Ушакова Н.П., Смольников А.В., Сус Г.Н. Когнитивные технологии системы поддержки принятия решений и управления борьбой за живучесть корабля, судна // Системы управления и обработки информации. 2019. № 3. С. 18–27.
6. Алексеев А.В., Куприянов Д.О. Автоматизированная система оценки готовности корабля к выходу в море в составе тактической группы // Актуальные проблемы профессиональной подготовки командиров кораблей и специалистов ВМФ. Применение сил и средств ВМФ в Арктике: матер. XVI науч.-практич. конф. 2022. С. 13–21.

7. Алексеев А.В. Модель инвариантной оценки качества и эффективности объектов морской техники // Морские интеллектуальные технологии. 2020. Т. 2. № 2. С. 53–60. doi: 10.37220/MIT.2020.48.2.026.
8. Helfrich E. Managing the military's big data challenge. *Military Embedded Systems*, 2020. URL: <https://militaryembedded.com/ai/big-data/managing-the-militarys-big-data-challenge> (дата обращения: 20.03.2024).
9. Whitney J. SWaP considerations for today's rugged computers. *Military & Aerospace Electronics*, 2019, vol. 30, no. 1. URL: <https://www.militaryaerospace.com/computers/article/16709683/swap-considerations-for-todays-rugged-computers> (дата обращения: 20.03.2024).
10. Алексеев П.Н. Проблемы и перспективы применения информационных технологий в деятельности органов военного управления // Военная мысль. 2021. № 11. С. 69–79.
11. Тарасов В.Б. Проблема понимания: настоящее и будущее искусственного интеллекта. Открытые семантические технологии проектирования интеллектуальных систем // OSTIS: матер. Междунар. науч.-технич. конф. 2015. С. 25–42.
12. Аксенова Е.И. Экспертный обзор развития технологий искусственного интеллекта в России и мире. Выбор приоритетных направлений развития искусственного интеллекта в России. М.: изд-во ГБУ «НИИОЗММ ДЗМ», 2019. 38 с.
13. Егоров Д.П., Зальмарсон А.Ф., Блынских Ю.С. Научно-методический подход к оценке информационных технологий в интересах их использования в перспективных автоматизированных системах управления // Состояние и перспективы развития современной науки по направлению «АСУ, информационно-телекоммуникационные системы»: сб. статей конф. 2021. Т. 1. С. 24–33.
14. Тарасов В.Б. Искусственный интеллект: от прошлого к будущему. О жизненном пути и научном наследии профессора Д. А. Поспелова // OSTIS: матер. Междунар. науч.-технич. конф. 2020. № 2. С. 12–42. doi: 10.14357/20718594200202.

Software &amp; Systems

doi: 10.15827/0236-235X.148.585-591

2024, 37(4), pp. 585–591

### Improving the information technology for automating Navy ships technical readiness monitoring

Mikhail I. Yeletsky <sup>1</sup>✉, Andrey F. Zalmarson <sup>1</sup>, Andrey A. Zakhariev <sup>2</sup><sup>1</sup> MESC of the Navy “Naval Academy”, St. Petersburg, 197045, Russian Federation<sup>2</sup> FRPC JSC “RPA “Mars”, Ulyanovsk, 432022, Russian Federation

#### For citation

Yeletsky, M.I., Zalmarson, A.F., Zakhariev, A.A. (2024) ‘Improving the information technology for automating Navy ships technical readiness monitoring’, *Software & Systems*, 37(4), pp. 585–591 (in Russ.). doi: 10.15827/0236-235X.148.585-591

#### Article info

Received: 08.04.2024

After revision: 03.06.2024

Accepted: 10.06.2024

**Abstract.** The article describes the urgent problem of increasing an automation level, efficiency and quality of functional task solution for ships technical readiness monitoring by developing information technology components. This requires a mathematical formulation of the problem based on logic-linguistic modeling methods. The need to improve the information technology for automating the monitoring of ships' technical readiness is related to solving some theoretical and practical issues. The analysis of the existing information technology has shown that the stages of processing information on technical readiness of ships are labor-intensive. Moreover, their qualitative performance requires saving and analyzing large arrays of information. However, the information technology currently implemented in ACS is presented in a fragmentary and non-systematic manner. To this end, a non-automated technology does not provide full and qualitative implementation of all information-processing stages identified during research. The solution of these problems requires the development of information technology components that provide comprehensive automation of ship technical readiness monitoring. The implementation of these requirements provides principally mathematical model development, which adequately and completely reflects the components of information technology, semantic and pragmatic properties of monitoring objects. The proposed approach should ensure creation of a high-quality information product from the information resource that meets operator's requirements.

**Keywords:** automation, information, information technology, technical readiness monitoring, logical-linguistic model, information processing

#### References

1. Mukhitov, E.I., Baboshin, V.A., Kolesnikov, A.V., Kuznetsov, A.V. (2023) ‘Automation of the navy maintenance management based on the transdisciplinary integration in the logistic information and intelligent decision support system’, *Automation of Management Processes*, (1), pp. 4–14 (in Russ.). doi: 10.35752/1991-2927\_2023\_1\_71\_4.

2. Tselykovskih, A.A., Mosendz, T.A., Dubovsky, V.A. (2019) 'Conceptual model of the operating condition monitoring subsystem within weapons, military and special equipment full life cycle control system structure', *Armament and Economics*, (2), pp. 36–42 (in Russ.).
3. Yeletsky, M.I., Zalmarson, A.F. (2023) 'Possibilities of digital technologies for monitoring the technical readiness of the fleet forces', *Proc. Sci.-Pract. Conf. Technologies. Innovation. Connection*, pp. 267–274 (in Russ.).
4. Eletsy, M.I., Salmarson, A.F. (2022) 'Solving the tasks of monitoring of technical readiness of ships and vessels of the navy using modern computer technologies', *Proc. Sci-Tech. Conf. Troops Tech. Support Problems in Modern Conditions*, pp. 294–299 (in Russ.).
5. Alekseev, A.V., Ushakova, N.P., Smolnikov, A.V., Sus, G.N. (2019) 'Cognitive technologies for decision support and damage control system of a vessel, ship', *Information Management and Processing Systems*, (3), pp. 18–27 (in Russ.).
6. Alekseev, A.V., Kupriyanov, D.O. (2022) 'Automated system for assessing the readiness of a ship to go to sea as part of a tactical group', *Proc. Conf. Urgent problems in professional training of ship commanders and Navy specialists. Application of Navy Forces and Means in the Arctic*, pp. 13–21 (in Russ.).
7. Alekseev, A.V. (2020) 'Model for invariant assessment of quality and efficiency for research design of marine engineering facilities', *Marine Intelligent Tech.*, 2(2), pp. 53–60 (in Russ.). doi: 10.37220/MIT.2020.48.2.026.
8. Helfrich, E. (2020) 'Managing the military's big data challenge', *Military Embedded Systems*, available at: <https://militaryembedded.com/ai/big-data/managing-the-militarys-big-data-challenge> (accessed March 20, 2024).
9. Whitney, J. (2019) 'SWaP considerations for today's rugged computers', *Military & Aerospace Electronics*, 30(1), available at: <https://www.militaryaerospace.com/computers/article/16709683/swap-considerations-for-todays-rugged-computers> (accessed March 20, 2024).
10. Alekseyev, P.N. (2021) 'Problems and prospects of applying information technologies in military control bodies activity', *Military Thought*, (11), pp. 69–79 (in Russ.).
11. Tarasov, V.B. (2015) 'Problem of understanding: Present and future of artificial Intelligence'. *Proc. Sci.-Pract. Conf. OSTIS*, pp. 25–42 (in Russ.).
12. Aksenova, E.I. (2019) *Expert Review of the Artificial Intelligence Technologies Development in Russia and the World. Selection of Priority Areas for the Artificial Intelligence Development in Russia*. Moscow, 38 p. (in Russ.).
13. Zalmarson, A.F., Egorov, D.P., Blynskikh, Yu.S. (2021) 'Scientific and methodological approach to the assessment of information technologies to use them in prospective automated control systems', *Proc. Conf. State and Prospects of Modern Science Development in the Field "ACS, information-telecommunication systems"*, 1, pp. 24–33 (in Russ.).
14. Tarasov, V.B. (2020) 'Artificial intelligence: from the past to future. life path stages and scientific heritage of prof. Dmitry A. Pospelov', *Proc. Sci.-Pract. Conf. OSTIS*, (2), pp. 12–42 (in Russ.). doi: 10.14357/20718594200202.

**Авторы**

**Елецкий Михаил Игоревич**<sup>1</sup>,  
адъюнкт, m.yeletsky@gmail.com  
**Зальмарсон Андрей Феликсович**<sup>1</sup>,  
к.в.н., старший научный сотрудник,  
zalmar@mail.ru  
**Захарьев Андрей Александрович**<sup>2</sup>, к.т.н.,  
ведущий научный сотрудник, andzah@bk.ru

**Authors**

**Mikhail I. Yeletsky**<sup>1</sup>, Adjunct,  
m.yeletsky@gmail.com  
**Andrey F. Zalmarson**<sup>1</sup>,  
Cand. of Sci. (Military Sciences),  
Senior Researcher, zalmar@mail.ru  
**Andrey A. Zakhariev**<sup>2</sup>, Cand. of Sci. (Engineering),  
Leading Researcher, andzah@bk.ru

<sup>1</sup> НИИ ОСИС ВМФ ВУНЦ ВМФ «Военно-морская академия», г. Санкт-Петербург, 197045, Россия

<sup>2</sup> ФНПЦ АО «НПО «Марс», г. Ульяновск, 432022, Россия

<sup>1</sup> MESC of the Navy "Naval Academy", St. Petersburg, 197045, Russian Federation

<sup>2</sup> FRPC JSC "RPA "Mars", Ulyanovsk, 432022, Russian Federation

## Визуализация данных для бизнес-аналитики уровня доступности финансовых услуг

В.С. Федько <sup>1</sup>✉, О.А. Ковалева <sup>1</sup>, С.В. Ковалев <sup>1</sup>

<sup>1</sup> Тамбовский государственный университет имени Г.Р. Державина,  
г. Тамбов, 392036, Россия

### Ссылка для цитирования

Федько В.С., Ковалева О.А., Ковалев С.В. Визуализация данных для бизнес-аналитики уровня доступности финансовых услуг // Программные продукты и системы. 2024. Т. 37. № 4. С. 592–599. doi: 10.15827/0236-235X.148.592-599

### Информация о статье

Группа специальностей ВАК: 2.3.1

Поступила в редакцию: 16.01.2024

После доработки: 29.04.2024

Принята к публикации: 14.05.2024

**Аннотация.** В статье проанализированы релевантные программные продукты и инструменты создания дашбордов, выделены преимущества и недостатки каждого из вариантов. Определены наиболее подходящие пути решения задачи визуализации данных о доступности финансовых услуг, которая включена в процесс работы Центрального банка России. В работе представлена IDEF0-модель последовательности этапов работы с данными для визуализации показателей доступности финансовых услуг. Модель включает внесение данных в хранилище, присвоение им кодов ОКТМО, расчет расстояний между населенными пунктами, визуализацию результатов, выгрузку отчетов в необходимых форматах. Описана архитектура разработанной программы для BI-системы уровня доступности финансовых услуг, которая состоит из модулей предобработки и визуализации данных и модуля выгрузки отчетов. Отличительной особенностью является возможность работы с географическими данными расположения объектов в полностью автономном режиме. Продемонстрирован алгоритм работы модуля предобработки данных, который заключается в присвоении им кодов ОКТМО и позволяет структурировать все имеющиеся данные о наличии инструментов доступности финансовых услуг в населенных пунктах. Описана база данных для хранения всей имеющейся информации о населенных пунктах, об уровне развития финансовой инфраструктуры. Показан процесс автоматического расчета расстояний между населенными пунктами с использованием формулы гаверсинусов. Описан модуль градиентной заливки графиков, позволяющий повысить наглядность данных благодаря использованию интуитивно понятной расцветки и выбору цвета в зависимости от конкретного значения. Разработанная программа позволяет существенно сократить время, затрачиваемое на обработку данных за один отчетный период, расширить возможности визуализации имеющейся информации.

**Ключевые слова:** визуализация данных, предобработка данных, дашборд, BI, доступность финансовых услуг

**Введение.** Определение уровня развития инфраструктуры предоставления финансовых услуг в *населенных пунктах* (НП) является одной из задач работы Центрального банка России. Поскольку процесс недостаточно автоматизирован, возникает потребность в ручной обработке информации сотрудниками банка. В свою очередь, это напрямую влияет на увеличение затрачиваемого времени, повышение риска ошибок, связанных с человеческим фактором. Поэтому разработка программы для *Business Intelligence* (BI) доступности финансовых услуг, позволяющая автоматизировать работу по предобработке данных, визуализации получаемых результатов, выгрузке аналитических отчетов, является актуальной задачей.

Одним из самых распространенных современных решений в области визуализации является применение BI-технологий и создание дашбордов. BI-технологии применяются в качестве инструмента работы для аналитики и используют методы визуализации данных для

получения информации из структурированных баз [1, 2]. В статьях [3, 4] продемонстрированы варианты использования BI в госсекторе, в области бюджетирования, например, открытый бюджет Санкт-Петербурга, Цифровая панель управления городом, сводная аналитика на Портале госпрограмм РФ и другие.

### Обзор релевантных BI-систем и инструментов создания дашбордов

В связи с широким использованием дашбордов в самых разных сферах существует большое количество примеров их реализации. С целью формирования параметров, которые необходимо учесть при создании дашборда определения уровня доступности финансовых услуг, проанализировано несколько существующих решений и коммерческих продуктов, применяемых в финансовой сфере.

Корпоративная система аналитики и визуализации данных *Almaz BI* обладает следую-

щими преимуществами: консолидация данных из различных источников и формирование OLAP-кубов, возможность публикации аналитических отчетов в различных информационных ресурсах. К недостаткам программы можно отнести невозможность удовлетворить потребности в работе с данными геолокации объектов и в предоставлении полного функционала в автономном режиме.

Плюсы масштабируемой платформы для бизнес-аналитики Microsoft Power BI: возможность интеграции с различными источниками данных, работа с предсказательным анализом и машинным обучением. Минусами использования являются высокий порог вхождения для конечных пользователей и проблемы с работой в облачной платформе для организаций, у которых есть строгие требования к безопасности и хранению данных.

Среди ключевых преимуществ онлайн-сервиса DataDeck можно выделить возможность создания настраиваемых информационных панелей, в том числе на базе встроенных шаблонов, и возможность совместной работы над отчетами для команды. К существенным недостаткам DataDeck относятся ограниченная масштабируемость и заметное снижение скорости работы на больших объемах данных.

В статье [5] приведено сравнение российских BI-систем по следующим критериям: взаимодействие с различными источниками данных, масштабируемость, защита данных, визуализация, простота использования. Данные критерии в числе прочих можно применить и к рассматриваемой задаче. Авторы статьи [6] описывают сравнительную характеристику BI-систем, в которой в качестве недостатков выступают проблемы с импортом и обработкой больших данных, трудности для пользователей без технической базы, низкий уровень графики.

В статье [7] описывается проектирование дашборда для повышения результатов работы производственных процессов. Практический пример реализован с поддержкой файла Excel, в котором представлены возможности ввода и обновления информации, вывода графических материалов визуализации. Файл выполнен с учетом требований к простоте в понимании и использовании любыми сотрудниками с базовыми навыками работы с компьютером. Также учтены требования к уменьшению затрачиваемого времени для обновления данных и генерации информации. Рабочее пространство разделено на множество листов с несколькими областями ввода информации, представ-

ления результатов обработки имеющихся данных, дополнительных информационных блоков. Авторы статьи [8] описывают процесс создания дашборда для поддержки принятия решений в области онлайн-банкинга. Продукт создан с помощью Tableau и содержит набор данных о транзакциях в 2016–2020 гг. На первом этапе работы дашборда происходит выборка данных на основе специализированных атрибутов. Далее следует предварительная обработка данных и их обогащение актуальной информацией, которая затем подвергается интеллектуальному анализу. Он предназначен для извлечения новых полезных показателей из больших объемов данных. В рамках исследования установлено, что применение разработанного дашборда повысило использование BI на 30 %.

Статья [9] посвящена разработке фреймворка для создания дашбордов больших данных о перемещениях. Практические примеры показывают возможные пути решения задач визуализации данных о географических местоположениях автоматизированным методом подсчета расстояний, прокладыванием маршрутов. Работа инструмента строится на обработке GPS-координат и графическом представлении полученных показателей.

Среди преимуществ рассмотренных программных решений можно выделить несколько основных пунктов. Во-первых, возможность предварительной автоматической обработки и изменения имеющихся данных перед началом работы инструментов визуализации. Это позволяет структурировать информацию и избавиться от лишних данных. Во-вторых, ориентированность программ на быстрое действие визуальных элементов при переключении фильтров и изменении иных вводных показателей. В-третьих, возможность работы в привычном для большинства пользователей Microsoft Excel. Полезным при разработке дашборда финансовой доступности будет пример, описанный в [7]. В-четвертых, возможность работы с географическими координатами и расчетом расстояний, которая необходима для определения транспортной доступности НП.

К недостаткам рассмотренных вариантов относятся невозможность использования в полном объеме имеющихся инструментов без подключения к Интернету. Например, без доступа к Сети не работают функции, связанные с географическими расположениями, с построением карт. Возможность функционирования в автономном режиме необходима при создании дашборда уровня доступности финансовых

услуг, так как большинство устройств конечных пользователей, которыми являются сотрудники ЦБ РФ, непосредственно задействованные в данном направлении работы, не имеют выхода в Интернет из соображений информационной безопасности. Другими минусами можно назвать недостаточную степень оценивания выходных данных, отсутствие сравнения результатов за различные отчетные периоды, слабое акцентирование внимания на лучшие и худшие показатели, на основе которых стоит принимать управленческие решения.

Компромиссом в данной ситуации является разработка собственного инструмента с учетом всех выявленных преимуществ и недостатков проанализированных решений.

Целью исследования является разработка программы для VI доступности финансовых услуг. Для достижения цели необходимо решить следующие задачи: определить последовательность этапов работы с данными, разработать модули визуализации имеющихся данных, выгрузки отчетов в необходимых форматах, автоматической предобработки данных, а именно, модули присвоения кодов ОКТМО и расчета транспортной доступности.

### Разработка модулей программы для VI-систем уровня доступности финансовых услуг

Не всегда возможно найти подходящую автоматизированную информационную систему, которая удовлетворяла бы всем требованиям [10]. И тогда возникает необходимость создать систему в соответствии с индивидуальным зака-

зом. Авторами показана зависимость этапа программирования от этапа проектирования. Отмечен также важный этап проектирования информационной системы.

Основные этапы работы с данными, используемыми в программе, представлены на рисунке 1.

Ключевые типы архитектуры VI-систем рассмотрены в статье [11]. Среди основных вопросов, которые возникают при ее разработке, выделяются сложность масштабирования, низкая производительность, проблемы в обеспечении безопасности и в управлении. На основе произведенного анализа создана архитектура программы для VI-систем уровня доступности финансовых услуг (рис. 2).

### Структура модулей предобработки данных

Предобработка данных начинается с присвоения имеющимся данным о наличии финансовых инструментов кодов ОКТМО, являющихся уникальными для каждого НП. С использованием возможностей Microsoft Excel и VBA разработан модуль, в основе которого заложен принцип динамических массивов, что позволяет избежать лишнего перебора и существенно сокращает время обработки (рис. 3). Все НП вносятся в соответствующие массивы своих районов, после чего среди адресов установки финансовых инструментов ищутся соответствия. При наличии совпадений в названиях района и НП происходит присвоение кода ОКТМО. При обычном переборе 3 000 НП присвоение всех кодов занимает около 70 секунд,

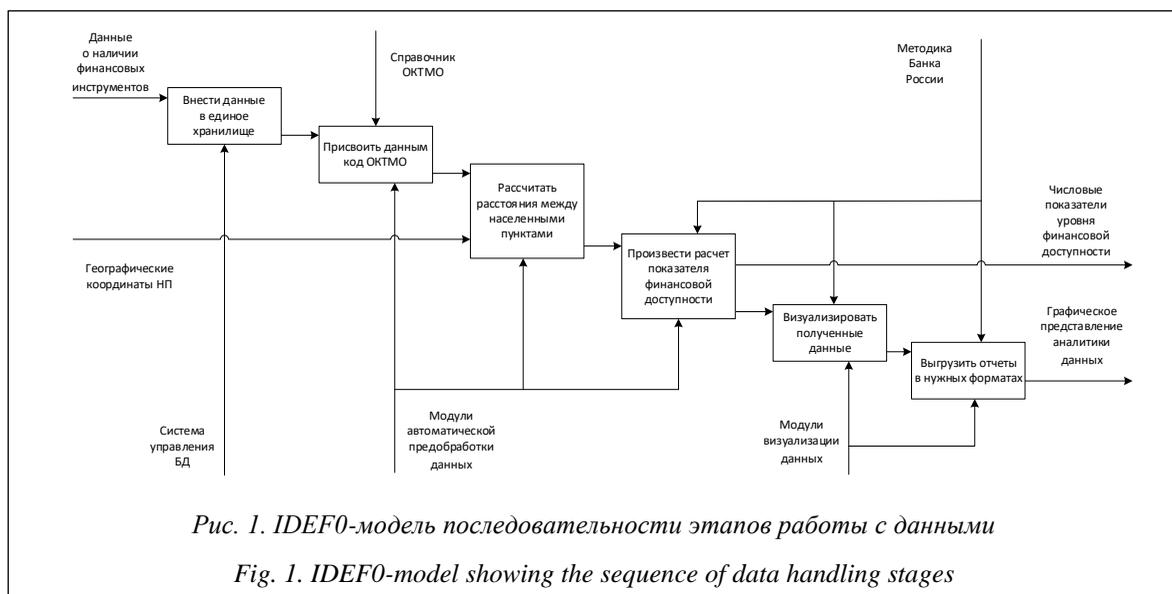


Рис. 1. IDEF0-модель последовательности этапов работы с данными

Fig. 1. IDEF0-model showing the sequence of data handling stages



а разработанный модуль справляется с этим менее чем за 2 секунды. Разница более заметна, если требуется присвоить коды одновременно нескольким десяткам или сотням тысяч инструментов доступности финансовых услуг.

После предобработки все данные помещаются в хранилище, реализованное в виде БД (рис. 4), состоящее из пяти ключевых таблиц и большого множества вспомогательных и служебных. В основной таблице ФД\_68 собраны все имеющиеся числовые показатели о наличии финансовых инструментов во всех исследуемых НП. Связь с остальными таблицами происходит по полю кода ОКТМО с таблицей ОКАТО, кода региона с таблицей Область, кода района с таблицей Районы, кода муниципального образования второго уровня с таблицей Горсоветы. В общей сложности таблица ФД\_68 содержит свыше 80 полей для хранения показателей.

Следующий этап предобработки данных – расчет расстояний между НП для определения транспортной доступности. Использование формулы гаверсинусов – возможный способ автоматического определения расстояния. Данная формула позволяет определить расстояние большого круга между двумя точками сферы исходя из их долготы и широты. Широко используемый в навигации частный случай закона гаверсинусов связывает стороны и углы сферических треугольников [12]:

$$\Delta\sigma = \arctan \frac{\sqrt{(\cos \phi_2 \sin(\Delta\lambda))^2}}{\sin \phi_1 \sin \phi_2 + \cos \phi_1 \cos \phi_2 \cos(\Delta\lambda)} + \frac{\sqrt{(\cos \phi_1 \sin \phi_2 - \sin \phi_1 \cos \phi_2 \cos(\Delta\lambda))^2}}{\sin \phi_1 \sin \phi_2 + \cos \phi_1 \cos \phi_2 \cos(\Delta\lambda)},$$

где  $\phi_1, \lambda_1, \phi_2, \lambda_2$  – широта и долгота двух точек в радианах;  $\Delta\lambda$  – разница координат по долготе;  $\Delta\sigma$  – угловая разница.

Результатом вычисления по представленной формуле является расстояние по прямой от центра одного НП до центра другого. Расчет по формуле гаверсинусов и расчет показателей финансовой доступности реализованы в системе хранения данных с использованием возможностей Microsoft Access и VBA.

### Структура модулей визуализации данных

Одним из разработанных модулей визуализации данных является паспорт НП. С использованием Microsoft Access создано несколько пользовательских форм, на одной из них можно увидеть список всех НП с возможностью фильтрации по регионам, районам, по наличию или отсутствию тех или иных финансовых инструментов. В лаконичном виде представлены основные показатели для каждого НП. При нажатии на один из них открывается

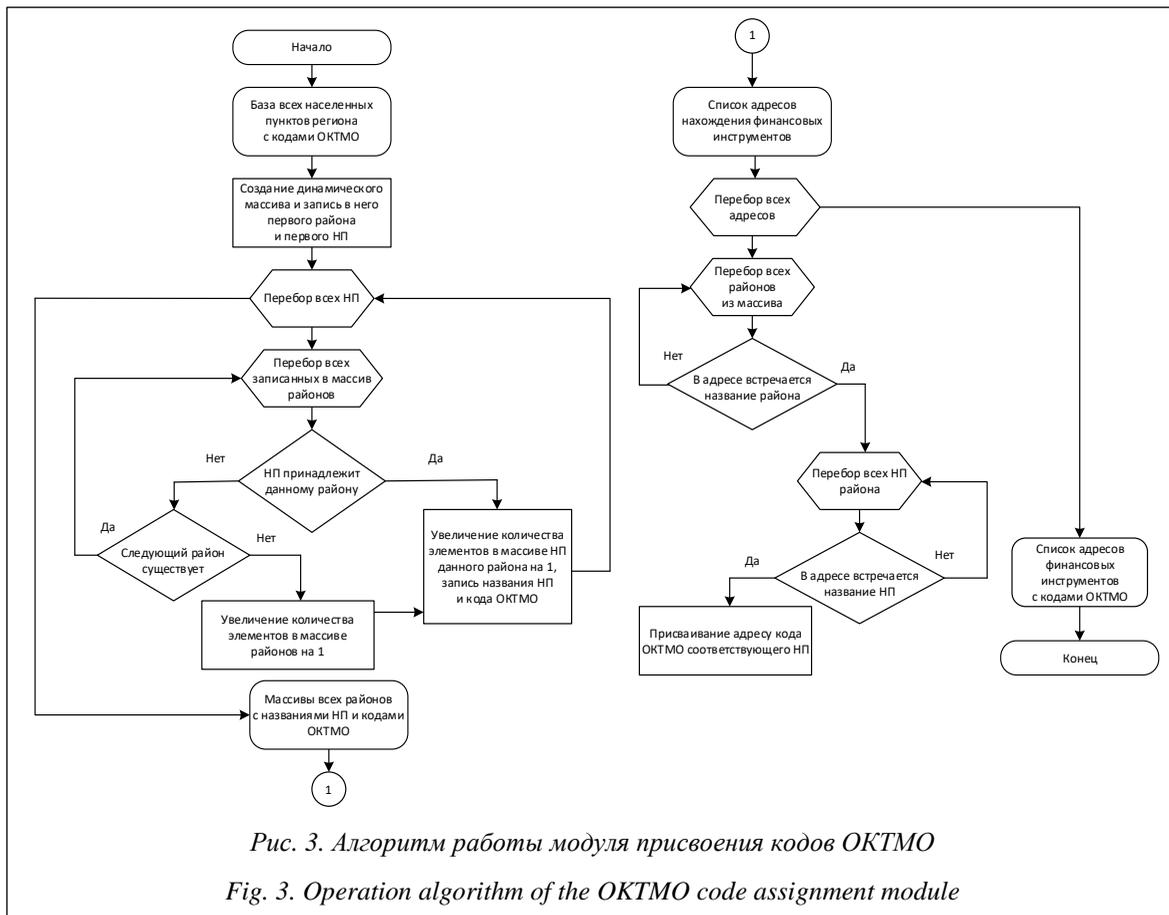


Рис. 3. Алгоритм работы модуля присвоения кодов ОКТМО

Fig. 3. Operation algorithm of the OKTMO code assignment module

форма, в которую подгружаются все имеющиеся по нему данные. В общей сложности на форме представлено более 50 показателей, сгруппированных в 4 вкладки.

Для создания аналитических срезов по целым муниципальным образованиям разработан модуль с использованием Microsoft Excel. Модуль реализован на принципе сводных таблиц, которые позволяют создать большое количество графиков, диаграмм по различным направлениям аналитической деятельности. С целью преобразования таблиц заложена возможность формирования списков НП по определенным критериям. С использованием VBA реализована автоматическая выгрузка визуализированных данных и списков НП в презентацию PowerPoint, что сильно облегчает и ускоряет процесс создания отчетов для задействованных в работе над финансовой доступностью органов.

По принципу сводных таблиц разработан и модуль сравнения показателей муниципальных образований за разные отчетные периоды. Для отслеживания изменений реализован с помощью VBA процесс выгрузки данных модуля в файл Microsoft Word с автоматической гене-

рацией текста, описывающего произошедшие изменения.

Еще одним важным инструментом повышения наглядности визуализированных данных является модуль градиентной заливки графиков. Для любой диаграммы в дашборде можно выбрать цвета в максимальной и минимальной точке. Остальные столбцы будут окрашены в соответствии с цветами значений. Данная технология позволяет использовать интуитивно понятные цвета в графике, например, переход от красного цвета у самых низких показателей к зеленому у самых высоких. Для каждого из трех цветов в палитре RGB высчитывается значение в зависимости от значения в столбце по формуле

$$color = If(cur \geq avg, clavg - (cur - avg) / (max - avg) * (clravg - clrmax), clrmin - (cur - min) / (avg - min) * (clrmin - clravg)),$$

где *color* – значение цвета в палитре RGB для текущего столбца; *cur* – значение в текущем столбце; *max*, *min*, *avg* – максимальное, минимальное, среднее значения ряда графика; *clrmax*, *clrmin*, *clravg* – максимальное, минимальное, среднее значения цвета в палитре RGB.

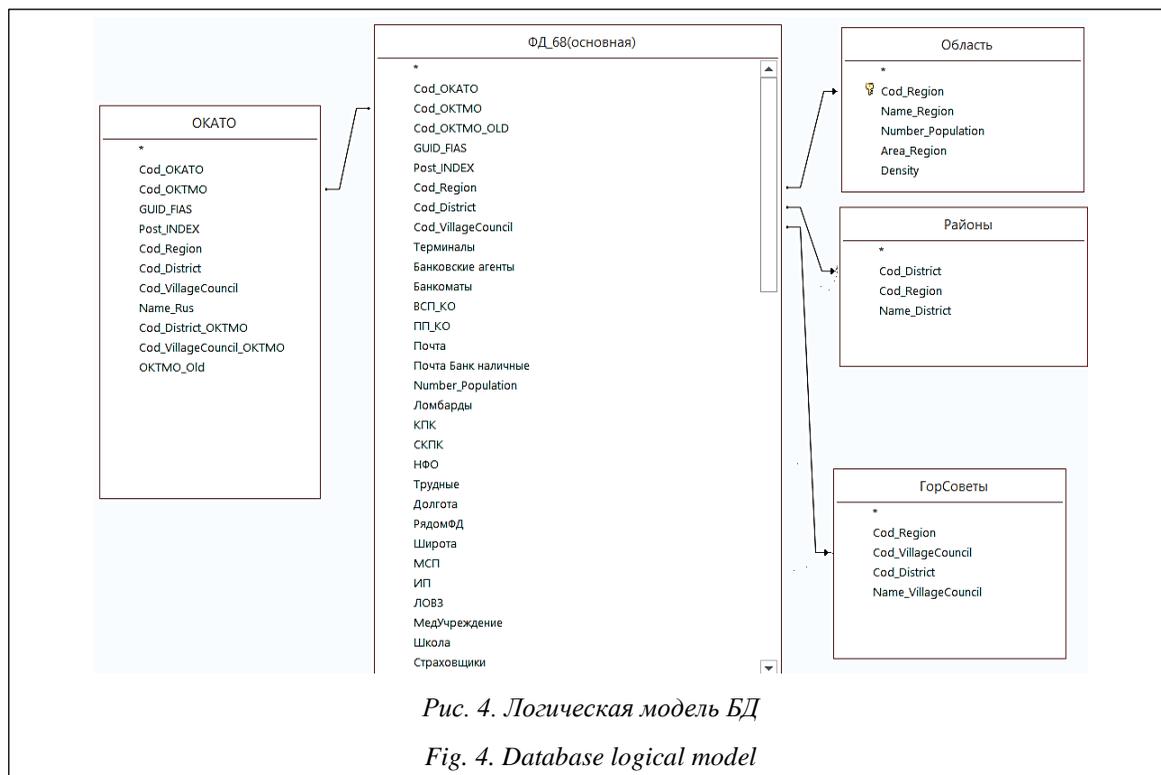


Рис. 4. Логическая модель БД

Fig. 4. Database logical model

### Анализ результатов практического применения разработанной программы

В хранилище программы содержатся обработанные данные обо всех НП Центрального федерального округа РФ за 4 отчетных периода – это около 240 тысяч записей с более чем 50 показателями в каждой из них. Для любой из записей возможно формирование паспорта НП, для каждого из отчетных периодов возможны генерация аналитических отчетов в разрезе муниципальных образований и их выгрузка в презентацию PowerPoint (<http://www.swsys.ru/uploaded/image/2024-4/8.jpg>). Для любых двух отчетных периодов возможны создание отчета с отслеживанием произошедших изменений и тенденций и их выгрузка в текстовый файл с генерацией текста описания. Программа показала хорошие результаты масштабируемости и быстродействия. Соблюдено условие полностью автономной работы без необходимости подключения к Интернету.

Разработанная программа для BI-системы уровня доступности финансовых услуг внедрена в процесс работы ЦБ РФ и используется для аналитики и создания отчетов для всех 18 регионов Центрального федерального округа.

### Выводы

В статье проанализированы релевантные программы для BI и способы создания дашбордов. Представлена архитектура программы, позволяющей хранить и визуализировать данные, производить их предобработку. Приведены результаты внедрения и практического использования разработанного продукта с использованием Microsoft Access, Microsoft Excel, VBA.

Основные преимущества данной программы – возможность работы с географическими данными расположения объектов, полностью автономный режим работы без подключения к Интернету, автоматическая выгрузка формируемых отчетов в различных форматах, в том числе в .docx, .pptx.

Дальнейшими направлениями исследования являются разработка модуля для создания тепловых карт муниципальных образований. Авторы планируют предусмотреть возможность визуализации графического расположения НП с отображением имеющихся в них инструментов доступности финансовых услуг.

### Список литературы

1. Prayitno D. Application of business intelligence for banking performance based on products analysis. IJPSAT, 2018, vol. 6, no. 2, pp. 554–569.

2. Ain N., Vaia G., DeLone W.H., Waheed M. Two decades of research on business intelligence system adoption, utilization and success – A systematic literature review. *Decision Support Systems*, 2019, vol. 125, art. 113113. doi: 10.1016/J.DSS.2019.113113.
3. Цуканова О.А., Ярская А.А. Сущность и роль BI-систем в современной экономике // Науч. журнал НИУ ИТМО. Сер.: Экономика и экологический менеджмент. 2021. № 2. С. 79–85. doi: 10.17586/2310-1172-2021-14-2-79-85.
4. Пешкова Г.Ю., Самарина А.Ю. Перспективы применения технологий Business Intelligence (BI) в Санкт-Петербурге // Экономика Северо-Запада: проблемы и перспективы развития. 2021. № 1. С. 69–74. doi: 10.52897/2411-4588-2021-1-69-74.
5. Салмина А.П., Крылов П.А. Сравнение BI-систем на российском рынке // Вуз и реальный бизнес: матер. конф. 2022. Т. 1. С. 178–184.
6. Петров Я.А., Степанов С.Ю., Сидоренко А.Ю., Глебова К.А. Business intelligence как современный инструмент бизнес-аналитики // Информационные технологии и системы: управление, экономика, транспорт, право. 2020. № 1. С. 135–140.
7. Vilarinho S., Lopes I., Sousa S. Design procedure to develop dashboards aimed at improving the performance of productive equipment and processes. *Procedia Manufacturing*, 2017, vol. 11, pp. 1634–1641. doi: 10.1016/j.promfg.2017.07.314.
8. Buananta S.E.A., Chowanda A. BI dashboard to support decision making on product promotion for payment/purchase transactions on e-banking. *JATIT*, 2021, vol. 99, no. 15, pp. 3713–3724.
9. Conrow L., Fu C., Huang H., Anrienko N., Andrienko G., Weibel R. A conceptual framework for developing dashboards for big mobility data. *CaGIS*, 2023, vol. 50, no. 5, pp. 495–514. doi: 10.1080/15230406.2023.2190164.
10. Гутгарц Р.Д. Особенности проектирования и программирования при создании информационных систем // Программные продукты и системы. 2020. Т. 33. № 3. С. 385–395. doi: 10.15827/0236-235X.131.385-395.
11. Сивов В.В. Проектирование комплексной архитектуры бизнес-аналитики // Вестн. ГГНТУ. Технич. науки. 2023. Т. XIX. № 1. С. 28–39.
12. Azdy R., Darnis F. Use of Haversine formula in finding distance between temporary shelter and waste end processing sites. *JPCS*, 2020, vol. 1500, art. 012104. doi: 10.1088/1742-6596/1500/1/012104.

Software &amp; Systems

doi: 10.15827/0236-235X.148.592-599

2024, 37(4), pp. 592–599

### Business analytics of a financial service availability level: Data visualization

Vasily S. Fedko <sup>1</sup>✉, Olga A. Kovaleva <sup>1</sup>, Sergey V. Kovalev <sup>1</sup><sup>1</sup> Derzhavin Tambov State University, Tambov, 392036, Russian Federation

#### For citation

Fedko, V.S., Kovaleva, O.A., Kovalev, S.V. (2024) 'Business analytics of a financial service availability level: Data visualization', *Software & Systems*, 37(4), pp. 592–599 (in Russ.). doi: 10.15827/0236-235X.148.592-599

#### Article info

Received: 16.01.2024

After revision: 29.04.2024

Accepted: 14.05.2024

**Abstract.** The paper analyzes relevant software products and tools for creating dashboards, highlights the advantages and disadvantages of each option. The authors identified the most appropriate ways to solve the problem of visualizing data on financial service availability that is a part of the work process of the Central Bank of Russia. The paper presents an IDEF0-model showing a sequence of data handling phases to visualize financial service availability indicators. The model includes data importing into the storage, assigning OCTMO codes (Russian classification of territories of municipal formations), calculating distances between settlements, visualizing results, and unloading reports in the required formats. The authors described the architecture of the developed program for BI-system of a financial service availability level. It consists of data preprocessing, visualization modules, and a report unloading module. A distinctive feature is the ability to work with object location data in a fully autonomous mode. The paper demonstrates the algorithm of the data-preprocessing module. It consists in assigning OCTMO codes and allows structuring all existing data on the availability of financial service accessibility tools in settlements. The authors describe the database for storing all available data about settlements and the level of financial infrastructure development. They also show the process of automatic calculation of distances between settlements using the haversine formula. They describe a gradient fill module for graphs. The module improves data clarity using intuitive coloring and selecting the color depending on a specific value. The developed program allows reducing the time spent on data processing for one reporting period, to expand visualization possibilities of the available data.

**Keywords:** data visualization, data preprocessing, dashboard, BI, availability of financial services

#### References

1. Prayitno, D. (2018) 'Application of business intelligence for banking performance based on products analysis', *IJPSAT*, 6(2), pp. 554–569.
2. Ain, N., Vaia, G., DeLone, W.H., Waheed, M. (2019) 'Two decades of research on business intelligence system adoption, utilization and success – A systematic literature review', *Decision Support Systems*, 125, art. 113113. doi: 10.1016/J.DSS.2019.113113.

3. Tsukanova, O.A., Yarskaya, A.A. (2021) 'The essence and role of BI systems in the modern economy', *Sci. J. NRU ITMO. Ser.: Economics and Environmental Management*, (2), pp. 79–85 (in Russ.). doi: 10.17586/2310-1172-2021-14-2-79-85.
4. Peshkova, G.Yu., Samarina, A.Yu. (2021) 'Prospects of business intelligence technologies use in Saint-Petersburg', *Economics of the North-West: Problems and Prospects of Development*, (1), pp. 69-74 (in Russ.). doi: 10.52897/2411-4588-2021-1-69-74.
5. Salmina, A.P., Krylov, P.A. (2022) 'Comparison of BI-systems on the Russian market', *Proc. Conf. University and Real Business*, 1, pp. 178–184 (in Russ.).
6. Petrov, Ya.A., Stepanov, S.Yu., Sidorenko, A.Yu., Glebova, K.A. (2020) 'Business intelligence as a modern tool of business analytics', *Inform. Tech. and Systems: Management, Economics, Transport, Law*, (1), pp. 135–140 (in Russ.).
7. Vilarinho, S., Lopes, I., Sousa, S. (2017) 'Design procedure to develop dashboards aimed at improving the performance of productive equipment and processes', *Procedia Manufacturing*, 11, pp. 1634–1641. doi: 10.1016/j.promfg.2017.07.314.
8. Buananta, S.E.A., Chowanda, A. (2021) 'BI dashboard to support decision making on product promotion for payment/purchase transactions on e-banking', *JATIT*, 99(15), pp. 3713–3724.
9. Conrow, L., Fu, C., Huang, H., Anrienko, N., Andrienko, G., Weibel, R. (2023) 'A conceptual framework for developing dashboards for big mobility data', *CaGIS*, 50(5), pp. 495–514. doi: 10.1080/15230406.2023.2190164.
10. Gutgartz, R.D. (2020) 'Design and programming features when creating information systems', *Software & Systems*, 33(3), pp. 385–395 (in Russ.). doi: 10.15827/0236-235X.131.385-395.
11. Sivov, V.V. (2023) 'Designing complex architecture of business analysis', *Herald of GSTOU. Tech. Sci.*, XIX(1), pp. 28–39 (in Russ.).
12. Azdy, R., Darnis, F. (2020) 'Use of Haversine formula in finding distance between temporary shelter and waste end processing sites', *JPCS*, 1500, art. 012104. doi: 10.1088/1742-6596/1500/1/012104..

**Авторы**

**Федько Василий Сергеевич**<sup>1</sup>, аспирант,  
fedko.vasily.fedko@yandex.ru

**Ковалева Ольга Александровна**<sup>1</sup>,  
д.т.н., доцент, профессор,  
solomina-oa@yandex.ru

**Ковалев Сергей Владимирович**<sup>1</sup>,  
д.т.н., доцент, профессор,  
sseedd@mail.ru

**Authors**

**Vasily S. Fedko**<sup>1</sup>, Postgraduate Student,  
fedko.vasily.fedko@yandex.ru

**Olga A. Kovaleva**<sup>1</sup>, Dr.Sci. (Engineering),  
Associate Professor, Professor,  
solomina-oa@yandex.ru

**Sergey V. Kovalev**<sup>1</sup>, Dr.Sci. (Engineering),  
Associate Professor, Professor,  
sseedd@mail.ru

<sup>1</sup> Тамбовский государственный университет имени  
Г.Р. Державина, г. Тамбов, 392036, Россия

<sup>1</sup> Derzhavin Tambov State University, Tambov,  
392036, Russian Federation

## Сценарий атаки на автоматизированную систему управления технологическим процессом с учетом уязвимости протокола Modbus TCP

А.А. Конев<sup>1</sup>, В.С. Репкин<sup>1</sup>✉, К.И. Цимбалов<sup>1</sup>

<sup>1</sup> Томский государственный университет систем управления и радиоэлектроники (ТУСУР), г. Томск, 634050, Россия

### Ссылка для цитирования

Конев А.А., Репкин В.С., Цимбалов К.И. Сценарий атаки на автоматизированную систему управления технологическим процессом с учетом уязвимости протокола Modbus TCP // Программные продукты и системы. 2024. Т. 37. № 4. С. 600–610. doi: 10.15827/0236-235X.148.600-610

### Информация о статье

Группа специальностей ВАК: 2.3.6

Поступила в редакцию: 19.03.2024

После доработки: 22.04.2024

Принята к публикации: 27.04.2024

**Аннотация.** Предметом исследования является компьютерное моделирование процесса реализации сценария атаки на автоматизированную систему управления (АСУ) технологическим процессом (ТП), в котором осуществляется эксплуатация уязвимостей промышленного протокола Modbus TCP. В данной публикации представлен аналитический обзор актуальных научных работ в области безопасности АСУ ТП, формализации и моделирования атак. Описан лабораторный стенд АСУ ТП, на основе которого сформирована компьютерная модель на языке программирования Python, включающая консоль управления (клиент Modbus), программируемый логический контроллер (сервер Modbus) и исполнительный модуль «Грузовой лифт» (графический интерфейс). Для визуализации пользовательского сценария и процессов между компонентами модели построена UML-диаграмма последовательности. Проверка адекватности модели осуществлялась путем сравнения результатов модели и стенда. Также был разработан и формально описан с помощью графической нотации MAL сценарий атаки, в котором эксплуатируются уязвимости протокола Modbus TCP, связанные с отсутствием встроенных механизмов аутентификации и шифрования. Сценарий атаки успешно реализован на модели и стенде с помощью Metasploit Framework. Для устранения возможности эксплуатации уязвимостей определены и протестированы защитные меры в виде протокола WireGuard, который благополучно справился с задачей. Результаты исследования могут быть использованы для обучения специалистов, например, в киберполигонах, для разработки, реализации и формального описания сценариев атак, для анализа уязвимостей протокола Modbus TCP, а также для тестирования программных средств защиты информации.

**Ключевые слова:** информационная безопасность, обучение специалистов, эксплуатация уязвимости, сценарий атаки, программируемый логический контроллер, Modbus TCP, Metasploit, Python, Meta Attack Language

**Благодарности.** Работа выполнена при финансовой поддержке Минобрнауки РФ в рамках базовой части госзадания ТУСУР на 2023 – 2025 гг. (проект № FEWM-2023-0015)

**Введение.** Актуальность кибербезопасности в области *информационных технологий* (ИТ) и промышленности обусловлена увеличением числа атак на данные отрасли и их критической значимостью для государства. Статистика компаний МТС RED показывает, что число кибератак на российские ИТ-компании во втором квартале 2023 года по сравнению с аналогичным периодом прошлого года увеличилось в четыре раза, достигнув отметки в 4 тысячи инцидентов ([https://www.tadviser.ru/index.php/Статья:Число\\_кибератак\\_в\\_России\\_и\\_в\\_мире](https://www.tadviser.ru/index.php/Статья:Число_кибератак_в_России_и_в_мире)). Наиболее атакуемыми отраслями в первом полугодии 2023 года стали ИТ – 35 % и АСУ технологическим процессом (ТП) – 21 %.

С увеличением зависимости от технологий и цифровых систем промышленные предприятия становятся объектами повышенного внимания злоумышленников, что подчеркивает важность разработки и применения эффективных мер противодействия. Для защиты АСУ ТП необходимо качественно обучать инженеров-

технологов и специалистов информационной безопасности, внедрять отечественные средства защиты и строго контролировать соблюдение организационных мер по защите информации [1]. Чтобы эффективно обучить сотрудников организации, нужно применять методики учебного процесса, направленные на получение практических навыков в сфере кибербезопасности. Например, использование тренировочных стендов, киберполигонов или пентест-лабораторий. Получив теоретические знания, специалист закрепляет их в условиях виртуальной имитации реальных событий, что помогает быстрее и качественнее осуществлять мониторинг инцидентов и реагирование на них в будущем. В этой связи в работе проводится моделирование сценария атаки на АСУ ТП в виртуальной среде. Атака реализуется с помощью эксплуатации уязвимостей промышленного протокола Modbus TCP и приводит к неправильной работе исполнительного модуля, а также к нарушению ТП.

## Обзор исследований

Совершенствование технологий и увеличение сложности систем управления требуют постоянной оптимизации методов контроля, управления и обеспечения безопасности [2]. Исследования и разработки в данной области становятся регулярной частью инженерной практики. В этой связи проводится обзор научных работ в области безопасности АСУ ТП и моделирования атак, а также рассматриваются утилиты для эксплуатации уязвимостей и способы формального описания сценариев атак.

### Информационная безопасность в АСУ ТП

Теме безопасности АСУ ТП в научном сообществе посвящено немало работ. Так, в исследовании [3] применялась нотация EPC (*Event-driven Process Chain*) при разработке моделей угроз для АСУ ТП. Модели могут быть использованы для оценки вероятности реализации угроз, определения связи между уязвимостями и последствиями от их эксплуатации, а также с целью принятия решений в выборе мер безопасности для систем управления. Авторы предлагают применять EPC-модели при разработке и модернизации АСУ ТП, чтобы обеспечить достаточный уровень информационной безопасности и не допустить реализацию описанных сценариев атак. Однако модели угроз не учитывают атаки внутренних нарушителей, поэтому в организации рабочей деятельности обязательно нужно вводить регламенты, политики безопасности и другие организационные меры защиты.

Исследование [4] посвящено описанию проблем, связанных с обеспечением безопасности в АСУ ТП, и анализу способов реализации атак на данные системы с помощью утилит Wireshark, Xerosploit, GoldenEye. При сравнении утилит автор особенно выделил сетевой анализатор Wireshark, так как в нем имеется возможность перехвата информации для последующей её модификации. Его недостатком является отсутствие функционала для расшифровки данных и реализации DoS-атак. Авторы выделяют три способа реализации атак: эксплуатация человеческого фактора, уязвимости сетевых протоколов или аппаратного комплекса. К человеческому фактору относятся социальная инженерия и ошибки рабочего персонала. При атаках на сетевые протоколы возможно вызвать отказ в обслуживании, перехватить сете-

вые пакеты, несанкционированно прочитать информацию с устройств или изменить информацию в устройстве, подменить сетевой пакет. В случае атак на оборудование используются специальные технические средства для электромагнитного воздействия или закладок в аппаратные комплексы ТП.

Автор публикации [5] анализирует способы нарушения информационной безопасности АСУ ТП, отмечает рекомендации по обеспечению защищенности таких систем и приводит пример атаки, осуществляемой с применением поискового инструмента Shodan. Целью атаки является несанкционированное управление ТП с помощью SCADA-системы, в которой не изменены стандартные логин и пароль, установленные производителем.

Из работ [3–5] можно сделать вывод, что атаки направлены на узлы управления, SCADA-системы, на *программируемые логические контроллеры* (ПЛК), протоколы и БД, а реализуются они с помощью уязвимостей, социальной инженерии или специального оборудования. В данной работе из перечисленных атакуемых объектов в компьютерной модели есть ПЛК и протокол Modbus. Социальную инженерию в рамках виртуальной системы осуществить не получится. Также отсутствует оборудование для проведения, например, электромагнитных атак. В лабораторном стенде используется современный ОВЕН ПЛК200 с последними обновлениями, поэтому в нем отсутствуют известные уязвимости. В некоторых публикациях обсуждаются и проводятся атаки на контроллеры [6, 7]. Первый вариант – это DDoS-атака [6] на основе стохастической сети процессов, а второй – это перегрузка линии электропередачи, к которой подключен ПЛК [7]. Следует отметить, что для защиты системы управления от определенных атак может быть использован сам ПЛК, если его модернизировать под данные цели [7]. В приведенных примерах при реализации атак необходимы определенные технологические решения, которые отсутствуют на момент проведения работ. Соответственно, рассматриваться будут уязвимости протокола Modbus.

Использованные в работах [3, 6] нотации для формального описания процессов атаки не подходят для обучения специалистов. В формализации с помощью EPC отсутствуют достаточный уровень строгости, защитные меры и логическое деление на этапы атаки. Кроме того, сложные сценарии имеют громоздкую EPC-модель. В [6] авторы построили стохастиче-

скую сеть DDoS-атаки, которая отражает алгоритм функционирования системы. Сеть сложна в понимании и не имеет структурных элементов сценария.

В статье [4] утилиты не позволяют использовать эксплойты уязвимостей из БД CVE (*Common Vulnerabilities and Exposures*) и не предоставляют полезной нагрузки, что ограничивает исследовательский потенциал. В [5] классическая атака методом подбора логина и пароля в i.LON SmartServer не сосредоточена на специфических уязвимостях, характерных для систем АСУ ТП. Аналогичная ситуация в публикации [6], где не уделяется достаточного внимания уникальным характеристикам безопасности промышленных систем, так как осуществлена традиционная DDoS-атака.

В рамках данного исследования нерационально применять симуляторы ПЛК наподобие LOGO! Soft Comfort [7] и SIMATIC S7-PLCSIM [3] для проектирования компонентов АСУ ТП. Это связано с отсутствием необходимости использовать сложные, специализированные симуляторы контроллеров, применяемые при изучении его детальной настройки, процессов и программирования. Для решения поставленных задач целесообразнее будет воспользоваться языком программирования Python при реализации виртуального стенда АСУ ТП. Это обусловлено тем, что разработка сценариев атак, уязвимых узлов и чекеров (скрипт, который проверяет наличие уязвимости на целевом хосте) в киберполигоне Amprige, в который будут интегрированы результаты исследования, осуществляется на языке Python. Соответственно, другим специалистам, работающим с Amprige, тоже будет понятна сформированная модель. Кроме того, у данного языка низкий порог вхождения, он достаточно популярен, что дает преимущество в вопросе количества готовых разработок по исследуемой теме. Например, для работы с протоколом Modbus на Python есть специальная библиотека pymodbus.

### Уязвимости промышленного протокола Modbus

Несмотря на широкое распространение в АСУ ТП, протокол Modbus слабо защищен. Он был создан во времена, когда вопросы безопасности сетей не получали должного внимания. В этой связи у протокола нет механизмов шифрования и аутентификации. Автор статьи [8] использует Wireshark для перехвата пакетов протокола Modbus TCP и отмечает, что при

использовании специального ПО можно модифицировать перехваченные пакеты. Для обеспечения безопасности передачи данных предлагается два решения VPN и DPI. DPI представляет собой метод анализа содержимого пакетов данных в сети. Он может использоваться для обнаружения и блокировки нежелательного трафика, а также для предотвращения атак. Применительно к протоколу Modbus DPI может помочь выявить аномалии в сетевом трафике, свидетельствующие о возможных атаках или модификации передаваемых данных.

В работе [9] разработан виртуальный тестовый стенд для проведения DoS-атак на протокол Modbus путем перегрузки сети, которая приводит к потере связи между устройствами, или с помощью передачи неверных команд. В контексте АСУ ТП, атака типа «отказ в обслуживании» представляет собой критическую угрозу. Такие атаки в состоянии привести к аварии или полной остановке ТП, что может не только нанести огромный ущерб организации или государству, но и представлять серьезную угрозу для природной среды.

С развитием сетевых технологий безопасность протокола Modbus стала более актуальной, так как до сих пор обнаруживаются новые уязвимости. По данным ФСТЭК России, последняя уязвимость, связанная с Modbus, была обнаружена в 2023 году. На момент написания работы в базе ФСТЭК находилось 54 510 уязвимостей, из которых 107 связаны с протоколом Modbus. Реализации атак на протокол Modbus посвящена работа [10]. Автор использует утилиту Metasploit Framework для внедрения в сетевой канал между SCADA-системой и ПЛК с целью несанкционированного чтения и модификации данных. Подлинный сервер Modbus не понимает, что команды отправляет нарушитель, так как отсутствуют механизмы аутентификации. Если установить межсетевой экран, то можно настроить получение пакетов только от определенного IP-адреса, но не стоит забывать, что существует IP-спуфинг. В качестве решения данной проблемы безопасности автор предлагает уделять особое внимание настройке сетевой конфигурации и использовать VPN при передаче информации.

Атака вида «человек посередине» проводится в работе [11] с помощью утилиты Scapy. Для проведения экспериментов была создана виртуальная система, состоящая из трех машин с операционными системами Ubuntu. Реализуется атака ARP poisoning, чтобы нарушитель стал посредником между двумя легальными

устройствами. То есть пакеты идут не напрямую, а через дополнительный сетевой узел. Атака ARP poisoning происходит в три этапа. На первом этапе нарушитель использует методы, такие как отравление ARP-таблиц, для установления доверия сети к предопределенному маршруту связи. На втором этапе при помощи функции перехвата трафика Scapy злоумышленник захватывает и модифицирует определенные пакеты между клиентом и сервером, оставаясь незамеченным. Третий этап – это удаление записей в ARP-таблицах, что маскирует следы атаки. В результате злоумышленник может изменять конфигурацию сервера Modbus, не раскрывая эти изменения его клиенту. Для защиты системы автор предлагает использовать защищенный канал связи и проверку ARP-запросов на время доставки.

В исследовании [12] представлен сценарий атаки с внедрением ложных команд в передаваемые пакеты для ПЛК. Для поиска открытых портов Modbus использовался NMAP, для анализа трафика – Wireshark. Реализована атака ARP-poisoning, которая проводилась на смоделированную виртуальную систему, состоящую из ScadaBR и OpenPLC, с использованием Python-скрипта. Для эффективной защиты от данной атаки автор предлагает использование специальных брандмауэров, которые разрабатывались для протокола Modbus. Политика безопасности таких брандмауэров работает с применением DPI [8].

В работе [10] для эксплуатации уязвимостей Modbus использовался Metasploit Framework – популярный инструмент для тестирования на проникновение и проверки безопасности систем. В нем есть не только перечень многочисленных эксплоитов, но и готовые полезные нагрузки [13]. В работе [14] утверждается, что Metasploit показал хорошие результаты при его использовании для оценки уязвимостей системы безопасности. Утилита позволяет выполнять сложные шаги, формирующие атаку. При этом с помощью созданной модели машинного обучения определялись наиболее подходящие эксплоиты для тестируемой системы. Для обучения модели использовались данные с площадки Hack the Box. Доля верных ответов модели – 33 %, авторы отмечают, что это хорошие результаты для небинарной классификации. В исследовании [15] на языке Python с помощью PyMetasploit была написана автоматизированная эксплуатация уязвимости с идентификационным номером CVE-2022-30781. Утилита позволила не только применить эксплоит

multi/http/gitea\_git\_fetch\_rce, но и доставить полезную нагрузку в виде metepreter-сессии.

Фреймворк Metasploit полностью соответствует потребностям в решении поставленных задач. Он бесплатный и обладает обширным функционалом, в том числе модулями для эксплуатации уязвимостей Modbus. Кроме того, существует множество обучающих материалов и документации, что делает его хорошим выбором для проводимого исследования.

### Особенности моделирования и формализации компьютерных атак

Во многих исследованиях упоминаются формализация и моделирование атак, при этом они тесно связаны, так как формальное описание существенно упрощает создание модели [16–18]. Формализация важна для создания точных и структурированных описаний атак и угроз в информационных системах. Она способствует наглядному восприятию сложных сценариев атак, позволяет понимать последствия эксплуатации уязвимостей и отражает последовательность действий злоумышленника [16]. Отсутствие формального описания приводит к тому, что специалисты по-разному интерпретируют один и тот же текст, каждый субъективно выделяет наиболее важные элементы и связи между ними [17].

Моделирование, в свою очередь, позволяет имитировать атаки на различные активы, например, на контроллер домена или SCADA-систему. Модель может использоваться для тестирования систем на проникновение [15, 18] и для обучения специалистов, например, в киберполигонах [15, 19, 20]. В исследовании [15] атака моделируется для дальнейшей интеграции в киберполигон. Как упоминалось ранее, для реализации автоматизированной эксплуатации уязвимости в системе, в которой используется ПО Gitea, применяются скрипт на языке Python и фреймворк Metasploit. Также в данной работе приводится формальное описание эксплуатации уязвимости с помощью графической нотации MAL: отмечены шаги злоумышленника и их причины, меры противодействия каждому шагу и используемые системы. MAL способствует понятному и подробному описанию атаки, при этом есть возможность модификации языка под свои потребности. Например, в публикации [16] авторы на основе MAL разработали язык vehicleLang для формализации сценариев атак на транспортные средства. Применимость данного языка оценивалась экс-

пертом в области автомобильной промышленности. Моделирование атак осуществлялось для оценки кибербезопасности систем и оценки времени, затраченного злоумышленником для достижения цели. В [18] создана методология на основе MAL для определения набора контрмер, оценки стоимости мер защиты и для затраченного времени на каждый шаг злоумышленника.

В работе [19] проводится сравнение четырех способов формализации: дерево атак, граф атак, MAL и Cyber Kill Chain. Авторы сделали вывод, что способ формализации зависит от решаемых задач. Дерево атак следует применять, если формальное описание используется для обучения специалистов. Для описания исключительно последовательности атакуемых узлов стоит воспользоваться графом атак. В случае подробного описания действий злоумышленника с указанием IP-адресов и конфигурации сети рекомендуется применить Cyber Kill Chain. Разработчикам сценариев атак будет удобен в использовании MAL, так как в нем учитываются используемые злоумышленником ресурсы, промежуточные активы и меры защиты от каждой уязвимости [20].

Исследователи не оставляют без внимания моделирование атак на сетевую инфраструктуру. Так, в [21] имитационная модель атаки типа Phishing реализована в среде AnyLogic, процесс атаки формально описан с помощью блок-схемы. В результате авторы предлагают использовать полученное решение для оценки защищенности сети и выявления зависимости вероятности атаки от времени в заданных условиях. В [22] для оценки среднего времени осуществления DDoS-атаки на сеть АСУ ТП применялась модель на основе стохастической сети, достоверность которой проверялась с применением экспериментального стенда «Информационная безопасность в промышленных системах».

В проводимом исследовании будет использован язык MAL как обеспечивающий точное и структурированное описание. В публикациях [16, 18, 20] отмечено, что MAL полноценно справился с задачей формализации. Кроме того, ранее была подчеркнута особенность языка MAL – возможность модификации под конкретные потребности. Еще одним важным преимуществом выбранной нотации является существование фреймворка, который применяется для автоматизированной генерации метаграфов [23]. Как и в статье [22], для оценки адекватности модели и применимости сценария

атаки будет задействован лабораторный стенд.

### Описание лабораторного стенда АСУ ТП и его виртуальной модели

Лабораторный стенд АСУ ТП, на основе которого формируется модель в виртуальной инфраструктуре, представлен на рисунке 1. В стенде присутствуют основные компоненты промышленности, например, грузовой лифт, ПЛК, печь, пневмопривод, двигатель и т.д. В данном исследовании в качестве исполнительного модуля и для дальнейшей работы взят грузовой лифт. На рисунке 2 представлена структурная схема, отражающая взаимодействие элементов стенда, которые управляют грузовым лифтом.

Для реализации системы управления грузовым лифтом была написана программа на языке ST (Structured Text). Представим исходный код:

```
// Логика Лифт
IF start THEN

    IF pos_1 THEN
        pos := 1;
        down := FALSE;
        start := FALSE;
    ELSIF pos_2 THEN
        pos := 2;
        down := FALSE;
        start := FALSE;
    ELSIF pos_3 THEN
        pos := 3;
        start := FALSE;
    ELSE
        down := TRUE;
    END_IF
ELSE

    IF call_1 OR call = 1 OR wCall_1
OR HMI_call_1 THEN
        call := 1;
        down := TRUE;
        IF pos_1 THEN
            down := FALSE;
            pos := 1;
            call := 0;
        END_IF
        ELSIF call_2 OR call = 2 OR wCall_2
OR HMI_call_2 THEN
            call := 2;
            IF pos = 1 THEN
                up := TRUE;
            ELSIF pos = 3 THEN
                down := TRUE;
            END_IF
            IF pos_2 THEN
                up := FALSE;
                down := FALSE;
```

```

        pos := 2;
        call := 0;
    END_IF

    ELSIF call_3 OR call = 3 OR wCall_3
    OR HMI_call_3 THEN
        call := 3;
        up := TRUE;
        IF pos_3 THEN
            up := FALSE;
            pos := 3;
            call := 0;
        END_IF
    END_IF
END_IF

```

Несмотря на то, что взят лифт, в создаваемой модели присутствует возможность ее модификации и масштабируемости. ПЛК соединен с АРМ, на котором установлена SCADA-система, через интерфейс Ethernet по протоколу Modbus TCP. Между ПЛК и исполнительными устройствами используется интерфейс RS-485, протокол Modbus RTU. Важно отметить основной принцип работы по протоколу Modbus: инициализировать команду может только Modbus-клиент [24].

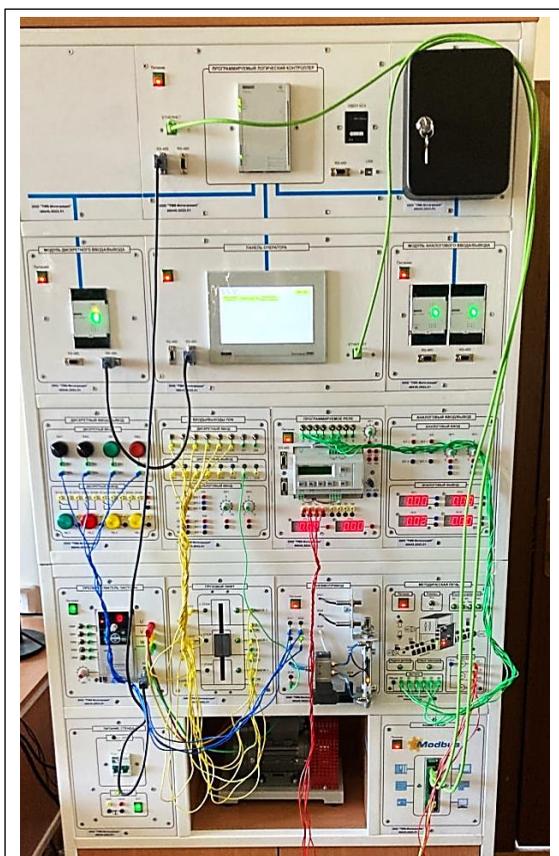


Рис. 1. Лабораторный стенд АСУ ТП

Fig. 1. APCS laboratory stand

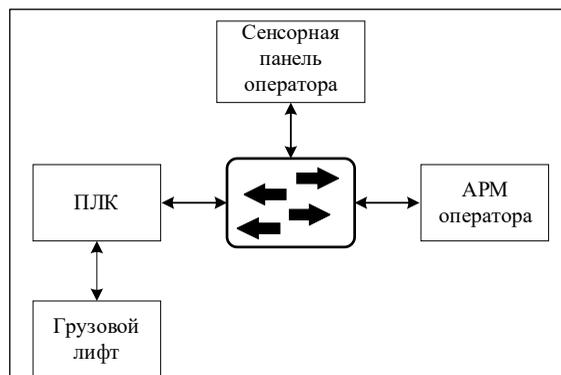


Рис. 2. Структурная схема стенда АСУ ТП

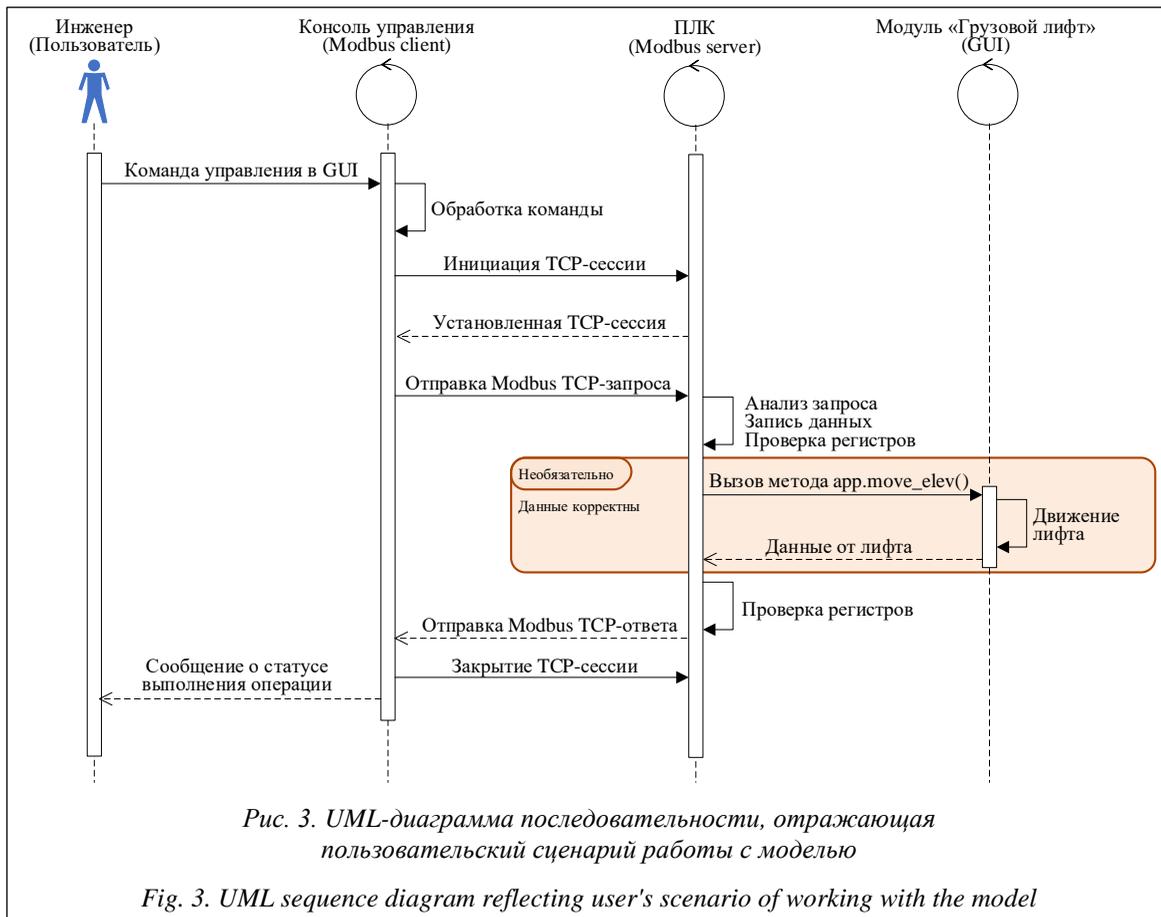
Fig. 2. Block diagram of the APCS stand

Для моделирования стенда в виртуальной среде был использован язык программирования Python 3.11 и библиотека rpymodbus 3.5.4. Созданная модель (<https://github.com/Volodyanoy/The-automated-control-system-model>) включает в себя консоль управления (клиент Modbus), ПЛК (сервер Modbus) и модуль «Грузовой лифт» (графический интерфейс). Консоль управления и ПЛК передают данные по протоколу Modbus TCP. Между ПЛК и лифтом вместо Modbus RTU были использованы возможности объектно-ориентированного программирования – модуль работает путем вызова определенных методов класса. Для визуализации пользовательского сценария и процессов между компонентами модели построена UML-диаграмма последовательности (рис. 3).

Модель сформирована в первую очередь для имитации основных процессов АСУ ТП в виртуальной среде с применением протокола Modbus. Проверка адекватности работы модели осуществлялась на основе 20 сетевых пакетов с различным содержанием. Результаты поведения модели аналогичны результатам лабораторного стенда, что говорит о применимости ее в рамках данного исследования.

### Формирование и реализация сценария атаки

Сценарий атаки представляет собой последовательность шагов, которые злоумышленник предпринимает для осуществления недопустимого события в инфраструктуре компании. Недопустимое событие – инцидент информационной безопасности в организации, который приводит к нарушению операционной деятельности [25]. В данной работе недопустимое событие – это нарушение работы ТП из-за некорректной работы исполнительного модуля «Гру-



зовой лифт». Условия реализации сценария: нарушитель преодолел сетевой периметр компании и получил доступ к рабочей станции, которая находится в подсети АСУ ТП.

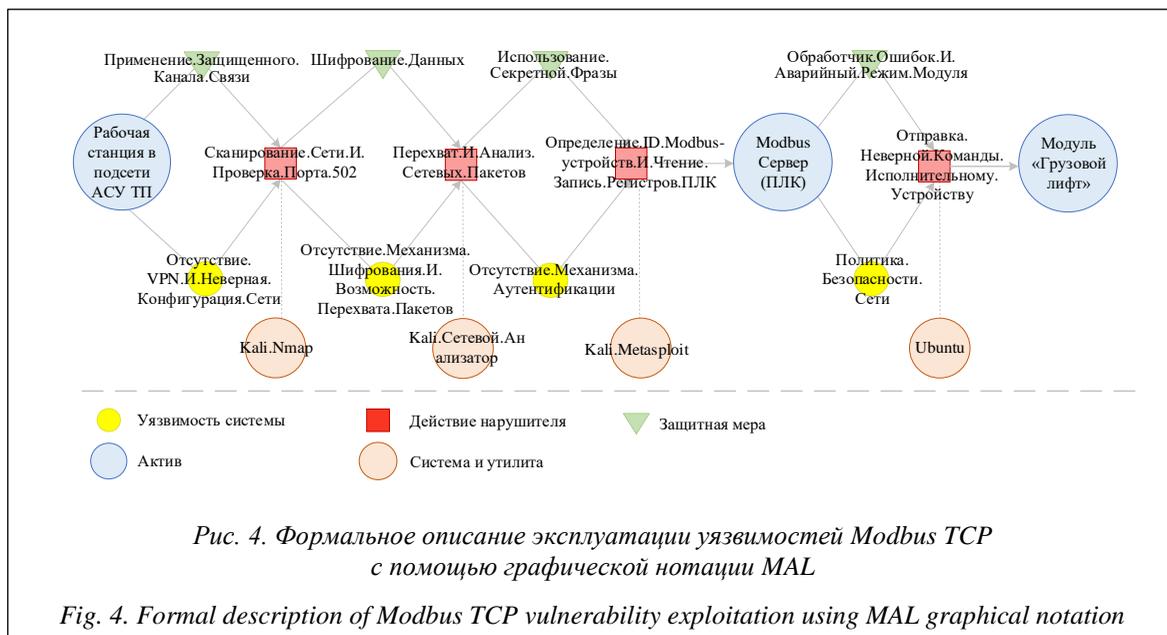
Для проведения атак применяется Metasploit Framework. Эксплуатация уязвимостей протокола Modbus осуществляется с помощью следующих модулей фреймворка: modbusclient.rb (имитация Modbus-клиента), modbusdetect.rb (обнаружение Modbus-устройств и их ID) и modbus\_banner\_grabbing.rb (сбор информации об устройствах Modbus). Для просмотра сетевых пакетов используется Wireshark, а для сканирования сети утилита Nmap.

Текстовое описание сценария атаки: нарушитель проводит сканирование сети с помощью Nmap, чтобы найти активные хосты в подсети. У каждого найденного хоста проверяется, открыт ли стандартный Modbus-порт 502. Также нарушитель использует сетевой анализатор трафика для просмотра пакетов, отправителей и получателей. В результате злоумышленник определяет IP-адреса Modbus-устройств и наличие протокола Modbus TCP. Далее применяется утилита Metasploit (modbus\_banner\_grabbing), чтобы собрать информацию об устрой-

ствах, например, сведения о производителе. С помощью модуля modbusdetect определяется ID Modbus-устройств. Теперь нарушитель знает IP-адрес сервера, порт и ID, что позволяет ему использовать модуль «modbusclient» для несанкционированной имитации действий Modbus-клиента. Злоумышленник считывает карту регистров ПЛК, что приводит к нарушению конфиденциальности информации. Затем производится запись данных в регистр ПЛК, то есть нарушается целостность информации. Из-за изменения регистра контроллер отправил исполнительному устройству неверные команды, что привело к нарушению ТП.

Формализация сценария атаки проведена с помощью графической нотации MAL (рис. 4). Текстовое и формальное описания дополняют друг друга, обеспечивая более полное и понятное представление о сценарии.

В результате были сконфигурированы виртуальные машины и успешно реализован сценарий атаки на модели и стенде. Применяв защитные меры в виде VPN, которые рекомендуются в [10–12], удалось закрыть уязвимости протокола Modbus TCP, связанные с отсутствием механизмов аутентификации и шифро-



вания. Использовался протокол WireGuard, так как он поддерживается в ОВЕН ПЛК 200. Благодаря VPN использование модулей Metasploit безуспешно, а данные в пакетах теперь зашифрованы, соответственно, если нарушитель перехватит пакеты, то не сможет понять их содержимое. Также при сканировании сети не получается обнаружить сервер, а значит, нарушитель не сможет понять, какой хост является Modbus-сервером и открыт ли там порт 502.

Поведение модели соответствует поведению реальной системы при осуществлении атаки как с учетом защитных мер, так и без них. Это свидетельствует об адекватности модели и ее применимости для исследований в области информационной безопасности и для обучения специалистов.

### Заключение

В современном промышленном мире безопасность АСУ ТП становится все более важной из-за ее массового применения. Взлом систем управления может привести к серьезным последствиям, включая нарушение производственных процессов, потерю данных, финансовый ущерб или природную катастрофу. Поэтому

предприятия должны обеспечивать надежную защиту систем от различных угроз, следовать лучшим практикам безопасности и поддерживать высокий уровень знаний трудового коллектива в данной области.

Качественная подготовка персонала промышленных предприятий, часть из которых относится к субъектам критической информационной инфраструктуры, способствует снижению рисков нарушения информационной безопасности. Представленные в работе компьютерная модель лабораторного стенда АСУ ТП и сценарий атаки, в котором эксплуатируются уязвимости протокола Modbus TCP, связанные с отсутствием встроенных механизмов аутентификации и шифрования, могут быть применены для обучения специалистов практическим навыкам выявления и реагирования на инциденты информационной безопасности. Также результаты исследования можно использовать для анализа уязвимостей протокола Modbus TCP и тестирования программных средств защиты информации. В рамках данной работы были протестированы защитные меры в виде протокола WireGuard, который обеспечил безопасность передачи данных между клиентом и сервером.

### Список литературы

1. Гарбук С.В., Правиков Д.И., Полянский А.В., Самарин И.В. Обеспечение информационной безопасности АСУ ТП с использованием метода предиктивной защиты // Вопросы кибербезопасности. 2019. № 3. С. 63–71. doi: 10.21681/2311-3456-2019-3-63-71.
2. Будников С.А., Коваленко С.М., Бочарова А.И. Методика оценки эффективности систем безопасности автоматизированных систем управления // Вопросы кибербезопасности. 2023. Т. 55. № 3. С. 2–12.

3. Машкина И.В., Гарипов И.Р. Разработка EPC-моделей угроз нарушения информационной безопасности автоматизированной системы управления технологическими процессами // Безопасность информационных технологий. 2019. Т. 26. № 4. С. 6–20. doi: 10.26583/bit.2019.4.01.
4. Цимбалов К.И., Брагин Д.С. Анализ способов нарушения информационной безопасности автоматизированной системы управления технологическими процессами // Электронные средства и системы управления: матер. докладов XIX Междунар. науч.-практич. конф. 2021. № 1-2. С. 137–139.
5. Грачков И.А. Информационная безопасность АСУ ТП: возможные вектора атаки и методы защиты // Безопасность информационных технологий. 2018. Т. 25. № 1. С. 90–98. doi: 10.26583/bit.2018.1.09.
6. Богер А.М., Соколов А.Н. Математическая модель воздействия DDoS-атаки на программируемые логические контроллеры АСУ ТП // Безопасность информационного пространства: сб. науч. тр. XXI Всерос. науч.-практич. конф. 2023. № 4. С. 196–198.
7. Ibraheem A., Ibrahim M., Shanshal A. PLC based overcurrent protection of three-phase transmission line. Proc. IMDC-SDSP, 2020. URL: <https://eudl.eu/pdf/10.4108/eai.28-6-2020.2298248> (дата обращения: 02.03.2024). doi: 10.4108/eai.28-6-2020.2298248.
8. Fedotov A.A. A research into the vulnerabilities of the Modbus protocol // Современные направления в истории, культуре, науке и технике: матер. Междунар. науч.-практич. конф. 2021. С. 95–97.
9. Rahman A., Mustafa G., Khan A.Q., Abid M., Durad M.H. Launch of denial of service attacks on the modbus/TCP protocol and development of its protection mechanisms. IJCIP, 2022, vol. 39, art. 100568. doi: 10.1016/j.ijcip.2022.100568.
10. Цимбалов К.И., Мосейчук В.А., Брагин Д.С. Атака, нарушающая конфиденциальность и целостность информации в АСУ ТП на основе открытого порта при использовании протокола Modbus TCP // ТУСУР: сб. статей. 2022. № 1-2. С. 161–164.
11. Parian C., Guldemann T., Bhatia S. Fooling the master: Exploiting weaknesses in the Modbus protocol. Procedia Comput. Sci., 2020, vol. 171, pp. 2453–2458. doi: 10.1016/j.procs.2020.04.265.
12. Alsabbagh W., Amogbonjaye S., Urrego D., Langendörfer P. A stealthy false command injection attack on Modbus based SCADA systems. Proc. CCNC, 2023, pp. 1–9. doi: 10.1109/CCNC51644.2023.10059804.
13. Ромейко Д.А., Паюсова Т.И. Обзор возможностей среды Metasploit Framework // Математ. и информ. моделирование: матер. Всерос. конф. молодых ученых. 2022. № 20. С. 318–325.
14. Valea O., Oprisa C. Towards pentesting automation using the Metasploit Framework. Proc. ICCP, 2020, pp. 171–178. doi: 10.1109/ICCP51029.2020.9266234.
15. Конец А.А., Коваленко А.С., Репкин В.С., Семенов Г.Ю. Уязвимость «Gitea Git Fetch Remote Code Execution»: анализ, формализация автоматизированной эксплуатации, меры защиты // Вестн. УрФО. Безопасность в информационной сфере. 2023. № 2. С. 67–73. doi: 10.14529/secu230207.
16. Katsikeas S., Johnson P., Hacks S., Lagerström R. Probabilistic modeling and simulation of vehicular cyber attacks: An application of the Meta attack language. Proc. ICISSP, 2019, vol. 1, pp. 175–182. doi: 10.5220/0007247901750182.
17. Novokhrestov A., Konev A., Shelupanov A., Buymov A. Computer network threat modelling. JPCS, 2020, vol. 1488, art. 012002. doi: 10.1088/1742-6596/1488/1/012002.
18. Widel W., Mukherjee P., Ekstedt M. Security countermeasures selection using the Meta attack language and probabilistic attack graphs. IEEE Access, 2022, vol. 10, pp. 89645–89662. doi: 10.1109/ACCESS.2022.3200601.
19. Якимук А.Ю., Устинов С.А., Лазарев Т.П., Коваленко А.С. Методы формализации описания сценариев кибератак // Электронные средства и системы управления: матер. докладов. 2022. № 1-2. С. 73–76.
20. Конец А.А., Репкин В.С., Семенов Г.Ю., Сермавкин Н.И. Формирование уязвимого узла «Adobe coldfusion Deserialization of Untrusted Data vulnerability» // Вопросы кибербезопасности. 2024. № 1. С. 75–81. doi: 10.21681/2311-3456-2024-1-75-81.
21. Добрышин М.М., Закалкин П.В. Модель компьютерной атаки типа «Phishing» на локальную компьютерную сеть // Вопросы кибербезопасности. 2021. № 2. С. 17–25. doi: 10.21681/2311-3456-2021-2-17-25.
22. Богер А.М., Соколов А.Н. Математическая модель вектора DDoS-атаки на сетевую инфраструктуру АСУ ТП с использованием метода топологического преобразования стохастических сетей // Вопросы кибербезопасности. 2023. № 4. С. 72–79. doi: 10.21681/2311-3456-2023-4-72-79.
23. Widel W., Hacks S., Ekstedt M., Johnson P., Lagerström R. The meta attack language – a formal description. Computers & Security, 2023, vol. 130, art. 103284. doi: 10.1016/j.cose.2023.103284.
24. Gäitan V.G., Zagan I. Modbus protocol performance analysis in a variable configuration of the physical fieldbus architecture. IEEE Access, 2022, vol. 10, pp. 123942–123955. doi: 10.1109/ACCESS.2022.3224720.

25. Alhaj T.A., Siraj M.M., Zainal A. et al. An effective attack scenario construction model based on identification of attack steps and stages. *Int. J. of Inform. Security*, 2023, vol. 22, pp. 1481–1496. doi: 10.1007/s10207-023-00701-2.

Software &amp; Systems

doi: 10.15827/0236-235X.148.600-610

2024, 37(4), pp. 600–610

### Scenario of an attack on an automated process control system taking into account Modbus TCP protocol vulnerability

Anton A. Konev<sup>1</sup>, Vladimir S. Repkin<sup>1</sup>✉, Kirill I. Tsimbalov<sup>1</sup><sup>1</sup> Tomsk State University of Control Systems and Radioelectronics (TUSUR), Tomsk, 634050, Russian Federation

#### For citation

Konev, A.A., Repkin, V.S., Tsimbalov, K.I. (2024) ‘Scenario of an attack on an automated process control system taking into account Modbus TCP protocol vulnerability’, *Software & Systems*, 37(4), pp. 600–610 (in Russ.). doi: 10.15827/0236-235X.148.600-610

#### Article info

Received: 19.03.2024

After revision: 22.04.2024

Accepted: 27.04.2024

**Abstract.** The paper focuses on computer modeling of a scenario of an attack on an automated process control system (APCS); the scenario recreates vulnerabilities in the industrial Modbus TCP protocol. The paper presents an analytical review of current scientific works related to automated process control system security, formalization and attack modeling. The authors describe the ACS laboratory stand, which became a base for a computer model in the Python programming language. The model includes a control console (Modbus client), a programmable logic controller (Modbus server) and an executive module “Freight elevator” (graphical interface). They is a special UML sequence diagram that visualizes the user scenario and the processes between model components. The authors also verified the adequacy of the model by comparing the model and bench results. In addition, they have developed and formally described the attack scenario using graphical notation MAL. It exploits Modbus TCP protocol vulnerabilities related to the lack of built-in authentication and encryption mechanisms. The attack scenario is successfully implemented on the model and bench using Metasploit Framework. To eliminate the possibility of exploiting the vulnerabilities, the authors defined and tested a defense measure in the form of the WireGuard protocol. The protocol safely accomplished the task. The study results can be used to train specialists in cyber range for developing, implementing and formal description of attack scenarios, for analyzing vulnerabilities in the Modbus TCP protocol, as well as for testing software information security measures.

**Keywords:** information security, specialist training, vulnerability exploitation, attack scenario, programmable logic controller, Modbus TCP, Metasploit, Python, Meta Attack Language

**Acknowledgements.** The work was carried out with the financial support of the Ministry of Science and Higher Education of the Russian Federation within the framework of the basic part of the TUSUR state task for 2023-2025 (project no FEWM–2023-0015)

#### References

1. Garbuk, S.V., Pravikov, D.I., Polyansky, A.V., Samarin, I.V. (2019) ‘Ensurin APCS information security using the predictive protection method’, *Cybersecurity Issues*, (3), pp. 63–71 (in Russ.). doi: 10.21681/2311-3456-2019-3-63-71.
2. Budnikov, S.A., Kovalenko, S.M., Bocharova, A.I. (2023) ‘Methodology for assessing the effectiveness of security systems of automated control systems’, *Cybersecurity Issues*, (3), pp. 2–12 (in Russ.).
3. Mashkina, I.V., Garipov, I.R. (2019) for ‘Development of EPC-Models of threats to information security of the automated process control system’, *IT Security*, 26(4), pp. 6–20 (in Russ.). doi: 10.26583/bit.2019.4.01.
4. Tsimbalov, K.I., Bragin, D.S. (2021) ‘Analysis of methods for violating the information security of automated control systems for technological processes’, *Proc. XIX Int. Sci. and Pract. Conf. Electronic Means and Control Systems*, (1-2), pp. 137–139 (in Russ.).
5. Grachkov, I.A. (2018) ‘Information security of industrial control systems: Possible attack vectors and protection methods’, *IT Security*, 25(1), pp. 90–98 (in Russ.). doi: 10.26583/bit.2018.1.09.
6. Boger, A.M., Sokolov, A.N. (2023) ‘Mathematical model of the impact of DDoS attack on programmable logic controllers of automated process control systems’, *Proc. XXI All-Russ. Sci. and Pract. Conf. Security of Information Space*, (4), pp. 196–198 (in Russ.).

7. Ibraheem, A., Ibrahim, M., Shanshal, A. (2020) 'PLC based overcurrent protection of three-phase transmission line', *Proc. IMDC-SDSP*, available at: <https://eudl.eu/pdf/10.4108/eai.28-6-2020.2298248> (accessed March 02, 2024). doi: 10.4108/eai.28-6-2020.2298248.
8. Fedotov, A.A. (2021) 'A research into the vulnerabilities of the Modbus protocol', *Proc. Int. Sci. and Pract. Conf. Current Trends in History, Culture, Sci. and Tech.*, pp. 95–97.
9. Rahman, A., Mustafa, G., Khan, A.Q., Abid, M., Durad, M.H. (2022) 'Launch of denial of service attacks on the modbus/TCP protocol and development of its protection mechanisms', *IJCIP*, 39, art. 100568. doi: 10.1016/j.ijcip.2022.100568.
10. Tsimbalov, K.I., Moseychuk, V.A., Bragin, D.S. (2022) 'Attack violating the confidentiality and integrity of information in automated control systems for technological processes using the MODBUS TCP protocol', *Proc. TUSUR*, (1-2), pp. 161–164 (in Russ.).
11. Parian, C., Guldemann, T., Bhatia, S. (2020) 'Fooling the master: Exploiting weaknesses in the Modbus protocol', *Procedia Comput. Sci.*, 171, pp. 2453–2458. doi: 10.1016/j.procs.2020.04.265.
12. Alsabbagh, W., Amogbonjaye, S., Urrego, D., Langendörfer, P. (2023) 'A stealthy false command injection attack on Modbus based SCADA systems', *Proc. CCNC*, pp. 1–9. doi: 10.1109/CCNC51644.2023.10059804.
13. Romeyko, D.A., Payusova, T.I. (2022) 'Overview of capabilities of the Metasploit Framework environment', *Proc. All-Russ. Conf. Math. and Inform. Modeling*, (20), pp. 318–325 (in Russ.).
14. Valea, O., Oprisa, C. (2020) 'Towards pentesting automation using the Metasploit Framework', *Proc. ICCP*, pp. 171–178. doi: 10.1109/ICCP51029.2020.9266234.
15. Konev, A.A., Kovalenko, A.S., Repkin, V.S., Semenov, G.Yu. (2023) 'Vulnerability «Gitea Git Fetch Remote Code Execution»: Analysis, automated exploitation formalization, and mitigation measures', *Bull. of the UrFD. Security in the Information Sphere*, 48(2), pp. 67–73 (in Russ.). doi: 10.14529/secur230207.
16. Katsikeas, S., Johnson, P., Hacks, S., Lagerström, R. (2019) 'Probabilistic modeling and simulation of vehicular cyber attacks: An application of the Meta attack language', *Proc. ICISSP*, 1, pp. 175–182. doi: 10.5220/0007247901750182.
17. Novokhrestov, A., Konev, A., Shelupanov, A., Buymov, A. (2020) 'Computer network threat modelling', *JPCS*, 1488, art. 012002. doi: 10.1088/1742-6596/1488/1/012002.
18. Wideł, W., Mukherjee, P., Ekstedt, M. (2022) 'Security countermeasures selection using the Meta attack language and probabilistic attack graphs', *IEEE Access*, 10, pp. 89645–89662. doi: 10.1109/ACCESS.2022.3200601.
19. Yakimuk, A.Yu., Ustinov, S.A., Lazarev, T.P., Kovalenko, A.S. (2022) 'Methods of formalization of cyber attack scenarios description', *Proc. Conf. Electronic Means and Control Systems*, (1-2), pp. 73–76 (in Russ.).
20. Konev, A.A., Repkin, V.S., Semenov, G.Yu., Sermavkin, N.I. (2024) 'Formation of vulnerable node «Adobe cold-fusion Deserialization of Untrusted Data vulnerability»', *Cybersecurity Issues*, (1), pp. 75–81 (in Russ.). doi: 10.21681/2311-3456-2024-1-75-81.
21. Dobryshin, M.M., Zakalkin, P.V. (2021) 'Model of a "Phishing" type of computer attack on a local computer network', *Cybersecurity Issues*, (2), pp. 17–25 (in Russ.). doi: 10.21681/2311-3456-2021-2-17-25.
22. Boger, A.M., Sokolov, A.N. (2023) 'Mathematical model of the vector of a DDoS attack on the ICS using the method of topological transformation of stochastic networks', *Cybersecurity Issues*, (4), pp. 72–79 (in Russ.). doi: 10.21681/2311-3456-2023-4-72-79.
23. Wideł, W., Hacks, S., Ekstedt, M., Johnson, P., Lagerström, R. (2023) 'The meta attack language – a formal description', *Computers & Security*, 130, art. 103284. doi: 10.1016/j.cose.2023.103284.
24. Găitan, V.G., Zagan, I. (2022) 'Modbus protocol performance analysis in a variable configuration of the physical fieldbus architecture', *IEEE Access*, 10, pp. 123942–123955. doi: 10.1109/ACCESS.2022.3224720.
25. Alhaj, T.A., Siraj, M.M., Zainal, A. et al. (2023) 'An effective attack scenario construction model based on identification of attack steps and stages', *Int. J. of Inform. Security*, 22, pp. 1481–1496. doi: 10.1007/s10207-023-00701-2.

**Авторы**

**Конеv Антон Александрович**<sup>1</sup>, к.т.н.,  
доцент, [kaa@fb.tusur.ru](mailto:kaa@fb.tusur.ru)  
**Репкин Владимир Сергеевич**<sup>1</sup>,  
студент, [repkin\\_vova@mail.ru](mailto:repkin_vova@mail.ru)  
**Цимбалов Кирилл Игоревич**<sup>1</sup>,  
аспирант, [cki@nti.tusur.ru](mailto:cki@nti.tusur.ru)

**Authors**

**Anton A. Konev**<sup>1</sup>, Cand. of Sci. (Engineering),  
Associate Professor, [kaa@fb.tusur.ru](mailto:kaa@fb.tusur.ru)  
**Vladimir S. Repkin**<sup>1</sup>,  
Student, [repkin\\_vova@mail.ru](mailto:repkin_vova@mail.ru)  
**Kirill I. Tsimbalov**<sup>1</sup>, Postgraduate Student,  
[cki@nti.tusur.ru](mailto:cki@nti.tusur.ru)

<sup>1</sup> Томский государственный университет систем управления и радиоэлектроники (ТУСУР), г. Томск, 634050, Россия

<sup>1</sup> Tomsk State University of Control Systems and Radioelectronics (TUSUR), Tomsk, 634050, Russian Federation

## Программная реализация математической модели расчета технологических параметров электротермических реакторов

А.А. Быков <sup>1</sup>✉, В.И. Бобков <sup>1</sup>, В.А. Орехов <sup>1</sup>, С.В. Незамаев <sup>2</sup>

<sup>1</sup> Национальный исследовательский университет «МЭИ» в г. Смоленске,  
г. Смоленск, 214013, Россия

<sup>2</sup> Первая горнорудная компания, г. Москва, 109004, Россия

### Ссылка для цитирования

Быков А.А., Бобков В.И., Орехов В.А., Незамаев С.В. Программная реализация математической модели расчета технологических параметров электротермических реакторов // Программные продукты и системы. 2024. Т. 37. № 4. С. 611–621. doi: 10.15827/0236-235X.148.611-621

### Информация о статье

Группа специальностей ВАК: 1.2.2

Поступила в редакцию: 23.04.2024

После доработки: 20.05.2024

Принята к публикации: 27.05.2024

**Аннотация.** В работе представлена математическая модель, описывающая возможный метод расчета тепломассовых процессов. На основе модели разработан итерационный алгоритм и создана информационная система, которая осуществляет моделирование расчета технологических параметров электротермических реакторов, используемых для переработки рудного фосфатного сырья. Особенность разработанной модели в том, что она учитывает взаимосвязь зонального представления происходящих в руднотермических печах процессов. А на базе параметров теплогидравлического режима можно получить решение задачи идентификации поступления сырья в зону реакции. В дальнейшем это даст возможность разработки системы автоматизации управления процессами в данной зоне с целью удаления примесей. Созданная математическая модель позволяет учитывать конфигурацию внутренней структуры исследуемых установок для понимания природы происходящих в них процессов и явлений. При поиске решения была также использована модель алгоритма автоматического построения сетки конечных элементов, предложен алгоритм реализации конечно-элементных уравнений с оптимизацией ширины матрицы и использованием свойств симметрии. Созданный программный комплекс позволяет выполнять автоматизированные расчеты режимов функционирования электротермических рудовосстановительных реакторов любых конструкций, в том числе и применяемых при выплавке фосфатного сырья. База данных информационной системы дает возможность пользователю выбрать модель рудовосстановительного реактора и произвести расчет тепломассообменных процессов, а также задать систему эффективных коэффициентов для расчета разработанной экспериментальной модели восстановительного реактора. При этом в программном комплексе можно выполнить не только комплексный расчет руднотермического реактора, но и отдельных его характеристик. Многодокументный интерфейс наглядно представляет результаты расчетов в виде полей изолиний и векторов скоростей тепломассовых процессов.

**Ключевые слова:** математическая модель, информационная система, программный комплекс, дифференциальные уравнения, итерационный алгоритм, компьютерное моделирование, численные методы

**Благодарности.** Работа выполнена в рамках госзадания, проект № FSWF-2023-0012

**Введение.** Математическое моделирование процессов в рудовосстановительных электротермических реакторах для получения сплавов фосфатного сырья позволяет выявить условия оптимального функционирования и снижения затрат на единицу получаемой продукции [1, 2]. Работ, отвечающих всем требованиям физики процессов, протекающих в реакторах подобного типа, весьма немного, например [3, 4].

Российскими и зарубежными исследователями предпринимались различные попытки моделирования процессов, происходящих в реакторах такого типа, но представленные модели являлись либо однофакторными, либо настолько математическими сложными, что при их реализации в компьютерной программе возникали сложности. В работах российских ученых [5, 6] представлены одномерные модели и ПО для расчета процессов, происходя-

щих только в реакторной зоне руднотермических печей, используемых для получения желтого фосфора. Сами авторы отмечают, что их разработка может быть использована как составляющая часть программного комплекса расчета технологического процесса. Зарубежными учеными также в основном рассматриваются однофакторные модели, учитывающие только один параметр реактора, в частности, температуру рабочей поверхности электродов, электрический потенциал или тепловые коэффициенты восстановительного процесса [7, 8]. Между тем необходимы технологический расчет реактора в целом и реализация возможности анализа руднотермических печей, используемых в различных отраслях металлургической промышленности. Рассматриваемые технологические установки характеризуются протеканием сложных тепломассообменных явлений,

для анализа которых требуется применение многофакторных математических моделей [9]. Таким образом, создание адекватной математической модели даст возможность не только более основательно обследовать процессы, происходящие внутри данных реакторов, но и разработать системы автоматизации расчетов, способствующие модернизации данных установок с целью повышения их энергетической эффективности.

Разработанная модель тепломассопереноса в исследуемых реакторах содержит уравнения движения фаз, тепломассообмена и распределения электрического потенциала. Реализация данного метода расчета параметров функционирования восстановительных реакторов дает возможность решения системы уравнений тепломассопереноса свести к решению уравнений теплопроводности и диффузии, что существенно облегчает реализацию поставленной задачи.

Таким образом, в зоне реакции и образования шлака уравнения тепломассопереноса можно представить в следующем виде:

$$\begin{aligned} \operatorname{div}(\lambda_T \operatorname{grad} T_f) + q_{ve} - q_{cr} \omega_{cr} &= 0, \\ \operatorname{div}(D_T \operatorname{grad} C_f) - \omega_{cr} &= 0, \end{aligned}$$

а уравнение распределения электрического потенциала как

$$\operatorname{div}((1/\rho_e) \operatorname{grad} \varphi_e) = 0.$$

Процессы теплообмена в шихтовой зоне описываются в приближении Дарси уравнениями энергопереноса [10]:

$$\begin{aligned} (1 - \varepsilon) c_m \rho_m v_m \operatorname{grad} T_m - \operatorname{div}(\lambda_m \operatorname{grad} T_m) &= \alpha_v (T_g - T_m), \\ \varepsilon c_g \rho_g v_g \nabla T_g - \operatorname{div}(\lambda_g \operatorname{grad} T_g) &= \alpha_v (T_m - T_g), \\ \operatorname{div}(\operatorname{grad} j_m) &= 0, \quad v_m = -\operatorname{grad} j_m, \\ \operatorname{div}(k \operatorname{grad} j_g) &= \gamma \omega_{cr}, \quad v_g = -k \operatorname{grad} j_g, \end{aligned}$$

где  $\lambda_T$  – эффективная теплопроводность расплава;  $q_{cr}$  – тепловой эффект восстановительного процесса;  $T_f$  – температура расплава;  $C_f$  – концентрация реагента.

Для углеродистой зоны граничные условия можно задать следующим образом:

$$\Gamma \in \Gamma_{melt}, T_f = T_{melt},$$

на стенках реактора:

$$\Gamma \in \Gamma_w, \lambda_T \frac{\partial T_f}{\partial n} = -k_{he} (T_f - T_o),$$

для переноса массы:

$$\Gamma \in \Gamma_{melt}, C = C_m.$$

Условие для стенок (непроницаемость):

$$\Gamma \in \Gamma_w, \frac{\partial C_f}{\partial n} = 0.$$

При наличии многослойной зоны для электрического потенциала граничные условия определяются для каждого компонента

на поде:

$$\Gamma \in \Gamma_d, \varphi_e = 0,$$

на поверхности электрода:

$$\Gamma \in \Gamma_e, \varphi_e = \varphi_{e3},$$

на границе расплав–шихта:

$$\Gamma \in \Gamma_{melt}, \frac{\partial \varphi_e}{\partial n} = 0,$$

на границе раздела  $k$ -й и  $k+1$ -й электропроводящих зон расплава:

$$\Gamma \in \Gamma_{k,k+1} \varphi_{ek} = \varphi_{ek+1} \text{ и } \frac{1}{\rho_{ek}} \frac{\partial \varphi_{ek}}{\partial n} = \frac{1}{\rho_{ek+1}} \frac{\partial \varphi_{ek+1}}{\partial n}.$$

На верхней границе температурные граничные условия шихты имеют вид

$$\Gamma \in \Gamma_m, T_m = T_{env},$$

в зоне плавления минеральных компонентов шихты:

$$\Gamma \in \Gamma_{melt}, T_m = T_{melt}.$$

На выходе в зоне реакции температуру газа можно задать как среднюю температуру расплава в надлежащем сечении:

$$T_g = T_f.$$

Таким образом, на фронте плавления сход шихты определяется следующим граничным условием:

$$\begin{aligned} \Gamma \in \Gamma_{melt}, \frac{\partial \varphi_m}{\partial n} &= \\ &= - \left( \lambda_T \frac{\partial T_f}{\partial n} - \lambda_m \frac{\partial T_m}{\partial n} \right) \frac{1}{(\rho_m Q_{melt})}, \end{aligned}$$

Учитывая непроницаемость стенок реактора, получаем следующее граничное условие:

$$\Gamma \in \Gamma_w, \frac{\partial \varphi_m}{\partial n} = 0,$$

где  $T_o$  – температура охладителя реактора;  $T_{melt}$  – температура плавления шихты;  $T_{env}$  – температура верхней границы шихтовой зоны;  $k_{he}$  – коэффициент теплопередачи, в случае идеальной изоляции градиент температуры равен нулю;  $Q_{melt}$  – количество теплоты, необходимое для плавления шихты;  $C_m$  – концентрация реагента в шихте;  $\rho_m$  – насыпная плотность шихты.

### Построение математической модели итерационного алгоритма расчета руднотермической установки

В представленной модели используются элементы четырехугольной формы с базисными функциями с непрерывностью первого порядка, поскольку они наиболее предпочтительны для решения уравнений конвективного теплообмена [11].

Искомая непрерывная величина в области четырехугольного конечного элемента представляется в виде

$$\Phi = \sum_{i=1}^4 N_i \Phi_i = [N][\Phi],$$

где  $\Phi_i$  – неизвестные значения искомой величины в узлах конечного элемента;  $N_i$  – базисные функции элемента. На конечном элементе вводится локальная система координат  $(\xi, \eta)$ , где  $-1 \leq \xi \leq 1, -1 \leq \eta \leq 1$ , которая позволяет использовать элементы нестандартной формы, а также упрощает интегрирование уравнений. Преобразование этой системы в декартовую систему координат  $(x, y)$  выглядит следующим образом:

$$x = \sum_{i=1}^4 N_i(\xi, \eta) x_i = [N][x],$$

$$y = \sum_{i=1}^4 N_i(\xi, \eta) y_i = [N][y],$$

где  $x_i$  и  $y_i$  – координаты узлов конечного элемента.

Базисные функции элемента с использованием локальной системы координат имеют вид

$$N_1 = \frac{1}{4}(1-\xi)(1-\eta), \quad N_2 = \frac{1}{4}(1+\xi)(1-\eta),$$

$$N_3 = \frac{1}{4}(1+\xi)(1+\eta), \quad N_4 = \frac{1}{4}(1-\xi)(1+\eta).$$

Для вычисления производных в координатах  $(\xi, \eta)$  используется соотношение

$$\begin{bmatrix} \frac{\partial \Phi}{\partial \xi} \\ \frac{\partial \Phi}{\partial \eta} \end{bmatrix} = [J] \begin{bmatrix} \frac{\partial \Phi}{\partial x} \\ \frac{\partial \Phi}{\partial y} \end{bmatrix},$$

где якобиан  $[J]$  вычисляется по формуле

$$[J] \equiv \begin{bmatrix} \frac{\partial x}{\partial \xi} & \frac{\partial y}{\partial \xi} \\ \frac{\partial x}{\partial \eta} & \frac{\partial y}{\partial \eta} \end{bmatrix}.$$

Для решения дифференциальных уравнений совместно с методом конечных элементов используется метод Галеркина, заключающийся в том, что искомое решение будет наилучшим, если интеграл взвешенной невязки равен нулю:

$$\int_S L(\Phi)[N] \, dx dy = 0,$$

где  $L(x) = 0$  – интегрируемое дифференциальное уравнение.

Для вычисления интегралов используется численное интегрирование с использованием квадратурных формул Ньютона–Котеса в локальных координатах элемента:

$$\int_S f(\xi, \eta) \, dx dy = \det[J] \sum_{k=1}^n H_k f(\xi_k, \eta_k),$$

где  $n = 4$  – количество точек интегрирования;  $H_k = 1$  – весовые коэффициенты, заданные для точек с локальными координатами  $\xi_k = \pm 0,577350$ ;  $\eta_k = \pm 0,577350$ ;  $f$  – интегрируемая функция. Полученные в результате интегрирования матрицы систем уравнений для каждого элемента объединяются в одну систему уравнений для всей области.

Для генерации сетки конечных элементов используется метод суперэлементов. При использовании этого метода в исходной области вводится нерегулярная сетка криволинейных восьмиузловых четырехугольников, которые называются суперэлементами  $S$  с узлами  $P_i = P(x_i^s, y_i^s), I = 1, 2, \dots, 8$ . При помощи отображения

$$x(\xi\eta) = x_i^s N_i, \quad y(\xi\eta) = y_i^s N_i(\xi\eta)$$

квадрат  $Q$  отображается в  $S$ , где  $N_i(\xi, \eta)$  – квадратичные базисные функции на суперэлементе:

$$N_1 = -\frac{1}{4}(1-\xi)(1-\eta)(\xi+\eta+1), \quad N_2 = \frac{1}{2}(1-\xi^2)(1-\eta),$$

$$N_3 = \frac{1}{4}(1+\xi)(1-\eta)(\xi-\eta-1), \quad N_4 = \frac{1}{2}(1+\xi)(1-\eta^2),$$

$$N_5 = \frac{1}{4}(1+\xi)(1+\eta)(\xi+\eta-1), \quad N_6 = \frac{1}{2}(1-\xi^2)(1+\eta),$$

$$N_7 = -\frac{1}{4}(1-\xi)(1+\eta)(\xi-\eta+1), \quad N_8 = \frac{1}{2}(1-\xi)(1-\eta^2).$$

Построенная в  $Q$  равномерная сетка отображается в криволинейную сетку суперэлемента  $S$ . При этом расположение узловых точек  $P_i$  на сторонах позволяет осуществлять локальное сгущение сетки. Объединение сеток, построенных таким образом на каждом суперэлементе, дает сетку на всей исходной области, как показано на рисунке 1.

На рисунке 2 представлена сетка четырехугольных конечных элементов, сгенерированная для модели электротермического рудовостановительного реактора. При этом границы суперэлементов выбирались на границах расчетных зон. Границы электродов аппроксимировались параболой.

Уравнения математической модели реактора можно привести к двум следующим формам:

$$\operatorname{div}(k \operatorname{grad} \Phi) - q = 0,$$

$$V \operatorname{grad} \Phi - \operatorname{div}(k \operatorname{grad} \Phi) + q = 0. \quad (1)$$

Аппроксимирующую систему уравнений можно представить в следующем виде:

$$[A][\Phi] = [F], \quad (2)$$

где  $[F]$  – вектор свободных членов,  $[A]$  – матрица системы уравнений.

На основе теоремы Грина и квадратурных формул численного интегрирования для уравнения (2) матрица системы уравнений задается как

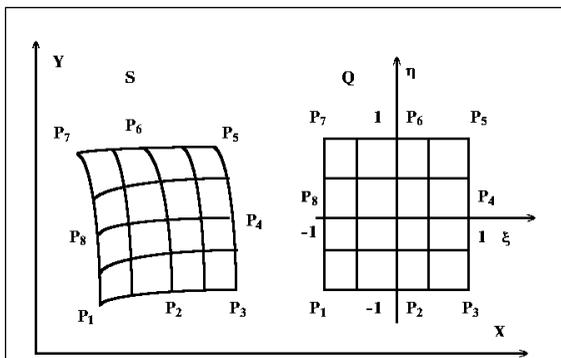


Рис.1. Криволинейный суперэлемент  $S$  исходной области и его образ  $Q$

Fig. 1. Curvilinear superelement  $S$  of the source region and its image  $Q$

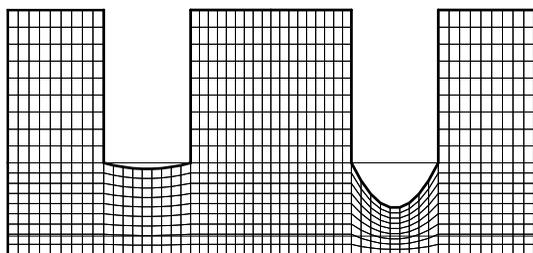


Рис. 2. Сетка конечных четырехугольных элементов

Fig. 2. Mesh of finite quadrilateral elements

$$[A] = \det [J] \sum_{k=0}^n [B]^T [K] [B],$$

где

$$[B] = \begin{bmatrix} \frac{\partial [N]}{\partial x} \\ \frac{\partial [N]}{\partial y} \end{bmatrix} - \text{матрица производных функций формы;}$$

матрица свойств.

$$[K] = \begin{bmatrix} k & 0 \\ 0 & k \end{bmatrix} - \text{матрица свойств.}$$

Вектор свободных членов определяется следующим образом:

$$[F] = q \cdot \det [J] \sum_{k=0}^n [N]^T.$$

При этом матрицы определяются в соответствующих точках  $(\xi_k, \eta_k)$  конечного элемента.

Для аппроксимации уравнений типа (1) при больших значениях отношения  $v/k$  используется модифицированный метод Галеркина, известный под названием метода Петрова-Галеркина, являющийся конечноэлементным аналогом метода конечных разностей с разностями вверх по потоку [12, 13]. При этом в качестве

весовых функций используются функции, отличные от базисной (так называемые асимметричные весовые функции). При этом условие наилучшего решения выглядит как

$$\int_S L(\Phi)[W] \, dx dy = 0,$$

здесь  $[W]$  – вектор асимметричных весовых функций, принимаемых в виде

$$W_1 = \frac{1}{4}(1-\xi)(1+1,5\alpha_{12}(1+\xi))(1-\eta)(1+1,5\alpha_{14}(1+\eta)),$$

$$W_2 = \frac{1}{4}(1+\xi)(1-1,5\alpha_{12}(1-\xi))(1-\eta)(1+1,5\alpha_{23}(1+\eta)),$$

$$W_3 = \frac{1}{4}(1+\xi)(1-1,5\alpha_{34}(1-\xi))(1+\eta)(1-1,5\alpha_{23}(1-\eta)),$$

$$W_4 = \frac{1}{4}(1-\xi)(1+1,5\alpha_{34}(1+\xi))(1+\eta)(1-1,5\alpha_{14}(1-\eta)),$$

где  $\alpha_{ij}$  – коэффициенты степени асимметричности (индексы определяют номера узлов конечного элемента, соединяемых соответствующими сторонами), оптимальное значение которых определяется из выражения

$$\alpha_{ij} = \frac{\exp\left(\frac{Pe}{2}\right) + \exp\left(-\frac{Pe}{2}\right)}{\exp\left(\frac{Pe}{2}\right) - \exp\left(-\frac{Pe}{2}\right)} - \frac{2}{Pe}.$$

Сеточное число Пекле  $Pe$  определяется по формуле

$$Pe = \frac{h_{ij} V_{cp}}{2k},$$

где  $h_{ij}$  – длина стороны конечного элемента, соединяющей узлы  $i, j$ ;  $V_{cp}$  – средняя скорость вдоль этой стороны, определяемая как

$$V_{cp} = \frac{(V_i + V_j) I_{ij}}{2},$$

где  $V_i$  и  $V_j$  – значения вектора скорости в узлах  $i$  и  $j$ ;  $I_{ij}$  – вектор направления стороны  $ij$ .

При использовании асимметричных базисных функций матрица системы линейных уравнений, аппроксимирующей дифференциальное уравнение, будет определяться следующим образом:

$$[A] = \det [J] \sum_{k=0}^n [W]^T [V] [B] + [D]^T [K] [B],$$

$$\text{где } [D] = \begin{bmatrix} \frac{\partial [W]}{\partial x} \\ \frac{\partial [W]}{\partial y} \end{bmatrix} - \text{матрица производных функций формы; } [V] - \text{вектор скоростей.}$$

Вектор свободных членов:

$$[F] = q \cdot \det [J] \sum_{k=0}^n [W]^T.$$

Результирующая матрица системы уравнений в методе конечных элементов является разреженной. Для решения таких систем уравнений в течение продолжительного времени использовались прямые методы типа метода исключения Гаусса с выбором исключаемого элемента и с учетом разреженности и ленточной структуры матрицы [14]. При этом большое внимание уделялось использованию алгоритмов перенумерации узлов сетки, минимизирующих ширину ленты, позволяющих сократить объем памяти ЭВМ, необходимой для хранения матрицы, и время расчета [15].

Матрицы системы линейных уравнений, аппроксимирующие дифференциальные уравнения диффузии, уравнения типа (1), являются симметричными, положительно определенными и хорошо решаются прямыми методами. В отличие от них матрицы систем, аппроксимирующие дифференциальные уравнения с конвективными членами, уравнения типа (1), не имеют регулярной структуры, несимметричны, неположительно определены и являются плохо обусловленными, что ограничивает применение прямых методов для решения таких систем. Кроме того, для увеличения точности необходимо увеличивать количество узлов расчетной сетки конечных элементов и, следовательно, увеличивать размерность систем уравнений, для которых нельзя применять прямые методы решения из-за нарастающих ошибок округления, связанных с выполнением большого числа арифметических операций. Для устранения этих трудностей необходимо использовать итерационные методы.

Данный метод формирования математической модели позволяет на основе применения управляющих параметров определить закономерности изменения температуры газов, возникающих в подсводной области реактора.

#### Итерационный алгоритм реализации математической модели расчета электротермических установок

В основе построения алгоритма лежит метод сопряженных градиентов. Данный метод адаптирован для решения симметричных положительно определенных систем алгебраических уравнений [16], его обобщение – метод уменьшения размерности (*Induced Dimension Reduction* – IDR) пригоден как для несимметричных, так и для неположительно определенных систем уравнений. Приведем алгоритм метода сопряженных градиентов (IDR) для решения системы уравнений вида  $[A][X] = [B]$ .

Задаются начальное приближение  $[X]_0$  и точность расчета  $\varepsilon$ :

$$1) [F]_0 = [A][X]_0 - [B]; [dG]_0 = [dY]_0 = 0;$$

$$2) n = n + 1;$$

$$3) [S]_n = [F]_{n-1} + w_{n-1}[dG]_{n-1}; [T]_n = [A][S]_n;$$

$$4) \text{ если } n = 1 \text{ или } n \text{ четное } \alpha_n = \frac{([T]_n[S]_n)}{([T]_n[T]_n)},$$

переход к п. 7;

$$5) \alpha_n = \alpha_{n-1};$$

$$6) [dX]_n = w_{n-1}[dY]_{n-1} - \alpha_n [S]_n;$$

$$7) [dF]_n = w_{n-1}[dG]_{n-1} - \alpha_n [T]_n;$$

$$8) [X]_n = [X]_{n-1} + [dX]_n; [F]_n = [F]_{n-1} + [dF]_n;$$

$$9) \text{ если } n \text{ четное, } [dG]_n = [dG]_{n-1}; [dY]_n = [dY]_{n-1},$$

переход к п.12;

$$10) [dG]_n = [dF]_n; [dY]_n = [dX]_n;$$

$$11) w_n = -\frac{([F]_n[F]_n)}{([F]_n[dG]_n)};$$

$$12) \text{ если } \frac{|[dX]_n|}{|[X]_n|} > \varepsilon, \text{ переход к п. 3;}$$

13) окончание расчета с результатом  $[X]_n$ .

Системы алгебраических уравнений метода конечных элементов, как уже отмечалось, имеют ленточную структуру. При этом время их решения  $T \sim \beta^2 N$ , а объем памяти ЭВМ, необходимый для хранения матрицы,  $V \sim \beta N$ , где  $N$  – число неизвестных;  $\beta$  – ширина ленты. Величина  $\beta$  зависит от нумерации узлов сетки и может принимать значения в широком диапазоне:  $\sqrt{N} \leq \beta < N$ .

В настоящей работе при хранении матрицы системы уравнений исключаются почти все нулевые (незначимые) коэффициенты. Это достигается использованием двух одномерных массивов, один из которых хранит вещественные значения коэффициентов уравнений, а другой целые номера переменных, соответствующие этим коэффициентам. Для четырехугольных билинейных конечных элементов количество узлов, соседних заданному узлу, не превышает восьми. Таким образом, подобный способ хранения матрицы системы уравнений эквивалентен традиционному с шириной ленты, равной девяти. Этот способ позволяет сократить объем памяти ЭВМ, необходимой для хранения матрицы системы уравнений при  $N = 1\,000$  и использовании вещественных чисел с двойной точностью, почти в три раза по сравнению с традиционным и почти в девять раз уменьшить время расчета.

Массивы для хранения коэффициентов системы уравнения и номеров соответствующих переменных имеют размерность  $N \times M$ , где  $M$  – максимальное количество переменных, входя-

щих в уравнение. Приведем алгоритм расчета вектора  $[F] = [A][X]$  с использованием предложенного способа хранения матрицы  $[A]$ . При этом полагается, что коэффициенты матрицы хранятся в массиве *Matr*, а номера соответствующих переменных в массиве *Index*:

```
for(i=0; i<N; i++) {
    F[i]=0;
    for(j=0; j<M; j++) {
        F[i]+=Matr[i*M+j]*X[Index[i*M+j]];
    }
}
```

Для моделирования процессов в электро-термическом рудовосстановительном реакторе используется двумерная математическая модель. При этом рассматриваются сечения, проходящие через центры двух электродов. Расчет тепломассопереноса сводится к решению системы дифференциальных уравнений. Использование метода конечных элементов позволяет аппроксимировать эти уравнения системой алгебраических уравнений.

Увеличение количества узлов сетки в силу необходимости повысить точность расчетов и наличие большого количества уравнений, описывающих математическую модель, делает размерность общей системы алгебраических уравнений очень большой, что уменьшает вероятность ее сходимости. Кроме того, система получается нелинейной вследствие вхождения в уравнения величин, определяемых из других уравнений, и плохо обусловленной из-за наличия в исходной системе уравнений конвективного теплообмена. Декомпозиция общей системы на подсистемы, аппроксимирующие отдельные дифференциальные уравнения, и последовательное решение этих систем итерационным методом практически устраняют трудности в реализации задачи. При этом значения величин, входящие в уравнения, берутся из предыдущей итерации. Для получения распределения какой-либо величины сначала формируется система алгебраических уравнений, аппроксимирующее соответствующее дифференциальное уравнение, которая затем решается с использованием метода, описанного выше. Критерием окончания расчета является равенство единице количества итераций метода сопряженных градиентов при решении каждой системы уравнений.

Уравнение теплообмена в реакционной зоне печи, определяющее распределение температуры, имеет ярко выраженную нелинейность, обусловленную экспоненциальной зависимо-

стью скорости реакции восстановления от температуры. Сходимость системы, включающей такое уравнение, ограничена правильным выбором начального приближения. В настоящей работе в качестве начального приближения для определения распределения температуры в реакционной зоне печи используется постоянная температура, равная температуре плавления. Для повышения сходимости системы уравнений используется метод нижней релаксации. При этом распределение температуры на каждой итерации ищется по схеме

$$[T]_n = [T]_{n-1}(1 - \alpha_p) + [T]_n^p \alpha_p,$$

где  $[T]_{n-1}$  – распределение температуры, полученное на предыдущей итерации;  $[T]_n^p$  – расчетное распределение температуры на текущей итерации, полученное решением соответствующей системы уравнений;  $0 < \alpha_p < 1$  – коэффициент релаксации. На первых итерациях значение коэффициента релаксации выбирается малым  $\alpha_p \approx 0,01$  и в дальнейшем по мере сходимости системы увеличивается либо вручную, либо автоматически с использованием зависимости

$$\alpha_p = \frac{k}{n_i},$$

где  $n_i$  – количество итераций метода сопряженных градиентов при решении системы уравнений, аппроксимирующей уравнение теплообмена в реакционной зоне печи;  $k$  – коэффициент пропорциональности. Использование в качестве начального приближения результатов предыдущих расчетов позволяет значительно сократить время решения совместной системы уравнений.

Для определения интегральных характеристик режима работы объема реактора, таких как полная активная мощность, выход газа и расход шихты, используется положение о том, что основное энерговыделение происходит в приэлектродных областях реакционной зоны. Там же происходит основная часть физико-химических процессов (образование газа и плавление шихты). При этом осуществляется интегрирование для каждого электрода по секторам величиной  $\pi$  с предположением, что тепловыделение на равноудаленной от электрода цилиндрической поверхности равно расчетному значению на том же расстоянии от электрода. Значение полной активной мощности в расчетном сечении ванны печи определяется выражением

$$P = \iint q \pi r_c dx dy,$$

где  $q$  – удельное объемное энерговыделение;  $r_e$  – расстояние до ближайшего электрода. Интегрирование для трех различных сечений печи позволяет получить полную мощность с учетом взаимодействия электродов между собой при их различных заглублениях и, следовательно, их положений относительно подины. Ее значение выводится из результатов, полученных для каждого сечения, и делением на два, так как влияние каждого электрода учитывается два раза. Аналогично определяются и другие интегральные характеристики режима работы.

Перед началом вычисления расчетная область печи разбивается на суперэлементы, а затем генерируется сетка конечных элементов на каждом из них. Путем объединения этих сеток получается сетка конечных элементов для всей расчетной области. После этого формируется массив для хранения систем уравнений.

Таким образом, разработанный алгоритм можно представить в следующем виде.

1. Создание системы суперэлементов для расчетной области.
2. Формирование сетки конечных элементов.
3. Формирование массивов для хранения систем уравнений.
4. Формирование системы уравнений для расчета распределения электрического потенциала.
5. Расчет распределения электрического потенциала.
6. Расчет распределения удельного энерговыделения.
7. Формирование системы уравнений для расчета распределения температуры в зоне расплава.
8. Расчет распределения температуры в зоне расплава.
9. Релаксация температуры в зоне расплава.
10. Формирование системы уравнений для расчета распределения концентрации реагента в зоне расплава.
11. Расчет распределения концентрации реагента в зоне расплава.
12. Формирование системы уравнений для расчета распределения потенциала газа.
13. Расчет распределения потенциала газа.
14. Расчет распределения скорости газа.
15. Формирование системы уравнений для расчета распределения потенциала шихты.
16. Расчет распределения потенциала шихты.
17. Расчет распределения скорости схода шихты.

18. Формирование системы уравнений для расчета распределения температуры газа в шихтовой зоне печи.

19. Расчет распределения температуры газа в шихтовой зоне печи.

20. Формирование системы уравнений для расчета распределения температуры шихты.

21. Расчет распределения температуры шихты.

Если при выполнении всех расчетов сходимость достигалась более чем за одну итерацию, переход к п. 3.

22. Расчет интегральных характеристик режима работы печи.

### **Информационная система расчета электротермических реакторов**

Представленная математическая модель и соответствующий ей алгоритм позволили создать на языке программирования C++ информационную систему моделирования и технологического расчета электротермических установок, применяемых в различных отраслях металлургии. Программный комплекс может функционировать под различными семействами операционных систем, кроме Mac OS, так как создавался с применением кроссплатформенного комплекса Visual Code, позволяющего компилировать программу под большой спектр операционных систем.

Информационная система обладает многозадачным интерфейсом, позволяющим моделировать различные компоненты исследуемых реакторов. Пользователь может либо выбрать расчет только одного конкретного параметра, либо осуществить комплексный анализ электротермической установки. Для ввода исходных данных используется специальная форма (<http://www.swsys.ru/uploaded/image/2024-4/3.jpg>).

В меню начальные данные включают основные характеристики рудновосстановительных реакторов. При этом от пользователя не требуется введения всех необходимых параметров. Можно ввести только те, которые знает исследователь, или только те, которые ему нужны. В результате система выполнит расчеты на основе введенных данных и выдаст сообщение о невозможности проведения всех запрограммированных вычислений. При выборе расчета отдельного параметра можно отыскать необходимый параметр, в случае соответствия введенных данных система произведет расчет, в противном случае она укажет на недостаточность введенных данных. Система содержит



## Список литературы

1. Tesser R., Santacesaria E. Revisiting the role of mass and heat transfer in gas–solid catalytic reactions. *Processes*, 2020, vol. 8, no. 12, art. 1599. doi: 10.3390/pr8121599.
2. Hu H.-P. Theoretical study of convection heat transfer and fluid dynamics in microchannels with arrayed micro-grooves. *Math. Problems in Engineering*, 2021, vol. 2021, art. 601509. doi: 10.1155/2021/3601509.
3. Антипина Е.В., Мустафина С.А., Антипин А.Ф. Программная реализация алгоритма поиска оптимального температурного режима каталитического процесса // Программные продукты и системы. 2022. Т. 35. № 1. С. 106–112. doi: 10.15827/0236-235X.137.106-112.
4. Пучков А.Ю., Лобанева Е.И., Кулыгин О.П. Алгоритм прогнозирования параметров системы переработки отходов апатит-нефелиновых руд // Прикладная информатика. 2022. Т. 17. № 1. С. 55–68. doi: 10.37791/2687-0649-2022-17-1-55-68.
5. Пучков А.Ю., Панченко С.В., Черновалова М.В. Программа моделирования теплофизических процессов в руднотермических печах при изменении управляющих воздействий // Программные продукты и системы. 2020. Т. 33. № 3. С. 523–529. doi: 10.15827/0236-235X.131.523-529.
6. Пучков А.Ю., Соколов А.М., Федотов В.В. Нейросетевой метод анализа процессов термической обработки окомкованного фосфатного рудного сырья // Прикладная информатика. 2022. Т. 17. № 5. С. 62–76. doi: 10.37791/2687-0649-2022-17-5-62-76.
7. Pedro A.A. Thermal monitoring of the electrode and bath in an ore-roasting furnace during downtime. *Steel in Translation*, 2013, vol. 43, pp. 110–112. doi: 10.3103/S0967091213020149.
8. Tian H., Pan J., Zhu D., Wang D., Xue Y. Utilization of ground sinter feed for oxidized pellet production and its effect on pellet consolidation and metallurgical properties. In: *Minerals, Metals and Materials Series. Proc. Int. Symposium on High-Temperature Metallurgical Processing*, 2020, pp. 857–866. doi: 10.1007/978-3-030-36540-0\_76.
9. Орехов В.А., Бобков В.И., Панченко С.В. Разработка информационной системы расчета, накопления информации и паспортизации теплофизических свойств фосфоритов // Программные продукты и системы. 2022. Т. 35. № 3. С. 502–508. doi: 10.15827/0236-235X.139.502-508.
10. Орехов В.А. Математическое моделирование процессов образования шлама в рудно-термических печах при переработке фосфатного рудного сырья // Современные наукоемкие технологии. 2023. № 7. С. 78–86. doi: 10.17513/snt.39698.
11. Борисов В.В., Курилин С.П., Луферов В.С. Нечёткие реляционные когнитивные темпоральные модели для анализа и прогнозирования состояния сложных технических систем // Прикладная информатика. 2022. Т. 17. № 1. С. 27–38. doi: 10.37791/2687-0649-2022-17-1-27-38.
12. Курилин С.П., Соколов А.М., Прокимов Н.Н. Компьютерная программа для эксплуатационной диагностики электромеханических систем на основе топологического подхода // Прикладная информатика. 2021. Т. 16. № 4. С. 62–73. doi: 10.37791/2687-0649-2021-16-4-62-73.
13. Кувшинова А.Н., Цыганов А.В. Программный комплекс для компьютерного моделирования процессов параметрической идентификации математических моделей конвективно-диффузионного переноса // Программные продукты и системы. 2021. Т. 34. № 4. С. 639–648. doi: 10.15827/0236-235X.136.639-648.
14. Свириденко А.Б. Прямые мультипликативные методы для разреженных матриц. Несимметричные линейные системы // Компьютерные исследования и моделирование. 2016. Т. 8. № 6. С. 833–860. doi: 10.20537/2076-7633-2016-8-6-833-860.
15. Пирова А.Ю., Кудрявцев Н.Ю., Мееров И.Б. Экспериментальное сравнение алгоритмов в параллельном многоуровневом методе вложенных сечений // Вестн. ЮУрГУ. Сер.: Вычислительная математика и информатика. 2017. Т. 6. № 1. С. 38–55. doi: 10.14529/cmse170103.
16. Лебедева А.В., Рябов В.М. О численном решении систем линейных алгебраических уравнений с плохо обусловленными матрицами // Вестн. СПбГУ. Математика. Механика. Астрономия. 2019. Т. 6. № 4. С. 619–626. doi: 10.21638/11701/spbu01.2019.407.

**Mathematical model for calculating electrothermal reactor technological parameters:  
Software implementation****Alexander A. Bykov <sup>1</sup>✉, Vladimir I. Bobkov <sup>1</sup>, Vladimir A. Orekhov <sup>1</sup>, Sergey V. Nezamaev <sup>1</sup>**<sup>1</sup> National Research University “MEI” in Smolensk, Smolensk, 214013, Russian Federation<sup>2</sup> First Mining Company, Moscow, 109004, Russian Federation**For citation**Bykov, A.A., Bobkov, V.I., Orekhov, V.A., Nezamaev, S.V. (2024) ‘Mathematical model for calculating electrothermal reactor technological parameters: Software implementation’, *Software & Systems*, 37(4), pp. 611–621 (in Russ.). doi: 10.15827/0236-235X.148.611-621

**Article info**

Received: 23.04.2024

After revision: 20.05.2024

Accepted: 27.05.2024

**Abstract.** The paper presents a mathematical model describing a possible method for calculating thermal mass processes. The model has become a basis for developing an iterative algorithm and creating an information system. The system simulates calculating technological parameters of electrothermal reactors used for ore phosphate raw material processing. The peculiarity of the developed model is that it takes into account the interrelation of zonal representation of the processes occurring in ore thermal furnaces. On the basis of thermal-hydraulic mode parameters it is possible to solve the problem of identifying raw materials entering the reaction zone. In the future, it will be possible to develop a system of process control automation in this zone in order to remove impurities. The created mathematical model allows taking into account the configuration of the internal structure of the investigated installations to understand the processes and phenomena occurring in them. When searching for a solution, the authors also used the model algorithm for the automatic construction of the finite element grid. They also proposed an algorithm for the finite element equation implementation with the matrix width optimization and the use of symmetry properties. The created software package allows automated calculations of operation modes of electrothermal ore-reduction reactors of any design, including those used in smelting phosphate raw materials. The information system database allows selecting an ore-reduction reactor model. It also allows calculating heat and mass exchange processes, as well as setting the system of effective coefficients for calculating the developed experimental reactor model. In this case, the program complex can calculate not only a complex of the ore thermal reactor, but also its individual characteristics. The multi-document interface visualizes calculation results in the field of isolines and velocity vectors of thermal mass processes.

**Keywords:** mathematical model, information system, software package, differential equations, iterative algorithm, computer modeling, numerical methods

**Acknowledgements.** The paper was carried out under the government assignment, project no. FSWF-2023-0012

**References**

1. Tesser, R., Santacesaria, E. (2020) 'Revisiting the role of mass and heat transfer in gas–solid catalytic reactions', *Processes*, 8(12), art. 1599. doi: 10.3390/pr8121599.
2. Hu, H.-P. (2021) 'Theoretical study of convection heat transfer and fluid dynamics in microchannels with arrayed microgrooves', *Math. Problems in Engineering*, 2021, art. 601509. doi: 10.1155/2021/3601509.
3. Antipina, E.V., Antipin, A.F., Mustafina, S.A. (2022) 'Software implementation of the algorithm for finding the optimal temperature condition of the catalytic process', *Software & Systems*, 35(1), pp. 106–112 (in Russ.). doi: 10.15827/0236-235X.137.106-112.
4. Puchkov, A.Yu., Lobaneva, E.I., Kultygin, O.P. (2022) 'Algorithm for predicting the parameters of a system for processing waste apatite-nepheline ores', *J. of Applied Inform.*, 17(1), pp. 55–68 (in Russ.). doi: 10.37791/2687-0649-2022-17-1-55-68.
5. Puchkov, A.Yu., Panchenko, S.V., Chernovalova, M.V. (2020) 'The simulation program for thermophysical processes in ore-thermal furnaces when changing the control actions', *Software & Systems*, 33(3), pp. 523–529 (in Russ.). doi: 10.15827/0236-235X.131.523-529.
6. Puchkov, A.Yu., Sokolov, A.M., Fedotov, V.V. (2022) 'Neural network analysis method of heat treatment processes of pelletized phosphate ore raw materials', *J. of Applied Inform.*, 17(5), pp. 62–76 (in Russ.). doi: 10.37791/2687-0649-2022-17-5-62-76.
7. Pedro, A.A. (2013) 'Thermal monitoring of the electrode and bath in an ore-roasting furnace during downtime', *Steel in Translation*, 43, pp. 110–112. doi: 10.3103/S0967091213020149.
8. Tian, H., Pan, J., Zhu, D., Wang, D., Xue, Y. (2020) 'Utilization of ground sinter feed for oxidized pellet production and its effect on pellet consolidation and metallurgical properties', in *Minerals, Metals and Materials Series. Proc. Int. Symposium on High-Temperature Metallurgical Processing*, pp. 857–866. doi: 10.1007/978-3-030-36540-0\_76.
9. Orekhov, V.A., Bobkov, V.I., Panchenko, S.V. (2022) 'Information system for calculation, information accumulation and certification of phosphorite thermophysical properties', *Software & Systems*, 35(3), pp. 502–508 (in Russ.). doi: 10.15827/0236-235X.139.502-508.
10. Orekhov, V.A. (2023) 'Mathematical modeling of the processes of sludge formation in ore thermal furnaces during processing of phosphate ore raw', *Modern High Technolog.*, (7), pp. 78–86 (in Russ.). doi: 10.17513/snt.39698.
11. Borisov, V.V., Kurilin, S.P., Lufarov, V.S. (2022) 'Fuzzy relational cognitive temporal models for analyzing and state prediction of complex technical systems', *J. of Applied Informatics*, 17(1), pp. 27–38 (in Russ.). doi: 10.37791/2687-0649-2022-17-1-27-38.
12. Kurilin, S.P., Sokolov, A.M., Prokimnov, N.N. (2021) 'A computer program for electromechanical system operational diagnostics based on the topological approach', *J. of Applied Inform.*, 16(4), pp. 62–73 (in Russ.). doi: 10.37791/2687-0649-2021-16-4-62-73.
13. Kuvshinova, A.N., Tsyganov, A.V. (2021) 'A software package for computer modeling of parametric identification processes for mathematical models of convection-diffusion transfer', *Software & Systems*, 34(4), pp. 639–648 (in Russ.). doi: 10.15827/0236-235X.136.639-648.
14. Sviridenko, A.B. (2016) 'Direct multiplicative methods for sparse matrices. Unbalanced linear systems', *Computer Research and Modeling*, 8(6), pp. 833–860 (in Russ.).

15. Pirova, A.Yu., Kudryavstev, N.Yu., Meerov, I.B. (2017) 'Experimental evaluation of algorithms in the parallel multilevel nested dissection method', *Bull. of YuUrGU. Ser. Computational Math. and Software Eng.*, 6(1), pp. 38–55 (in Russ.). doi: 10.14529/cmse170103.

16. Lebedeva, A.V., Ryabov, V.M. (2019) 'On the numerical solution of system of linear algebraic equations with ill-conditioned matrices', *Bull. of SPbU. Math. Mechanics. Astronomy*, 6(4), pp. 619–626 (in Russ.). doi: 10.21638/11701/spbu01.2019.407.

#### Авторы

**Быков Александр Александрович**<sup>1</sup>,

к.п.н, доцент, alex1by@mail.ru

**Бобков Владимир Иванович**<sup>1</sup>, д.т.н.,

доцент, профессор, vovabobkoff@mail.ru

**Орехов Владимир Александрович**<sup>1</sup>,

старший преподаватель, fundukoff@mail.ru

**Неземаев Сергей Владимирович**<sup>2</sup>,

генеральный директор, pgrk@armz.ru

#### Authors

**Alexander A. Bykov**<sup>1</sup>, Cand. of Sci. (Education),

Associate Professor, alex1by@mail.ru

**Vladimir I. Bobkov**<sup>1</sup>, Dr.Sci. (Engineering), Associate

Professor, Professor, vovabobkoff@mail.ru

**Vladimir A. Orekhov**<sup>1</sup>, Senior Lecturer,

fundukoff@mail.ru

**Sergey V. Nezamaev**<sup>2</sup>, Director General,

pgrk@armz.ru

<sup>1</sup> Национальный исследовательский университет «МЭИ» в г. Смоленске, г. Смоленск, 214013, Россия

<sup>2</sup> Первая горнорудная компания, г. Москва, 109004, Россия

<sup>1</sup> National Research University "MEI" in Smolensk, Smolensk, 214013, Russian Federation

<sup>2</sup> First Mining Company, Moscow, 109004, Russian Federation

## Программная реализация диагностики непрерывных многостадийных производств с использованием динамической экспертной системы

С.А. Меркурьев <sup>1</sup>✉, Б.В. Палюх <sup>1</sup>, А.Н. Ветров <sup>1</sup>

<sup>1</sup> Тверской государственный технический университет,  
г. Тверь, 170026, Россия

### Ссылка для цитирования

Меркурьев С.А., Палюх Б.В., Ветров А.Н. Программная реализация диагностики непрерывных многостадийных производств с использованием динамической экспертной системы // Программные продукты и системы. 2024. Т. 37. № 4. С. 622–628. doi: 10.15827/0236-235X.148.622-628

### Информация о статье

Группа специальностей ВАК: 2.3.3

Поступила в редакцию: 15.04.2024

После доработки: 08.05.2024

Принята к публикации: 14.05.2024

**Аннотация.** В данной статье описана архитектура разрабатываемого программного комплекса для диагностики многостадийных непрерывных производств. Актуальность работы обусловлена тем, что дефект или отказ оборудования на таких производствах зачастую представляет большую опасность. Целью является проектирование архитектуры и разработка программного комплекса динамической экспертной системы для анализа текущего состояния технологического процесса и поддержки принятия решения для выявления и устранения дефектов в технологических цепях. В работе кратко рассмотрена проблема прикладной области. Описаны проблемы существующих программных реализаций и их основные недостатки. Приведены структурная и интеграционная схемы разрабатываемого программного комплекса. Представлена блок-схема и описаны разрабатываемые модули. Первый модуль реализует задачи обнаружения аномалий и определения момента бифуркаций. Эти две задачи решаются на основе анализа многомерного потока диагностических переменных с помощью оригинального метода двойного скользящего окна. Второй модуль предназначен для формирования гипотез о причинах кризисных состояний с помощью функции доверия. Третий модуль реализует логический вывод по обнаружению дефектов в технологических цепях и по способам их устранения. В заключении описаны используемые инструментальные средства для реализации диагностической динамической экспертной системы. Практическая значимость работы заключается в разработке информационного, алгоритмического и программного обеспечения на языке C# для повышения эксплуатационной надежности непрерывных производств.

**Ключевые слова:** прикладные системы искусственного интеллекта, динамические экспертные системы, диагностика, точка бифуркации

**Введение.** В данной статье поднимается проблема диагностики непрерывных многостадийных производств. К ним можно отнести процессы химической промышленности, атомной энергетики, металлургии и т.п. Многие из этих процессов потенциально опасны. Одной из важных задач управления такими процессами является обеспечение их стабильности. Для этого производится непрерывный мониторинг состояния технологического процесса путем периодического измерения диагностических переменных и сравнения их с нормативными значениями. При штатной работе оборудования технологический процесс функционирует в стационарном режиме, но со временем оборудование физически и морально изнашивается. В ходе этого процесса технологический процесс постепенно переходит в нестационарный режим и становится неуправляемым. Время обнаружения аномального поведения технологического процесса существенно влияет на способность системы к восстановлению.

Проблема обнаружения аномалий в сложных динамических системах является развивающейся областью научных исследований и активно обсуждается в литературе (например, [1, 2]). В проведенных исследованиях в области обнаружения и диагностики аномальных событий подчеркивается, что вопросы раннего предупреждения об опасной ситуации являются очень актуальными [3, 4]. Традиционно обнаружение аномалий относится к проблеме выявления паттернов в данных, полученных на основе мониторинга деятельности системы, которые не соответствуют ожидаемому поведению. Эти несоответствующие паттерны часто называют аномалиями. Системы обнаружения аномалий во временных рядах используются в самых разных приложениях [5, 6]. Проблема поиска аномалий интерпретируется как проблема классификации. Для ее решения используются методы классификации, такие как метод  $k$ -ближайшего соседа, кластерный анализ, классификаторы, построенные с использовани-

ем искусственных нейронных сетей, метод случайного леса и т.п. [7, 8]. Хотя существующие системы поиска аномалий во временных рядах доказали свою эффективность, существует ряд проблем, связанных с их практическим применением: высокий уровень ложных срабатываний, ограниченные типы аномалий, которые система может обнаружить. Кроме того, такие системы не могут выполнять обнаружение аномалий в реальном времени. Наиболее существенным недостатком с точки зрения определения бифуркаций в непрерывных многостадийных технологических процессах является необходимость предварительного обучения рассмотренных систем поиска аномалий. Поскольку аварийные ситуации относятся к классу редких событий, формирование БД для обучения проблематично.

В настоящее время на непрерывных многостадийных производствах зачастую, используются только базовые системы мониторинга технологического процесса путем периодического измерения диагностических переменных и сравнения их с нормативными значениями.

Решению проблемы диагностики многостадийных производств и прогнозирования отказов и аварийных ситуаций посвящены работы ряда российских исследователей. Так, в [9] предложена композиционная модель прогнозирования аварийных ситуаций. Она осуществляет прогнозирование аварийных ситуаций, включая долгосрочный и текущий прогнозы. Долгосрочный прогноз основывается на термодинамической модели, а текущий – на лингвистической и нейросетевой моделях, при этом в ситуациях, не предусмотренных обучающими выборками, предпочтение отдается нейросетевой модели. Предлагаемая экспертная система позволяет совместить долгосрочный и текущий прогнозы, обучение системы осуществляется автоматически, что повышает эффективность прогнозирования диагностических процедур.

В статье [10] поднимается вопрос совершенствования процесса управления региональной безопасностью за счет развития методов анализа данных о чрезвычайных ситуациях. Авторами предложена многомерная классификация данных в области чрезвычайных ситуаций на основе нечеткой логики, которую нельзя применить для отдельного технологического процесса.

В данной статье рассмотрен разрабатываемый программный комплекс экспертной системы ExpSystem, которая совмещает в себе методы экспертного и количественного ана-

лиза состояния технологического процесса в условиях неопределенности. Экспертная система ExpSystem позволяет не только обнаруживать и диагностировать переход технологического процесса в нестационарное состояние, но и определять причину возникновения аномалии, находить дефектное оборудование, являющееся источником критической ситуации. Отличительной особенностью ExpSystem является использование метода двойного скользящего окна для определения момента бифуркаций и формирование гипотез о причинах кризисных состояний технологического процесса с помощью функции доверия.

### Назначение программного комплекса

Объектом диагностики программного комплекса ExpSystem является непрерывный многостадийный процесс. Данный комплекс состоит из трех программных модулей (рис. 1).

В качестве входных параметров рассматривается матрица  $D(d, k)$  потоковых данных на момент времени  $T$ . Наблюдения осуществляются через равные промежутки времени  $\Delta t$  (или с частотой  $f = 1 / \Delta t$ ):

$$D = \begin{pmatrix} y_{11} & \dots & y_{1k} \\ \vdots & \ddots & \vdots \\ y_{d1} & \dots & y_{dk} \end{pmatrix},$$

где  $d$  – количество датчиков в системе;  $k$  – количество временных срезов до момента времени  $T$ .

Далее эта матрица с данными поступает на вход первого модуля динамической экспертной системы ExpSystem. Задачей этого модуля является обнаружение аномалий и определение момента бифуркации на основе метода двойного скользящего окна по каждому дат-

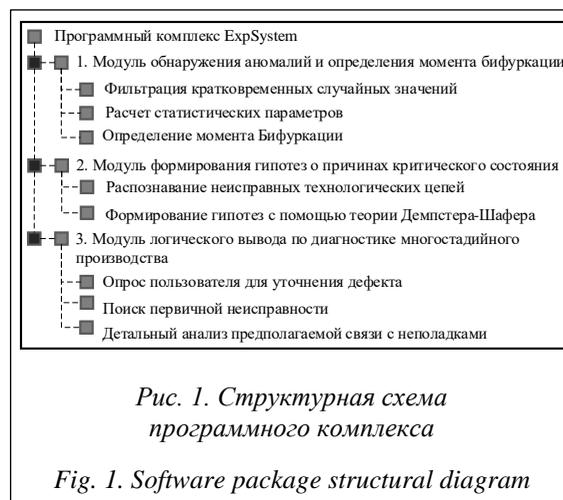


Рис. 1. Структурная схема программного комплекса

Fig. 1. Software package structural diagram

чику. На выходе получаем массив  $R(d)$  с логическими переменными, которые обозначают факт наличия точки бифуркации в потоке данных с датчика  $d$ .

Задачей второго модуля является формирование гипотез о причинах кризисных состояний с помощью функций доверия на основе массива  $R(d)$ . Далее эта информация поступает в третий модуль, где агент-координатор технологического процесса с помощью полученной информации определяет неработоспособную цепь производства, и с помощью наводящих вопросов экспертная система определяет место дефекта и возможные действия для его устранения. Результат этого анализа передается агентам (операторам), целью которых будет принятие решений функционирования приведения оборудования технологической цепи в стационарный режим.

Интеграционная схема программного комплекса представлена на рисунке 2.

### Модуль обнаружения аномалий и определения момента бифуркации

Перед первым модулем ставится задача обнаружения аномалий и определения момента бифуркации на основе анализа многомерного

потока данных. На вход первого модуля подается матрица  $D(d, k)$ . Для определения точки бифуркации предлагается метод двойного скользящего окна, суть которого заключается в проверке условий стационарности на основе выборочных данных для коротких временных рядов [11].

Блок-схема алгоритма программного комплекса представлена на рисунке 3.

Работа алгоритма заключается в определении наличия точки бифуркации для каждого потока данных с датчика  $D(d, k)$ , где  $d$  – количество датчиков в системе;  $k$  – количество временных срезов. С этой целью для каждого потока данных каждого датчика необходимо выполнить определенные этапы.

Этап 1. Входной массив данных делится на  $N$  равных кортежей длины  $h$ .

Этап 2. Для каждого получившегося кортежа выполняются следующие действия:

- применяется алгоритм первого скользящего окна, которое используется как низкочастотный фильтр, подавляющий кратковременные случайные выбросы значений диагностических переменных;
- вычисляются средние  $\bar{y}_n$  и среднеквадратичные  $\overline{y_n^2}$  значения по анализируемому кортежу;

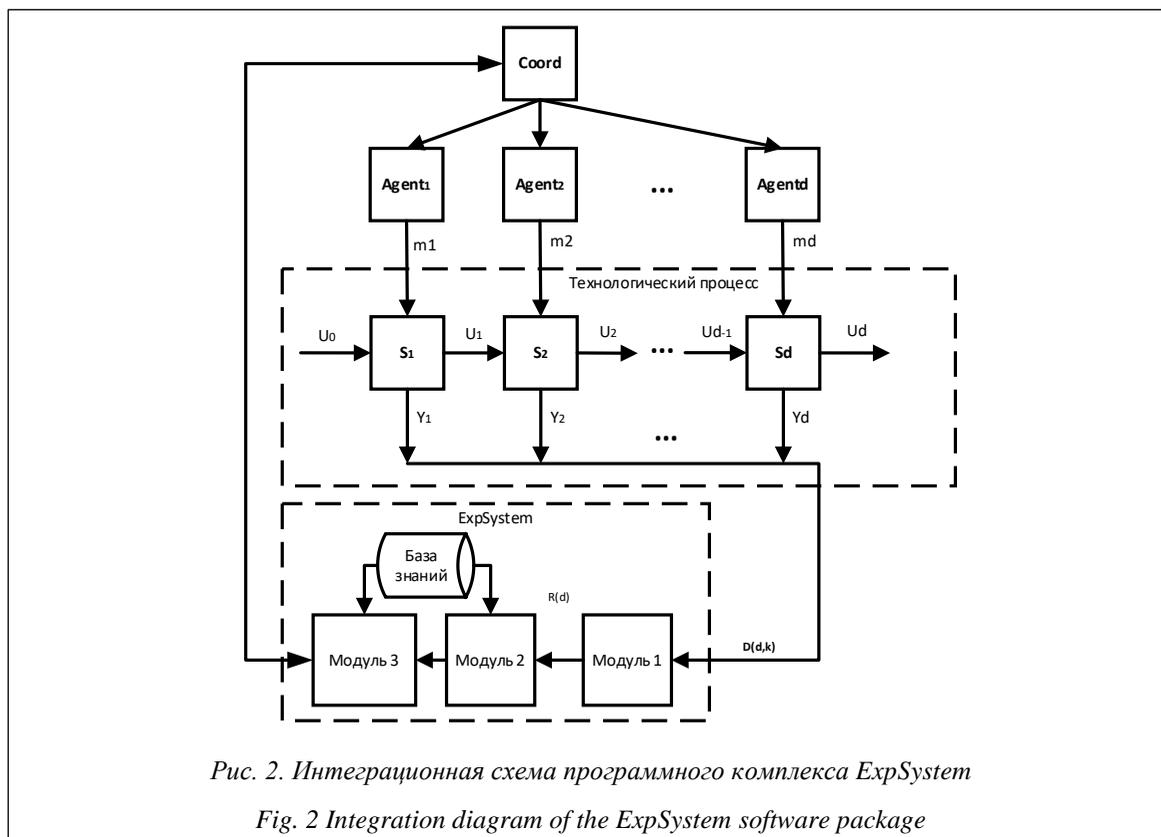


Рис. 2. Интеграционная схема программного комплекса ExpSystem

Fig. 2 Integration diagram of the ExpSystem software package

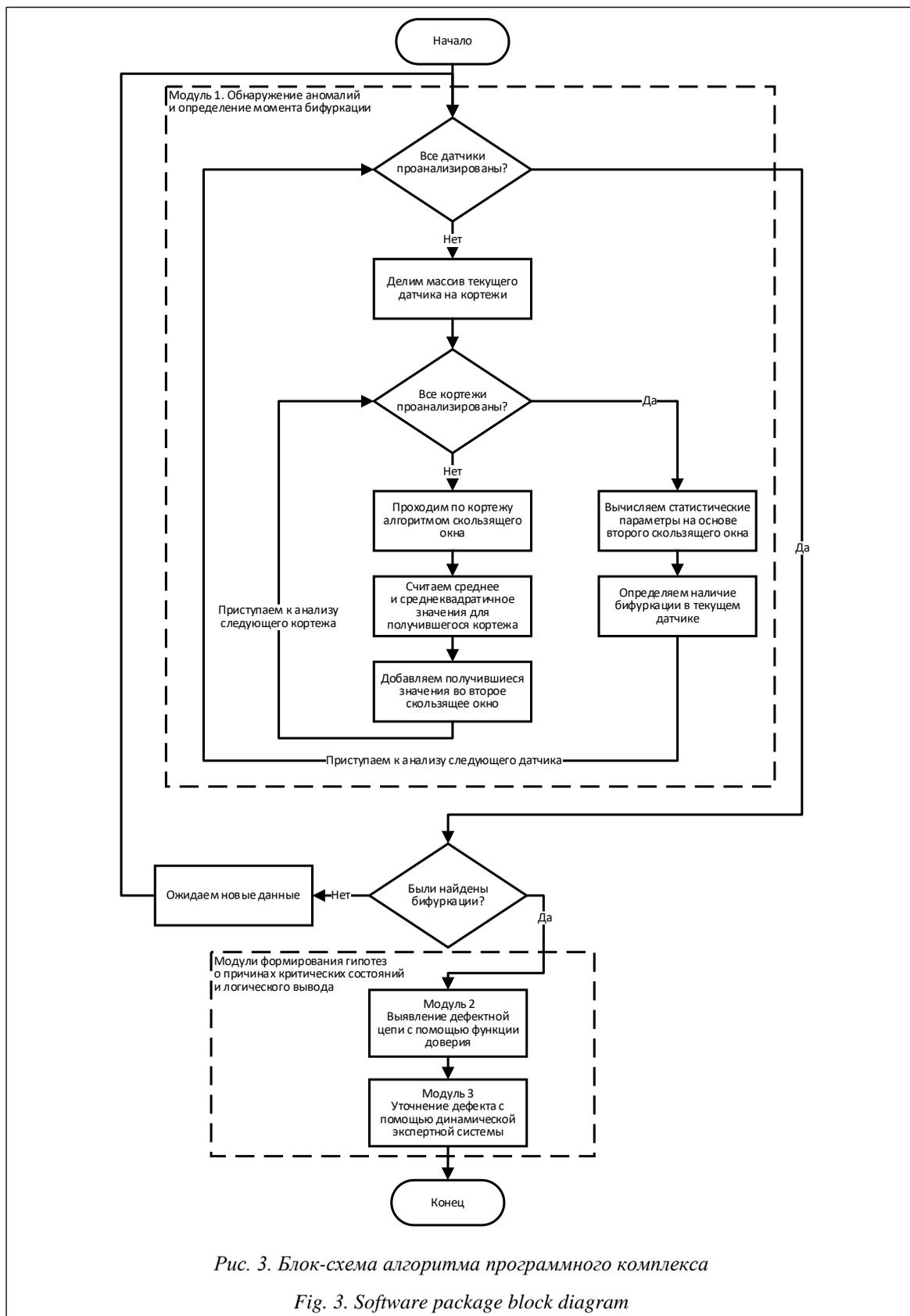


Рис. 3. Блок-схема алгоритма программного комплекса

Fig. 3. Software package block diagram

- получившиеся значения записываются во второе скользящее окно.

Этап 3. После анализа всех кортежей считаются статистические параметры. Для этого за-

писанные во второе скользящее окно значения проверяются на наличие тренда с помощью модификации критерия Фостера–Стюарда и рассчитываются множества  $u_n$  и  $v_n$  по формулам

$$u_n = \begin{cases} 1 \leftarrow \text{if } \overline{y_n} > \overline{y_{n-1}}, \overline{y_{n-2}}, \dots, \overline{y_1}, \\ 0 \leftarrow \text{else} \end{cases}$$

$$v_n = \begin{cases} 1 \leftarrow \text{if } \overline{y_n} \leq \overline{y_{n-1}}, \overline{y_{n-2}}, \dots, \overline{y_1}. \\ 0 \leftarrow \text{else} \end{cases}$$

Затем вычисляются две статистики  $W$  и  $F$ :

$$W = \sum_{n=2}^N (u_n - v_n); \quad F = \sum_{n=2}^N (u_n + v_n)$$

и значения  $t_W$  и  $t_F$  по формулам

$$t_W = \frac{W}{\sigma_W}; \quad t_F = \frac{F - \mu}{\sigma_F},$$

где

$$\sigma_W = \left(2 \sum_{n=2}^N \frac{1}{n}\right)^{0.5}; \quad \sigma_F = \left(\mu - 4 \sum_{n=2}^N \frac{1}{n^2}\right)^{0.5};$$

$$\mu = 2 \sum_{n=2}^N \frac{1}{n}.$$

После проверяется гипотеза об отсутствии стационарности. При отсутствии тренда нормализованные значения статистик приближенно описываются распределением Стьюдента с числом степеней свободы  $df = N$ .

После анализа временных рядов в диагностических переменных от каждого датчика проверяем, были ли обнаружены бифуркации. Если они не были найдены, то система ожидает поступления новых данных. При нахождении бифуркаций система приступает к последующим этапам анализа по второму и третьему модулям.

### Модули формирования гипотез о причинах критических состояний и логического вывода

Второй модуль основан на применении теории Демпстера–Шафера с целью формирования гипотез о причинах кризисных состояний технологического процесса. Теория свидетельств, или теория Демпстера–Шафера, является общей основой для рассуждений с неопределенностью и позволяет объединить свидетельства из разных источников и прийти к некоторой степени уверенности в наличии того или иного события [12–14]. Логика данного модуля подробно рассмотрена в [15]. Для его работы используются данные, полученные из предыдущего модуля, и данные из базы знаний.

В третьем модуле логического вывода по диагностике многостадийного производства реализована экспертная система [16]. С помощью системы и сформированных гипотез агент-

координатор выявляет неисправную технологическую цепь и уточняет дефект. После этого он передает полученную информацию агентам-технологам для корректирования состояния или остановки технологического процесса.

### Используемые инструментальные средства

Данный программный комплекс разработан с помощью объектно-ориентированного языка программирования C#. Известно, что язык C# является мощным и гибким объектно-ориентированным языком программирования с поддержкой современных технологий, таких как LINQ и асинхронное программирование. Он также предлагает высокую производительность, что важно для эффективного выполнения сложных операций. Это делает его идеальным выбором для разработки сложных программных комплексов.

Моделирование данных в первом модуле реализовано с помощью сторонней программной системы AnyLogic, в которой была построена соответствующая модель.

Формирование гипотез в третьем модуле реализовано с помощью программной среды для разработки экспертных систем CLIPS и набора продукционных правил, загруженных из базы знаний.

### Заключение

Фундаментальная проблема обеспечения устойчивости системы управления эволюцией непрерывных многостадийных процессов к возникающим критическим ситуациям актуальна. В ней недостаточно проработаны методы раннего обнаружения момента перехода технологического процесса из штатного режима в предаварийное состояние. В статье рассмотрен программный комплекс, который использует новый подход для выявления начала переходного процесса в системе управления, а также моделирует состояние технологического процесса в ближайшем будущем.

Это позволяет не только узнавать о критическом состоянии технологического процесса в текущий момент времени, но и предсказывать его состояние, благодаря чему появляется возможность заблаговременно проанализировать ситуацию и принять правильные решения. В конечном итоге это позволит избежать не только затрат на восстановление производства в случае аварийной ситуации, но и простоев в работе производства.

## Список литературы

1. Zhang Y., Chen Y., Wang J., Pan Z. Unsupervised deep anomaly detection for multi-sensor time-series signals. *IEEE Transactions on Knowledge and Data Engineering*, 2021, vol. 35, no. 3, pp. 2118–2132. doi: 10.1109/TKDE.2021.3102110.
2. He Y., Zhao J. Temporal convolutional networks for anomaly detection in time series. *JPCS*, 2019, vol. 1213, art. 042050. doi: 10.1088/1742-6596/1213/4/042050.
3. Kim D., Yang H., Chung M. et al. Squeezed Convolutional Variational AutoEncoder for unsupervised anomaly detection in edge device industrial Internet of Things. *Proc. ICICT*, 2018, pp. 67–71. doi: 10.1109/INFOCT.2018.8356842.
4. Zhao H., Wang Y., Duan J. et al. Multivariate time-series anomaly detection via graph attention network. *ICDM*, 2020, pp. 841–850. doi: 10.1109/ICDM50108.2020.00093.
5. Li D., Chen D., Jin B. et al. MAD-GAN: Multivariate anomaly detection for time series data with generative adversarial networks. In: *ICANN. LNTCS*, 2019, vol. 11730, pp. 703–716. doi: 10.1007/978-3-030-30490-4\_56.
6. Lin S., Clark R., Birke R., Schonborn S., Trigoni N., Roberts S. Anomaly detection for time series using VAE-LSTM hybrid model. *Proc. ICASSP*, 2020, pp. 4322–4326. doi: 10.1109/ICASSP40776.2020.9053558.
7. Su Y., Zhao Y., Niu C., Liu R., Sun W., Pei D. Robust anomaly detection for multivariate time series through stochastic recurrent neural network. *Proc. KDD*, 2019, pp. 2828–2837. doi: 10.1145/3292500.3330672.
8. Chen Z., Chen D., Yuan Z., Cheng X., Zhang X. Learning graph structures with transformer for multivariate time series anomaly detection in IoT. *ИТЖ*, 2022, vol. 9, no. 12, pp. 9179–9189. doi: 10.1109/ИИОТ.2021.3100509.
9. Ткалич С.А. Инструментальная методика прогнозирования аварийных ситуаций технологических процессов // *Вестн. РГРТУ*. 2022. № 79. С. 101–113.
10. Лабинский А.Ю., Бородушко И.В., Таранцев А.А. Поддержка принятия управленческих решений при рисках чрезвычайных ситуаций на основе применения методов анализа многомерных статистических данных // *Инженерный вестник Дона*. 2023. № 10. URL: [http://www.ivdon.ru/uploads/article/pdf/IVD\\_20\\_\\_10\\_Labinskiy\\_Borodushko\\_Tarantsev.pdf\\_a6205847c7.pdf](http://www.ivdon.ru/uploads/article/pdf/IVD_20__10_Labinskiy_Borodushko_Tarantsev.pdf_a6205847c7.pdf) (дата обращения 10.04.2024).
11. Ветров А.Н., Мальков А.А., Палюх Б.В. Детектирование бифуркаций многомерных случайных процессов в динамических экспертных системах // *КИИ: тр. конф.* 2023. С. 173–182.
12. Ullah I., Youn J., Han Y.-H. Multisensor data fusion based on modified belief entropy in Dempster–Shafer theory for smart environment. *IEEE Access*, 2021, vol. 9, pp. 37813–37822. doi: 10.1109/ACCESS.2021.3063242.
13. Jiroušek R., Kratochvíl V., Shenoy P.P. Entropy for evaluation of Dempster–Shafer belief function models. *IJAR*, 2022, vol. 151, pp. 164–181. doi: 10.1016/j.ijar.2022.09.009.
14. Zhang C., Chen K., Zeng X., Xue X. Misbehavior detection based on support vector machine and Dempster–Shafer theory of evidence in VANETs. *IEEE Access*, 2018, pp. 59860–59870. doi: 10.1109/ACCESS.2018.2875678.
15. Иванов В.К., Виноградова Н.В., Палюх Б.В., Сотников А.Н. Современные направления развития и области приложения теории Демпстера–Шафера // *ИИиПР*. 2018. № 4. С. 32–42. doi: 10.14357/20718594180403.
16. Shabanov B., Sotnikov A., Palyukh B., Vetrov A., Alexandrova D. Expert system for managing policy of technological security in uncertainty conditions: Architectural, algorithmic, and computing aspects. *Proc. ElConRus*, 2019, pp. 1716–1721. doi: 10.1109/ElConRus.2019.8657114.

Software &amp; Systems

doi: 10.15827/0236-235X.148.622-628

2024, 37(4), pp. 622–628

**Continuous multistage production diagnostics: Software implementation using a dynamic expert system**Sergey A. Merkurev <sup>1</sup>✉, Boris V. Palyukh <sup>1</sup>, Alexander N. Vetrov <sup>1</sup><sup>1</sup> Tver State Technical University, Tver, 170026, Russian Federation**For citation**

Merkurev, S.A., Palyukh, B.V., Vetrov, A.N. (2024) 'Continuous multistage production diagnostics: Software implementation using a dynamic expert system', *Software & Systems*, 37(4), pp. 622–628 (in Russ.). doi: 10.15827/0236-235X.148.622-628

**Article info**

Received: 15.04.2024

After revision: 08.05.2024

Accepted: 14.05.2024

**Abstract.** The paper describes the architecture of the developed software system for diagnostics of multistage continuous productions. The work is relevant due to a great danger of an equipment failure or defect in such productions. The authors aim to design an architecture and to develop a software package of a dynamic expert system in order to analyze the current state of a technological process and to support decision making for identifying and eliminating defects in technological circuits. The paper summarizes the application domain problem. It also describes the problems of existing software imple-

mentations and their main shortcomings. The authors give the structural and integration schemes of the developed software system. They present the block diagram and describe modules. The first module implements two tasks: anomaly detection and bifurcation moment detection. The authors solve these two tasks based on the analysis of multivariate flow of diagnostic variables using the original double sliding window method. The second module aims to form hypotheses about the causes of crisis states using a confidence function. The third module implements logical inference on detection of defects in technological circuits and methods of their elimination. The conclusion describes the used tools for implementing the diagnostic dynamic expert system. The practical significance of the work consists in the development of information, algorithmic and software tools in C# language to improve the operational reliability of continuous productions.

**Keywords:** applied artificial intelligence systems, dynamic expert systems, diagnostics, bifurcation point

### References

- Zhang, Y., Chen, Y., Wang, J., Pan, Z. (2021) 'Unsupervised deep anomaly detection for multi-sensor time-series signals', *IEEE Transactions on Knowledge and Data Engineering*, 35(3), pp. 2118–2132. doi: 10.1109/TKDE.2021.3102110.
- He, Y., Zhao, J. (2019) 'Temporal convolutional networks for anomaly detection in time series', *JPCS*, 1213, art. 042050. doi: 10.1088/1742-6596/1213/4/042050.
- Kim, D., Yang, H., Chung, M. et al. (2018) 'Squeezed Convolutional Variational AutoEncoder for unsupervised anomaly detection in edge device industrial Internet of Things', *Proc. ICICT*, pp. 67–71. doi: 10.1109/INFOCT.2018.8356842.
- Zhao, H., Wang, Y., Duan, J. et al. (2020) 'Multivariate time-series anomaly detection via graph attention network', *ICDM*, pp. 841–850. doi: 10.1109/ICDM50108.2020.00093.
- Li, D., Chen, D., Jin, B. et al. (2019) 'MAD-GAN: Multivariate anomaly detection for time series data with generative adversarial networks', in *ICANN. LNTCS*, 11730, pp. 703–716. doi: 10.1007/978-3-030-30490-4\_56.
- Lin, S., Clark, R., Birke, R., Schonborn, S., Trigoni, N., Roberts, S. (2020) 'Anomaly detection for time series using VAE-LSTM hybrid model', *Proc. ICASSP*, pp. 4322–4326. doi: 10.1109/ICASSP40776.2020.9053558.
- Su, Y., Zhao, Y., Niu, C., Liu, R., Sun, W., Pei, D. (2019) 'Robust anomaly detection for multivariate time series through stochastic recurrent neural network', *Proc. KDD*, pp. 2828–2837. doi: 10.1145/3292500.3330672.
- Chen, Z., Chen, D., Yuan, Z., Cheng, X., Zhang, X. (2022) 'Learning graph structures with transformer for multivariate time series anomaly detection in IoT', *IITJ*, 9(12), pp. 9179–9189. doi: 10.1109/IJOT.2021.3100509.
- Tkalich, S.A. (2022) 'Instrumental methodology for predicting emergency situations of technological processes', *Vestn. of RSREU*, (79), pp. 101–113 (in Russ.).
- Labinskiy, A.Yu., Borodushko, I.V., Tarantsev, A.A. (2023) 'Support for managerial decision-making in case of emergency risks based on the use of methods of analysis of multidimensional statistical data', *Engineering J. of Don*, (10), available at: [http://www.ivdon.ru/uploads/article/pdf/IVD\\_20\\_10\\_Labinskiy\\_Borodushko\\_Tarantsev.pdf\\_a6205847c7.pdf](http://www.ivdon.ru/uploads/article/pdf/IVD_20_10_Labinskiy_Borodushko_Tarantsev.pdf_a6205847c7.pdf) (accessed April 10, 2024) (in Russ.).
- Vetrov, A.N., Malkov, A.A., Palyukh, B.V. (2023) 'Detection of bifurcations of multidimensional random processes in dynamic expert systems', *Proc. CAI*, pp. 173–182 (in Russ.).
- Ullah, I., Youn, J., Han, Y.-H. (2021) 'Multisensor data fusion based on modified belief entropy in Dempster-Shafer theory for smart environment', *IEEE Access*, 9, pp. 37813–37822. doi: 10.1109/ACCESS.2021.3063242.
- Jiroušek, R., Kratochvíl, V., Shenoy, P.P. (2022) 'Entropy for evaluation of Dempster-Shafer belief function models', *IJAR*, 151, pp. 164–181. doi: 10.1016/j.ijar.2022.09.009.
- Zhang, C., Chen, K., Zeng, X., Xue, X. (2018) 'Misbehavior detection based on support vector machine and Dempster-Shafer theory of evidence in VANETs', *IEEE Access*, pp. 59860–59870. doi: 10.1109/ACCESS.2018.2875678.
- Ivanov, V.K., Vinogradova, N.V., Palyukh, B.V., Sotnikov, A.N. (2018) 'Current trends and applications of Dempster-Shafer theory (review)', *AIDT*, (4), pp. 32–42 (in Russ.). doi: 10.14357/20718594180403.
- Shabanov, B., Sotnikov, A., Palyukh, B., Vetrov, A., Alexandrova, D. (2019) 'Expert system for managing policy of technological security in uncertainty conditions: Architectural, algorithmic, and computing aspects', *Proc. ElConRus*, pp. 1716–1721. doi: 10.1109/ElConRus.2019.8657114.

### Авторы

**Меркурьев Сергей Алексеевич**<sup>1</sup>,  
аспирант, serj.merkur@yandex.ru  
**Палюх Борис Васильевич**<sup>1</sup>, д.т.н.,  
профессор, зав. кафедрой, pboris@tstu.tver.ru  
**Ветров Александр Николаевич**<sup>1</sup>, к.т.н.,  
доцент, профессор, vetrov\_48@mail.ru

### Authors

**Sergey A. Merkurev**<sup>1</sup>, Postgraduate Student,  
serj.merkur@yandex.ru  
**Boris V. Palyukh**<sup>1</sup>, Dr.Sci. (Engineering),  
Professor, Head of Chair, pboris@tstu.tver.ru  
**Alexander N. Vetrov**<sup>1</sup>, Cand. of Sci. (Engineering),  
Associate Professor, Professor, vetrov\_48@mail.ru

<sup>1</sup> Тверской государственный технический университет, г. Тверь, 170026, Россия

<sup>1</sup> Tver State Technical University, Tver, 170026, Russian Federation

## Применение концепции цифрового двойника на этапах проектирования, моделирования и управления химическим процессом

М.И. Кузьмин<sup>1,2</sup>✉, Д.И. Кушнирук<sup>1</sup>, А.В. Аникин<sup>1</sup>,  
С.В. Верба<sup>1</sup>, Д.В. Зубов<sup>2</sup>

<sup>1</sup> Государственный научно-исследовательский и проектный институт редкометаллической промышленности «Гиредмет» имени Н.П. Сагина, г. Москва, 111524, Россия

<sup>2</sup> Российский химико-технологический университет имени Д.И. Менделеева, г. Москва, 125047, Россия

### Ссылка для цитирования

Кузьмин М.И., Кушнирук Д.И., Аникин А.В., Верба С.В., Зубов Д.В. Применение концепции цифрового двойника на этапах проектирования, моделирования и управления химическим процессом // Программные продукты и системы. 2024. Т. 37. № 4. С. 629–637. doi: 10.15827/0236-235X.148.629-637

### Информация о статье

Группа специальностей ВАК: 1.2.2

Поступила в редакцию: 25.03.2024

После доработки: 15.05.2024

Принята к публикации: 03.06.2024

**Аннотация.** В настоящей работе рассмотрено развитие актуальной концепции цифрового двойника, сопряженной с интенсивным развитием информационно-коммуникационных технологий и с формированием цифрового производства. Приведены некоторые варианты классификации по различным признакам. Непосредственная практическая реализация концепции цифрового двойника продемонстрирована на примере установки, в которой протекает химический процесс, контролируемый изменением температуры. Рассмотрен жизненный цикл изделия от этапа проектирования обособленной виртуальной копии до получения физической копии, связанной с виртуальной двухсторонним обменом данными посредством промежуточного программного обеспечения. Предложен вариант структурно-организационного исполнения цифровой модели с использованием информации о химической кинетике процесса в качестве взаимозаменяемого ядра. Описана внутренняя логика работы цифровой модели в формате конечного автомата. Логика направлена на оптимальное управление в контексте достижения максимального уровня конверсии по полупродукту на каждой промежуточной стадии и в результате по целевому продукту на конечной стадии. Применение концепции цифрового двойника на этапе проектирования изделия показано с позиции проведения виртуальных испытаний для нивелирования конструкторских изъянов и проверки проектных решений. По итогам работы сформулированы основные достоинства использования концепции цифрового двойника по сравнению с традиционным подходом. Результаты настоящей работы могут быть использованы как отправная точка при проектировании и построении аналогичных систем в области химических процессов, управление которыми осуществляется изменением какого-либо технологического параметра. Правообладателем результатов интеллектуальной деятельности является частное учреждение «Наука и инновации».

**Ключевые слова:** цифровой двойник, оптимальное управление, проектирование, математическое моделирование, химический процесс

**Благодарности.** Работа выполнена в рамках проекта ЕОТП-МТ-454 (договор от 14.12.2021 № 774/409-Д)

**Введение.** Стремительный прогресс в области информационно-коммуникационных технологий за последние десятилетия ознаменовал собой переход различных промышленных отраслей к новому технологическому укладу – цифровому производству. Такие автоматизированные средства, как CAD, CAE, CAM, PDM и другие классы систем прочно вошли в обиход и фактически являются стандартом при разработке нового продукта либо процесса. Большой потенциал также представляет использование технологий Big Data, IoT (*Internet of Things*), AI (*Artificial Intelligent*), облачных и граничных вычислений, высокоскоростных беспроводных сетей. Все перечисленные средства и технологии предоставляют огромные возможности для интеграции физического и цифрового мира.

Подобная интеграция позволяет бизнесу оперативно решать возникающие задачи и своевременно подстраиваться под динамично меняющиеся требования рынка, тем самым повышая свою конкурентоспособность. Таким образом, говоря о цифровом производстве, чаще всего подразумевают использование некоторой киберфизической системы, интегрированной в производственные процессы предприятия.

Ярким примером подобной системы может являться цифровой двойник некоторого процесса либо изделия. Концепция была сформулирована исследователями в начале 2000-х (например, [1]) и впоследствии значительно развита к концу 2010-х годов (например, [2]).

Данная область продолжает динамично развиваться и охватывать различные смежные

сферы, что подтверждает постоянно растущая публикационная активность. Несмотря на это наблюдается отсутствие единой системы определений и общепринятой классификации. В литературных источниках можно обнаружить множество ее различных вариантов с соответствующей аргументацией. Например, предлагается подразделять цифровые двойники, в частности, по таким признакам: используемая технология [2], выполняемые функции [3], уровень интеграции данных [4], уровень сложности применяемых решений [5]. Такое многообразие нередко приводит к различной интерпретации и своему пониманию возможностей применения концепции на практике многими исследователями.

Применительно к химическим процессам следует отметить широко используемые варианты реализации: построение цифрового двойника постфактум для уже спроектированного и введенного в эксплуатацию объекта исходя из исследовательских целей [6], построение цифрового двойника как отдельной трехмерной копии объекта с целью визуализации течения процесса и отображения телеметрии [7], создание полноценных связей объекта и цифровой копии для достижения оптимального качества выпускаемой продукции [8, 9]. В целом детальная специфика построения подобных систем для химических процессов в общем плане рассматривается в достаточно малом количестве работ [10, 11], и зачастую авторы сфокусированы на тонкостях конкретного технологического процесса. Основным же недостатком является применение концепции цифрового двойника вне контекста жизненного цикла изделия. Таким образом, многими авторами обозревается только стадия эксплуатации и не затрагиваются этапы проектирования, виртуальных испытаний, автоматизации, обслуживания и другие, что не позволяет задействовать потенциал концепции в полной мере.

Научная значимость настоящей работы заключается в описании систематизированного подхода, полученного на основе компиляции совокупного практического опыта авторов в области создания автоматизированных установок для проведения химических процессов различной природы и знаний, представленных в актуальных литературных источниках.

Практическая значимость исследования обосновывается возможностью использования как отправной точки предложенного подхода специалистами, работающими в сфере проектирования и построения систем эффективного

управления процессами и оборудованием химической промышленности.

### **Цифровой двойник на этапе эксплуатации**

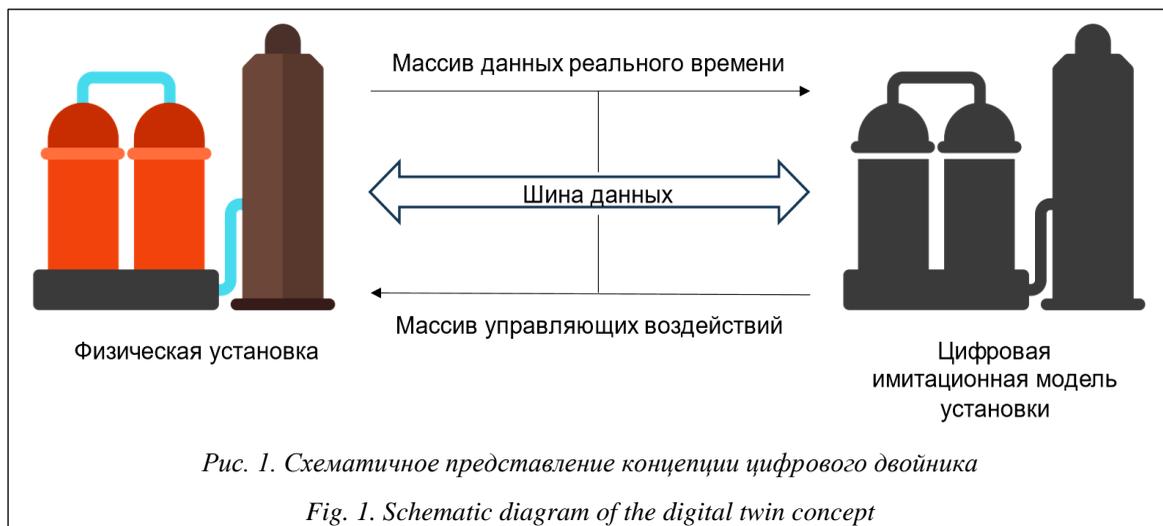
Интенсивное развитие концепции цифрового двойника, как и любой относительно новой области знания, не обходится без разночтений и спорных аспектов. Для их нивелирования начиная с 1 января 2022 года в Российской Федерации введен в действие новый государственный стандарт для цифровых двойников изделий (ГОСТ Р 57700.37–2021), целью которого является определение общих положений и создание единого унифицированного теоретического базиса. В настоящее время этот стандарт применяется преимущественно в области машиностроения, где он регламентирует создание цифровых двойников изделий на всех стадиях их жизненного цикла.

Основываясь на описании концепции и деталях стандарта, реализация цифрового двойника некой установки, в которой протекает химический процесс, может быть представлена схематично (рис. 1).

В качестве физической установки на рисунке 1 может выступать, например, емкостный либо трубчатый реактор, снабженный обогреваемой внешней частью в виде рубашки или печи с резистивным нагревом соответственно. Совместно с контуром автоматизации, состоящим из различных справочных датчиков, управляющих устройств, контроллеров и других средств, установка представляет собой так называемую физическую копию.

Задача реализации цифрового двойника подразумевает создание виртуальной копии, способной с некоторой точностью эмулировать протекание процессов, происходящих в различных частях физической установки. В данном случае в виде виртуальной копии может выступать имитационная цифровая модель или соответствующая модель машинного обучения. При этом возможен последовательный переход с первой на вторую по мере накопления архивных данных о проводимом процессе и формировании обучающей выборки. В обоих случаях основным условием является функционирование в режиме реального времени.

В настоящей работе в качестве виртуальной копии рассматривается имитационная цифровая модель. Ядром модели применительно к единичному химическому процессу являются



кинетическая модель и триплет (или же несколько триплетов, если процесс протекает через ряд стадий различного характера: параллельных, конкурирующих, последовательных, автокаталитических и проч.), определяющие химизм процесса в локальной точке пространства.

Кинетическая модель и соответствующие триплеты могут быть получены путем исследования химического процесса с помощью методов физико-химического анализа. В зависимости от фазового состояния исходных реагентов применяемый подход к изучению несколько различается, но сохраняется основная суть – исследование природы конкретной стадии химического процесса абстрагировано от конструктивных особенностей установки.

Типовой вид кинетической модели может быть продемонстрирован на примере процесса сублимации кристаллического хлорида аммония, который широко применяется для его очистки:

$$\frac{dm_A}{dT} = -\frac{A_1}{\beta} e^{-\frac{E_{a1}}{RT}} f_1(m_A, m_B), \quad (1)$$

$$\frac{dm_B}{dT} = \frac{A_1}{\beta} e^{-\frac{E_{a1}}{RT}} f_1(m_A, m_B) - \frac{A_2}{\beta} e^{-\frac{E_{a2}}{RT}} f_2(m_B, m_C), \quad (2)$$

$$\frac{dm_C}{dT} = \frac{A_2}{\beta} e^{-\frac{E_{a2}}{RT}} f_2(m_B, m_C) - \frac{A_3}{\beta} e^{-\frac{E_{a3}}{RT}} f_3(m_C, m_D), \quad (3)$$

$$m_D = 1 - m_A - m_B - m_C, \quad (4)$$

$$\Delta m = \Delta m_0 - \delta m \left[ c_1(1 - m_A) + c_2(m_C + m_D) + (1 - c_1 - c_2)m_D \right], \quad (5)$$

где  $m_A, m_B, m_C, m_D$  – массовые доли соответствующих псевдокомпонентов;  $f_1(m_A, m_B), f_2(m_B, m_C), f_3(m_C, m_D)$  – модели реакций, описывающие превращение исходного соединения в продукт;  $A_1, A_2, A_3$  – предэкспоненциальные множители,  $c^{-1}$ ;  $E_{a1}, E_{a2}, E_{a3}$  – энергии активации, Дж/моль;  $\beta$  – скорость линейного нагрева, К/с;  $T$  – температура, К;  $R$  – универсальная газовая постоянная, равная 8.314 Дж/(моль·К);  $\Delta m$  – остаточная массовая доля образца относительно начальной, %;  $\Delta m_0$  – начальная массовая доля образца (обычно принимаемая за 100 %), %;  $\delta m$  – изменение массовой доли образца, %;  $c_1, c_2$  – вклады соответствующих стадий в общее изменение массовой доли образца.

Система уравнений (1)–(5) описывает изменение массовой доли образца хлорида аммония при произвольном температурном режиме. Совокупность предэкспоненциального множителя  $A_i$ , модели  $f_i(m_k, m_j)$  и энергии активации  $E_{a_i}$  представляет собой кинетический триплет, описывающий отдельную стадию многостадийного процесса.

Кинетическая модель в дальнейшем встраивается в общую математическую обертку цифровой модели. Математическая обертка может содержать внутри отдельные системы дифференциальных уравнений, обособленные математические выражения для описания конкретных частей установки (например, материальный баланс участка абсорбции или тепловой баланс конденсатора), и вспомогательные численные методы, применяемые для нахождения решений. В кинетической модели (1)–(5) на уровне математической обертки требуется реализация какого-либо численного метода для решения системы обыкновенных дифференциальных уравнений, например, метода Дорманда–

Принса [12]. На данном этапе также может быть реализован ряд моделей, позволяющих прогнозировать естественное старение и износ деталей конструкции [13–15]. Это дает возможность планировать ремонтные работы и осуществлять профилактическое обслуживание установки.

Дальнейшая имплементация совокупности математических конструкций с использованием технологий программной разработки и конкретного языка программирования позволяет получить подключаемый модуль, представляющий собой цифровую имитационную модель.

Следует отметить, что полученная виртуальная копия установки обладает некоторыми особенностями. Во-первых, она применима только к соответствующей ей физической копии, что следует из использования в процессе построения сопряжения информации о химизме процесса с особенностями конструктивной реализации. Во-вторых, использование кинетической модели в ядре виртуальной копии дает возможность создавать серии цифровых моделей для однотипных химических процессов, в которых, например, варьируется один из двух реагентов, а другой остается неизменным. При этом заменяется только кинетическая модель, а остальная часть цифровой модели сохраняется.

Как следует из рисунка 1, основной задачей цифровой модели является генерация массива управляющих воздействий в ответ на входящий массив данных реального времени, содержащий информацию о текущих значениях технологических параметров химического процесса. Глобальной целью в данном случае является оптимальное управление в контексте достижения максимального уровня конверсии исходных реагентов в целевой продукт и по возможности минимизация протекания побочных реакций.

Внутренняя логика работы цифровой модели строится по принципу конечного автомата. На примере многостадийного химического процесса, контролируемого изменением температуры и состоящего из  $n$  стадий, можно записать:

$$S \rightarrow A_1 \rightarrow B_1 \rightarrow C_1 \rightarrow \dots \rightarrow A_n \rightarrow B_n \rightarrow C_n \rightarrow E, \quad (6)$$

где  $S$  – начальное состояние цифровой модели;  $E$  – конечное состояние цифровой модели;  $A_1 \dots A_n$  – состояния инициализации процесса на отдельных стадиях;  $B_1 \dots B_n$  – состояния контроля текущего протекания процесса на отдельных стадиях;  $C_1 \dots C_n$  – состояния завершения процесса на отдельных стадиях.

Задачей цифровой модели является достижение конечного состояния  $E$  из начального  $S$  путем последовательного перехода через промежуточные состояния  $A_1 \dots A_n, B_1 \dots B_n, C_1 \dots C_n$ .

Переход  $S \rightarrow A_1$  осуществляется только после задания пользователем запрашиваемых начальных параметров (например, скорости желаемого линейного нагрева шихты).

Переход  $A_1 \rightarrow B_1$  следует после предварительного расчета траектории химической реакции на первой стадии химического процесса и выработки стартового массива управляющих воздействий. Для этого цифровая модель определяет оптимальную температуру протекания реакции, при которой скорость химического превращения максимальна. Далее следует выработка такого управляющего воздействия, которое позволит достичь рассчитанного значения с учетом инертности системы и необходимости избегания возникновения больших температурных градиентов внутри содержимого реактора.

Переход  $B_1 \rightarrow C_1$  становится возможен только после того, как на текущей стадии будет достигнуто значение конверсии по промежуточному продукту, близкое к 1. В остальное время цифровая модель, находясь в состоянии  $B_1$ , осуществляет контроль и выработку корректирующих воздействий в случае необходимости, основываясь на предварительном расчете, сделанном в состоянии  $A_1$ . Перейдя в состояние  $C_1$ , цифровая модель контролирует корректное завершение работы и ожидает разрешения пользователя на переход к следующей стадии химического процесса.

Немаловажным звеном в структуре цифрового двойника является ПО. Его роль заключается в обеспечении взаимосвязанной работы физической установки с ее виртуальной копией путем непрерывного двухстороннего обмена данными. Со стороны физической копии они представлены массивом технологических параметров, агрегируемых с датчиков, расположенных на установке и связанных в единый контур. Со стороны виртуальной копии – массивом параметров, характеризующих управляющее воздействие. Таким образом, ПО должно, во-первых, реализовывать общий интерфейс для приема и отправки данных в соответствии с общепринятыми протоколами (семейство Modbus, OPC DA, OPC UA и др.). Во-вторых, оно должно предоставлять единый контракт для подключения цифровых моделей или моделей машинного обучения. Дополнительно в нем может быть реализован различный функцио-

нал для удобства пользователя: отслеживание и предотвращение аварийных ситуаций, уведомления различного характера, агрегация, задача, визуализация данных и т.д.

Общий вид главного окна подобной системы, разработанной на языке программирования C# с использованием платформы .NET, приведен на рисунке 2. В целом данное ПО можно отнести к классу систем SCADA с некоторыми модификациями, позволяющими гибко интегрироваться в современные технологии. Его функционал включает в себя менеджер проектов, позволяющий настраивать параметры автоматизации под конкретные установку и процесс, модуль подключения и взаимодействия с БД посредством ORM-технологии Entity Framework Core, модуль логирования событий и отслеживания аварийных случаев. Разработанное ПО позволяет взаимодействовать с подключаемыми устройствами с помощью единого интерфейса, поддерживающего наиболее распространенные протоколы промышленной связи. Главной особенностью является возможность управления химическим процессом в одном из двух режимов – технологический или цифровой двойник. Технологический режим предоставляет пользователю возможность самому определять последовательность операций и их параметры в масштабе времени, формируя связанную структуру, аналогичную диаграмме Ганта. Далее в соответствии с ней ПО

осуществляет планирование операций с использованием фреймворка Quartz.NET. Исполнение распланированной последовательности осуществляется с учетом того, что единичная операция может не иметь четких временных отметок начала, продолжительности и конца. В таком случае моменты начала и конца единичной операции заменяются триггерами по событию (например, операция подачи газа в реактор начнется только после завершения операции его нагрева до определенной температуры). Управление в режиме цифрового двойника обеспечивается путем выбора пользователем модуля цифровой имитационной модели, представленной в виде заранее созданного динамически подгружаемого DLL-файла. В таком случае от пользователя при необходимости требуется только задать значения параметров, постоянных на протяжении всего процесса и обеспечить корректное связывание массивов входных и выходных параметров с соответствующими подключенными устройствами. Подключение моделей машинного обучения взамен цифровых имитационных моделей осуществляется посредством библиотеки ML.NET. Таким образом, модель машинного обучения может быть построена с использованием возможностей Python, основываясь на накопленном массиве данных о процессе в формате вход-выход за некоторый промежуток времени, и подключена к разработанному ПО.

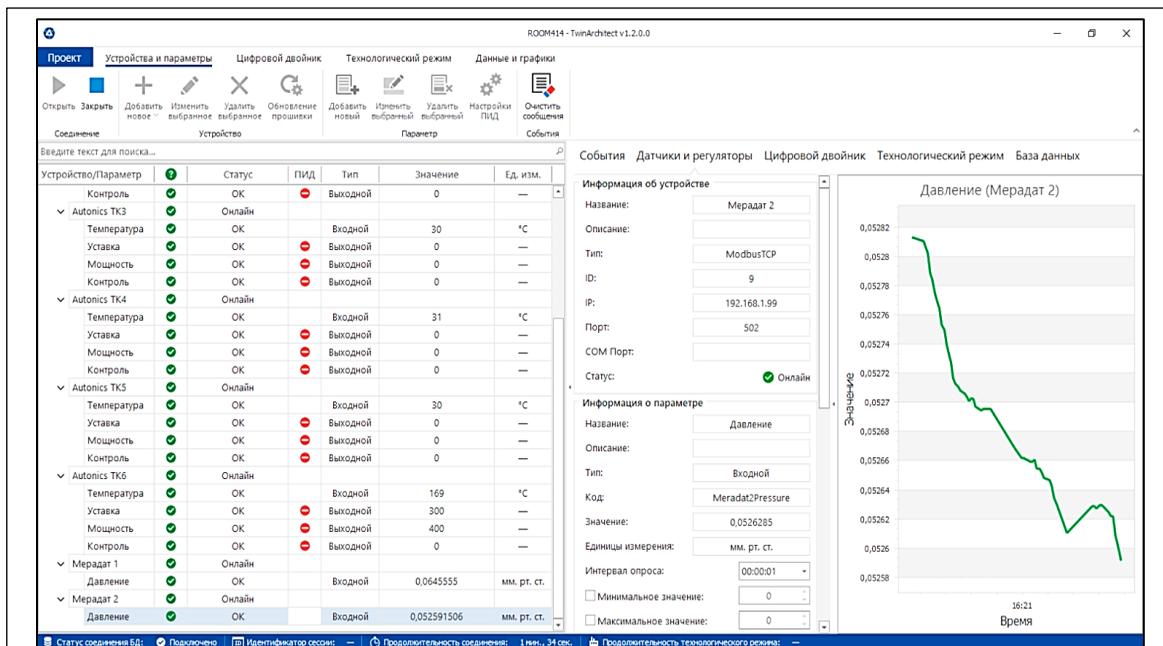


Рис. 2. Главное окно системы связывания физической и виртуальной копии установки

Fig. 2. Main window of the system for linking physical and virtual installation copy

## Предцифровой двойник на этапе проектирования

Исходя из сути концепции цифрового двойника, а также одного из вариантов классификации [5] можно выделить отдельную стадию жизненного цикла изделия – предцифровой двойник. Его главной особенностью является полное отсутствие физической копии. В данном случае построение виртуальной копии предваряет ее непосредственную физическую реализацию. Основной задачей этого этапа применительно к рассматриваемой в настоящей работе установке является проектирование с акцентом на технологию проводимого химического процесса, на выявление различных технических и конструкционных рисков. Для этого с помощью разработанной в САД-системе 3D-модель в упрощенной форме осуществляют постановку и решение сопряженной тепло-гидродинамической задачи в САЕ-системе с учетом химической кинетики и особенностей проведения целевого процесса.

Процесс моделирования проводился в программном комплексе конечно-элементного анализа COMSOL Multiphysics. Использовался функционал модулей гидродинамики и теплопередачи в сопряженном расчете с кинетической моделью, аналогичной системе (1)–(5), для описания поведения химической системы. Полученные результаты (<http://www.swsys.ru/uploaded/image/2024-4/2.jpg>) позволяют проанализировать с некоторой точностью, как будет протекать процесс с учетом текущей конфигурации экспериментальной установки. Например, рассматривая распределение конверсии в объеме шихты внутри реактора, можно сказать, что из-за возникших в результате конструкционных решений больших температурных гра-

диентов наблюдается сильная неравномерность протекания реакции по продольному профилю реактора. Таким образом, реакция на концах реактора имеет значительно меньшую скорость по сравнению с его центром из-за повышенного теплорассеяния и недостатка теплоизоляции. Аналогично могут быть выявлены множественные критические конструкторские изъяны, проверена эффективность тех или иных проектных решений без проведения множества натурных испытаний.

## Заключение

Таким образом, концепция цифрового двойника предоставляет бизнесу и различным отраслям производства мощный и современный инструмент для реализации своих потребностей и решения возникающих задач. В целом можно выделить следующие основные достоинства по сравнению с традиционным подходом применительно к рассматриваемой в настоящей работе предметной области.

Большое число натурных испытаний может быть заменено с некоторой точностью проведением моделирования в виртуальной среде при условии использования корректных данных о химической кинетике процесса. Снижение числа натурных испытаний ведет к уменьшению материальных и временных затрат. Агрегация и последующий анализ массива данных позволяют проводить предиктивную аналитику и впоследствии разрабатывать модели машинного обучения взамен имитационных.

Кроме того, оптимальное управление процессом на каждой стадии многостадийного химического процесса обосновано теоретически, что исключает влияние человеческого фактора.

## Список литературы

1. Grieves M.W. Product lifecycle management: the new paradigm for enterprises. *IJPD*, 2005, vol. 2, no. 1-2, pp. 71–84. doi: 10.1504/IJPD.2005.006669.
2. Wong C.Y., McFarlane D., Zaharudin A.A., Agarwal V. The intelligent product driven supply chain. *Proc. IEEE Int. Conf. SMC*, 2002, vol. 4, pp. 6. doi: 10.1109/ICSMC.2002.1173319.
3. Hribernik K.A., Rabe L., Thoben K.-D., Schumacher J. The product avatar as a product-instance-centric information management concept. *IJPLM*, 2006, vol. 1, no. 4, pp. 367–379. doi: 10.1504/IJPLM.2006.011055.
4. Grieves M., Vickers J. Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems. In: *Transdisciplinary Perspectives on Complex Systems*, 2017, pp. 85–113. doi: 10.1007/978-3-319-38756-7\_4.
5. Grieves M.W. Digital twins: Past, present, and future. In: *The Digital Twin*, 2023, pp. 97–121. doi: 10.1007/978-3-031-21343-4\_4.
6. Kosacka-Olejnik M., Kostrzewski M., Marczevska M., Mrówczyńska B., Pawlewski P. How digital twin concept supports internal transport systems? – Literature review. *Energies*, 2021, vol. 14, no. 16, art. 4919. doi: 10.3390/en14164919.
7. Feasibility of an immersive digital twin: The definition of a digital twin and discussions around the benefit of immersion. *HVM Catapult*, 2018. URL: [https://www.amrc.co.uk/files/document/219/1536919984\\_HVM\\_CATAPULT\\_DIGITAL\\_TWIN\\_DL.pdf](https://www.amrc.co.uk/files/document/219/1536919984_HVM_CATAPULT_DIGITAL_TWIN_DL.pdf) (дата обращения: 10.03.2024).

8. Kritzinger W., Karner M., Traar G., Henjes J., Sihn W. Digital Twin in manufacturing: A categorical literature review and classification. *Ifac-PapersOnline*, 2018, vol. 51, no. 11, pp. 1016–1022. doi: 10.1016/j.ifacol.2018.08.474.
9. Madni A.M., Madni C.C., Lucero S.D. Leveraging digital twin technology in model-based systems engineering. *Systems*, 2019, vol. 7, no. 1, art. 7. doi: 10.3390/systems7010007.
10. Yu W., Patros P., Young B., Klinac E., Walmsley T.G. Energy digital twin technology for industrial energy management: Classification, challenges and future. *Renewable and Sustainable Energy Reviews*, 2022, vol. 161, art. 112407. doi: 10.1016/j.rser.2022.112407.
11. Kichatov K.G., Prosochkina T.R., Vorobyova I.S. Principles of creating a digital twin prototype for the process of alkylation of benzene with propylene based on a neural network. *Fine Chemical Tech.*, 2023, vol. 18, no. 5, pp. 482–497. doi: 10.32362/2410-6593-2023-18-5-482-497.
12. Dli M.I., Vlasova E.A., Sokolov A.M., Morgunova E.V. Creation of a chemical-technological system digital twin using the Python language. *J. of Applied Inform.*, 2021, vol. 16, no. 1, pp. 22–31. doi: 10.37791/2687-0649-2021-16-1-22-31.
13. Gerogiorgis D.I., Castro-Rodriguez D. A digital twin for process optimisation in pharmaceutical manufacturing. In: *Computer Aided Chemical Engineering*, 2021, vol. 50, pp. 253–258. doi: 10.1016/B978-0-323-88506-5.50041-3.
14. Zhu X., Ji Y. A digital twin-driven method for online quality control in process industry. *The Int. J. of Advanced Manufacturing Tech.*, 2022, vol. 119, pp. 3045–3064. doi: 10.1007/s00170-021-08369-5.
15. Gao L., Jia M., Liu D. Process digital twin and its application in petrochemical industry. *JSEA*, 2022, vol. 15, no. 8, pp. 308–324. doi: 10.4236/jsea.2022.158018.
16. Oliveira L.M., Dias R., Rebello C.M., Martins M.A., Rodrigues A.E., Ribeiro A.M., Nogueira I.B. Artificial intelligence and cyber-physical systems: A review and perspectives for the future in the chemical industry. *AI*, 2021, vol. 2, no. 3, pp. 429–443. doi: 10.3390/ai2030027.
17. Dormand J.R., Prince P.J. A family of embedded Runge–Kutta formulae. *J. of Computational and Applied Math.*, 1980, vol. 6, no. 1, pp. 19–26. doi: 10.1016/0771-050X(80)90013-3.
18. Бутов А.А., Самохвалов М.В. Математические модели многостадийного износа продуктивных систем // Ученые записки УлГУ. Сер.: Математика и информационные технологии. 2020. № 1. С. 17–24.
19. Поляков Д.А., Холмов М.А., Плотников Д.И., Никитин К.И., Полякова У.В. Математическое моделирование срока службы полимерной изоляции кабелей // Омский науч. вестн. 2020. № 6. С. 69–73. doi: 10.25206/1813-8225-2020-174-69-73.
20. Острейковский В.А., Сорочкин А.В. Физико-математические модели закономерностей процессов естественного старения конструкционных материалов в задачах долговечности оборудования структурно и функционально сложных критически важных систем с длительными сроками активного существования // Вестн. кибернетики. 2021. № 4. С. 22–27.

Software &amp; Systems

doi: 10.15827/0236-235X.148.629-637

2024, 37(4), pp. 629–637

### Applying the digital twin concept to chemical process design, modeling and control

Maksim I. Kuzmin <sup>1,2</sup>✉, David I. Kushniruk <sup>1</sup>, Andrey V. Anikin <sup>1</sup>,  
Sergey V. Verba <sup>1</sup>, Dmitry V. Zubov <sup>2</sup>

<sup>1</sup> N. Sazhin State Research and Design Institute of Rare Metal Industry “Giredmet”,  
Moscow, 111524, Russian Federation

<sup>2</sup> D. Mendeleev University of Chemical Technology of Russian Federation,  
Moscow, 125047, Russian Federation

#### For citation

Kuzmin, M.I., Kushniruk, D.I., Anikin, A.V., Verba, S.V., Zubov, D.V. (2024) ‘Applying the digital twin concept to chemical process design, modeling and control’, *Software & Systems*, 37(4), pp. 629–637 (in Russ.). doi: 10.15827/0236-235X.148.629-637

#### Article info

Received: 25.03.2024

After revision: 15.05.2024

Accepted: 03.06.2024

**Abstract.** The paper considers the development and increasing relevance of the digital twin concept. It is relevant due to the intensive development of information and communication technologies, as well as and the digital production. The

authors give some classification options according to various features. They demonstrate the practical realization of the digital twin concept on the example of an installation in which a chemical process controlled by temperature change takes place. The authors consider the product life cycle from the design stage of a stand-alone virtual copy to the physical copy, which is associated with the virtual twin by means of middleware. They propose the structural and organizational design version of a digital model using information about the process chemical kinetics as an interchangeable core. There is a description of the internal logic of the numerical model in the finite automaton format. The logic aims at optimal control in terms of achieving the maximum level of conversion by a semi-product at each intermediate stage and by the target product at the final stage. The authors apply the digital twin concept at the product design stage in terms of conducting virtual tests to level design flaws and verify design solutions. Based on the results of the work, the authors state the main advantages of using the digital twin concept in comparison with the traditional approach. The results can be used as a starting point when designing and constructing similar systems in the field of chemical processes, which are controlled by changing any technological parameter. The copyright holder for the project results of the is a private enterprise "Science and Innovations".

**Keywords:** digital twin, chemical process, optimal control, design, mathematical modeling; chemical process

**Acknowledgements.** The work was carried out as part of the EOTP-MT454 project (contract no. 774/409-D, dated 12/14/2021)

### References

1. Grieves, M.W. (2005) 'Product lifecycle management: the new paradigm for enterprises', *IJPD*, 2(1-2), pp. 71–84. doi: 10.1504/IJPD.2005.006669.
2. Wong, C.Y., McFarlane, D., Zaharudin, A.A., Agarwal, V. (2002) 'The intelligent product driven supply chain', *Proc. IEEE Int. Conf. SMC*, 4, pp. 6. doi: 10.1109/ICSMC.2002.1173319.
3. Hribernik, K.A., Rabe, L., Thoben, K.-D., Schumacher, J. (2006) 'The product avatar as a product-instance-centric information management concept', *IJPLM*, 1(4), pp. 367–379. doi: 10.1504/IJPLM.2006.011055.
4. Grieves, M., Vickers, J. (2017) 'Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems', in *Transdisciplinary Perspectives on Complex Systems*, pp. 85–113. doi: 10.1007/978-3-319-38756-7\_4.
5. Grieves, M.W. (2023) 'Digital twins: Past, present, and future', in *The Digital Twin*, pp. 97–121. doi: 10.1007/978-3-031-21343-4\_4.
6. Kosacka-Olejnik, M., Kostrzewski, M., Marczevska, M., Mrówczyńska, B., Pawlewski, P. (2021) 'How digital twin concept supports internal transport systems? – Literature review', *Energies*, 14(16), art. 4919. doi: 10.3390/en14164919.
7. (2018) 'Feasibility of an immersive digital twin: The definition of a digital twin and discussions around the benefit of immersion', *HVM Catapult*, available at: [https://www.amrc.co.uk/files/document/219/1536919984\\_HVM\\_CATAPULT\\_DIGITAL\\_TWIN\\_DL.pdf](https://www.amrc.co.uk/files/document/219/1536919984_HVM_CATAPULT_DIGITAL_TWIN_DL.pdf) (accessed March 10, 2024).
8. Kritzinger, W., Karner, M., Traar, G., Henjes, J., Sihm, W. (2018) 'Digital Twin in manufacturing: A categorical literature review and classification', *Ifac-PapersOnline*, 51(11), pp. 1016–1022. doi: 10.1016/j.ifacol.2018.08.474.
9. Madni, A.M., Madni, C.C., Lucero, S.D. (2019) 'Leveraging digital twin technology in model-based systems engineering', *Systems*, 7(1), art. 7. doi: 10.3390/systems7010007.
10. Yu, W., Patros, P., Young, B., Klinac, E., Walmsley, T.G. (2022) 'Energy digital twin technology for industrial energy management: Classification, challenges and future', *Renewable and Sustainable Energy Reviews*, 161, art. 112407. doi: 10.1016/j.rser.2022.112407.
11. Kichatov, K.G., Prosochkina, T.R., Vorobyova, I.S. (2023) 'Principles of creating a digital twin prototype for the process of alkylation of benzene with propylene based on a neural network', *Fine Chemical Tech.*, 18(5), pp. 482–497. doi: 10.32362/2410-6593-2023-18-5-482-497.
12. Dli, M.I., Vlasova, E.A., Sokolov, A.M., Morgunova, E.V. (2021) 'Creation of a chemical-technological system digital twin using the Python language', *J. of Applied Inform.*, 16(1), pp. 22–31. doi: 10.37791/2687-0649-2021-16-1-22-31.
13. Gerogiorgis, D.I., Castro-Rodriguez, D. (2021) 'A digital twin for process optimisation in pharmaceutical manufacturing', in *Computer Aided Chemical Engineering*, 50, pp. 253–258. doi: 10.1016/B978-0-323-88506-5.50041-3.
14. Zhu, X., Ji, Y. (2022) 'A digital twin-driven method for online quality control in process industry', *The Int. J. of Advanced Manufacturing Tech.*, 119, pp. 3045–3064. doi: 10.1007/s00170-021-08369-5.
15. Gao, L., Jia, M., Liu, D. (2022) 'Process digital twin and its application in petrochemical industry', *JSEA*, 15(8), pp. 308–324. doi: 10.4236/jsea.2022.158018.
16. Oliveira, L.M., Dias, R., Rebello, C.M., Martins, M.A., Rodrigues, A.E., Ribeiro, A.M., Nogueira, I.B. (2021) 'Artificial intelligence and cyber-physical systems: A review and perspectives for the future in the chemical industry', *AI*, 2(3), pp. 429–443. doi: 10.3390/ai2030027.
17. Dormand, J.R., Prince, P.J. (1980) 'A family of embedded Runge–Kutta formulae', *J. of Computational and Applied Math.*, 6(1), pp. 19–26. doi: 10.1016/0771-050X(80)90013-3.
18. Butov, A.A., Samokhvalov, M.V. (2020) 'Mathematical models of multistage wear of productive systems', *Sci. Notes of UIGU. Ser.: Math. and Inform. Tech.*, (1), pp. 17–24 (in Russ.).
19. Polyakov, D.A., Kholmov, M.A., Plotnikov, D.I., Nikitin, K.I., Polyakova, U.V. (2020) 'Mathematical modeling of service life of cables polymer insulation', *Omsk Sci. Bull.*, (6), pp. 69–73 (in Russ.). doi: 10.25206/1813-8225-2020-174-69-73.

20. Ostreikovskiy, V.A., Sorochkin, A.V. (2021) 'Physico-mathematical models of the regularities of the processes of natural aging of structural materials in the problems of durability of equipment of structurally and functionally complex critical systems with long periods of active existence', *Bull. of Cybernetics*, (4), pp. 22–27 (in Russ.).

**Авторы**

**Кузьмин Максим Игоревич**<sup>1,2</sup>, научный сотрудник, аспирант, mimikatz@mail.ru

**Кушнирук Давид Ильич**<sup>1</sup>, начальник группы, DIKushniruk@rosatom.ru

**Аникин Андрей Витальевич**<sup>1</sup>, младший научный сотрудник, anikinwk@gmail.com

**Верба Сергей Владимирович**<sup>1</sup>, главный конструктор, SVVerba@rosatom.ru

**Зубов Дмитрий Владимирович**<sup>2</sup>, к.т.н., доцент, dvzubov@gmail.com

<sup>1</sup> Государственный научно-исследовательский и проектный институт редкометаллической промышленности «Гиредмет» имени Н.П. Сажина, г. Москва, 111524, Россия

<sup>2</sup> Российский химико-технологический университет имени Д.И. Менделеева, г. Москва, 125047, Россия

**Authors**

**Maksim I. Kuzmin**<sup>1,2</sup>, Associate Research, Postgraduate Student, mimikatz@mail.ru

**David I. Kushniruk**<sup>1</sup>, Head of Group, DIKushniruk@rosatom.ru

**Andrey V. Anikin**<sup>1</sup>, Junior Researcher, anikinwk@gmail.com

**Sergey V. Verba**<sup>1</sup>, Chief Designer, SVVerba@rosatom.ru

**Dmitry V. Zubov**<sup>2</sup>, Cand. of Sci. (Engineering), Associate Professor, dvzubov@gmail.com

<sup>1</sup> N. Sazhin State Research and Design Institute of Rare Metal Industry "Giredmet", Moscow, 111524, Russian Federation

<sup>2</sup> D. Mendeleev University of Chemical Technology of Russian Federation, Moscow, 125047, Russian Federation

## СО Д Е Р Ж А Н И Е

<b>Баранов А.В., Кузнецов Е.Е.</b>	
Управление пользовательскими заданиями в сети суперкомпьютерных центров с применением федеративной аутентификации .....	461
<b>Аладышев О.С., Захарченко А.В., Огарышев В.Ф., Шабанов Б.М.</b>	
Ожидания от производительности вычислительного кластера при выборе параллельной файловой системы .....	472
<b>Котенко И.В., Саенко И.Б., Паращук И.Б., Десницкий В.А., Виткова Л.А.</b>	
Аналитическая обработка больших массивов данных о событиях кибербезопасности с применением суперкомпьютерных вычислений.....	487
<b>Абрамов А.Г.</b>	
Высокопроизводительный сервис сбора и анализа файлов журналов сетевого и серверного оборудования в национальной исследовательской компьютерной сети.....	495
<b>Балабаев С.А., Лупин С.А., Телегин П.Н., Шабанов Б.М.</b>	
Повышение вычислительной мощности персонального компьютера за счет интеграции с распределенной системой из смартфонов.....	504
<b>Власова С.А., Каленов Н.Е.</b>	
Диалоговый программный комплекс для формирования онтологии Единого цифрового пространства научных знаний .....	514
<b>Погорелко К.П., Савин Г.И.</b>	
Организация поиска в базе данных со связанными сущностями.....	524
<b>Долгов Н.В., Ильин В.А.</b>	
Особенности функционального моделирования сложных иерархических систем при различных способах организации управления исследуемыми процессами с целью выбора рациональной структуры системы .....	532
<b>Еремеев А.П., Панявин Н.А., Маренков М.А.</b>	
Реализация темпоральной базы данных для интеллектуальных систем поддержки принятия решений реального времени на примере образовательного процесса вуза .....	539
<b>Сидоркина И.Г., Михалищев С.В.</b>	
Совершенствование метода оценки стойкости пароля аутентификации пользователя компьютерных систем на основе использования известных уязвимостей.....	547
<b>Соколов А.М., Борисов В.В.</b>	
Рекуррентные нечеткие ситуационно-прецедентные модели для оперативного управления сложными техническими объектами.....	554
<b>Игнатьева А.С.</b>	
Нечеткая кластеризация как способ повышения эффективности управления в автоматических системах.....	566
<b>Абу Абед Ф.Н., Жиронкин С.А.</b>	
Разработка программного комплекса многоканального распознавания и коррекции речевых сообщений на основе алгоритмов машинного обучения в структуре импортозамещения .....	576
<b>Елецкий М.И., Зальмарсон А.Ф., Захарьев А.А.</b>	
Совершенствование информационной технологии автоматизации мониторинга технической готовности кораблей Военно-Морского Флота .....	585

---

<b>Федько В.С., Ковалева О.А., Ковалев С.В.</b> Визуализация данных для бизнес-аналитики уровня доступности финансовых услуг .....	592
<b>Конов А.А., Репкин В.С., Цимбалов К.И.</b> Сценарий атаки на автоматизированную систему управления технологическим процессом с учетом уязвимости протокола Modbus TCP .....	600
<b>Быков А.А., Бобков В.И., Орехов В.А., Незамаев С.В.</b> Программная реализация математической модели расчета технологических параметров электротермических реакторов .....	611
<b>Меркурьев С.А., Палюх Б.В., Ветров А.Н.</b> Программная реализация диагностики непрерывных многостадийных производств с использованием динамической экспертной системы .....	622
<b>Кузьмин М.И., Кушнирук Д.И., Аникин А.В., Верба С.В., Зубов Д.В.</b> Применение концепции цифрового двойника на этапах проектирования, моделирования и управления химическим процессом .....	629

# C O N T E N T

<b>Baranov, A.V., Kuznetsov, E.E.</b> The user job managing in a HPC network using federated authentication.....	461
<b>Aladyshev, O.S., Zakharchenko, A.V., Ogaryshev, V.F., Shabanov, B.M.</b> Computing cluster performance expectations when selecting a parallel file system .....	472
<b>Kotenko, I.V., Saenko, I.B., Parashchuk, I.B., Desnitsky, V.A., Vitkova, L.A.</b> Analytical processing of large data sets of cybersecurity events using supercomputing .....	487
<b>Abramov, A.G.</b> High-performance service for collecting and analyzing network and server hardware log files on a National Research Computer Network.....	495
<b>Balabaev, S.A., Lupin, S.A., Telegin, P.N., Shabanov, B.M.</b> Increasing the PC computing power: Integration with a distributed smartphone system.....	504
<b>Vlasova, S.A., Kalenov, N.E.</b> Dialogue software complex for forming Common Digital Space of Scientific Knowledge ontology .....	514
<b>Pogorelko, K.P., Savin, G.I.</b> Search organization in a database with related entities .....	524
<b>Dolgov, N.V., Ilin, V.A.</b> Features of functional modeling of complex hierarchical systems with different ways of organizing the control of the study processes for the purpose of selecting a rational system structure.....	532
<b>Eremeev, A.P., Panyavin, N.A., Marenkov, M.A.</b> Temporal database in intelligent real-time decision support systems: The case of a university educational process.....	539
<b>Sidorkina, I.G., Mikhailishchev, S.V.</b> Assessing the strength of a computer system user authentication password: Improving the method based on using known vulnerabilities .....	547
<b>Sokolov, A.M., Borisov, V.V.</b> Recurrent fuzzy situational-precedent models for operational control of complex technical objects .....	554
<b>Ignatyeva, A.S.</b> Fuzzy clustering for improving control efficiency in automatic systems.....	566
<b>Abu-Abed, F.N., Zhironkin, S.A.</b> Software complex for multi-channel recognition and correction of speech messages based on machine learning algorithms in terms of import substitution .....	576
<b>Yeletsky, M.I., Zalmarson, A.F., Zakhariev, A.A.</b> Improving the information technology for automating Navy ships technical readiness monitoring .....	585
<b>Fedko, V.S., Kovaleva, O.A., Kovalev, S.V.</b> Business analytics of a financial service availability level: Data visualization .....	592
<b>Konev, A.A., Repkin, V.S., Tsimbalov, K.I.</b> Scenario of an attack on an automated process control system taking into account Modbus TCP protocol vulnerability.....	600
<b>Bykov, A.A., Bobkov, V.I., Orekhov, V.A., Nezamaev, S.V.</b> Mathematical model for calculating electrothermal reactor technological parameters: Software implementation.....	611
<b>Merkuryev, S.A., Palyukh, B.V., Vetrov, A.N.</b> Continuous multistage production diagnostics: Software implementation using a dynamic expert system.....	622
<b>Kuzmin, M.I., Kushniruk, D.I., Anikin, A.V., Verba, S.V., Zubov, D.V.</b> Applying the digital twin concept to chemical process design, modeling and control .....	629

## *Внимание подписчиков*

Научный журнал  
«Программные продукты и системы»  
распространяется по подписке  
через подписное агентство «УП УРАЛ-ПРЕСС»  
Индекс издания  
в электронном каталоге агентства 70799

*Телефон редакции: (4822) 39-91-49*  
*Электронная почта: red@cps.tver.ru*  
*Сайт журнала: www.swsys.ru*