

ISSN 2078-9181

DOI 10.15622/sp.54

РОССИЙСКАЯ АКАДЕМИЯ НАУК
Отделение нанотехнологий и информационных технологий

САНКТ-ПЕТЕРБУРГСКИЙ
ИНСТИТУТ ИНФОРМАТИКИ И АВТОМАТИЗАЦИИ РАН

ТРУДЫ СПИИРАН

proceedings.spiras.nw.ru



ВЫПУСК 5(54)



Санкт-Петербург
2017

18+

SPIIRAS PROCEEDINGS

Issue № 5(54), 2017

Scientific, educational, and interdisciplinary journal primarily specialized
in computer science, automation, and applied mathematics

Trudy SPIIRAN ♦ Founded in 2002 ♦ Труды СПИИРАН

Founder and Publisher

St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences

Editor-in-Chief

R. M. Yusupov, Prof., Dr. Sci., Corr. Member of RAS, St. Petersburg, Russia

Editorial Board Members

A. A. Ashimov ,	Prof., Dr. Sci., Academician of the National Academy of Sciences of the Republic of Kazakhstan, Almaty, Kazakhstan
S. N. Baranov ,	Prof., Dr. Sci., St. Petersburg, Russia
N. P. Veselkin ,	Prof., Dr. Sci., Academician of RAS, St. Petersburg, Russia
V. I. Gorodetski ,	Prof., Dr. Sci., St. Petersburg, Russia
O. Yu. Gusikhin ,	Ph. D., Dearborn, USA
V. Delic ,	Prof., Dr. Sci., Novi Sad, Serbia
A. Dolgui ,	Prof., Dr. Habil., St. Etienne, France
M. Zelezny ,	Assoc. Prof., Ph.D., Plzen, Czech Republic
I. A. Kalyaev ,	Prof., Dr. Sci., Academician of RAS, Taganrog, Russia
D. A. Ivanov ,	Prof., Dr. Habil., Berlin, Germany
G. A. Leonov ,	Prof., Dr. Sci., Corr. Member of RAS, St. Petersburg, Russia
K. P. Markov ,	Assoc. Prof., Ph.D., Aizu, Japan
Yu. A. Merkurjev ,	Prof., Dr. Habil., Academician of the Latvian Academy of Sciences, Riga, Latvia
R. V. Meshcheryakov ,	Prof., Dr. Sci., Tomsk, Russia
N. A. Moldovian ,	Prof., Dr. Sci., St. Petersburg, Russia
V. E. Pavlovskiy ,	Prof., Dr. Sci., Moscow, Russia
A. A. Petrovsky ,	Prof., Dr. Sci., Minsk, Belarus
V. A. Putilov ,	Prof., Dr. Sci., Apatity, Russia
V. K. Pshikhopov ,	Prof., Dr. Sci., Taganrog, Russia
A. L. Ronzhin	(Deputy Editor-in-Chief), Prof., Dr. Sci., St. Petersburg, Russia
A. I. Rudskoi ,	Prof., Dr. Sci., Academician of RAS, St. Petersburg, Russia
V. Sgurev ,	Prof., Dr. Sci., Academician of the Bulgarian academy of sciences, Sofia, Bulgaria
V. Skormin ,	Prof., Ph.D., Binghamton, USA
A. V. Smirnov ,	Prof., Dr. Sci., St. Petersburg, Russia
B. Ya. Sovetov ,	Prof., Dr. Sci., Academician of RAE, St. Petersburg, Russia
V. A. Soyfer ,	Prof., Dr. Sci., Academician of RAS, Samara, Russia
B. V. Sokolov ,	Prof., Dr. Sci., St. Petersburg, Russia
L. V. Utkin ,	Prof., Dr. Sci., St. Petersburg, Russia
A. L. Fradkov ,	Prof., Dr. Sci., St. Petersburg, Russia
L. B. Sheremetov ,	Assoc. Prof., Dr. Sci., Mexico, Mexico

Editor: E. P. Miroshnikova

Technical editor: A. I. Motienko

Translator: P. N. Emeleva

Editorial Board's address

14-th line VO, 39, SPIIRAS, St. Petersburg, 199178, Russia,
e-mail: publ@iias.spb.su, web: <http://www.proceedings.spiiras.nw.ru/>

The journal is indexed in Scopus

© St. Petersburg Institute for Informatics and Automation
of the Russian Academy of Sciences, 2017

ТРУДЫ СПИИРАН

Выпуск № 5(54), 2017

Научный, научно-образовательный, междисциплинарный журнал с базовой специализацией в области информатики, автоматизации и прикладной математики
Журнал основан в 2002 году

Учредитель и издатель

Федеральное государственное бюджетное учреждение науки
Санкт-Петербургский институт информатики и автоматизации Российской академии наук
(СПИИРАН)

Главный редактор

Р. М. Юсупов, чл.-корр. РАН, д-р техн. наук, проф., С-Петербург, РФ

Редакционная коллегия

- А. А. Ашимов**, академик национальной академии наук Республики Казахстан д-р техн. наук, проф., Алматы, Казахстан
С. Н. Баранов, д-р физ.-мат. наук, проф., С.-Петербург, РФ
Н. П. Веселкин, академик РАН, д-р мед. наук, проф., С.-Петербург, РФ
В. И. Городецкий, д-р техн. наук, проф., С.-Петербург, РФ
О. Ю. Гусихин, Ph.D., Диаборн, США
В. Делич, д-р техн. наук, проф., Нови-Сад, Сербия
А. Б. Долгий, Dr. Habil., проф., Сент-Этьен, Франция
М. Железны, Ph.D., доцент, Пльзень, Чешская республика
Д. А. Иванов, д-р экон. наук, проф., Берлин, Германия
И. А. Каляев, академик РАН, д-р техн. наук, профессор, Таганрог, РФ
Г. А. Леонов, член-корр. РАН, д-р физ.-мат. наук, проф., С.-Петербург, РФ
К. П. Марков, Ph.D., доцент, Аизу, Япония
Ю. А. Меркурьев, академик Латвийской академии наук, Dr. Habil., проф., Рига, Латвия
Р. В. Мещеряков, д-р техн. наук, профессор, Томск, РФ
Н. А. Молдовян, д-р техн. наук, проф., С.-Петербург, РФ
В. Е. Павловский, д-р физ.-мат. наук, профессор, Москва, РФ
А. А. Петровский, д-р техн. наук, проф., Минск, Беларусь
В. А. Путилов, д-р техн. наук, проф., Апатиты, РФ
В. Х. Пшихопов, д-р техн. наук, профессор, Таганрог, РФ
А. Л. Ронжин (зам. главного редактора), д-р техн. наук, проф., С.-Петербург, РФ
А. И. Рудской, академик РАН, д-р техн. наук, проф., С.-Петербург, РФ
В. С. Стурев, академик Болгарской академии наук, д-р техн. наук, проф., София, Болгария
В. А. Скормин, Ph.D., проф., Бингемптон, США
А. В. Смирнов, д-р техн. наук, проф., С.-Петербург, РФ
Б. Я. Советов, академик РАН, д-р техн. наук, проф., С.-Петербург, РФ
В. А. Сойфер, академик РАН, д-р техн. наук, проф., Самара, РФ
Б. В. Соколов, д-р техн. наук, проф., С.-Петербург, РФ
Л. В. Уткин, д-р техн. наук, проф., С.-Петербург, РФ
А. Л. Фрадков, д-р техн. наук, проф., С.-Петербург, РФ
Л. Б. Шереметов, д-р техн. наук, Мехико, Мексика

Редактор: Е. П. Мирошникова

Технический редактор: А. И. Мотиечко

Переводчик: П. Н. Емелева

Адрес редакции

199178, Санкт-Петербург, 14-я линия, д. 39,
e-mail: publ@iias.spb.su, сайт: <http://www.proceedings.spiiras.nw.ru/>

Журнал индексируется в международной базе данных Scopus

Журнал входит в «Перечень ведущих рецензируемых научных журналов и изданий, в которых должны быть опубликованы основные научные результаты диссертации на соискание ученой степени доктора и кандидата наук»

© Федеральное государственное бюджетное учреждение науки

Санкт-Петербургский институт информатики и автоматизации Российской академии наук, 2017
Разрешается воспроизведение в прессе, а также сообщение в эфир или по кабелю опубликованных в составе печатного периодического издания-журнала «Труды СПИИРАН» статей по текущим экономическим, политическим, социальным и религиозным вопросам с обязательным указанием имени автора статьи и печатного периодического издания-журнала «Труды СПИИРАН»

CONTENTS

Information Security

- I.V. Kotenko, A.A. Kuleshov, I.A. Ushakov
A SYSTEM FOR COLLECTING, STORING AND PROCESSING SECURITY INFORMATION
AND EVENTS BASED ON ELASTIC STACK TOOLS 5
- A.N. Oreshin, I.Yu. Lisanov
A NEW METHOD FOR AUTOMATION OF THE PERSONNEL AUTHENTICATION PROCESS
USING A VIDEO STREAM 35
- E.S. Novikova, I.N. Murenin
THE TECHNIQUE OF THE VISUAL ANALYSIS OF THE ORGANIZATION EMPLOYEES
ROUTES FOR ANOMALY DETECTION 57

Methods of Information Processing and Management

- T.M. Sukhov, S.Yu. Strakhov, L.B. Kochin, E.A. Belyaev
APPLICATION AND IMPLEMENTATIONS OF THE PRINCIPLE OF CONSTANT COLOR
LUMINANCE IN DIGITAL VIDEO CODING SYSTEMS 84
- E.P. Minakov, B.S. Sokolov, S.E. Shaldaev
INVESTIGATION OF THE CHARACTERISTICS OF THE NEAR-MOON SYSTEM FOR HITTING
ASTEROIDS 106
- S.A. Kabanov, A.I. Krivushov, F.V. Mitin
MODELING OF JOINT DEPLOYMENT OF UNITS OF THE LARGE-SIZED TRANSFORMABLE
REFLECTOR OF SPACE BASING 130

Algorithms and Software

- A.V. Ponomarev
QUALITY CONTROL METHODS IN CROWD COMPUTING: LITERATURE REVIEW 152
- S.A. Krasnov, A.S. Ilatovsky, A.D. Khomonenko, V.N. Arseniev
ASSESSMENT OF SEMANTIC SIMILARITY OF DOCUMENTS ON THE BASIS OF THE
LATENT SEMANTIC ANALYSIS WITH THE AUTOMATIC CHOICE OF RANK VALUES 185
- A.A. Voevoda, D.O. Romannikov
SYNTHESIS OF NEURAL NETWORK FOR SOLVING LOGICAL-ARITHMETIC PROBLEMS 205

Theoretical and Applied Mathematics

- N.A. Balonin, M.B. Sergeev, V.S. Suzdal
DYNAMIC GENERATORS OF THE QUASIORTOGONAL HADAMARD MATRIX FAMILY 224
- I.S. Polansky, Yu.S. Pehov
BARYCENTRIC METHOD IN THE SOLUTION OF SINGULAR INTEGRAL EQUATIONS OF
ELECTRODYNAMIC THEORY OF REFLECTOR ANTENNAS 244

СОДЕРЖАНИЕ

Информационная безопасность

И.В. Котенко, А.А. Кулешов, И.А. Ушаков
СИСТЕМА СБОРА, ХРАНЕНИЯ И ОБРАБОТКИ ИНФОРМАЦИИ И СОБЫТИЙ
БЕЗОПАСНОСТИ НА ОСНОВЕ СРЕДСТВ ELASTIC STACK 5

А.Н. Орешин, И.Ю. Лысанов
НОВЫЙ МЕТОД АВТОМАТИЗАЦИИ ПРОЦЕССОВ АУТЕНТИФИКАЦИИ ПЕРСОНАЛА
С ИСПОЛЬЗОВАНИЕМ ВИДЕОПОТОКА 35

Е.С. Новикова, И.Н. Муренин
МЕТОДИКА ВИЗУАЛЬНОГО АНАЛИЗА МАРШРУТОВ СОТРУДНИКОВ
ОРГАНИЗАЦИИ ДЛЯ ОБНАРУЖЕНИЯ АНОМАЛИЙ 57

Методы управления и обработки информации

Т.М. Сухов, С.Ю. Страхов, Л.Б. Кочин, Е.А. Беляев
ПРИМЕНЕНИЕ И РЕАЛИЗАЦИИ ПРИНЦИПА ПОСТОЯННОЙ ЦВЕТОВОЙ ЯРКОСТИ В
ЦИФРОВЫХ СИСТЕМАХ КОДИРОВАНИЯ ВИДЕОИНФОРМАЦИИ 84

Е.П. Минаков, Б.В. Соколов, С.Е. Шалдаев
ИССЛЕДОВАНИЕ ХАРАКТЕРИСТИК И ВАРИАНТОВ ПРИМЕНЕНИЯ ОКОЛОЛУННОЙ
СИСТЕМЫ ПОРАЖЕНИЯ АСТЕРОИДОВ 106

С.А. Кабанов, А.И. Кривушов, Ф.В. Митин
МОДЕЛИРОВАНИЕ СОВМЕСТНОГО РАСКРЫТИЯ ЭЛЕМЕНТОВ
КРУПНОГАБАРИТНОГО ТРАНСФОРМИРУЕМОГО РЕФЛЕКТОРА КОСМИЧЕСКОГО
БАЗИРОВАНИЯ 130

Алгоритмы и программные средства

А.В. Пономарев
МЕТОДЫ ОБЕСПЕЧЕНИЯ КАЧЕСТВА В СИСТЕМАХ КРАУД-ВЫЧИСЛЕНИЙ:
АНАЛИТИЧЕСКИЙ ОБЗОР 152

С.А. Краснов, А.С. Илатовский, А.Д. Хомоненко, В.Н. Арсеньев
ОЦЕНКА СЕМАНТИЧЕСКОЙ БЛИЗОСТИ ДОКУМЕНТОВ НА ОСНОВЕ ЛАТЕНТНО-
СЕМАНТИЧЕСКОГО АНАЛИЗА С АВТОМАТИЧЕСКИМ ВЫБОРОМ РАНГОВЫХ
ЗНАЧЕНИЙ 185

А.А. Воевода, Д.О. Романников
СИНТЕЗ НЕЙРОННОЙ СЕТИ ДЛЯ РЕШЕНИЯ ЛОГИКО-АРИФМЕТИЧЕСКИХ ЗАДАЧ 205

Теоретическая и прикладная математика

Н.А. Балонин, М.Б. Сергеев, В.С. Суздаль
ДИНАМИЧЕСКИЕ ГЕНЕРАТОРЫ КВАЗИОРТОГОНАЛЬНЫХ МАТРИЦ СЕМЕЙСТВА
АДАМАРА 224

И.С. Полянский, Ю.С. Пехов
БАРИЦЕНТРИЧЕСКИЙ МЕТОД В РЕШЕНИИ СИНГУЛЯРНЫХ ИНТЕГРАЛЬНЫХ
УРАВНЕНИЙ ЭЛЕКТРОДИНАМИЧЕСКОЙ ТЕОРИИ ЗЕРКАЛЬНЫХ АНТЕНН 244

И.В. Котенко, А.А. Кулешов, И.А. Ушаков
**СИСТЕМА СБОРА, ХРАНЕНИЯ И ОБРАБОТКИ
ИНФОРМАЦИИ И СОБЫТИЙ БЕЗОПАСНОСТИ НА ОСНОВЕ
СРЕДСТВ ELASTIC STACK**

Котенко И.В., Кулешов А.А., Ушаков И.А. Система сбора, хранения и обработки информации и событий безопасности на основе средств Elastic Stack.

Аннотация. В статье рассматривается подход к построению системы сбора, хранения и обработки информации и событий безопасности на основе средств Elastic Stack. Анализируются задачи мониторинга и управления инцидентами безопасности, исследуются архитектуры систем мониторинга, выявляются требования к ним, и предлагается архитектура системы сбора, хранения и обработки информации и событий безопасности. Описывается разработанный программный прототип системы и представляются результаты экспериментов с разработанным прототипом.

Ключевые слова: мониторинг и управление инцидентами компьютерной безопасности, SIEM-системы, Elastic Stack, Elasticsearch, Logstash, Kibana.

1. Введение. Технологический прогресс не стоит на месте, и системы защиты информации развиваются и эволюционируют вместе с ним. Системы управления информацией и событиями безопасности (Security Information and Event Management, SIEM) не являются исключением. Ранее функционал классического SIEM-решения больших и средних компаний более-менее удовлетворял имеющимся требованиям. Однако в настоящее время необходимы новые механизмы и функции, способные своевременно и адекватно выявлять, обрабатывать и анализировать текущие потоки информации и событий безопасности и управлять инцидентами для гораздо большего количества устройств с учетом существенно возросших объемов информации, в том числе данных о пользователях, трафике, сервисах, событиях и так далее [1-4].

Проблема заключается в том, что современные SIEM-системы в недостаточной степени адаптированы к своевременной обработке больших объемов информации и событий безопасности, необходимой для оценки текущего состояния, управления инцидентами и выработки контрмер. В работе ставится задача разработки архитектуры и реализации исследовательского прототипа системы сбора, хранения и обработки информации и событий безопасности, базирующейся на технологиях больших данных как основы будущей SIEM-системы, а также проведения предварительного анализа параметров функционирования данной системы. Специфика поставленной задачи заключается в необходимости применения технологий больших данных для задач мониторинга и выборе наиболее производительной архитектуры. Предлагаемое в статье решение отличается от

существующих комплексированием программных продуктов Elastic stack, Nginx и Docker для решения задачи сбора, хранения и обработки большого количества данных в целях мониторинга событий информационной безопасности. Данное решение направлено на обеспечение высокой производительности обработки сообщений с учетом возможных перегрузок и дальнейшего расширения системы.

Статья имеет следующую структуру. Во *втором разделе* представляется краткий анализ релевантных работ и наиболее представительных SIEM-продуктов, рассматриваются их достоинства и недостатки. В *третьем разделе* выявляются технические требования к перспективной системе сбора, хранения и обработки информации и событий безопасности, а также на их основании формируется обобщенная архитектура предлагаемого решения, и с учетом характеристик продуктов, имеющихся на рынке, производится выбор средств реализации. В *четвертом разделе* описан предлагаемый подход к реализации системы сбора, хранения и обработки информации и событий безопасности, удовлетворяющий заданным требованиям, описана его реализация. В *пятом разделе* представлены результаты экспериментов, и проведено сравнение разрабатываемого прототипа с несколькими решениями, предложенными в научных публикациях. В *заключении* сделаны выводы и определены направления будущих исследований.

2. Релевантные решения. Выделим два направления обзора релевантных решений — исследовательские работы и программные реализации коммерческих продуктов и продуктов с открытым исходным кодом.

Рассмотрим вначале несколько исследовательских работ. В [5] описывается обобщенная архитектура SIEM-системы нового поколения, которую можно разбить на уровни сети, данных, событий и приложений. Выделяются следующие основные компоненты системы: коллектор, универсальный транслятор событий, высоконадежная шина данных, масштабируемый процессор событий, репозиторий, система принятия решений и реагирования, компонент моделирования атак и анализа защищенности, прогностический анализатор безопасности и система визуализации.

В [6] определены три архитектурных уровня построения SIEM-систем: анализа данных, управления данными и сбора данных. Данный подход дает возможность проанализировать сложность обработки и количество обрабатываемых событий на каждом из уровней. Делается вывод, что самым нагруженным и требовательным к вычислительным мощностям является уровень сбора данных.

Исследования, приведенные в [4], направлены на анализ центрального компонента любой SIEM-системы — системы хранения данных. Представлены преимущества хранения информации в гибридном репозитории. Предложенный подход к хранению обеспечивает удобный и надежный обмен данными между разнородными системами хранения и позволяет реализовать достаточно высокую производительность.

Системы, способные работать с большими данными, интегрируются во все большее количество различных продуктов, и SIEM-продукты не являются исключением [7, 8]. Представителем такой системы является платформа Hadoop [9]. Она работает по принципу пакетной обработки, когда непосредственный анализ не зависит от потока данных. Платформа предоставляет средства обработки структурированных и неструктурированных файлов больших размеров. За данную функциональность отвечает компонент управления ресурсами MapReduce [10]. Основная идея MapReduce заключается в распределении задач с использованием большого количества узлов, организованных в кластер. В [11] представлен результат реализации такого подхода (скорость — до 30 000 МВ обработанных данных в секунду). Такая производительность возможна на основе применения большого количества средств для вычисления, в частности, использовалось 1800 узлов, каждый из которых включал два процессора 2ГГц Intel Xeon и 4Гб оперативной памяти. В указанной работе рассмотрена технология пакетной обработки, однако для обработки данных в потоке используют потоковую обработку, позволяющую отслеживать сообщения в реальном времени. Такую задачу выполняет, например, система Apache Storm [12]. Компания Cisco представила анализ данного подхода [13], способного обрабатывать свыше миллиона пакетов за секунду на одном узле системы.

Компания Gartner [14] выделяет несколько наиболее продвинутых SIEM-систем, в том числе HP ArcSight, IBM Qradar (из коммерческих) и AlienVault OSSIM (с открытым исходным кодом).

SIEM-система HP ArcSight [15] обладает модулями мониторинга событий, поведенческого анализа и системой правил обработки событий безопасности. Из-за закрытого исходного кода затруднено добавление нового функционала. HP Arcsight в качестве СУБД использует собственную разработку CORR. В системе по умолчанию не реализована возможность хранения событий, поступающих без определенного шаблона либо маски, то есть для добавления принципиально новой информации необходимо дополнительное вмешательство в систему. В системе реализовано централизованное хранение данных. Для географически

распределенных организаций необходимо передавать все события в центральную базу данных, где и выполняется вся обработка. У ArcSight ESM ядро системы лицензируется по объему логов в день. Кроме ядра необходимо лицензировать набор различных параметров и опций, например количество пользователей, разработка собственных коннекторов, количество источников событий (считается раздельно по типам источников), модули соответствия требованиям, управление логами и так далее. Из чего следует, что ArcSight ориентируется на крупные корпорации.

Конкурентом ArcSight является IBM SIEM Qradar [16]. Данное средство представляет собой единое унифицированное решение для управления информацией и событиями безопасности, в том числе управления системными журналами, выявления аномалий, управления конфигурациями и устранения уязвимостей. Эта система позволяет выявлять признаки наиболее критичных инцидентов. Qradar использует преимущества единой архитектуры для анализа системных журналов, потоков данных, уязвимостей, данных пользователей и информационных ресурсов. Вместе с тем данный подход приводит к технически сложному процессу увеличения ресурсов, используемых системой. Кроме того, отсутствует программно-аппаратная реализация. Стоимость приобретения SIEM IBM Qradar складывается из многих факторов комплекта поставки и конфигурации самой системы. Цена лицензии зависит также от возможностей по обработке событий в секунду. Высокая цена не позволяет применять решение на малых и средних предприятиях.

OSSIM (Open Source Security Information Management) [17] — это SIEM-система на основе открытого исходного кода от компании AlienVault. OSSIM реализует функции сбора, анализа и корреляция событий, а также обнаружения вторжений. Она включает хостовую систему обнаружения вторжений (HIDS), сетевую систему обнаружения вторжений (NIDS), систему обнаружения вторжений для беспроводных сетей (WIDS), компоненты мониторинга узлов сети, анализа сетевых аномалий, сканер уязвимостей, систему обмена информацией об угрозах между пользователями, набор плагинов для синтаксического анализа и корреляции записей системных журналов с различных внешних устройств и служб. Основным недостатком перечисленных решений является ограниченная функциональность по агрегации получаемых сообщений.

Поскольку SIEM-продукты развивались эволюционно, то с приходом концепции больших данных в сферу информационной безопасности большинство из них были вынуждены внедрить

дополнительный набор методов Hadoop по работе с большим потоком информации от компании Apache Software Foundation. На сегодняшний день ситуация не изменилась, и Hadoop используют как основное решение. Стоит учитывать, что проект Hadoop был разработан с целью построения программной инфраструктуры распределенных вычислений. Выход осенью 2013 года модуля YARN, делающего технологию универсальной для обработки данных, считается большим развитием для проекта, однако, по нашему мнению, является избыточным для решения задач по сбору, индексации и распределению данных.

3. Требования, обобщенная архитектура и выбор средств для реализации системы сбора, хранения и обработки информации событий безопасности. С учетом результатов анализа релевантных работ в настоящей статье ставится цель разработать обобщенную архитектуру и прототип системы сбора, хранения и обработки информации и событий безопасности, базирующуюся на технологиях больших данных, как основу будущей SIEM-системы.

Перечислим базовые требования к системе сбора, хранения и обработки информации и событий безопасности: (1) обеспечение сбора информации от множества различных источников; (2) эластичность системы — обеспечение оптимального (рационального) распределения нагрузки и низкой зависимости показателей функционирования компонентов при изменении отдельных компонентов; (3) возможность эффективной интеграции собственных алгоритмов в подсистему аналитики для последующего развития SIEM-системы; (4) минимизация времени на развертывание и настройку инфраструктуры, сервисов и других системных задач; (5) реализация эффективных механизмов обеспечения отказоустойчивости и высокой доступности на программном уровне; (6) поддержка горизонтального масштабирования — способности разделения системы на отдельные компоненты и разнесения их по отдельным физическим машинам; (7) распространенность средств разработки системы на рынке и активное сообщество пользователей.

Архитектура системы должна быть микросервисной, что обеспечит системе виртуализацию, отказоустойчивость и быстрое горизонтальное масштабирование. Это также позволит автоматизировать и упростить развертывание. Для эффективной интеграции алгоритмов аналитики необходима поддержка развитого API на популярных языках программирования, таких как JAVA, C++ или Python. Не должно быть жесткой настройки прототипа к определенному источнику события, то есть прием информации должен осуществляться по маске или универсальному образцу. Должна быть обеспечена возможность использования нескольких сетевых потоков

или одинаковых компонентов для распределения нагрузки, во избежание перегрузок на стороне сервера.

Основным компонентом архитектуры SIEM-системы является набор средств, осуществляющий функции сбора, хранения, обработки информации и событий безопасности и функции аналитики, что в совокупности представляет систему мониторинга и управления инцидентами [1-4].

Рассмотрим представленную в [5] архитектуру SIEM-системы нового поколения (рисунок 1).

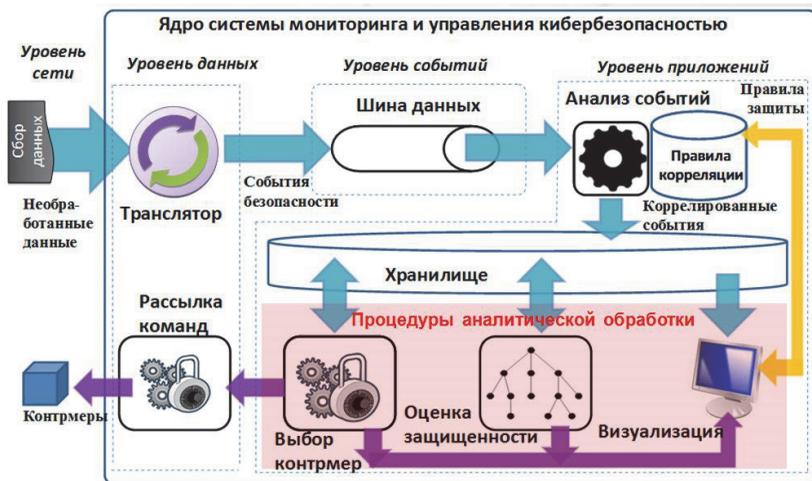


Рис. 1. Архитектура системы мониторинга и управления безопасностью [5]

В данной архитектуре можно выделить четыре уровня: (1) уровень сети, к которому относятся компонент сбора данных с устройств в сети и компонент реализации контромер; (2) уровень данных, к которому относится транслятор, необходимый для надежной передачи сообщений и компонент рассылки команд; (3) уровень событий, представляющий шину данных, служащий для оперативного хранения и передачи поступающей информации; (4) уровень приложений, состоящий из компонента корреляции событий, их анализа, долговременного хранилища и специальных процедур аналитической обработки.

Указанная архитектура является абстрактной и не раскрывает особенностей обработки большого потока входных данных. Предлагаемая обобщенная архитектура разрабатываемой SIEM-системы нового поколения, основанная на технологиях больших данных, представлена на рисунке 2.

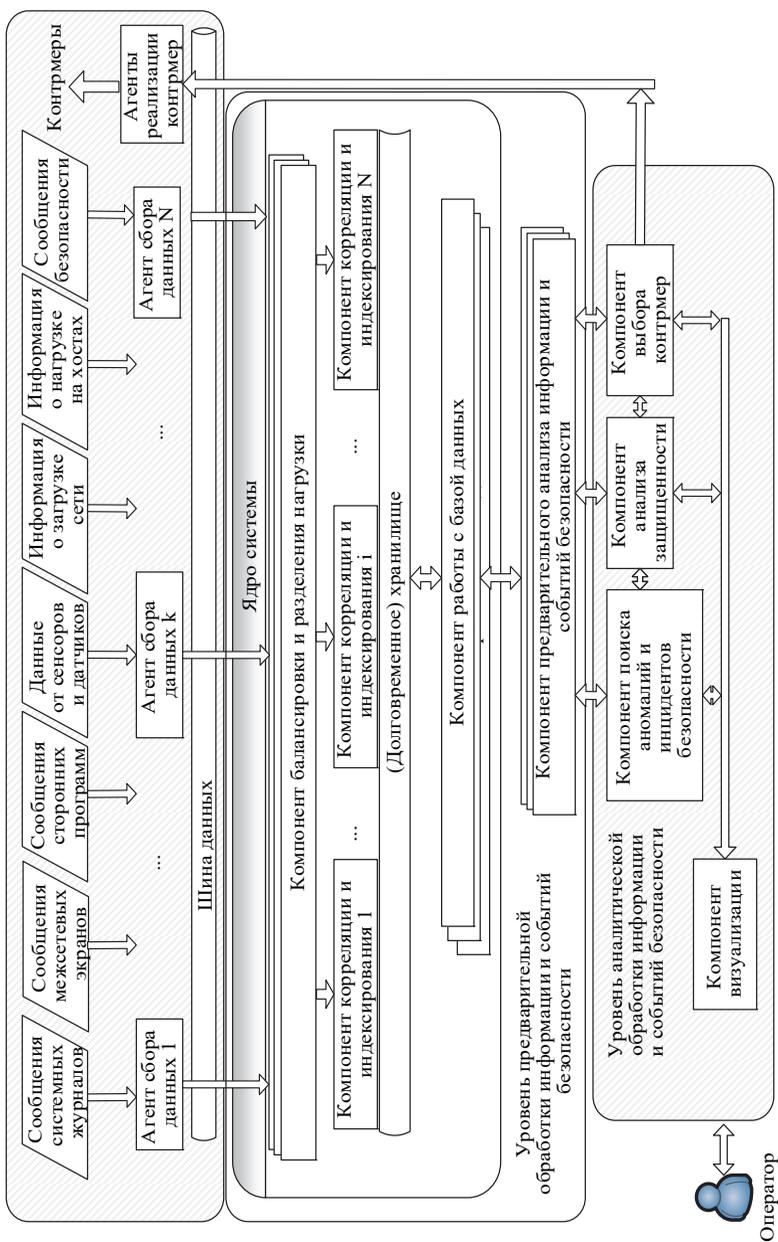


Рис. 2. Обобщенная архитектура SPIEM-системы нового поколения, основанной на технологии больших данных

Архитектура включает три уровня обработки информации: (1) уровень сети и данных (данные от источников, специальные агенты сбора и отправки данных, агенты реализации контрмер); (2) уровень предварительной обработки информации и событий безопасности (компонент балансировки и разделения нагрузки, компоненты корреляции и индексирования, хранилище, компонент работы с базой данных, компонент предварительного анализа информации и событий безопасности); (3) уровень аналитической обработки информации и событий безопасности (компонент поиска аномалий и инцидентов безопасности, компонент анализа защищенности, компонент выбора контрмер, компонент визуализации).

Для обеспечения сбора информации от множества источников различной информации на уровне сети и данных используются агенты сбора данных, обеспечивающие отправку и сбор информации с устройств в сети. Далее полученная информация отправляется на шину данных, предназначенную для распространения информации о событиях безопасности и их гарантированной доставки требуемым компонентам SIEM-системы. Из-за большого количества входных данных возможны перегрузки системы. Для предотвращения и рационального распределения нагрузки в ядре системы весь входной поток информации принимается компонентами балансировки и разделения нагрузки, и корреляции и индексирования, которые находятся на уровне предварительной обработки информации и событий безопасности. После этого данные попадают в хранилище. Для обеспечения быстрого поиска и выборки из хранилища используются компоненты работы с базой данных. Компонент предварительного анализа информации и событий безопасности позволяет собрать необходимую выборку событий и произвести первичный анализ полученной информации для дальнейшего предоставления компонентам уровня аналитической обработки информации и событий безопасности.

Уровень аналитической обработки информации и событий безопасности включает компоненты поиска аномалий и инцидентов, анализа защищенности, выбора контрмер и визуализации. Компонент поиска аномалий и инцидентов отправляет запросы компоненту предварительного анализа информации и событий безопасности на предоставление необходимых данных. Далее происходит оценка защищенности данных компонентом анализа защищенности и в случае обнаружения угроз безопасности данная информация передается компоненту выбора контрмер. Компонент выбора контрмер, в свою очередь, используется для выбора необходимых контрмер и отправляет инструкции агентам реализации контрмер. Оператор взаимодействует с

компонентами аналитической обработки информации и событий безопасности с помощью компонента визуализации.

Представленная архитектура позволяет реализовать комплексный подход к задаче мониторинга инцидентов информационной безопасности. В задачу ядра системы входит сбор и хранение как можно большего объема информации в реальном времени, чтобы впоследствии произвести комплексный анализ предоставленных данных на уровне аналитической обработки. В случае обнаружения инцидентов имеется возможность учесть при анализе ситуации и выработке контрмер изменения системы за длительное время.

Рассмотрим ниже практические решения по реализации предложенной архитектуры. Одним из возможных решений для реализации данной архитектуры является коммерческий программный продукт Splunk Enterprise [18]. Данный продукт — платформа аналитики для сбора и анализа машинных данных с автоматической балансировкой нагрузки. Он обладает возможностью наращивать производительность за счет добавления типовых серверов, что обеспечивает горизонтальную масштабируемость и отказоустойчивость. Splunk имеет документированный RESTful API и SDK для популярных языков программирования. Платформа обрабатывает данные любого формата, включая динамические данные программных приложений, серверов приложений, веб-серверов, операционных систем и многих других источников. Однако Splunk является коммерческим продуктом с закрытым исходным кодом, что ограничивает и замедляет развитие API и делает масштабирование системы платным.

Еще одно коммерческое решение — ManageEngine EventLog Analyzer [19], имеющее схожие достоинства и недостатки. Этот продукт поддерживает распределение нагрузки с использованием внутреннего сервиса. Алгоритмы, методы и основа платформы не доступны на программном уровне, а вся настройка и работа производится с помощью API, который развивается в зависимости от нужд потребителя, что уменьшает гибкость системы.

В 2014 году компания Cisco Systems опубликовала исходные коды системы OpenSOC [20], используемой для создания центра мониторинга киберугроз и основанной на системе Apache Big Data open source. Данные исходные коды стали основой проекта Apache Metron [21]. Система имеет архитектуру, изначально ориентированную на обработку массивов больших данных от множества распределенных источников. На текущий момент проект Apache Metron не является активно развиваемым открытым сообществом — статистика поиска Google и данные GitHub [22] по добавлению кода в проект показывают

слабый интерес со стороны потенциальных пользователей. Развитием проекта сейчас занимается практически единолично компания Hortonworks, предлагающая коммерческое внедрение данного продукта. К сожалению, адаптированная под задачи архитектура Apache Metron имеет плохо реализованную систему обратного масштабирования, и для построения простейшей системы мониторинга необходимо развертывание большого количества компонентов и ресурсоемких сервисов, которые большей частью не будут использоваться. В условиях среднестатистических ресурсов порой достаточно сложно обеспечить загрузку данными от источников, под которую рассчитана данная система. Одновременно с этим для функционирования компонентов системы требуются значительные человеческие ресурсы на развертывание и конфигурирование. Проект Apache Metron заслуживает внимания в случае необходимости решения задач, ориентированных именно на обслуживание высоконагруженных и распределенных информационных систем.

Одним из самых производительных из открытых программных продуктов, специализирующихся на решении обозначенных задач, является Graylog [23]. Это бесплатная система централизованного сбора, хранения и анализа информации, часто используемая в среде DevOps-команд. Эта система использует функции Elasticsearch. Стоит учесть, что, несмотря на частые обновления Graylog и развитое сообщество пользователей, интеграция актуальных версий Elasticsearch в проект требует много времени. Доказательство этому то, что на 2017 год последняя версия Graylog 2.2.1 работает только с Elasticsearch версии 2.4.4 [24], являющейся устаревшей.

В связи с частичной невозможностью удовлетворить всем требованиям, сформулированным выше, в работе был предложен следующий подход к реализации архитектуры системы.

Основываясь на результатах анализа имеющихся в индустрии средств выбран комплекс программного обеспечения Elastic Stack [25] с открытым исходным кодом, так как он подходит для удовлетворения ранее сформированных требований, является бесплатным и популярным решением, сумевшим доказать свою эффективность в большом количестве внедрений по всему миру. Основными программными компонентами являются: Elasticsearch — поисковое и аналитическое ядро системы; Logstash — программный конвейер обработки данных; Kibana — средство визуализации и навигации по системе; Beats — набор программ, необходимых для сбора и транспортировки системных журналов и файлов.

Программный стек, состоящий из Elasticsearch, Logstash и Kibana (далее ELK), специально разработан для решения задач сбора, хранения и обработки системных журналов. Необходимо отметить, что решение на базе программного стека ELK является достаточно зрелым продуктом, и на его базе реализован коммерческое решение Elastic X-Pack, объединяющее вышеперечисленные технологии с добавлением дополнительных функций аналитики.

Рассмотрим основные составляющие Elastic Stack, которые используются при реализации предлагаемой системы мониторинга.

Elasticsearch — распределенное, поисковое и аналитическое ядро системы, поддерживающее REST API и передачу данных через JSON. Elasticsearch централизованно хранит поступающие в него данные, поддерживает кластерную архитектуру, позволяя масштабировать систему, и является прозрачной и надежной системой, обладающая средствами обнаружения сбоев, что, в свою очередь, обеспечивает высокую доступность. Ядро Elasticsearch выполняет в реальном времени поиск по большим объемам разнотипных структур данных — документов. Документ — базовая единица информации, которая может быть проиндексирована. Документы специфицируются в формате JSON. Система имеет развитый API, в список поддерживаемых языков для взаимодействия входят Java, Python, C++ и другие.

Logstash — программный конвейер обработки данных, он одновременно собирает данные из множества различных источников, обрабатывает их и отправляет в подсистему хранения. Logstash имеет встроенный синтаксический анализатор, позволяющий нормализовать разнотипные данные, производить определение географических координат по IP, обрабатывать информацию различных источников независимо от формата и структуры.

Kibana — программный компонент, реализующий функции визуализации и навигации в комплексе Elastic Stack. Kibana представляет данные в виде настраиваемой интерактивной панели индикаторов в реальном времени. Он реализует большое количество встроенных настраиваемых виджетов (гистограммы, графы, карты и другие стандартные инструменты) и имеет развитый API.

Beats — набор программ — коллекторов данных с низкими требованиями к ресурсам, которые устанавливаются на клиентских устройствах для сбора системных журналов и файлов. Имеется широкий выбор коллекторов, а также возможность написать свой коллектор. Filebeat транслируют на сервер информацию из динамических обновляемых журналов системы и файлов, содержащих текстовую информацию. Для аналогичных действий с журналами Windows-систем используется Winlogbeat. Packetbeat — сетевой

анализатор пакетов, который отправляет информацию о сетевой активности между серверами приложений. Он перехватывает сетевой трафик, декодирует протоколы и извлекает необходимые данные. Metricbeat собирает метрики операционных систем, характеризующие, например, использование CPU и памяти, количество переданных пакетов в сети, состояние системы и сервисов, запущенных на сервере.

4. Архитектура и реализация прототипа системы сбора, хранения и анализа информации и событий безопасности. Реализованный прототип имеет архитектуру, представленную на рисунке 3. Данные, которые собираются с помощью Winlogbeat и Metricbeat XML-документом, отправляются по сети Интернет на сервер, где развернут ELK-сервер. Данные принимает экземпляр Logstash, настроенный на сбор данных, который выполняет функцию коллектора. Затем специальный экземпляр Logstash индексирует поступающую информацию с коллекторов и отправляет полученные данные Elasticsearch для последующего сохранения и обработки. Вся аналитическая обработка происходит посредством программного компонента Kibana. Чтобы реализовать функцию просмотра представления Kibana по доменному имени сети Интернет, необходим сервер Nginx, на котором информация с порта вывода Kibana проксируется на 80-ый порт. Полученные команды пользователя с помощью API от Kibana передаются Elasticsearch, где обрабатываются и получают результирующие данные.

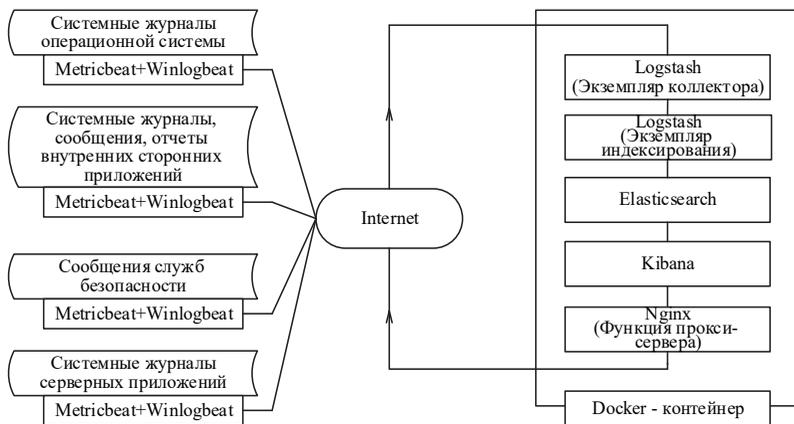


Рис. 3. Архитектура прототипа системы сбора, хранения и анализа информации и событий безопасности

Исходя из описания уровней обобщенной архитектуры, архитектура прототипа соответственно разделяется на уровень сети и данных, предварительной обработки информации и событий

безопасности, а также аналитической обработки. Агентами сбора данных являются Metricbeat и Winlogbeat, а вывод результатов анализа и мониторинга осуществляется с помощью сервера Nginx. Функцию балансировки и разделения нагрузки выполняют экземпляры коллектора и индексирования Logstash. Elasticsearch включает методы поиска и корреляции событий, индексирования перед хранением и хранение входной информации. С помощью открытого API Elastic Stack возможно внедрить скрипты, в том числе по машинному обучению, которые уже на основе полученной выборки будут проводить дополнительные расчеты.

Для полного описания и последующего анализа предлагаемых решений по созданию прототипа системы сбора, хранения и обработки информации и событий безопасности условно разделим прототип на следующие компоненты: (1) отправки данных с клиентских устройств; (2) конвейерной обработки и доставки данных; (3) обеспечения отказоустойчивости и балансировки нагрузки; (4) поискового и аналитического ядра и хранения; (5) визуализации.

4.1. Подсистема отправки данных с клиентских устройств.

Система сбора, хранения и обработки информации и событий безопасности разрабатывается как ядро SIEM-системы, поэтому необходимо обеспечить в прототипе широкий охват доступной информации для анализа. На данный момент поддерживается сбор событий протокола syslog, журналов событий Windows, телеметрии оборудования, ОС и сервисов, а также информации о потоках сетевого трафика из протокола netflow/sflow с помощью filebeat, winlogbeat, metricbeat, packetbeat соответственно. В будущем, при необходимости отправки специфичных данных, планируется написание собственных Beat-коллекторов на основе библиотеки libbeat и развитого API.

4.2. Подсистема конвейерной обработки и доставки данных.

Для обработки и доставки данных в Elastic Stack используется конвейер Logstash, показанный на рисунке 4.

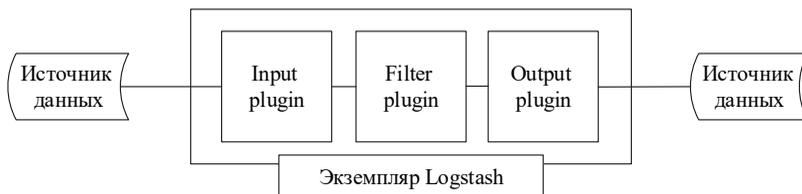


Рис. 4. Схема работы Logstash

Он разделен на три основных функциональных блока, реализованных в виде трех расширений (plugins): input; filter; output.

В функциональном блоке Input plugin указывается конкретный источник событий, который будет считываться Logstash-конвейером. В прототипе этими источниками являются beats. Используются документы JSON-формата, содержащие данные из журналов системы, метрики системы, информацию из протоколов и другие доступные данные в соответствии с выбранным коллектором.

Блок Filter plugin выполняет промежуточную обработку события. Это позволяет структурировать данные, извлекая только необходимую информацию, например, дату, время, IP-адрес, код ошибки и так далее, и, сохраняя их в структуры данных, отправлять далее информацию в output plugin для последующей передачи в Elasticsearch. Используемый фильтр выбирается в зависимости от характеристик события. Следует учитывать, что этот блок является ресурсоемким, в связи с этим Filter plugin активно использует распараллеливание вычислений.

В Output plugin указывается дальнейший маршрут обработки документов в JSON-формате. Это окончательный этап функционирования конвейера. В прототипе данные передаются в подсистему аналитики и хранения — Elasticsearch.

4.3. Механизм обеспечения отказоустойчивости и балансировки нагрузки. Архитектура конвейера обработки и доставки данных в прототипе показана на рисунке 5. При превышении скорости входящих событий над скоростью обработки данных конвейер Logstash начинает отбрасывать события. Для предотвращения потерь в качестве буфера был использован брокер сообщений Redis (возможно также использовать стандартные для индустрии решения, например Redis/Kafka/RabbitMQ).

В реализованном прототипе подсистема конвейерной обработки на базе Logstash была разделена на два отдельных программных сервиса: (1) сервис приема данных от источников событий beats, с последующей отправкой в буфер message broker; (2) сервис приема из буфера, дальнейшей обработки и отправки в Elasticsearch. Использование нескольких экземпляров объектов Logstash с разделением функциональной роли позволяет выполнять балансировку нагрузки между источником данных и кластером Logstash.

Чтобы избежать невозможности ввода данных определенного типа, когда экземпляр Logstash данного типа не доступен, используются специально настроенные конвейеры Logstash, поддерживающие множество модулей input plugin. Например, если бы в подсистеме имелся только один экземпляр объекта Logstash с file input plugin, то при его сбое, исчезла бы возможность принимать данные от Filebeat. Увеличение количества модулей input plugin позволяет масштабировать горизонтально, а разделенные параллельные принимающие конвейеры увеличивают надежность системы и устраняют единую точку отказа.

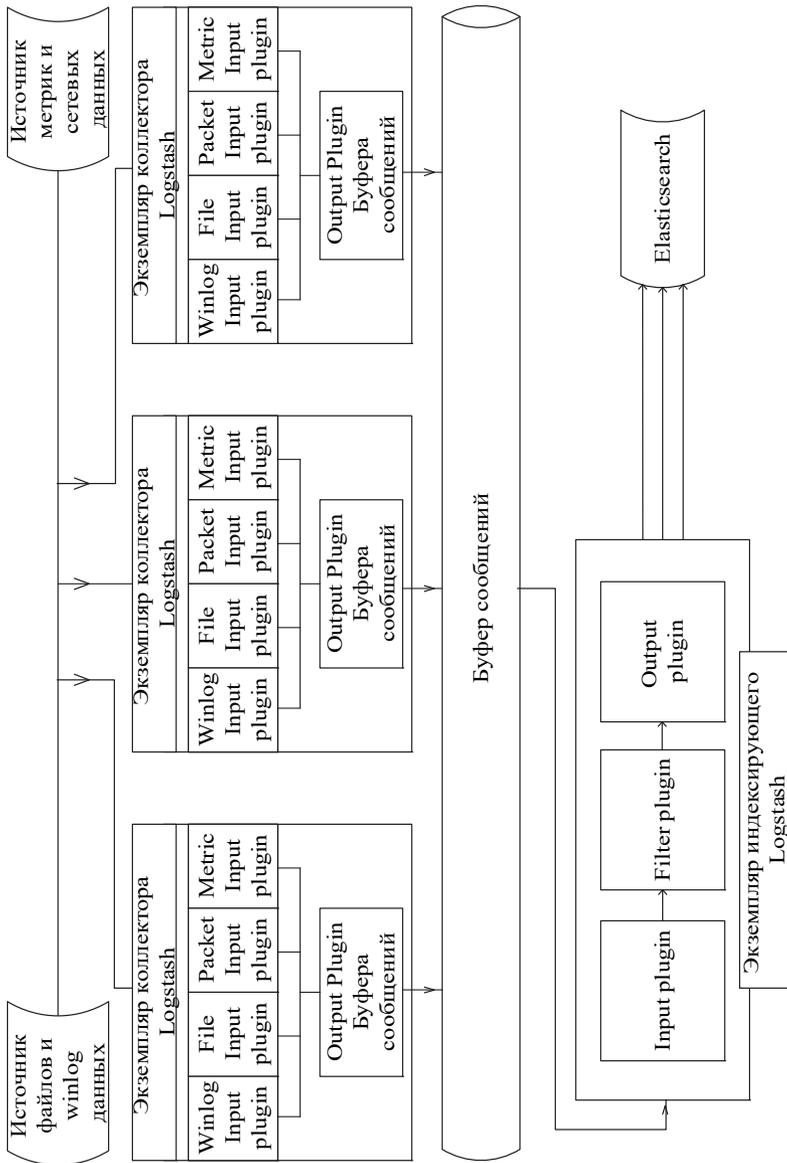


Рис. 5. Архитектура конвейера обработки и доставки данных

Модуль Elasticsearch output plugin также настроен на автоматическую балансировку нагрузки с использованием множества узлов в кластере Elasticsearch. При отказе одного из узлов поток данных не прерывается, что позволяет исключить единую точку отказа. Это обеспечивает высокую доступность кластера и маршрутизацию трафика к активным узлам кластера.

Использование нескольких экземпляров объектов Logstash с разделением функциональной роли позволяет использовать балансировку нагрузки между источником данных и кластером Logstash. Чтобы избежать невозможности ввода данных определенного типа, когда экземпляр Logstash данного типа не доступен, используются специально настроенные конвейеры Logstash, поддерживающие множество модулей input plugin. Например, если бы в подсистеме имелся только один экземпляр объекта Logstash с file input plugin, то при его сбое исчезла бы возможность принимать данные от Filebeat. Увеличение количества модулей input plugin позволяет выполнять горизонтальное масштабирование, а разделенные параллельные принимающие конвейеры увеличивают надежность системы и устраняют единую точку отказа.

4.4. Подсистема поискового и аналитического ядра, совмещенная с подсистемой хранения. В прототипе развернут кластер Elasticsearch. В терминологии Elastic Stack кластер представляет собой набор узлов (серверов), которые хранят всю информацию и предоставляют возможность индексирования и поиска по всем узлам. В терминологии ELK набор структур-документов, имеющих какие-либо похожие характеристики, называется *индексом*. Для удобства индексы разделяют на типы — документы, имеющие общие поля. Потенциально индекс может увеличиваться до больших размеров, превышающих физические возможности узла. Для этого индекс делят на несколько частей называемых *осколками* (shards). Это позволяет распределять данные на несколько узлов, а также распределить и распараллелить операции с осколками, что увеличивает производительность и пропускную способность. Для предотвращения сбоев и обеспечения отказоустойчивости, Elasticsearch позволяет делать копии осколков индекса, которые называются *репликами*. Осколок и его реплика никогда не располагаются на одном узле.

Элементы системы развернуты в контейнерах Docker, что позволяет автоматизировать функции развертывания системы и управления. Для ускорения разработки и тестирования, а также легкой модернизации до следующей версии сервисов в будущем, выбраны микросервисный подход и развертывание в среде контейнерной

виртуализации. Это разделило подсистему на модули, что упрощает обновление отдельных сервисов, а также тестирование их на совместимость друг с другом. В то же время контейнеры решают задачи резервирования инфраструктурных сервисов, позволяют инженерам сфокусироваться на логике решения и абстрагироваться от инфраструктурных проблем.

5. Результаты экспериментов. С целью тестирования предложенных решений был проведен ряд экспериментов.

Стенд для исследований имел следующие характеристики: 16 Гбайт RAM, использующей технологию ddr3; 4-х ядерный процессор частотой 1.9 ГГц; версии Elasticsearch 5.1.1, Kibana 5.1.1, Logstash 5.1.1, Winlogbeat 5.1.1, Nginx 1.10.2, Java(TM) SE Runtime Environment (build 1.8.0_73-b02); операционная система CentOS 7 на ядре Linux 3.10.0-327.36.3.el7.x86_64.

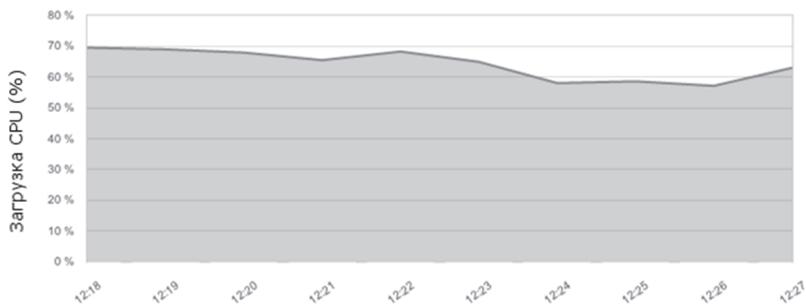
Одним из важных критериев работы системы мониторинга является пропускная способность системы (определяемая как количество обработанных пакетов от устройств за определенный промежуток времени). Когда пакет поступает на порт сервера ELK, Elasticsearch принимает его и начинает индексировать информацию. После индексирования происходит сохранение полученных данных в базе и обновление API. Обновление происходит не сразу, а через интервал времени, прописанный в настройках как `refresh_interval`, который по умолчанию равен 1 сек. На обновление API требуются вычислительные ресурсы, поэтому при загрузке большого количества пакетов, и при допустимости задержки мониторинга, интервал увеличивают, что повышает скорость индексирования. Все эксперименты проводились при `refresh_interval`, равном 1 сек.

Для проведения экспериментов к серверу по сети WLAN с максимальной пропускной способностью 10Мбит/с подсоединялся компьютер, с которого отправлялись данные из заранее подготовленных системных журналов. Во время экспериментов отправлялись данные без дополнительных сведений об источнике в виде одного сообщения. При этом средняя скорость индексирования и обработки составила 13000 пакетов в секунду при средней загрузке процессора 65%. На рисунке 6 показана зависимость загрузки CPU в процентном соотношении от времени. На рисунке 7 представлена зависимость количества обрабатываемых пакетов от времени.

Для того чтобы система могла отслеживать, от какого узла пришло сообщение, какой процесс отправитель, а также время происшествя и так далее, сообщения упаковываются в JSON-пакеты, содержащие дополнительную информацию. Во время экспериментов

отправлялась библиотека JSON-пакетов с дополнительной информацией. В результате размер пакета по сравнению с первым экспериментом увеличился и в среднем стал равен 330 байтам. Это сказалось на скорости обработки, которая в среднем стала равна 4000 пакетов в секунду, а загрузка процессора в среднем составила 40%, что показано на рисунках 8 и 9.

Важной функцией SIEM-системы является аналитическая обработка информации и событий безопасности, для которой критична возможность быстрого извлечения необходимых данных из базы данных. Чтобы проверить время поиска, системе экспериментально были посланы запросы выгрузки проиндексированных и сохраненных сообщений, содержащие слово из 5 букв. Из полученных результатов, представленных на рисунке 10 (где ось Y задает время, мс, а ось X — номер испытания), выявлено, что среднее время поиска сообщений по фразе составляет 1.28 мс.



Время проведения эксперимента (часы:минуты)

Рис. 6. Зависимость загрузки CPU в процентном соотношении от времени при обработке пакетов

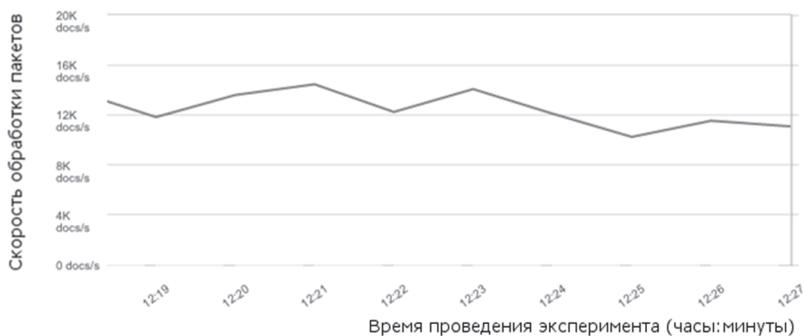


Рис. 7. Зависимость скорости обработки пакетов от времени



Рис. 8. Зависимость загрузки CPU при обработке пакетов с учетом отправки дополнительной информации от времени

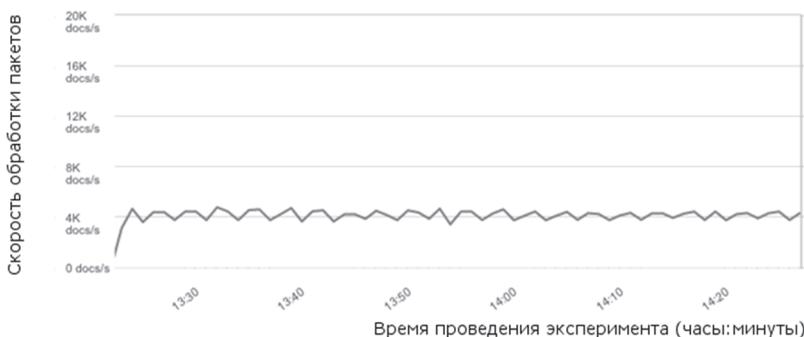


Рис. 9. Зависимость скорости обработки пакетов с учетом отправки дополнительной информации от времени

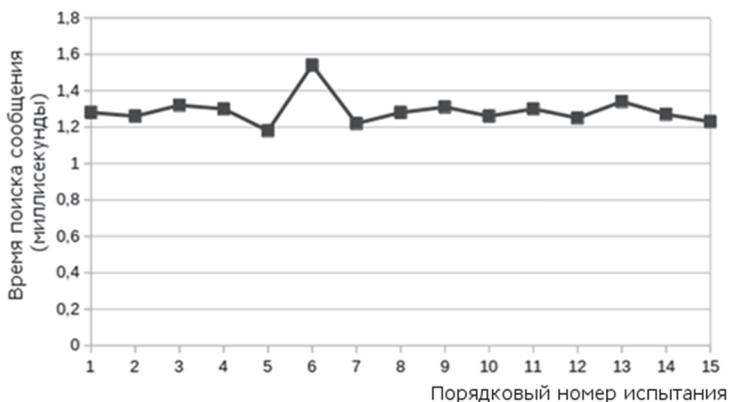


Рис. 10. Зависимость времени поиска сообщений по фразе от номера испытания

Эксперименты демонстрируют, что вычислительной мощности системы хватает для увеличения скорости обработки данных, однако увеличения скорости не происходит. Причина этому — ограничения пропускной способности сети. Однако поставленная цель обработки и представления большого количества пакетов в реальном времени достигнута. С учетом того, что в Elastic Stack есть поддержка протокола netflow, обеспечивается возможность анализа сетевых протоколов в реальном времени.

Рассмотрим несколько решений, предлагаемых в научных публикациях, и сравним их с разрабатываемым прототипом.

Сравнительные характеристики рассмотренных прототипов представлены в таблице 1.

Рассмотрим ниже показатели, использованные для сравнения.

1. *Объем входных данных.* В анализируемых работах [26-30] приведены результаты тестов и экспериментов, в которых предоставлены сведения о тестировании разрабатываемых прототипов с использованием различных объемов данных. В сравнении с этими работами наш прототип использует 1 Гб данных, обрабатываемых в потоковом режиме. Это несколько выше, чем в работе [26], но значительно ниже, чем в остальных работах [27-29], но достаточно для подтверждения работоспособности системы и позволяет сделать вывод об удовлетворении требованиям по обработке и анализу данных в реальном времени.

2. *Количество серверов/узлов,* используемых для построения прототипа. Большинство рассмотренных прототипов [26-29] используют распараллеливание и распределение нагрузки на несколько потоков. Данный параметр определяет способность системы распределять нагрузку между различными узлами сети. В нашем прототипе использовался один физический сервер с представленными в таблице характеристиками, однако различные компоненты системы находились на разных виртуальных машинах.

3. *Время обработки данных.* Данный параметр показывает скорость индексирования и анализа загружаемых в систему данных. Провести сравнительный количественный анализ по данному параметру для работ [26-30] (ввиду отсутствия возможности проведения экспериментов на едином стенде) не представляется возможным. Однако с учетом используемых вычислительных ресурсов и времени обработки данных, полученных в результате исследования, можно предположить, что Elastic Stack является одним из самых производительных решений.

4. *Способ обработки данных.* Как видно из таблицы, большинство исследователей используют потоковую обработку данных [26-29]. Разрабатываемый нами прототип использует потоковую обработку данных, что является ключевым параметром для построения SIEM-системы нового поколения.

5. *Задачи, решаемые с помощью прототипа.* Каждое исследование, рассмотренное в таблице, направлено на решение конкретной задачи. Все представленные решения способны справиться с целью обработки больших потоков данных для оперативного выявления инцидентов информационной безопасности. Разрабатываемый нами прототип призван решить задачу создания распределенного поискового и аналитического ядра системы мониторинга и управления инцидентами, которое выполняет поиск по большому объему разнотипных данных безопасности.

Таблица 1. Сравнительные характеристики рассмотренных решений

Рассматриваемая система	Объем входных данных	Количество серверов/узлов	Время обработки	Способ обработки	Задачи, решаемые с помощью прототипа
Massive Distributed and Parallel Log Analysis For Organizational Security [27]	От 100 Мб до 500 Мб в тестовой выборке	До 8 slave-серверов в эксперименте	Время обработки составило 450 секунд при обработке 500 Мб трафика на 8 серверах. Авторы отмечают, что скорость работы прототипа может быть увеличена, если использовать несколько серверов хранения	Потоковая обработка данных, данные журналов загружаются на один сервер хранения информации	Разработка архитектуры распределенной обработки журналов инцидентов безопасности, которая позволяет анализировать большие массивы данных в облаке
VSS Monitoring. Leveraging a Big Data Model in the Network Monitoring Domain [28]	Более 3.5 экзбайт	Данные не предоставляются	Поддерживается скорость обработки трафика от 100 Мбит/с до 100 Гбит/с	Возможность обрабатывать данные на основе как пакетной, так и поточной обработки. Поддержка Hadoop/Spark	Возможность отделить сетевую аналитику от системы хранения и получения большей эффективности с помощью интеграции различных данных

Продолжение таблицы 1

<p>Toward a Standard Benchmark for Computer Security Research. The Worldwide Intelligence Network Environment (WINE) [29]</p>	<p>Обрабатываемая информация (более 100 Тб) включает: - до 10 миллионов доменов URL-репутаций (30 Терабайт); - 2,5 миллиона e-mail-спам аккаунтов; - данные телеметрии антивирусного ПО с 130 млн. хостов; - репутационные базы с 50 млн. хостов; - примеры вредоносного ПО с ПК 200 стран мира</p>	<p>240 000 сенсоров по всему миру</p>	<p>Данные не предоставлены</p>	<p>Возможность обрабатывать данные на основе как пакетной, так и поточной обработки. Поддержка Hadoop/Spark</p>	<p>Возможность обработки данных, собираемых с миллионов хостов по всему миру с использованием ключевых полей безопасности</p>
<p>Using Large Scale Distributed Computing to Unveil Advanced Persistent Threats [30]</p>	<p>Репрезентативная выборка составила 74 гигабайта данных о 144 миллионах событий</p>	<p>Один физический сервер с 16 ядрами процессора и распределением ресурсов</p>	<p>Время обработки информации - 1500 сек. при 15 ядрах. Авторы делают предположение о возрастании скорости обработки при использовании различных жестких дисков и распределении нагрузки между различными серверами</p>	<p>Пакетная обработка данных</p>	<p>Предполагаемая схема детектирования атак должна интегрировать все инциденты безопасности, собранные организацией с использованием контекста, и на основе различных алгоритмов, детектировать возможную подозрительную активность, применяя MapReduce</p>

Продолжение таблицы 1

RF (Random Forest) [26]	500 Гб	10 устройств, объединенных в кластер, используется 500 деревьев принятия решений	Эксперимент проводился для приложения в области медицины, в качестве входных данных использовалась информация по пациентам. Время обработки информации - 517,8 секунд	Классификация объектов проводится путем голосования, где каждое дерево принятия решений сопоставляет каждый объект определенному классу, в итоге выбирается то соотношение, которое соответствует наибольшему количеству деревьев	Алгоритм машинного обучения, способный обрабатывать данные с большим числом признаков и классов
Spark-MLRF [26]	500 Гб	10 устройств, объединенных в кластер, используется 500 деревьев принятия решений	Эксперимент проводился на примере приложения для медицины. В качестве входных данных использовалась информация о пациентах. Время обработки информации - 186,2 секунд	Основан на библиотеке Apache Spark Mllib	Метод распараллеливания работы алгоритма Random Forest

Продолжение таблицы 1

PRF (Parallel Random Forest) [26]	500 Гб	10 устройств, объединенных в кластер, использует 500 деревьев принятия решений	Эксперимент проводился на примере приложения для медицины. В качестве входных данных использовалась информация о пациентах. Время обработки информации - 101,3 секунд	Разработан на платформе Apache Spark. Для предварительной обработки и машинного обучения строятся деревья регрессии и классификации	Гибридный метод оптимизации распараллеливания работы алгоритма Random Forest
<i>Разрабатываемая система на базе комплексирования средств Elastic Stack</i>	Поток данных при индексировании - 1 Гб. Для обработки и анализа данных использовалась библиотека из 2,5 млн. записей объемом 768 Мб	Один физический сервер: 16 Гигабайт RAM ddr3; 4-х ядерный процессор частотой 1.9 ГГц; пропускная способность передачи данных - 10 Мбит/с	Отправка JSON-пакетов размером 330 байт. Время обработки потока из таких пакетов - 794 мс	Потоковая обработка данных	Распределенное, поисковое и аналитическое ядро должно выполнять поиск по большим объемам разнотипных структур данных

Продемонстрированный в экспериментах и представленный при сравнительном анализе большой потенциал предлагаемой системы при малых затратах вычислительных ресурсов, позволяет получить существенный выигрыш по сравнению с аналогами и дает возможность для проведения дальнейших исследований при больших объемах поступающей информации и событий безопасности.

7. Заключение. В статье представлена архитектура и реализованный прототип системы сбора, хранения и обработки информации и событий безопасности, основанной на технологиях больших данных. Для решения этой задачи проведен анализ релевантных работ и современных SIEM-продуктов, реализующих сбор,

хранение и анализ информации и событий безопасности. На основе этого предложена обобщенная архитектура системы мониторинга, удовлетворяющая представленным требованиям. Выбрано решение Elastic Stack, и на его базе реализован прототип системы для исследовательских целей. Проведены эксперименты, демонстрирующие работоспособность реализованного прототипа системы. Способность реализованного прототипа собирать, хранить, структурировать и анализировать разнородные данные с высокой производительностью, а также программная расширяемость системы предоставляют широкие возможности для дальнейшей разработки перспективной SIEM-системы.

Будущие исследования и разработки будут направлены на дальнейшее совершенствование архитектуры системы, исследование взаимодействия компонентов для обработки информации и событий безопасности, а также анализ, экспериментальную оценку параметров функционирования системы для поиска величин, при которых система сохраняет работоспособность, и выявление зависимости производительности прототипа от увеличения потока данных.

Литература

1. *Котенко И.В., Саенко И.Б.* Создание новых систем мониторинга и управления кибербезопасностью // Вестник Российской академии наук. 2014. Том 84. № 11. С. 993–1001.
2. *Котенко И.В., Саенко И.Б.* SIEM-системы для управления информацией и событиями безопасности // Защита информации. Инсайд. 2012. № 5(47). С. 54–65.
3. *Котенко И.В., Саенко И.Б.* Архитектура системы интеллектуальных сервисов защиты информации в критически важных инфраструктурах // Труды СПИИРАН. 2013. № 1(24). С. 21–40.
4. *Kotenko I., Polubelova O., Saenko I.* Data Repository for Security Information and Event Management in Service Infrastructures // Proceedings of the International Conference on Security and Cryptography (SECRYPT 2012). 2012. pp. 308–313.
5. *Котенко И.В., Саенко И.Б., Юсупов Р.М.* Новое поколение систем мониторинга и управления инцидентами безопасности // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. 2014. № 3(198). С. 7–18.
6. *Котенко И.В., Саенко И.Б.* Построение системы интеллектуальных сервисов для защиты информации в условиях кибернетического противоборства // Труды СПИИРАН. 2012. № 3(22). С. 84–100.
7. Big Data Analytics for Security Intelligence. Cloud Security Alliance. 2013. pp. 1–12.
8. *Zuech R., Khoshgoftaar T.M., Wald R.* Intrusion detection and Big Heterogeneous Data: a Survey // Journal of Big Data. 2015. pp. 1–42.
9. Apache Hadoop 2.7.2. URL: <http://hadoop.apache.org/docs/current/> (дата обращения: 20.02.2017).
10. *Dean J., Ghemawat S.* MapReduce: Simplified Data Processing on Large clusters // Communications of the ACM. 2008. vol. 51. no. 1. pp. 107–113.
11. *Shim K.* MapReduce algorithms for big data analysis // International Workshop on Databases in Networked Information Systems. 2013. LNCS 7813. pp. 44–48.
12. Apache Storm. URL: <http://storm.apache.org/> (дата обращения 20.02.2017).

13. Santos O. Network Security with NetFlow and IPFIX: Big Data Analytics for Information Security // Cisco Press. 2015. 320 p.
14. Kavanagh K.M., Rochford O., Bussa T. Magic Quadrant for SIEM // Gartner. 2016.
15. HPE Security ArcSight ESM. URL: <https://saas.hpe.com/en-us/software/siem-security-information-event-management> (дата обращения: 20.02.2017).
16. IBM Security QRadar SIEM. URL: <http://www-03.ibm.com/software/products/en/qradar-siem> (дата обращения: 20.02.2017).
17. Alienvault OSSIM. URL: <https://www.alienvault.com/products/ossim> (дата обращения: 20.02.2017).
18. Splunk Enterprise. URL: https://www.splunk.com/ru_ru/products/splunk-enterprise.html (дата обращения: 20.02.2017).
19. ManageEngine EventLog Analyzer. URL: <https://www.manageengine.com/products/eventlog/> (дата обращения: 20.02.2017).
20. Cisco Systems OpenSOC. URL: <https://github.com/OpenSOC/> (дата обращения: 20.02.2017).
21. Apache Metron. URL: <http://metron.incubator.apache.org/> (дата обращения: 20.02.2017).
22. GitHub Apache Metron. URL: <https://github.com/apache/incubator-metron/pulls> (дата обращения: 20.02.2017).
23. Graylog. URL: <https://www.graylog.org/> (дата обращения: 20.02.2017).
24. Documentation Graylog. URL: <http://docs.graylog.org/en/2.2/pages/configuration/elasticsearch.html> (дата обращения: 20.02.2017).
25. Elastic Stack. URL: <https://www.elastic.co/> (дата обращения: 20.02.2017).
26. Chen J. et al. A Parallel Random Forest Algorithm for Big Data in a Spark Cloud Computing Environment // IEEE Transactions on Parallel and Distributed Systems. 2017. vol. 28. no. 4. pp. 919–933.
27. Shu X., Smiy J., Yao D., Lin H. Massive Distributed and Parallel Log Analysis For Organizational Security // IEEE Globecom Workshops. December 2013. pp. 194–199.
28. Leveraging a Big Data Model in the Network Monitoring Domain. White Paper. VSS Monitoring. 2014. URL: <http://www.vssmonitoring.com/resources/whitepapers/Leveraging-a-BD-Model-Whitepaper.pdf> (дата обращения: 03.12.2016).
29. Dumitras T., Shou D. Toward a Standard Benchmark for Computer Security Research: the Worldwide Intelligence Network Environment (WINE) // Proceedings of the First Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS'11). 2011. pp. 89–96.
30. Giura P., Wang W. Using Large Scale Distributed Computing to Unveil Advanced Persistent Threats // Science Journal. 2013. vol. 1. no. 3. pp. 93–105.

Котенко Игорь Витальевич — д-р техн. наук, профессор, заведующий лабораторией проблем компьютерной безопасности, Федеральное государственное бюджетное учреждение науки Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН). Область научных интересов: безопасность компьютерных сетей, в том числе управление политиками безопасности, разграничение доступа, аутентификация, анализ защищенности, обнаружение компьютерных атак, межсетевые экраны, защита от вирусов и сетевых червей, анализ и верификация протоколов безопасности и систем защиты информации, защита программного обеспечения от взлома и управление цифровыми правами, технологии моделирования и визуализации для противодействия кибертерроризму. Число научных публикаций — 450. ivkote@comsec.spb.ru, <http://www.comsec.spb.ru>; 14-я линия В.О., 39, Санкт-Петербург, 199178; р.т.: +7-(812)-328-71-81, Факс: +7(812)328-4450.

Кулешов Артем Андреевич — бакалавр, Федеральное государственное бюджетное образовательное учреждение высшего образования «Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича» (СПбГУТ). Область научных интересов: информационная безопасность. Число научных публикаций — 3. gart9515@gmail.com; пр. Большевиков, 22, к. 1, Санкт-Петербург, 193382; р.т.: +79997509515.

Ушаков Игорь Александрович — старший преподаватель кафедры защищенных систем связи, Федеральное государственное бюджетное образовательное учреждение высшего образования «Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича» (СПбГУТ). Область научных интересов: большие данные, информационная безопасность, безопасность компьютерных сетей, технологии виртуализации. Число научных публикаций — 20. ushakovia@gmail.com; пр. Большевиков, 22, к. 1, Санкт-Петербург, 193382; р.т.: +79214106586.

Поддержка исследований. Работа выполнена при финансовой поддержке РФФИ (проект № 15-11-30029).

I.V. KOTENKO, A.A. KULESHOV, I.A. USHAKOV
**A SYSTEM FOR COLLECTING, STORING AND PROCESSING
 SECURITY INFORMATION AND EVENTS BASED ON ELASTIC
 STACK TOOLS**

Kotenko I.V., Kuleshov A.A., Ushakov I.A. A System for Collecting, Storing and Processing Security Information and Events based on Elastic Stack Tools.

Abstract. The paper considers an approach to the design of a system for data and security events collecting, storing and processing based on Elastic Stack tools. The tasks of monitoring and incident management are analyzed; architectural solutions for monitoring systems are studied; requirements to such systems are defined; and the architecture of systems for data and security events collecting, storing and processing is suggested. The developed software prototype of such system is described, and the results of experiments are specified.

Key words: security information and event management, Big Data, SIEM systems, Elastic Stack, Elasticsearch, Logstash, Kibana.

Kotenko Igor Vitalievich — Ph.D., Dr. Sci., professor, head of computer security problems Laboratory, St. Petersburg Institute for Informatics and Automation of Russian Academy of Sciences (SPIIRAS). Research interests: computer network security, including security policy management, access control, authentication, network security analysis, intrusion detection, firewalls, deception systems, malware protection, verification of security systems, digital right management, modeling, simulation and visualization technologies for counteraction to cyber terrorism. The number of publications — 450. ivkote@comsec.spb.ru, <http://www.comsec.spb.ru>; 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone: +7-(812)-328-71-81, Fax: +7(812)328-4450

Kuleshov Artem Andreevich — bachelor, Federal State Budget-Financed Educational Institution of Higher Education, The Bonch-Bruевич Saint - Petersburg State University of Telecommunications (SPbSUT). Research interests: Big Data, information security. The number of publications — 3. gart9515@gmail.com; 22 (b.1) pr. Bolshevnikov, St.Petersburg 193232 Russia; office phone: +79997509515.

Ushakov Igor Aleksandrovich — senior lecturer of secure communications system department, Federal State Budget-Financed Educational Institution of Higher Education, The Bonch-Bruевич Saint - Petersburg State University of Telecommunications (SPbSUT). Research interests: Big Data, information security, computer network security, virtualization technologies. The number of publications — 20. ushakovia@gmail.com; 22 (b.1) pr. Bolshevnikov, St.Petersburg 193232 Russia; office phone: +79214106586.

Acknowledgements. This research is supported by RSF (project № 15-11-30029).

References

1. Kotenko I.V., Saenko I.B. [Creating New Generation Cybersecurity Monitoring and Management Systems]. *Sozdanie novyh system monitoring I upravleniya kiberbezopasnostyu – Herald of the Russian Academy of Sciences*. 2014. vol. 84. no. 6. pp. 993–1001. (In Russ.).
2. Kotenko I.V., Saenko I.B. [SIEM-systems for security information and events management]. *SIEM sistemy dlya upravleniya informaciej I sobyitiami bezopasnosti – Protection of the information. Inside*. 2012. vol. 5. pp. 54–65. (In Russ.).

3. Kotenko I.V., Saenko I.B. [Architecture of the system of intelligent services to protect information in cyber warfare]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2012. vol. 1(24). pp. 21–40. (In Russ.).
4. Kotenko I., Polubelova O., Saenko I. Data Repository for Security Information and Event Management in Service Infrastructures. Proceedings of the International Conference on Security and Cryptography (SECRYPT 2012). 2012. pp. 308–313.
5. Kotenko I.V., Saenko I.B., Yusupov R.M. [New Generation of Security information and Event Management Systems]. *Novoe pokolenie system monitoring I upravleniya incidentami bezopasnosti – St. Petersburg State Polytechnical University Journal. Computer Science. Telecommunication and Control Systems*. 2014. vol. 3(198). pp. 7–18. (In Russ.).
6. Kotenko I.V., Saenko I.B. [Developing the system of intelligent services to protect information in cyber warfare]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2012. vol. 3(22). pp. 84–100. (In Russ.).
7. Big Data Analytics for Security Intelligence. Cloud Security Alliance. 2013. pp. 1–12.
8. Zuech R., Khoshgoftaar T.M., Wald R. Intrusion detection and Big Heterogeneous Data: a Survey. *Journal of Big Data. Springer*. 2015. pp. 1–42.
9. Apache Hadoop 2.7.2. Available at: <http://hadoop.apache.org/docs/current/> (accessed 20.02.2017).
10. Dean J., Ghemawat S. MapReduce: Simplified Data Processing on Large clusters. Google Inc. 2004. pp. 1–13.
11. Shim K. MapReduce algorithms for big data analysis. International Workshop on Databases in Networked Information Systems. 2013. LNCS 7813. pp. 44–48.
12. Apache Storm. Available at: <http://storm.apache.org/> (accessed 20.02.2017).
13. Santos O. Network Security with NetFlow and IPFIX: Big Data Analytics for Information Security. Cisco Press. 2015. 320 p.
14. Kavanagh K.M., Rochford O., Bussa T. Magic Quadrant for SIEM. Gartner. 2016.
15. HPE Security ArcSight ESM. Available at: <https://saas.hpe.com/en-us/software/siem-security-information-event-management> (accessed 20.02.2017).
16. IBM Security QRadar SIEM. Available at: <http://www-03.ibm.com/software/products/en/qradar-siem>. (accessed 20.02.2017).
17. Alienvault OSSIM. Available at: <https://www.alienvault.com/products/ossim> (accessed 20.02.2017).
18. Splunk Enterprise. Available at: https://www.splunk.com/ru_ru/products/splunk-enterprise.html (accessed 20.02.2017).
19. ManageEngine EventLog Analyzer [Official web site of EventLog Analyzer]. Available at: <https://www.manageengine.com/products/eventlog/> (accessed 20.02.2017).
20. Cisco Systems OpenSOC. Available at: <https://github.com/OpenSOC/> (accessed 20.02.2017).
21. Apache Metron. Available at: <http://metron.incubator.apache.org/> (accessed 20.02.2017).
22. GitHub Apache Metron. Available at: <https://github.com/apache/incubator-metron/pulls> (accessed 20.02.2017).
23. Graylog. Available at: <https://www.graylog.org/> (accessed 20.02.2017).
24. Documentation Graylog. Available at: <http://docs.graylog.org/en/2.2/pages/configuration/elasticsearch.html> (accessed 20.02.2017).
25. Elastic Stack. Available at: <https://www.elastic.co/> (accessed 20.02.2017).
26. Chen J. et al. A Parallel Random Forest Algorithm for Big Data in a Spark Cloud Computing Environment. *IEEE Transactions on Parallel and Distributed Systems*. 2017. vol. 28. no. 4. pp. 919–933.

27. Shu X., Smiy J., Yao D., Lin H. Massive Distributed and Parallel Log Analysis For Organizational Security. IEEE Globecom Workshops. 2013. pp. 194–199.
28. Leveraging a Big Data Model in the Network Monitoring Domain. White Paper. VSS Monitoring. 2014. Available at: <http://www.vssmonitoring.com/resources/whitepapers/Leveraging-a-BD-Model-Whitepaper.pdf> (accessed 03.12.2016).
29. Dumitras T., Shou D. Toward a Standard Benchmark for Computer Security Research: the Worldwide Intelligence Network Environment (WINE). Proceedings of the First Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS'11). 2011. pp. 89–96.
30. Giura P., Wang W. Using Large Scale Distributed Computing to Unveil Advanced Persistent Threats. *Science Journal*. 2013. vol. 1. no. 3. pp. 93–105.

А.Н. ОРЕШИН, И.Ю. ЛЫСАНОВ
**НОВЫЙ МЕТОД АВТОМАТИЗАЦИИ ПРОЦЕССОВ
АУТЕНТИФИКАЦИИ ПЕРСОНАЛА С ИСПОЛЬЗОВАНИЕМ
ВИДЕОПОТОКА**

Орешин А.Н., Лысанов И.Ю. **Новый метод автоматизации процессов аутентификации персонала с использованием видеопотока.**

Аннотация. На основе проведенных исследований автоматизированных систем управления промышленных предприятий в работе предложена математическая модель аутентификации персонала с использованием технологии формирования и обработки гетерогенного потока, содержащего с источников видеонаблюдения в трафике видеоданных фотографические идентификационные признаки персонала. Отличительной особенностью модели являются операторы переходов и выходов, направленные на формирование сигнала управления по результатам обработки агрегированного потока видеоданных, выделении кадров-вставок, несущих фотографические идентификационные признаки о персонале, сегментации пространства изображения, необходимого для аутентификации персонала в интегрированной системе контроля и управления доступом, а также применение технологий по восстановлению видеопотока. Разработана методика идентификации кадров-вставок в гетерогенном потоке видеоданных, базирующаяся на оценке суммарной дифференциальной экстремальной яркости кадров-сверток, анализе гистограмм изображений кадров по коэффициентам корреляции и полиномам Чебышевской аппроксимации, отличающаяся использованием динамического порога при идентификации кадров и принятия решения об их классе на основе мажоритарного способа обработки данных. Представлены результаты имитационного моделирования процесса аутентификации персонала на основе предложенных модели и методик, позволяющие оценить эффективность визуальной аутентификации персонала при выполнении требований достоверности принятия решения и ограничения пропускной способности сети передачи данных.

Ключевые слова: система контроля и управления доступом, система видеонаблюдения, модель, гетерогенный и агрегированный потоки данных, идентификационный признак (биометрический, кодовый и фотографический), кадр-вставка, аутентификация персонала.

1. Введение. Процедура проверки подлинности чего-либо (аутентификация), включая персонал предприятия, разносторонняя — от оценки введенного пароля или цифровой подписи по открытому ключу проверяемой личности до оценки ее органолептических признаков, биометрических показателей или географического местоположения. При этом нельзя, по-видимому, исключать не только варианты использования расширенной (многофакторной) взаимной аутентификации, сформированной на совместном применении нескольких факторов, но и варианты аутентификации совместно с оценкой возможностей использования субъектом предоставленных ему определенных прав (как сопутствующий процесс авторизации) или совместно с распознаванием субъекта по представленным им идентификационным (то есть совместно с дополнительной идентификацией) [1].

Несмотря на такие возможности аутентификации и обобщенность идеи организации взаимной проверки подлинности, в системах автоматизации промышленных предприятий выделяют компоненты контроля и управления доступом персонала с односторонней аутентификацией, когда только субъект априорно характеризуется некоторой степенью недоверия и требуется подтверждение санкционированного общения (доступа). Как правило, при создании таких подсистем контроля и управления доступом ориентируются на нормативные положения ГОСТ Р ИСО/МЭК [2], ведущая концепция которого, вообще говоря, базируется на двух видах аутентификации — простой, как проверки паролем заявленной идентичности, и строгой, как проверки удостоверения личности.

2. Актуальность решаемой технической задачи. В реальности организация санкционированного доступа сводится к разделам ввода идентификационных признаков; отображению оператору видео и фотографических данных субъекта доступа; визуальной идентификации, аутентификации персонала; сбору, хранению и обработке событий. Преимущество централизованного управления и детализация событий на объектах инфраструктуры промышленного предприятия приводят к возрастанию числа точек доступа (ТД), и, как правило, существенному росту интенсивности их использования при потребности соответствующего сокращения времени на аутентификацию персонала. В условиях ограниченного времени эти факторы принципиально сужают возможности оператора по исполнению функций с требуемыми показателями качества.

Известные разработки систем контроля и управления доступом [3, 4, 5] ведущих производителей PERCo (РФ), Parsec (РФ), Gate (РФ), Anviz Biometric Inc (США), APOLLO (США), Honeywell (США), Bosch Security Systems (Германия) и SYNEL (Израиль) не позволяют обеспечить решение вопросов визуальной аутентификации персонала оператором уже при количестве точек доступа, большим 8-12. Результаты исследований PSDB (Великобритания) показали, что восприятие человеком изображения на мониторах ухудшается с увеличением их количества (с 85% для одного, до 53% для девяти).

Кроме того, известный методический материал отечественных и зарубежных ученых В.А. Ворона, Р.Г. Магауенова, В.А. Тихонова, Г. Кругля, М. Гарсиа, В. Дамьяновски [4, 7] не касается вопросов централизованного контроля и управления доступом по результатам идентификации и аутентификации персонала с фотографических данных, а в большей степени затрагивает структуру системы видеонаблюдения совместно с техническими (реакция на код, электронный ключ, пластиковая карта, радиочастотная метка и т. п.) или биометрически-

ми (считыватель отпечатков пальцев, рисунка сетчатки глаза, геометрии ладони и лица и т. п.) идентификаторами на локальных участках точек доступа [7-10].

С другой стороны, применение систем видеонаблюдения позволяет оператору осуществлять контроль территориально удаленных объектов инфраструктуры, но не дает детального представления об их состояниях, о санкционированном пребывании и численности персонала в интересующей среде обитания в реальном масштабе времени. Причем влияние «человеческого фактора» (эмоциональная напряженность, снижение внимания, случайное отвлечение, несанкционированное прерывание деятельности) существенно уменьшает адекватность и оперативность принятия решений.

Каждый из представленных подходов (идентификация по фото или видео) анализирует ограниченный набор идентификационных признаков под определенным углом зрения. Совместное (гибридное) их использование генерирует новый источник информации, характеризующийся многофакторной оценкой идентификационных признаков, необходимый для аутентификации персонала. Поэтому разработка метода аутентификации персонала, используя видеопоток для передачи идентификационных признаков (в частности фотографических), не только актуальна, но и научно необходима и технически целесообразна.

3. Математическая модель аутентификации персонала в интегрированной системе контроля и управления доступом. Представим математическую модель [11] аутентификации персонала в виде множества (1):

$$Q_{СКВД} = \{T, S_{ПД}(t), X_{ПД}(t), Y_{ПД}(t), K_{ВП}(t), K_{ПКВ}(t), K_{ГП}(t), Q_{СПД}, Q_{ЦПП}, F_1, F_2, F_3, F_4\}, \quad (1)$$

элементами которого являются подмножества, характеризующие в момент времени $t \in T$: состояния пространства доступа (ПД) $S_{ПД}(t)$; входные воздействия на пространства доступа $X_{ПД}(t)$; реакции пространства доступа $Y_{ПД}(t)$; видеопоток $K_{ВП}(t)$; поток кадров-вставок $K_{ПКВ}(t)$; гетерогенный поток $K_{ГП}(t)$, а также подмножества, описывающие структуру сети передачи данных $Q_{СПД}$ [12] и цветовое пространство пикселя $Q_{ЦПП}$ модели *RGB* [13, 14]; механизм изменения состояний пространства доступа по входным воздействиям (F_1) и формирования выходного сигнала как реакцию пространства доступа на внутренние и внешние воздействия (F_2), а также

операторы формирования гетерогенного потока (F_3) и выделения кадров-вставок и восстановления видеопотока (F_4).

$$F_1: T \times X_{ПД}(t) \times S_{ПД} \rightarrow S_{ПД}; \quad (2)$$

$$F_2: T \times X_{ПД}(t) \times S_{ПД} \rightarrow Y_{ПД}; \quad (3)$$

$$F_3: K_{ВП}(t), K_{ПКВ}(t) \rightarrow K_{ГП}(t) = K_{ВП}(t) \cup K_{ПКВ}(t); \quad (4)$$

$$F_4: K_{ГП}(t) \rightarrow K_{ВП}(t), K_{ПКВ}(t). \quad (5)$$

Требуется реализовать операторы формирования гетерогенного потока (F_3), а также выделения кадров-вставок $K_{вст_k} \in K_{ПКВ}(t) = K_{вст_1}, K_{вст_2}, \dots, K_{вст_k}, \dots$ и восстановления видеопотока $K_{ВП}(t) = K_{вид_1}, K_{вид_2}, \dots, K_{вид_k}, \dots$ (F_4), обеспечивающие автоматизацию процесса аутентификации персонала и управления доступом на максимальном количестве точек доступа (ТД):

$$\sum_{m=1}^{N^{3Д}} \sum_{n=1}^{N^{ТД}} N_{m,n}^{ТД} \Rightarrow \max, \quad (6)$$

где $N^{3Д}$ — количество зон доступа в пространстве доступа; $N^{ТД}$ — количество точек доступа в n -ой зоне доступа; $N_{m,n}^{ТД}$ — точка доступа в m -ом пространстве доступа, n -ой зоны доступа, при ограничениях:

$$\forall K_{ГП}^n(t) \in K_{ГП}^{ПД}(t) \Rightarrow K_{ГП}^n = K_{ВП}^n, \quad (7)$$

$$\forall K_{ГП}^n(t) \in K_{ГП}^{ПД}(t) \Rightarrow K_{ПКВ}^n \leq 3, \quad (8)$$

$$Q_{СПД} = \text{const}, q_{СМО} \geq q_{СМО}^{mp}, \quad (9)$$

где $K_{ГП}^n$ — число кадров в одну секунду в гетерогенном потоке; $K_{ВП}^n$ — число кадров в одну секунду в видеопотоке; $K_{ПКВ}^n$ — число подряд идущих в гетерогенном потоке кадров-вставок; $Q_{СПД}$ — структура сети передачи данных [15]; $q_{СМО}$ — относительная пропускная способность системы контроля и управления доступом с визуальной аутентификацией персонала, рассматриваемой как система массового

обслуживания [16]; $K_{ГП}^{ПД}(t)$ — множество гетерогенных потоков в пространстве доступа.

Данная модель базируется на технологии формирования и обработки гетерогенного потока $K_{ГП}(t) = K_{пот1}, K_{пот2}, \dots, K_{потk}, \dots = K_{ВП}(t) \cup K_{ПКВ}(t)$, содержащего видеокадры $K_{видk'} \in K_{ВП}(t) = K_{вид1}, K_{вид2}, \dots, K_{видk'}, \dots$ и кадры-вставки $K_{вст} \in K_{ПКВ}(t) = K_{вст1}, K_{вст2}, \dots, K_{встk'}, \dots$

Математическая модель основывается на передаче кадров-вставок (растровых изображений), несущих фотографические данные о субъекте доступа, при котором k -ый кадр гетерогенного потока представляется в виде множества пикселей $\pi_{i,j}^k$ на плоскости изображения и математически описывается матрицей $\kappa_{nomk} = \left\| \pi_{i,j}^k \right\|, i = \overline{1, I}, j = \overline{1, J}$, где k — номер кадра, i и j — номера столбца и строки матрицы [17, 18, 19]. Изображение кадра гетерогенного потока представим в виде множества цветowych точек $c_{i,j}^k$, соответствующих каждому пикселю $\pi_{i,j}^k$ модели RGB : $I_k = \left\| c_{i,j}^k \right\|, i = \overline{1, I}, j = \overline{1, J}$.

В модели аутентификации персонала введены следующие обозначения (рисунок 1):

- К — контроллер;
- БФПВК — блок формирования потока видеокладов;
- БФПКВ — блок формирования потока кадров-вставок;
- БФГП — блок формирования гетерогенного потока;
- ОД — объект доступа;
- БАОД — блок аутентификации объектов доступа;
- БИКВ — блок идентификации кадров-вставок;
- БИОДБП — блок идентификации ОД по биометрическому

признаку;

- БИОДКП — блок идентификации ОД по кодовому признаку;
- входные воздействия на точки доступа (ТД)

$\vec{x} = (x_{УПУ}, x_{ГБР}, x_{БУ})$, $\vec{x} \in X(t)$, элементами которого являются воздействия на управляемое преграждающее устройство (УПУ) $x_{УПУ}$, группу быстрого реагирования (ГБР) $x_{ГБР}$ и блокирующее устройство (БУ) $x_{БУ}$;

– реакция точки доступа $\vec{y} = (y_{ОД}, y_{УПУ}, y_{ГБР}, y_{БУ}, y_{СС})$, $\vec{y} \in Y(t)$, элементами которого являются величины характеризующие состояния ОД $s_{ОД}$, УПУ $s_{УПУ}$, ГБР $s_{ГБР}$, БУ $s_{БУ}$ и системы сигнализации (СС) $s_{СС}$.

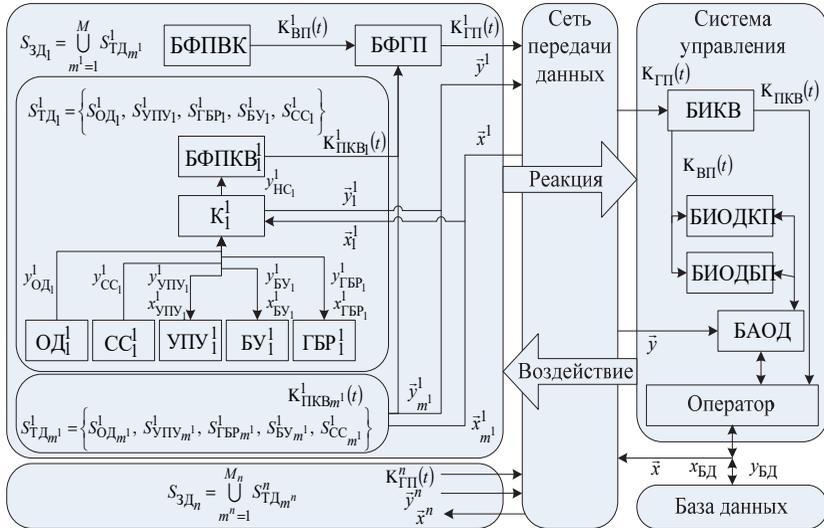


Рис. 1. Модель аутентификации персонала в интегрированной системе контроля и управления доступом

При этом состояние точки доступа $\vec{s}_{ТД} = (s_{ОД}, s_{УПУ}, s_{ГБР}, s_{БУ}, s_{СС}) \in S_{ТД}$ характеризуется в произвольный момент времени $t \in T$ набором переменных, изменяющихся под влиянием воздействий \vec{x} , формируемых контроллером. Состояние зоны доступа $S_{ЗД}$ определяется входящими в нее ТД. Взаимодействие с базой данных (БД) определяется воздействиями $x_{БД}$ и реакцией $y_{БД}$.

Множество состояний ТД описывается выражением $S_{ТД} = \{S_{ОД}, S_{УПУ}, S_{ГБР}, S_{БУ}, S_{СС}\}$, где:

– $S_{ОД} = \{S_{НОД}, S_{КП}, S_{БП}\}$ — множество состояний объекта доступа;

- $S_{\text{НОД}} = \{S_{\text{НОД}_1}, S_{\text{НОД}_2}\}$ — множество состояний ОД по его нахождению в ТД;
- $S_{\text{БП}} = \{S_{\text{БП}_1}, S_{\text{БП}_2}\}$ — множество состояний ОД по наличию у него биометрического признака;
- $S_{\text{КП}} = \{S_{\text{КП}_1}, S_{\text{КП}_2}\}$ — множество состояний объекта доступа по наличию у него кодового идентификатора;
- $S_{\text{БУ}} = \{S_{\text{БУ}_1}, S_{\text{БУ}_2}\}$ — множество состояний блокирующих устройств;
- $S_{\text{СС}} = \{S_{\text{ДСС}_i}\}$ — множество состояний системы сигнализации, характеризующееся состояниями датчиков;
- $S_{\text{УПУ}} = \{S_{\text{УПУ}_1}, S_{\text{УПУ}_2}\}$ — множество состояний УПУ;
- $S_{\text{ГБР}} = \{S_{\text{ГБР}_1}, S_{\text{ГБР}_2}\}$ — множество состояний ГБР;

Реакция на нестандартное событие $u_{\text{НС}}$ формируется в случаях, когда:

- идентификационный признак не соответствует субъекту доступа (отказ в допуске);
- необходимо осуществить допуск в ручном режиме;
- время пребывания субъекта доступа в ТД выходит за рамки установленного;
- аварии на линии между контролером и сервером или влияние на нее злоумышленника с целью скрытия реального события на ТД (при наличии связи между сервером и контроллером события автоматически транслируются на сервер системы контроля и управления доступом персонала [3, 4, 5]);
- срабатывают датчики систем сигнализации (пожарной, тревожной, охранной и т. д.).

Предложенная математическая модель автоматизации процессов аутентификации персонала позволяет определить стратегию развития систем контроля и управления доступом (СКУД) на основе технологии формирования и обработки гетерогенного потока при выполнении требований достоверности и адекватности принятия решения, а также ограничении пропускной способности сети передачи данных.

4. Методика идентификации кадров-вставок в гетерогенном потоке видеоданных, сегментации пространства изображения и восстановления видеопотока. В алгоритмическом виде методика представлена на рисунке 2.

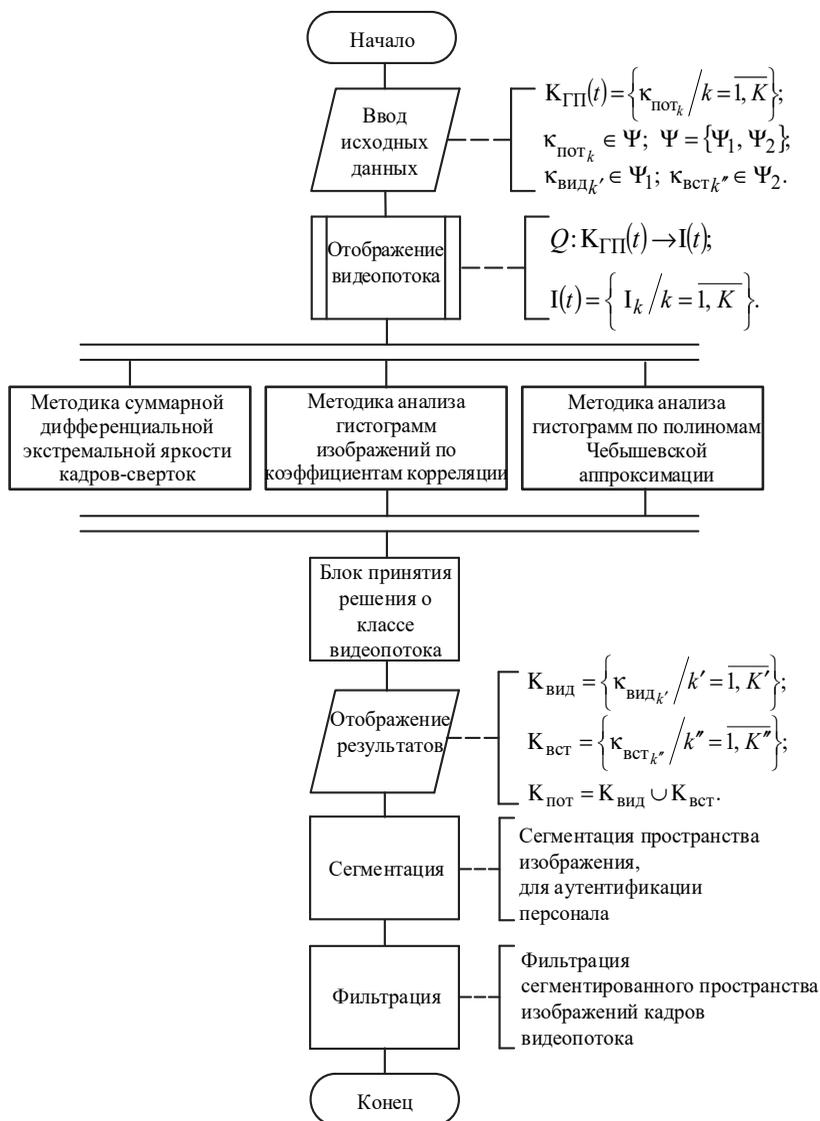


Рис. 2. Алгоритм идентификации кадров-вставок в гетерогенном потоке видеоданных, сегментации пространства изображения и восстановления видеопотока

Суть методики состоит в следующем:

1. На первом этапе осуществляется ввод гетерогенного потока видеоданных $K_{ГП}(t) = \kappa_{пот1}, \kappa_{пот2}, \dots, \kappa_{потk}, \dots$ — множество кадров в потоке видеоданных [20];

$\Psi_1, \Psi_2 \in \Psi$ — множество классов кадров в гетерогенном потоке;

Ψ_1 — класс кадров видеопотока $\kappa_{видk'} \in \Psi_1$;

Ψ_2 — класс кадров-вставок $\kappa_{встk''} \in \Psi_2$, несущих фотографические данные о персонале.

2. Второй этап заключается в отображении кадра гетерогенного потока в растровое изображение $Q: K_{ГП}(t) \rightarrow I(t)$, где

$$I(t) = \left\{ I_k / k = \overline{1, K} \right\}.$$

3. На третьем этапе формируется наблюдаемое окно $I_{окн} = \left\{ I_k / k = \overline{1, 8} \right\}$.

4. Четвертый этап сводится к идентификации кадра-вставки тремя подходами с разными порогами чувствительности на основе:

Оценки суммарной дифференциальной экстремальной яркости кадров-сверток (попиксельной разности изображений кадров гетерогенного потока) [21-23].

Оценка класса кадра $\Psi_1, \Psi_2 \in \Psi$ (рисунок 3) осуществляется способом сравнения экстремальной яркостной составляющей кадра-свертки $C_{k,k+1} \in C_{сверт}$ [13], где $k = \overline{4, 7}$ с динамически изменяющимся пороговым значением $C_{пор1}$ (10):

$$C_{пор1} = \frac{\sum_{k=1}^3 C_{k,k+1}}{3} + \Delta_1, \quad (10)$$

где Δ_1 — абсолютная погрешность экстремальной яркостной составляющей, определяемая по результатам статистических данных о видеопотоке; $C_{k,k+1} \in C_{сверт}$ — попиксельная разность по $r_{ij}^k, g_{ij}^k, b_{ij}^k$ составляющим цвета c_{ij}^k пикселя $\pi_{i,j}^k$ модели *RGB*:

$$C_{k,k+1} = \sum_{i=1}^I \sum_{j=1}^J \left| r_{i,j}^k - r_{i,j}^{k+1} \right| + \sum_{i=1}^I \sum_{j=1}^J \left| g_{i,j}^k - g_{i,j}^{k+1} \right| + \sum_{i=1}^I \sum_{j=1}^J \left| b_{i,j}^k - b_{i,j}^{k+1} \right|. \quad (11)$$

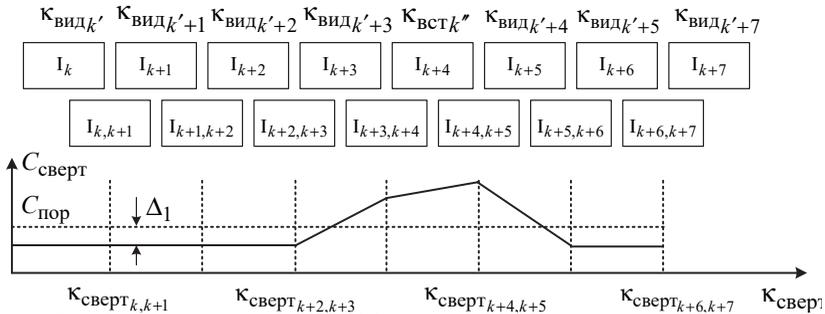


Рис. 3. Оценка суммарной дифференциальной экстремальной яркости изображений кадров-сверток

В случае, когда вторая граница $C_{k,k+1} \in C_{\text{сверт}}$, превышающая $C_{\text{пор}}$, выходит за границы исследуемого окна — регистрируется следующий сюжет, то есть $K_{\text{пот}k'} \in \Psi_1$.

Анализа гистограмм изображений кадров по коэффициентам корреляции [24].

Для анализа изображений кадров $K_{\text{пот}k}$ исследуемого окна используется закон распределения $p^k(\xi)$ величин r_{ij}^k , g_{ij}^k , b_{ij}^k цвета c_{ij}^k пикселя $\pi_{i,j}^k$, устанавливающий связь между значениями интенсивностями $\xi_{i,\varepsilon}^k$ цветов пикселя k -ого кадра и соответствующими им статистическими вероятностями $p^k\left(\xi_{i,\varepsilon}^k\right)$.

Решения о классе кадра $\Psi_1, \Psi_2 \in \Psi$ (рисунок 4) принимается по результатам сравнения оценок коэффициентов корреляции кадров декодированного гетерогенного потока видеоданных $\tilde{K}_{k,k+1}(\xi)$ со средними значениями оценок коэффициентов корреляции $\tilde{K}_{\text{средн}k}(\xi)$ [20, 24] по критерию (12):

$$\tilde{K}_{\text{средн}k+h}(\xi) > \tilde{K}_{k+h,k+h+1}(\xi) + \Delta_2, h = \overline{0,6}, \quad (12)$$

где:

$$\tilde{K}_{k,k+1}(\xi) = \frac{\sum_{i^\varepsilon=1}^{256} \left(\xi_{i^\varepsilon}^k - \tilde{m}_{p^k(\xi)} \right) \left(\xi_{i^\varepsilon}^{k+1} - \tilde{m}_{p^{k+1}(\xi)} \right)}{255 \times \tilde{\sigma}_{p^k(\xi)} \tilde{\sigma}_{p^{k+1}(\xi)}}, \quad (13)$$

$$\tilde{K}_{\text{средн}k}(\xi) = \frac{\tilde{K}_{k,k+1}(\xi) + \tilde{K}_{k+1,k+2}(\xi) + \tilde{K}_{k,k+3}(\xi)}{3}, \quad (14)$$

оценки математического ожидания интенсивности изображения кадра и среднеквадратичного отклонения:

$$\tilde{m}_{p^k(\xi)} = \sum_{i^{\xi}=1}^{256} p^k(\xi_{i^{\xi}}^k) \times \xi_{i^{\xi}}^k, \quad (15)$$

$$\tilde{\sigma}_{p^k(\xi)} = \sqrt{\tilde{D}_{p^k(\xi)}} = \sqrt{\frac{\sum_{i^{\xi}=1}^{256} \left(\xi_{i^{\xi}}^k - \tilde{m}_{p^k(\xi)} \right)^2}{255}}, \quad (16)$$

Δ_2 — абсолютная погрешность, определяемая эмпирически.

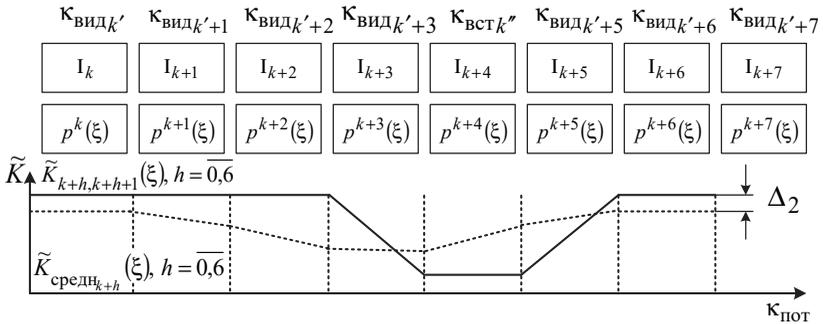


Рис. 4. Анализ гистограмм изображений кадров по коэффициентам корреляции

Анализа гистограмм изображений кадров по полиномам Чебышевской аппроксимации [20, 25].

Принимается решение о классе кадров входного потока (рисунок 5) по результатам поэлементного сравнения варьируемых параметров A_n и B_m дробей Чебышева $f(\xi)$ всех 24 гистограмм

$p^{k^o}(\xi)$, $k^o = \overline{1, 8}$, $\xi \in (r, g, b)$, множеств разностей матриц

$H_{k+h,k+h+1}$, $h = \overline{3, 6}$ в наблюдаемом окне с пороговым значением (17):

$$\sup \left\{ H_{k+h,k+h+1} / h = \overline{0, 2} \right\} + \Delta_3, \quad (17)$$

где:

$$f(\xi) = \frac{A_0 \xi^N + A_1 \xi^{N-1} + \dots + A_N}{\xi^M + B_1 \xi^{M-1} + \dots + B_M}, \quad (18)$$

$A_n, n = \overline{1, N}$ — параметры числителя дроби, а $B_m, m = \overline{1, M}$ — параметры знаменателя дроби,

$$\mathbf{H}_{k,k+1} = \left| \mathbf{H}_k - \mathbf{H}_{k+1} \right| = \begin{vmatrix} A_{n_0}^{k,k+1} & A_{n_1}^{k,k+1} & \dots & A_{n_N}^{k,k+1} & B_{n_0}^{k,k+1} & B_{n_1}^{k,k+1} & \dots & B_{n_M}^{k,k+1} \\ A_{g_0}^{k,k+1} & A_{g_1}^{k,k+1} & \dots & A_{g_N}^{k,k+1} & B_{g_0}^{k,k+1} & B_{g_1}^{k,k+1} & \dots & B_{g_M}^{k,k+1} \\ A_{b_0}^{k,k+1} & A_{b_1}^{k,k+1} & \dots & A_{b_N}^{k,k+1} & B_{b_0}^{k,k+1} & B_{b_1}^{k,k+1} & \dots & B_{b_M}^{k,k+1} \end{vmatrix}, \quad (19)$$

Δ_3 — абсолютная погрешность, определяемая по результатам экспериментов.

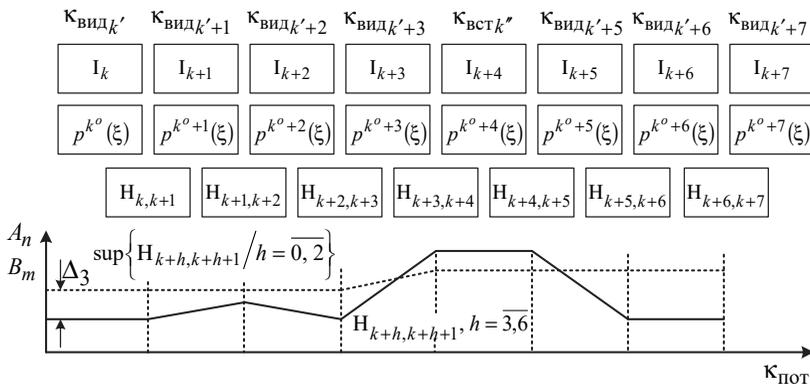


Рис. 5. Анализ гистограмм изображений кадров по полиномам Чебышевской аппроксимации

5. На пятом этапе принимается решения о классе кадров гетерогенного потока по мажоритарному правилу [20]. Параллельное применение представленных методик позволяет идентифицировать кадры-вставки, несущие фотографические данные о персонале, при следующих ограничениях:

- число кадров в визуальной вставке меньше трех;
- стартовые четыре кадра наблюдаемого окна не являются вставкой;

- кадр-вставка не граничит с разными сюжетными линиями.

6. На шестом этапе осуществляется сегментация пространства изображения, идентифицированного кадра-вставки, необходимого для аутентификации персонала. Математическое обоснование способов сегментации пространства изображения заключается в определении производных, представленных в виде дискретных приближений градиента для цифровых изображений [26]. В качестве градиентных способов выделяют операторы Собеля, Робертса, Гаусса, Лапласиана, Преветта и другие [21, 27, 28].

7. Заключительный этап направлен на восстановление и отображение видеопотока. Для фильтрации сегментированного пространства с идентифицированных кадров воспользуемся технологией «Морфинг» [29]. В качестве основы функционирования данного процесса определяются контрольные точки. Вокруг данных точек и происходит плавная трансформация одного набора пикселей в другой путем их корректной замены.

5. Моделирование и экспериментальное исследование процессов формирования и обработки гетерогенного потока с фотографическими данными о персонале. Для оценки эффективности визуальной аутентификации персонала разработана имитационная модель, представленная на рисунке 6.

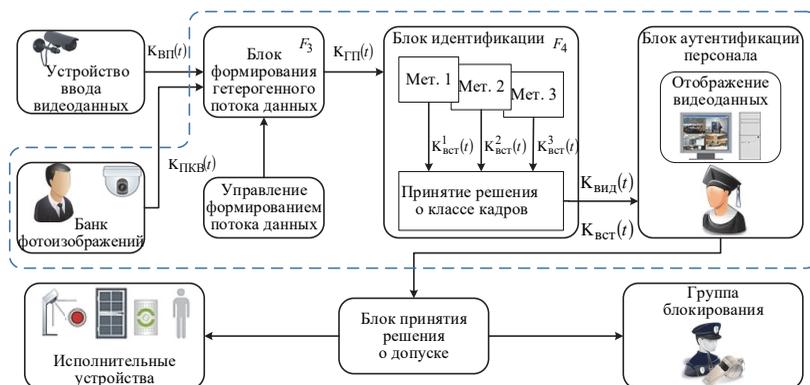


Рис. 6. Структура имитационной модели формирования гетерогенного потока видеоданных и идентификации кадров-вставок

Основными элементами имитационной модели являются:

- поток видеок кадров, поступающий с устройства ввода видеоданных $K_{ВП}(t) = \kappa_{вид_1}, \kappa_{вид_2}, \dots, \kappa_{вид_k}, \dots;$
- банк фотоизображений, используемый для формирования потока кадров-вставок, несущих фотографические данные о персонале $K_{ПКВ}(t) = \kappa_{вст_1}, \kappa_{вст_2}, \dots, \kappa_{вст_k}, \dots;$
- блок формирования гетерогенного потока $K_{ГП}(t)$, содержащего видеок кадры $\kappa_{вид_k'} \in \Psi_1$ и кадры-вставки $\kappa_{вст_k'} \in \Psi_2;$
- блок управления агрегированным потоком данных, определяющий структуру гетерогенного потока;
- блок, реализующий методику идентификации кадров-вставок в гетерогенном потоке видеоданных, сегментации пространства изображения и восстановления видеопотока;
- блок вывода результатов, отображающий на мониторе ЭВМ отдельно множество видеок кадров и множество кадров-вставок.

Имитационная модель реализована с использованием среды графического программирования LabVIEW [13, 19], языка программирования C++ и библиотеки компьютерного зрения OpenCV.

Имитируя с помощью ЭВМ процессы формирования гетерогенного потока и идентификации кадров-вставок четырьмя алгоритмами (рисунок 6) как в отдельности, так и в совокупности, получим оценки характеристик разработанного метода, представленные в таблице 1.

Таблица 1. Результаты экспериментального исследования алгоритмов идентификации кадров-вставок в гетерогенном потоке

Метод идентификации	Частота пропуска кадра-вставки	Частота ложного обнаружения кадра-вставки
Суммарной дифференциальной яркости (существующий)	0,21	0,54
Суммарной дифференциальной экстремальной яркости кадров-сверток с динамическим порогом	0,20	0,12
Анализа гистограмм изображений кадров по коэффициентам корреляции	0,34	0,18
Анализа гистограмм изображений кадров по полиномам Чебышевской аппроксимации	0,27	0,13
Алгоритм аналитической обработки выделенных кадров	0,13	0,07

Учитывая различные пороги чувствительности представленных подходов к цветовым характеристикам изображений кадров видеопотока, наименьшие частоты пропуска или ложного обнаружения кадров-вставок соответствуют разработанному новому методу с параллельным применением трех методик и дальнейшей аналитической обработкой выделенных кадров.

Для оценки эффективности функционирования интегрированной СКУД процесс визуальной аутентификации персонала оператором опишем системой массового обслуживания (СМО), где $\lambda_{СМО}$ — интенсивность входящего потока кадров-вставок на обслуживание; $\mu_{СМО}$ — интенсивность выходящего потока обслуженных кадров-вставок; $\rho_{СМО} = \lambda_{СМО} / \mu_{СМО}$ — приведенная интенсивность потока кадров-вставок; $q_{СМО}$ — относительная пропускная способность СМО (показатель эффективности функционирования).

Существующие в настоящее время СКУД с визуальной аутентификацией персонала можно рассматривать как СМО с отказами, при этом число каналов обслуживания $n_{СМО}$ равно количеству операторов.

Разработанный метод представляет новую концепцию визуальной аутентификации персонала в СКУД, изменив дисциплину обслуживания заявок (несущих фотографические данные о СД) с отказами на СМО с комбинированным способом обслуживания (ограничение по длине очереди), где $m_{СМО}$ — длина очереди.

Зависимости $q_{СМО}$ от $\rho_{СМО}$, отраженные на рисунке 7, свидетельствуют о повышении эффективности визуальной аутентификации персонала в интегрированной системе контроля и управления доступом.

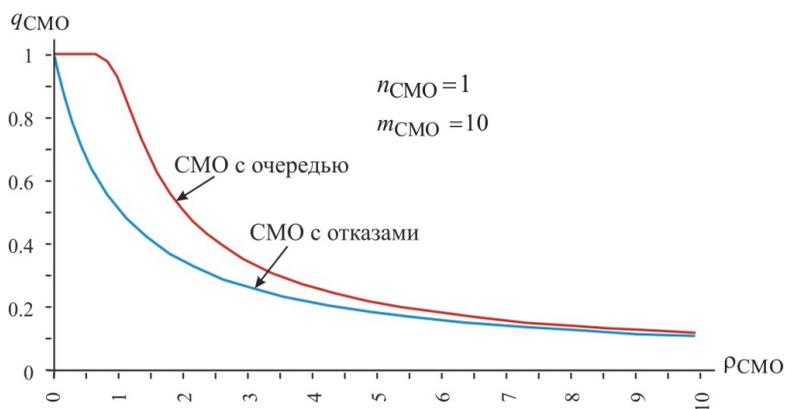


Рис. 7. Зависимость $q_{СМО} = \phi_1(\rho_{СМО}, m_{СМО})$, при $m_{СМО} = 10$

Кроме того, если оператор наблюдает за несколькими мониторами одновременно, точность визуальной аутентификации уменьшается и составляет: 85% при одном, 74% — четырех, 58% — шести и 53% при девяти мониторах. При обслуживании заявок, поступающих из очереди, оператору необходимо наблюдать всего за одним монитором, таким образом, при увеличении количества ТД вероятность правильного принятия решения остается постоянной величиной.

Формирование гетерогенного потока видеоданных для передачи статических изображений несущих информацию о персонале позволяет увеличить количество ТД при заданной пропускной способности сети передачи данных. Зависимость количества точек доступа от интенсивности появления персонала в ТД показана на рисунке 8.

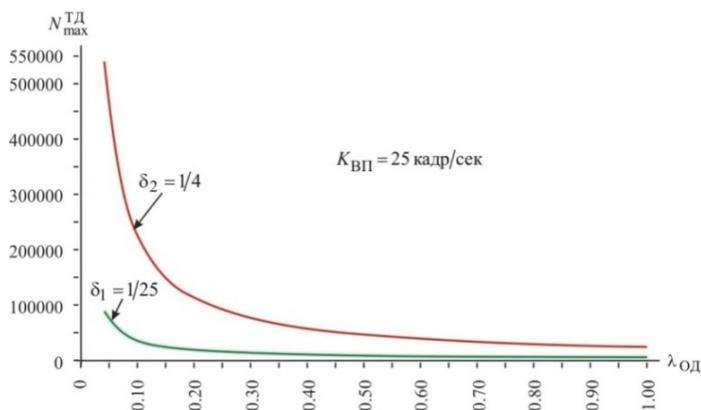


Рис. 8. Зависимость $N_{\max}^{\text{ТД}} = \phi_2(\tilde{\lambda}_{\text{ОД}}, \delta)$, при $K_{\text{ВП}} = 25$ кадр/сек

В данных зависимостях были использованы следующие обозначения: $N_{\max}^{\text{ТД}}$ — потенциальные возможности зоны доступа (максимальное количество точек доступа); $K_{\text{ВП}}$ — кадровая частота видеопотока (количество кадров в единицу времени в видеопотоке), входящего в гетерогенный поток; $\tilde{\lambda}_{\text{ОД}} = \frac{1}{N^{\text{ТД}}} \sum_{m=1}^{N^{\text{ТД}}} \tilde{\lambda}_{\text{ОД}m}$ — средняя интенсивность появления персонала в точках доступа анализируемой зоны доступа; $\delta = K_{\text{ПВК}}/K_{\text{МП}}$ — относительное значение количества кадров-вставок к числу кадров гетерогенного потока.

Поскольку имитационная модель описывает процесс функционирования проектируемой интегрированной СКУД, которая в данный

момент не существует, проверка адекватности в классическом понимании не реализуема. В работе проверка адекватности осуществлялась путем проведения натурального эксперимента. Гетерогенный поток видеоданных сформирован на основе реальных устройств системы видеонаблюдения. Проанализировано 45 реализаций, включающих 270 тысяч кадров, из которых 5,5 тысяч кадров-вставок.

6. Заключение. По результатам анализа известных технологий аутентификации персонала разработана математическая модель, использующая кадр-вставку для передачи статических изображений, базирующаяся на технологии формирования и обработки гетерогенного потока видеоданных, характеризующаяся структурой сети передачи данных, механизмом изменения состояний точек доступа по входным воздействиям и их реакции на внутренние и внешние воздействия, позволяющая значительно увеличить количество наблюдаемых объектов при заданных вероятности правильного принятия решения и пропускной способности сети передачи данных.

Разработана методика идентификации кадров-вставок в потоке видеоданных, базирующаяся на оценке суммарной дифференциальной экстремальной яркости кадров-сверток, анализе гистограмм изображений кадров по коэффициентам корреляции и полиномам Чебышевской аппроксимации, отличающаяся использованием динамического порога при идентификации кадров и принятия решения об их классе на основе мажоритарного способа обработки данных.

Разработана имитационная модель процесса формирования гетерогенного потока и идентификации кадров-вставок на основе предложенной математической модели и методики, позволяющая провести ее верификацию и валидацию на предмет подтверждения адекватности применения теоретических положений. Частоты пропуска и ложного обнаружения кадров-вставок уменьшились минимум в 1,5 раза. Эффективность визуальной аутентификации персонала в интегрированной СКУД в диапазоне от 0,1 до 10 приведенной интенсивности потока кадров-вставок, несущих фотографические данные о персонале, повышена не менее чем на 10%.

Представлен новый метод аутентификации персонала, позволяющий передавать через видеопоток идентификационные признаки субъекта доступа, необходимые для одновременного отображения реальных и эталонных фотографических данных в интегрированных СКУД на промышленных предприятиях.

Литература

1. *Козачок А.В., Бочков М.В., Фаткиева Р.Р., Туан Л.М.* Аналитическая модель защиты файлов документальных форматов от несанкционированного доступа // Труды СПИИРАН. 2015. Вып. 6(43). С. 228–252.
2. ГОСТ Р ИСО/МЭК 9594-8-98 Информационная технология. Взаимосвязь открытых систем. Справочник. Часть 8. Основы аутентификации // М.: Госстандарт России. 1998.
3. *Novak T., Gerstinger A.* Safety and Security-Critical Services in Building Automation and Control Systems // *Industrial Electronics. IEEE.* 2010. vol. 57. pp. 3614–3621.
4. Официальный сайт компании PERCo. URL: www.perco.ru (дата обращения: 17.08.2015).
5. Официальный сайт компании Anviz Biometric Inc. URL: www.anviz.com (дата обращения: 6.05.2015).
6. *Ibrahim R., Zin Z.* Study of automated face recognition system for office door access control application // *IEEE 3rd International Conference on Communication Software and Networks (ICCSN).* 2011. pp. 132–136.
7. *Лалина Т.И., Енушеев Н.Н., Лалин Д.В.* Способ биометрической аутентификации пользователя в компьютеризованных системах контроля доступа // Труды СПИИРАН. 2013. Вып. 4(27). С. 189–199.
8. *Park J., Lin S.* A random effect model for reconstruction of spatial chromatin structure // *Biometrics journal of the International Biometric Society.* 2017. vol. 73. no. 1. pp. 52–62.
9. *Shen C. et al.* Identification of subpopulations with distinct treatment benefit rate using the Bayesian tree // *Biometrical Journal.* 2016. vol. 58. no. 6. pp. 1357–1375.
10. *Бутрина А.В., Бугаева А.П.* Биометрическая идентификация личности // Актуальные проблемы авиации и космонавтики. 2012. № 8. С. 389–390.
11. *Цеханоский В.В., Чертовской В.Д.* Математическая модель многоуровневой адаптивной автоматизированной системы управления производством // Вестник Нижневартского государственного университета. 2014. № 3. С. 86–95.
12. *Орешин А.Н., Орешин Н.А., Шумилин В.С., Панчук П.В.* Методика оптимального распределения каналов и трактов на основе определения продуктивных путей с помощью автоматизированных систем управления связью // Телекоммуникации. 2016. № 10. С. 44–48.
13. *Визильтер Ю. и др.* Обработка и анализ цифровых изображений с примерами на LabVIEW и IMAQ Vision // М.: ДМК Пресс. 2008. 464 с.
14. *Харинов М.В.* Модель цифрового изображения с виртуальной памятью на основе псевдотройной системы счисления // Труды СПИИРАН. 2007. Вып. 4. С. 126–135.
15. *Трегубов Р.Б., Алексиков Ю.Г., Саитов С.И.* Представление иерархических много-уровневых маршрутизирующих систем в теоретико-множественном базисе // Труды СПИИРАН. 2016. № 1(44). С. 139–168.
16. *Трегубов Р.Б., Саитов И.А.* Теоретические основы анализа и оптимизации иерархических многоуровневых маршрутизирующих систем // Орел: Академия ФСО России. 2017. 585 с.
17. *Сальников И.И., Мартенс-Атюшев Д.С.* Методы цифровой обработки изображений // Международный студенческий научный вестник. 2015. № 3-2. С. 276–277. URL: www.eduherald.ru/article/view?id=12478 (дата обращения: 04.03.2017).
18. *Гонсалес Р., Вудс Р., Чоуи П.А.* Мир цифровой обработки. Цифровая обработка изображений // М.: Техносфера. 2006. 1072 с.
19. *Shapiro L., Stockman G.C.* *Computer Vision* // New Jersey. Prentice-Hall. 2001. pp. 279–325.

20. *Орешин А.Н., Саитов И.А., Орешин Н.А.* Стратегия повышения качества услуг видеосвязи на основе фильтрации видеопотока, содержащего кадры-вставки с информационным шумом // Труды СПИИРАН. 2015. № 4(41). С. 57–81.
21. *Харинов М.В., Ханьков И.Г.* Комбинированный метод улучшения сегментации изображения // Вестник Бурятского государственного университета. Серия. Математика и информатика. 2015. Вып. 9. С. 118–124.
22. *Кормен Т., Лейзерсон Ч., Ривест Р., Штайн К.* Алгоритмы: построение и анализ. 3-е издание // М.: Издательский дом "Вильямс". 2013. 1323 с.
23. *Орешин А.Н., Скурнович А.В., Кирухин Д.А.* Способ идентификации кадров-вставок в потоке мультимедийных данных. Патент на изобретение № 2506640. 2014.
24. *Орешин А.Н., Лысанов И.Ю.* Методика идентификации кадра-вставки в потоке видеоданных на основе корреляционного анализа гистограмм изображений кадров с динамическим порогом принятия решения // Информационные системы и технологии. 2014. № 6. С. 129–138.
25. *Лысанов И.Ю., Орешин А.Н.* Идентификация кадра-вставки в мультимедийном потоке на основе сравнения коэффициентов полиномов, аппроксимируемых гистограммы изображений кадров с динамическим порогом принятия решения // Информационные системы и технологии. 2015. №1. С. 79–87.
26. *Grady L.* Random Walks for Image Segmentation // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2006. vol. 28. no. 11. pp. 1768–1783.
27. *Белим С.В., Ларионов С.Б.* Сегментация изображений на основе алгоритма выделения сообществ на графе // Математические структуры и моделирование. 2016. № 3(39). С. 74–85.
28. *Старожилова О.В.* Распознавание цифровых изображений методом построения поля направлений // Международный научно-исследовательский журнал. 2017. № 2(56). С. 135–137.
29. Official web site FantaMorph. URL: www.fantomorph.com/index.html (дата обращения: 15.02.2015).

Орешин Андрей Николаевич — к-т техн. наук, доцент, сотрудник, Академия Федеральной службы охраны Российской Федерации. Область научных интересов: теория информации, теория графов, теория массового обслуживания, теория вероятностей, теория принятия решений, применение методов математического и имитационного моделирования в телекоммуникациях. Число научных публикаций — 67. strongnuts@mail.ru; Приборостроительная, 35, Орел, 302034; р.т.: +7(4862)549500.

Лысанов Иван Юрьевич — сотрудник, Академия Федеральной службы охраны Российской Федерации. Область научных интересов: теория вероятностей, теория графов, теория массового обслуживания, теория принятия решений, применение методов математического моделирования в телекоммуникациях, имитационное моделирование, компьютерная обработка изображений и цифровых сигналов, прикладное программирование. Число научных публикаций — 27. IvanLisanov@yandex.ru; Приборостроительная, 35, Орел, 302034; р.т.: +7(4862)549500

A.N. ORESHIN, I.YU. LISANOV
**A NEW METHOD FOR AUTOMATION OF THE PERSONNEL
 AUTHENTICATION PROCESS USING A VIDEO STREAM**

Oreshin A.N., Lisanov I.Yu. A New Method for Automation of the Personnel Authentication Process Using a Video Stream.

Abstract. On the basis of research carried out on automated control systems for industrial enterprises (ICS), the work proposes a mathematical model for the authentication of personnel using the technology of generation and processing of a heterogeneous stream containing photographic identification signs of personnel from video surveillance sources in the traffic of video data. A distinctive feature of the model are navigation operators and outputs to signal processing according to aggregate stream video management, allocation of frame-inserts bearing photographic identification signs, segmentation of space images required for authentication of staff in integrated access control system, as well as the use of technologies to restore the video stream. A technique for identifying frames-inserts in heterogeneous video stream has been developed, which is based on the estimation of total differential extreme brightness of frame-folds and an analysis of the histograms of image frames according to the correlation coefficients and polynomials of Chebyshevskii approximation. This technique uses dynamic threshold for authentication of frames and making a decision about their class on the basis of a majority method of data processing. The paper presents the results of simulation of the process of personnel authentication on the basis of the proposed model and methodologies, allowing one to estimate efficiency of visual personnel authentication, meeting the requirements of the reliability of decision making and limiting the transmission capacity of the data network.

Keywords: access control system, video surveillance system, model, heterogeneous and aggregated flows of data, identification sign (biometric, code and photographic), frame-insertion, authentication of personnel.

Oreshin Andrey Nikolaevich — Ph.D., associate professor, researcher, The Academy of Federal Security Guard Service of the Russian Federation. Research interests: information theory, graph theory, a waiting-line theory, theory of chances, decision principle, methods of mathematic and imitational modeling application in telecommunication. The number of publications — 67. strongnuts@mail.ru; 35, Priborostroitel'naya str., Orel, 302034, Russia; office phone: +7(4862)549500.

Lisanov Ivan Yurevich — сотрудник, Академия Федеральной службы охраны Российской Федерации. Research interests: probability theory, graph theory, queueing theory, decision theory, decision principle, application of mathematical simulation methods in telecommunications, simulation modeling, computer processing of images and digital signals, application programming. The number of publications — 27. IvanLisanov@yandex.ru; 35, Priborostroitel'naya Street, Orel, 302034, Russia; office phone: +7(4862)549500.

References

1. Kozachok A.V., Bochkov M.V., Fatkueva R.R., Tuan L.M. [Analytical model for protecting documentary file formats from unauthorized access]. *Trudy SPIIRAN – Proceedings of SPIIRAS*. 2015. vol. 6(43). pp. 228–252. (In Russ.).

2. GOST R ISO/IEC 9594-8-98 [Information technology. Open Systems Interconnection. Reference book. Part 8. Authentication basics]. M.: Gosstandart Rossii. 2001. (In Russ.).
3. Novak T., Gerstinger A. Safety and Security-Critical Services in Building Automation and Control Systems. *IEEE Industrial Electronics*. 2010. vol. 57. pp. 3614–3621.
4. Oficial'nyj sajt kompanii PERCo [Official web site of the company PERCo]. Available at: www.perco.ru. (accessed 17.08.2015). (In Russ.).
5. Oficial'nyj sajt kompanii Anviz Biometric Inc [Official web site of the company Anviz Biometric Inc] Available at: www.anviz.com. (accessed 6.05.2015). (In Russ.).
6. Ibrahim R., Zin Z. Study of automated face recognition system for office door access control application. IEEE 3rd International Conference on Communication Software and Networks (ICCSN). 2011. pp. 132–136.
7. Lapina T.I., Epishev N.N., Lapin D.V. [Method for biometric authentication of the user in access control systems]. *Trudy SPIIRAN – Proceedings of SPIIRAS*. 2013. vol. 4(27). pp. 189–199. (In Russ.).
8. Park J., Lin S. A random effect model for reconstruction of spatial chromatin structure. *Biometrics journal of the International Biometric Society*. 2017. vol. 73. no. 1. pp. 52–62.
9. Shen C. et al. Identification of subpopulations with distinct treatment benefit rate using the Bayesian tree. *Biometrical Journal*. 2016. vol. 58. no. 6. pp. 1357–1375.
10. Bytrina A.V., Bygaeva A.P. [Biometric identification]. *Aktual'nie problemi aviatsii i kosmonavtika – Actual problems of aviation and cosmonautics*. 2012. vol. 8. pp. 389–390. (In Russ.).
11. Cekhanovskiy V.V., Chertovsky V.D. [Mathematical modeling of local regulation of blood flow]. *Vestnik Nizhnegorodckogo gosydarstvennogo univertsiteta – Bulletin Nizhnevartovsk State University*. 2014. vol. 3. pp. 86–95. (In Russ.).
12. Oreshin A.N., Oreshin N.A., Shumilin V.S., Panchyuk P.V. [Methodology of optimal channels distribution and highways based on the determination of productive ways by using the automated control systems of communication]. *Telekommunikatsii – Telecommunications*. 2016. vol. 10. pp. 44–48. (In Russ.).
13. Vizil'ter Yu. et al. *Obrabotka i analiz tsifrovih izobrazhenii c primerami na LabVIEW i IMAQ Vision* [Digital images processing and analysis with examples in LabVIEW and IMAQ Vision]. Moscow: DMK Press. 2008. 464 p. (In Russ.).
14. Kharinov M.V. [Model of Digital Image Attributed with Virtual Memory Basing on Pseudo-Ternary Number System]. *Trudy SPIIRAN – Proceedings of SPIIRAS*. 2007. vol. 4. pp. 126–135. (In Russ.).
15. Tregubov R.B., Aleksikov Yu.G., Saitov S.I. [Representation of the hierarchical multilayer routing systems in a set-theoretic base]. *Trudy SPIIRAN – Proceedings of SPIIRAS*. 2016. vol. 1(44). pp. 139–168. (In Russ.).
16. Tregubov R.B., Saitov I.A. *Teoreticheskie osnovy analiza i optimizatsii ierarkhicheskikh mnogourovnevnykh marshrutiziruyushchikh system* [The theoretical framework for the analysis and optimization of hierarchical multilayer routing systems]. Orel: Akademiya FSO Rossii. 2016. 585 p. (In Russ.).
17. Salnikov I.I., Martens-Atushev D.S. [Digital image processing techniques]. *Mezhdynarodnii stydencheskii naychnii vestnik – International student research bulletin*. 2015. vol. 3–2. pp. 276–277. Available at: www.eduherald.ru. (accessed 04.03.2017). (In Russ.).
18. Gonsales R., Vuds R., Chochia P.A. *Mir tsifrovoi obrabotki. Tsifrovaia obrabotka izobrazhenii* [World of the digital processing. Digital Image Processing]. Moscow: Tehnosfera Publ. 2006. 1072 p. (In Russ.).
19. Shapiro L.G., Stockman G.C. *Computer Vision*. New Jersey. Prentice-Hall. 2001. pp. 279–325.

20. Oreshin A.N., Saitov I.A., Oreshin N.A. [Strategy of the Video Communication Services Quality Enhancement Based on the Filtration of a Video Stream Containing Snap-Insertions with Information Noise]. *Trudy SPIIRAN – Proceedings of SPIIRAS*. 2015. vol. 4(41). pp. 57–81. (In Russ.).
21. Kharinov M.V., Hanikov I.G. [The combined method of improving image segmentation]. *Vestnik Byriatskogo gosydarstvennogo univversiteta. Seria. Matematika i informatika – Bulletin Buryat State University. Mathematic and informatization series*. 2015. vol. 9. pp. 118–124. (In Russ.).
22. Kormen T., Lejzerson Ch., Rivest R., Shtajn K. *Algoritmi: postroenie i analiz. 3 izdanie* [Algorithms: design and analysis]. Moscow: "Vil'jams". 2013. 1323 p. (In Russ.).
23. Oreshin A.N., Skurnovich A.V., Kiryuhin D.A. [Method of frame authentication of multimedia data flow based on cross-correlation analysis of histograms of frame images]. Patent RF. no. 2506640. 2014. (In Russ.).
24. Oreshin A.N., Lysanov I.Yu. [Snap-insertion identification procedure in a video data stream based on the correlative analysis of the histograms of snap images with a dynamic threshold of decision making]. *Informacionnye sistemy i tehnologii – Information systems and technologies*. 2014. vol. 6. pp. 129–138. (In Russ.).
25. Lisanov I.Yu., Oreshin A.N. [Authentication of frame-insertion in a multimedia stream based on comparison of coefficients of the polynomials approximated histograms of frame images with the dynamic making decision threshold]. *Informacionnye sistemy i tehnologii – Information systems and technologies*. 2015. vol. 1. pp. 79–87. (In Russ.).
26. Grady L. Random Walks for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2006. vol. 28. no. 11. pp. 1768–1783.
27. Belim S.V., Larionov S.B. [Segmentation of images the basis of algorithm of communities separation on the graph]. *Matematicheskie stryktury i modelirovanie – Mathematical structures and modeling*. 2016. vol. 3(39). pp. 74–85. (In Russ.).
28. Starozhilova O.V. [Recognition of digital images by means of directional field construction]. *Mezhdynarodnii naychno-issledovatel'skii zhyrnal – International research journal*. 2017. vol. 2(56). pp. 135–137. (In Russ.).
29. Official web site FantaMorph. Available at: www.fantomorph.com/index.html (accessed 15.02.2015).

Е.С. НОВИКОВА, И.Н. МУРЕНИН
**МЕТОДИКА ВИЗУАЛЬНОГО АНАЛИЗА МАРШРУТОВ
СОТРУДНИКОВ ОРГАНИЗАЦИИ ДЛЯ ОБНАРУЖЕНИЯ
АНОМАЛИЙ**

Новикова Е. С., Муренин И. Н. Методика визуального анализа маршрутов сотрудников организации для обнаружения аномалий.

Аннотация. Обнаружение аномалий в перемещениях сотрудников является важной задачей, которая связана с обеспечением киберфизической безопасности предприятий, включая критические инфраструктуры. В работе представлен подход к анализу перемещений сотрудников критической инфраструктуры, отличающийся сочетанием алгоритмов интеллектуального анализа данных и интерактивных методик визуализации. Он включает в себя два этапа — определение групп сотрудников с похожим поведением и обнаружение аномалий. Группировка пользователей по их поведению осуществляется с помощью самоорганизующихся карт Кохонена; для отображения пространственно-временных шаблонов поведения используется разработанная авторами модель визуализации BandView. Для обнаружения аномалий в поведении сотрудников предлагается механизм оценки значений пространственно-временных атрибутов движения. Отображение отклонений осуществляется с помощью тепловой карты, позволяющей аналитику с легкостью определить зону и интервал времени с подозрительной активностью. Подход апробирован на наборе данных, предоставленном в рамках конкурса VASTMiniChallenge-2 2016, который описывает перемещения сотрудников внутри здания организации.

Ключевые слова: выявление аномалий в траекториях, визуальная аналитика, паттерны поведения, оценка отклонений в поведении, тепловые карты.

1. Введение. В настоящее время данные, описывающие перемещения движущихся точек в пространстве, являются наиболее распространенным типом пространственных данных, и их анализ имеет множество практических применений. Они могут быть получены с помощью датчиков мобильных телефонов, автомобилей, считывателей карт доступа, видеокамер, расположенных в местах наблюдений за объектами. Значения атрибутов движения объектов могут быть получены как напрямую от наблюдаемых объектов, так и от специализированных сенсоров, регистрирующих их появление. Обычно такие наборы данных называют траекториями. Исследование траекторий помогает определить шаблоны поведения объектов, выявить ограничения, существующие в исследуемой среде, например, правила или политики безопасности, регулирующие права доступа к определенным зонам, или социально значимые объекты, например, банкоматы, кафе, аптеки и так далее [1-3]. Другим важным применением анализа траекторий является создание шаблонов поведения объектов для последующего обнаружения возможных аномалий в их маршрутах. В этом случае анализ траекторий может использоваться для обеспечения безопасно-

сти движения воздушного, наземного и водного транспорта путем мониторинга местоположения, траектории, скорости движения транспортного средства и обнаружения неожиданных препятствий на маршруте. Мониторинг перемещений сотрудников критических инфраструктур обеспечивает контроль соблюдения мер и политик безопасности, установленных на предприятии [4].

В настоящее время на рынке программных продуктов представлено большое количество решений, осуществляющих мониторинг перемещений сотрудников внутри организации, а также отслеживающих их действия на рабочих местах, контролируя список запущенных процессов, набираемый текст, делая аудио- и видеозапись [5-8]. Основной задачей таких систем является контроль активности сотрудников организации на рабочих местах, заключающийся в расчете рабочего времени, оценке динамики опозданий, продуктивности работы. Вместе с тем, согласно [9], анализ поведения пользователей должен позволять строить паттерны поведения, на основе которых в дальнейшем возможно выявление различных аномалий, которые могут свидетельствовать о таких угрозах, как внутренний нарушитель, финансовые мошенничества и целенаправленные атаки.

В работе представлен подход к выявлению пространственно-временных шаблонов перемещений персонала организации и аномалий в них, отличающийся сочетанием автоматических методик анализа данных и графического представления данных. Множество интерактивных моделей визуализации не только осуществляет графическое представление выявленных шаблонов в перемещениях сотрудников и потенциально-аномальных ситуаций, но и позволяет контролировать результаты применения автоматических методов анализа данных. Паттерны поведения отображаются с помощью разработанной авторами модели визуализации BandView, отражающей последовательность контролируемых зон, посещаемых сотрудником. Отклонения в перемещениях сотрудников отображаются в виде тепловой карты. Для уменьшения возможных шумов на тепловой карте предлагается механизм оценки отклонений, позволяющий сфокусироваться на отдельных подозрительных выбросах. Оценка аномалий осуществляется на основе пространственно-временных атрибутов отклонений, таких как место, продолжительность пребывания и время посещения. Сценарий использования методики и ее эффективность продемонстрирована на множестве тестовых данных, предложенных в рамках конкурса VAST Challenge 2016 [10].

Основной вклад авторов заключается в разработке подхода к визуальному анализу движений сотрудников организации, оперирующему разнородными данными, включающими должность сотрудников,

расположение их рабочих мест и временные атрибуты их перемещений для обеспечения киберфизической безопасности.

2. Методики анализа и визуализации траекторий. В случае отсутствия каких-либо априорных данных о типичном поведении (движении) объектов для анализа траекторий чаще всего применяются алгоритмы кластеризации данных. Полученные кластеры используются в дальнейшем для описания модели нормального поведения объектов, на основе которой производится последующее обнаружение аномалий. Для поиска кластеров применяются подходы, основанные на центроидах и методах, иерархических моделях, а также на основе плотности распределения объектов в пространстве признаков [11]. Автоматические методы могут обнаруживать интересные шаблоны поведения и аномалии, но для понимания и объяснения полученных результатов необходимо использовать дополнительные методики графического представления данных [12, 13].

Метод нейронной кластеризации, также известный как самоорганизующаяся карта Кохонена (self-organizing maps, SOM-карта), сочетает многомерную кластеризацию и методику проецирования признаков объектов из многомерного пространства в двумерное с сохранением расстояния между ними, обеспечивая при этом визуализацию полученного распределения объектов по кластерам. Шрек и другие применили нейронную сеть SOM для исследования траекторий и предложили инструмент визуального анализа данных, позволяющий контролировать результаты кластеризации, полученные с помощью сети SOM [14]. Основной акцент в работе сделан на анализ пространственных атрибутов траекторий. В [15] с помощью сети SOM исследуются как временные, так и пространственные атрибуты, однако их анализ выполняется путем выделения всех временных атрибутов для каждого географического местоположения или всех пространственных атрибутов для определенного интервала времени.

В общем случае существующие модели графического представления траекторий могут быть разделены на 3 группы: статические или интерактивные географические карты, часто дополняемые глифами, отвечающими за отображение атрибутов перемещений (1); пространственно-временные кубы (2) и вложенные графики с временной шкалой (3) [16].

Географические карты являются наиболее очевидным способом графического представления данных, описывающим изменения в местоположении объектов с течением времени. Траектории объектов или группы объектов отображаются в виде линий [17-19]. Атрибуты перемещений, такие как время, скорость, тип движущегося объекта кодируются цветом линии или отображаются с помощью специальных

глифов. В случаях, когда отдельные траектории не значимы, используются карты потоков [18, 19]. Карты потоков представляют собой модели визуализации, акцентирующие внимание на направлении движения, начальной и конечной точках маршрутов. Количественные характеристики потоков обычно отображаются с помощью цвета линии потока и ее насыщенности. Интересная модификация карт потоков используется в [19] для анализа миграций популяции. Географические регионы, выступающие в качестве источника или пункта назначения миграционных потоков, располагаются по окружности в виде сегментов кольца. Потоки отображаются в виде линий, соединяющих соответствующие сегменты кольца. Следует отметить, что для всех моделей визуализации траекторий на основе географических карт характерен один общий недостаток — они не могут отображать пространственные и временные атрибуты движения одновременно.

Пространственно-временной куб — это трехмерная модель визуализации, разработанная для одновременного представления пространственных и временных характеристик движущихся объектов. В этой модели точки траекторий отображаются в трехмерное пространство, где одна из осей, чаще всего вертикальная, обозначает время. В [20] представлено расширение пространственно-временного куба, названное *стеной траекторий (trajectory wall)*. В этой модели визуализации вертикальная ось трехмерного куба используется для графического представления некоторого подмножества траекторий, обладающих схожими пространственными характеристиками. Так как они могут быть упорядочены по времени в третьем измерении, такая модель может рассматриваться в качестве трехмерного пространственно-временного куба, где абсолютное время преобразовано в порядок следования траекторий. Как и многие трехмерные модели визуализации, трехмерные пространственно-временные кубы могут быть неэффективны из-за большого нагромождения траекторий и недостатка пространства для отображения.

Вложенный график с временной осью представляет собой традиционный линейный график, одна ось которого обозначает время. По оси Y откладываются значения пространственных атрибутов точек траектории. Поскольку на графике обычно отображается сразу несколько характеристик траектории или несколько траекторий, то график называется *вложенным*. Каждый маршрут представляется в виде кривой или полосы [21-23]. Полоса (площадь под кривой) может быть разделена на сегменты, окрашенные в соответствии со значениями атрибутов движения объекта.

В современных системах контроля доступа анализ перемещений сотрудников выполняется путем построения маршрутов их перемещений и тепловых карт движения, осуществляется расчет пройденных дистанций и контроль соблюдения временного регламента перемещений на основе заданных графиков посещения контролируемых зон [5, 6, 8]. В некоторых случаях возможны выявления фактов обмана, связанных с одномоментной отметкой нескольких человек или отметкой присутствия за коллег [5]. Задача формирования паттернов движения сотрудников в общем случае не решается, выявление аномалий формируется в основном на основе правил, построенных в соответствии с описанием должностных инструкций. Кроме того, отсутствуют механизмы визуального анализа данных, процесс исследования инцидентов безопасности заключается в работе с исходными данными, представленными в табличном виде или в формате видео.

Представленная методика анализа траекторий сотрудников наиболее близка к методике, представленной в [14], поскольку в ней также применяются нейронные сети SOM для выявления траекторий, обладающих схожими характеристиками. Однако в отличие от [14], в работе оцениваются как пространственные, так и временные атрибуты маршрутов сотрудников. Графическое представление сети SOM усилено использованием специального глифа, дающего краткую характеристику объектов, принадлежащих одному кластеру. Кроме того, в работе представлен механизм оценки обнаруженных аномалий, учитывающий периодичность их появления и ранжирующий их по значимости.

3. Подход к обнаружению аномалий в перемещениях пользователей. Разработанная авторами методика анализа перемещений сотрудников внутри организации позволяет ответить на следующие вопросы:

- 1) Если ли группы сотрудников с похожим поведением?
- 2) Есть ли какие-то особенности в перемещениях сотрудников, зависящие от времени, то есть имеет ли место периодичность в перемещениях в зависимости от роли сотрудника?
- 3) Каков общий шаблон перемещений сотрудников, принадлежащих к одной группе? Как он меняется в зависимости от дня недели? Как это соотносится с положением работника в организации?
- 4) Есть ли какие-либо значимые отклонения в перемещениях сотрудников?
- 5) Каков характер аномалий, то есть как часто, где и когда они происходят?

Последовательно отвечая на данные вопросы, аналитик формирует в первую очередь общее понимание существующих шаблонов

перемещений, а затем фокусируется на деталях, описывающих возможные аномалии. Таким образом, авторы считают, что эти вопросы определяют основные этапы анализа траекторий и в целом согласуются с мантрой поиска информации, сформулированной Б. Шнейдерманом: «*общее представление данных → масштабирование и фильтрация → детали по требованию*» [24]. Модели визуализации и лежащие в их основе методики анализа данных, применяемые в предложенном подходе, поддерживают описанную схему аналитического процесса.

Ключевыми элементами методики являются нейронные сети SOM, использующиеся для формирования групп сотрудников, имеющих похожие шаблоны перемещений, и тепловые карты, которые применяются для обнаружения периодов аномального поведения. Они дополняются двумя моделями визуализации — графом контролируемых зон и вложенной моделью визуализацией *BandView*, отражающей последовательность посещенных зон и длительность пребывания в них.

Методика анализа состоит из следующих основных этапов:

1) этап подготовки данных, преобразующий записи считывателей контроля доступа в формат, необходимый для их последующей обработки;

2) формирование групп пользователей с одинаковым поведением и их отображение с помощью нейронных сетей Кохонена;

3) определение периодичности поведения пользователей одной группы путем выявления дней недели с одинаковым поведением;

4) вычисление и формирование графического представления паттернов поведения внутри одной группы или для заданного дня недели;

5) выявление аномалий в траекториях путем статистической оценки отклонений от паттерна траектории и их графическое представление с помощью тепловой карты;

6) детальный анализ аномального участка маршрута сотрудника с помощью модели визуализации *BandView*.

На рисунке 1 представлен макет основного окна прототипа программного обеспечения, реализующего предложенную методику. Первая SOM-карта «Сотрудники» (А) осуществляет разбиение персонала на группы с похожими траекториями перемещений; вторая SOM-карта «Дни» оценивает периодичность в траекториях сотрудников, принадлежащих одной группе. Панель (С) имеет две вкладки, предназначенные для отображения шаблонов маршрутов и аномалий — *Pattern View* и *Anomaly View* соответственно. В обоих случаях используется модель визуализации *BandView*.

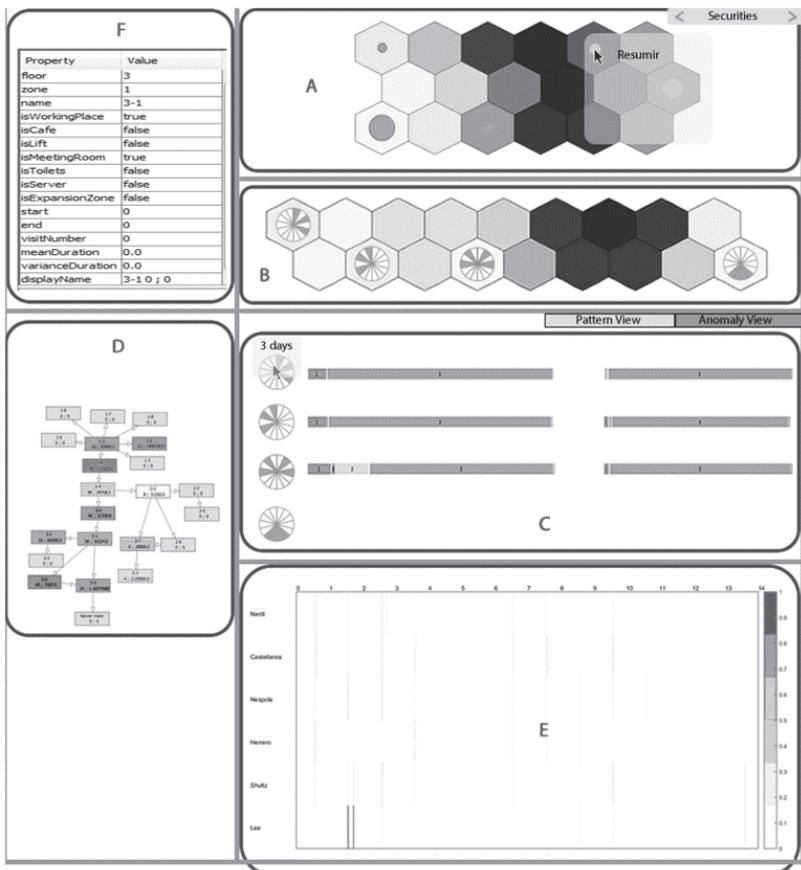


Рис. 1. Макет основного окна программного обеспечения, реализующего предложенную методику

Граф контролируемых зон (D) отражает маршрут выбранного сотрудника или группы сотрудников. Тепловая карта (E) демонстрирует отклонения в поведении для групп сотрудников или групп дней с похожими шаблонами перемещений. Свойства объектов, представленных различными графическими элементами моделей визуализации, отображаются с помощью таблицы свойств (F). Все модели графического представления данных интерактивны и связаны между собой. Выбирая мышкой элемент любой модели визуализации, аналитик обновляет информацию на остальных моделях, связанных с текущей. Например, при выборе элемента карты SOM «Сотрудники» происходит обновление остальных компонент графического интерфейса: таб-

лица свойств (F) наполняется подробной информацией о выбранной группе, SOM-карта «Дни» (B) отображает кластеры дней с похожими траекториями перемещений для выбранной группы, пространственно-временные шаблоны маршрутов для каждой группы дней представляются с помощью графа (D) и панели (C). Отклонения в траекториях внутри выбранной группы отображаются на тепловой карте (E). Выбор элемента на SOM-карте «Дни» позволяет сфокусироваться на определенном подмножестве дней, обновляя соответствующим образом компоненты (C)-(F). Следует отметить, что тепловая карта E непосредственно связана с вкладкой Anomaly View: выбор любой области тепловой карты приводит к обновлению модели BandView, расположенной на вкладке AnomalyView. Таким образом, SOM-карты могут рассматриваться как элементы графической фильтрации; а тепловая карта — как навигационная панель для выбора наиболее интересных для аналитика отклонений.

В следующих разделах описаны этапы предварительной обработки данных, а также приводятся подробные пояснения к предложенным методикам визуализации и анализа данных.

4. Этап подготовки исходных данных. В представленном подходе исходными данными являются журналы, формируемые датчиками контроля доступа и содержащие информацию только о времени посещения контролируемой зоны определенным сотрудником. Особенностью подхода является работа с журналами датчиков контроля доступа, и результат применения методики не зависит от технологии, используемой для реализации считывателей системы контроля доступа; это могут быть считыватели на основе rfid/nfc меток [6] или Bluetooth-маяков Beacon [5]. Важным моментом является возможность извлечения из логов системы контроля доступа следующей информации о нахождении сотрудника: время посещения контролируемой зоны и идентификатор данной зоны. Данные о должностях сотрудников и расположении контролируемых зон позволяют более точно описать паттерны передвижения, добавив к ним семантическую составляющую, например, указав, что в данной зоне находится рабочее место сотрудника.

Отличительная особенность логов, формируемых датчиками контроля доступа в том, что регистрация сотрудников происходит непосредственно в момент их появления в контролируемой зоне, вследствие чего записи в журнале носят нерегулярный характер и интервал между ними для определенного сотрудника может варьироваться от нескольких секунд до нескольких часов. Кроме того, некоторые сотрудники могут совершать множество перемещений по зданию организации в соответствии с их должностными обязанностями, в то

время как другие значительно реже покидают свои рабочие места. Таким образом, логи датчиков можно рассматривать как неравноотстоящие временные ряды переменной длины.

Одним из способов предварительной обработки неравноотстоящих временных рядов для их дальнейшего анализа является их дискретизация, позволяющая преобразовать их в векторы конечной длины.

Пусть $E = \{e_i\}_{i=1}^n$ — множество сотрудников, $Z = \{z_j\}_{j=1}^m$ — множество контролируемых зон, $T = \{t_k : t_i < t_j; i < j\}_{k=1}^p$ — упорядоченное множество временных меток, соответственно, записи из журнала считывателей контролируемых зон имеют вид (e_i, z_j, t_k) . Общий интервал времени, представленный первой и последней записью журнала, обозначим как T_0 , а множество записей обозначим как $LOGS = \{(e_i, z_j, t_k)\}, i = 1 \div n, j = 1 \div m, k = 1 \div p$. Интервал времени T_0 разбивается на последовательность одинаковых временных интервалов Δt , то есть:

$$T_0 = \{\Delta t_l : \Delta t_i = \Delta t_j; \Delta t_i = [t_i; t_{i+1}); \Delta t_{i+1} = [t_{i+1}; t_{i+2}); i \neq j; i, j \leq l; \}_{l=1}^r.$$

Для каждого интервала времени Δt_l и для каждого сотрудника e_i вычисляется число посещений $n_{z_j}^{\Delta t_l}$ и длительность пребывания $\Delta t_{z_j}^{\Delta t_l}$ в каждой контролируемой зоне z_j . Таким образом, множество LOGS можно представить в виде множества упорядоченных во времени пар $LOGS = \{(n_{z_j}^{\Delta t_l}; \Delta t_{z_j}^{\Delta t_l})\}_{e_i}, i = 1 \div n, j = 1 \div m, l = 1 \div r$, вычисленных для каждого сотрудника $e_i, i = 1 \div n$. Это множество упорядоченных пар формирует множество атрибутов траектории пользователя, позволяя совместить пространственно-временные характеристики траектории сотрудников организации. Атрибуты вектора упорядочиваются сначала по наблюдаемому временному интервалу, а затем по идентификатору контролируемой зоны.

Длительность интервала времени Δt_l по умолчанию составляет 4 часа — эксперименты показали, что этого достаточно для того, чтобы обнаружить даже незначительные временные отклонения длительностью в 1 минуту.

5. Модель визуализации на основе нейронной сети SOM. К основным методам визуальной кластеризации следует отнести методы

главных компонент, многомерного шкалирования, метод t-SNE и другие. Данные алгоритмы осуществляют проекцию исходного многомерного пространства в пространство меньшей размерности, чаще всего двумерное, для последующего построения графического представления данных с помощью графиков рассеивания. Другим примером визуальной кластеризации является методика RadViz, которая также осуществляет проекцию многомерного пространства в двумерное, в основе которой лежит физическая метафора: каждый объект соединяется с n координатными узлами, обозначающими атрибуты объекта, n пружинами, жесткость которых зависит от значения соответствующих координат. Данные методики обладают общим свойством — снижая размерность признакового пространства объектов, они не снижают объем исходной выборки. Таким образом, они подходят как средство контроля результатов применения автоматических методов кластеризации и определения исходного числа кластеров. Однако для выявления некоторого шаблона данных, описываемого, например, центроидом кластера, они не подходят. По этой причине для обнаружения групп сотрудников с похожим поведением и индивидуальных шаблонов поведения для отдельных сотрудников используется нейронная сеть SOM (и ее графическое представление — SOM-карта). Ее основная задача — это разведывательный анализ данных при отсутствии информации о структуре исходных данных. Данный тип нейронной сети использует обучение без учителя для кластеризации многомерных данных [25]. Сеть SOM состоит из узлов или нейронов, с соответствующими весовыми векторами, размерность которых определяется исходными данными. В ходе итерационного обучения входные векторы сравниваются с весовыми векторами каждого нейрона, а веса нейрона, наиболее соответствующие входному вектору, и веса соседних нейронов регулируются так, чтобы быть ближе к входному вектору. Это обеспечивает возможность нейронной сети сохранять отношения между соседними нейронами, что означает, что кластеры, ассоциированные с узлами сети, расположенными рядом друг с другом, имеют большую степень сходства, чем кластеры, нейроны которых расположены далеко друг от друга. Одной из главных проблем сетей SOM является необходимость в непрерывной обработке данных без пропусков отдельных значений для каждого атрибута. Однако в предложенной методике этап предварительной обработки данных гарантирует получение векторов с явно определенными значениями для каждого атрибута.

Нейронные сети SOM могут рассматриваться в качестве методики визуализации данных, так как они обеспечивают графическое отображение многомерных данных в двумерное пространство. Для

представления структуры сети SOM часто используется U-матрица [26]. Она отражает структуру данных с помощью отображения среднего расстояния между весовыми векторами соседних нейронов. В предложенном подходе используется шестиугольная сетка, поэтому рассматривается вектор с 6 соседними узлами. Чем темнее цвет узла — тем дальше он находится от своих соседей. Узлы, окрашенные цветами с похожей интенсивностью, схожи друг с другом. Узлы, содержащие центры кластеров, отмечены глифом в форме круга, размер которого пропорционален количеству объектов в кластере.

В предложенном подходе нейронная сеть SOM используется дважды. Цель SOM-карты «Сотрудники» заключается в выявлении групп сотрудников с похожим поведением. Вектор признаков сотрудника описывает его активность на протяжении всего периода времени работы датчиков контроля доступа, в результате однократные или редкие отклонения в перемещениях не влияют на результат кластеризации. Это позволяет предположить, что SOM отображает различия в траекториях сотрудников в соответствии с особенностями их ролей в организации.

Уровень детализации анализа траекторий сотрудников пропорционален количеству узлов на SOM-карте. Выбор карты большего размера приводит к увеличению общего числа отображаемых кластеров. В таком случае ячейки сети SOM, соответствующие работникам с похожим поведением, будут сосредоточены в регионе карты, раскрашенном схожими оттенками. Эксперименты показали, что разница в поведении сотрудников, принадлежащих к соседним узлам, незначительна и объясняется, как правило, небольшими отклонениями в продолжительности пребывания в пределах контролируемой зоны, в то время как число посещенных зон и даже последовательности посещений сохраняются. Такие различия во времени редко превышают 5-10 минут, таким образом, есть возможность использовать одну общую пространственно-временную траекторию для всех сотрудников, принадлежащих к одному региону SOM. Для объединения таких сотрудников в один кластер рекомендуется настраивать размер карты SOM примерно равным их количеству. Однако разрабатываемый программный инструмент предусматривает возможность задавать различные размеры SOM-карты, чтобы исследовать возможные варианты группировки сотрудников.

SOM-карта для сотрудников на рисунке 2 содержит результаты кластеризации сотрудников одной должности. Из него следует, что данное множество сотрудников можно разбить на 5 групп, исходя из подобия их маршрутов в течение рабочего дня. Одна из них довольно многочисленна (группа №1), в то время как остальные состоят из одного-двух человек. Кроме того, траектории движения двух групп сотрудников

(группы №4 и №5), расположенные в правом верхнем углу SOM-карты, сильно отличаются, так как отделяются друг от друга множеством узлов, окрашенных в темный цвет. Дальнейший анализ их траекторий позволяет сделать вывод, что основные различия в их поведении объясняются расположением рабочих мест и должностными обязанностями.

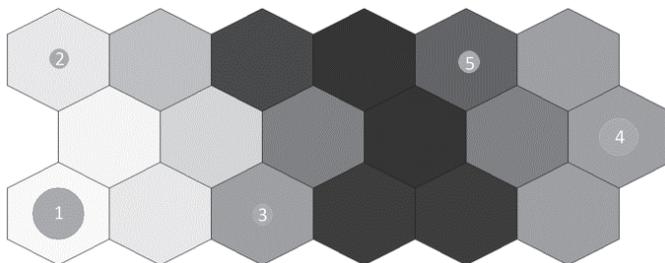


Рис. 2. SOM-карта «Сотрудники»

Авторы полагают, что сотрудники могут иметь обязанности, зависящие от дня недели, например, проведение еженедельного инструктажа, психологических тренингов и так далее. Подобные обязанности могут вызывать периодические изменения в перемещениях сотрудников. SOM-карта «Дни» позволяет выявить группы дней с похожими шаблонами перемещений сотрудников, принадлежащих к выбранному аналитическому кластеру, и определить таким образом периодичность в их маршрутах. Для построения данной SOM-карты используются центроиды кластеров SOM-карты «Сотрудники», преобразованные в векторы, соответствующие рабочим дням сотрудников, путем разбиения исходного вектора на векторы меньшей длины. Результат кластеризации нейронной сети также отображается с помощью U-матрицы, однако узлы SOM-карты дополнены специально разработанными глифами WeekCircle, отражающими распределение дней одного по дням недели. При проектировании глифа авторы учли, что во многих организациях деятельность сотрудников зависит от того, является ли неделя четной или нет. Глиф WeekCircle может быть использован для отображения шаблонов перемещений с периодической активностью в рамках одной или двух недель. В зависимости от заданной детализации он разделяется на 7 или 14 секторов, соответствующих дням недели. Модель глифа для двухнедельного периода перемещений представлена на рисунке 3. Правая половина глифа демонстрирует дни четной недели, а левая — нечетной. Понедельники обозначены верхними секторами, а выходные — нижними. На рисунке 3 изображены два глифа, которые отражают, какие дни включает группа дней. Левый глиф демонстрирует, что кластер дней состоит из понедельни-

ков, сред и пятниц четной недели, это означает, что выбранный сотрудник или группа сотрудников имеет определенные должностные обязанности, выполняемые каждые понедельник, среду и пятницу. Правый глиф показывает, что в группу дней вошли только выходные дни. Следует отметить, что SOM-карта «Дни» позволяет обнаружить дни с аномальным поведением, если эти аномалии имеют достаточно продолжительный характер, например, длятся более часа, поскольку дни с подобной активностью формируют отдельный кластер, расположенный по соседству с остальными.

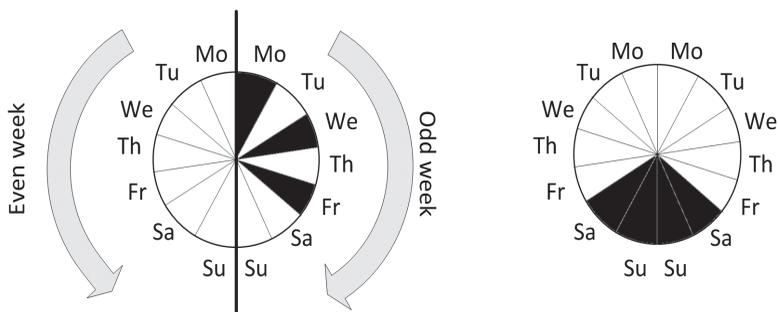


Рис. 3. Глиф WeekCircle, демонстрирующий разбиение множества дней на группы в соответствии с днем недели

6. Модель визуализации траекторий сотрудников BandView. Модель визуализации BandView предназначена для выявления связи между пространственными и временными атрибутами перемещений. Она представляет собой вложенную столбиковую диаграмму, горизонтальная ось которой обозначает время. Маршрут работника представлен с помощью последовательности блоков-сегментов, каждый из которых соответствует временному интервалу, в течение которого работник находится в данной зоне. Цвет сегмента используется для кодирования атрибутов зоны. В настоящее время цветовая схема кодирования контролируемых зон построена следующим образом: каждому этажу соответствует определенный цвет, а цвета для зон, расположенных на этом этаже создаются путем изменения яркости соответствующего цвета: чем больше значение идентификатора контролируемой зоны — тем темнее цвет соответствующего сегмента. На рисунке 4 представлены маршруты сотрудников отдела, выполняющих обслуживание здания в течение одного дня. Из него следует, что сотрудники данного отдела работают в три смены, выходя в ночное, дневное и вечернее время.

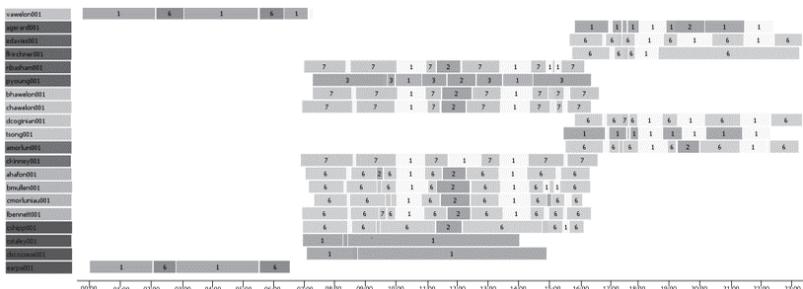


Рис. 4. Модель визуализации маршрутов сотрудников BandView

Очевидно, что продолжительность пребывания в зонах сильно варьируется — от нескольких секунд до нескольких часов, что приводит к тому, что длины сегментов полос BandView могут быть или очень маленькими и практически незаметными, или такими длинными, что для их просмотра требуется полоса прокрутки. Для того чтобы обеспечить аналитику возможность просматривать как короткие, так и длинные сегменты, реализован механизм масштабирования, построенный на нелинейном преобразовании шкалы времени, увеличивающем короткие интервалы времени и сокращающем длительные. Каждый сегмент модели может быть выбран с помощью мышки, подробная информация о нем, такая как продолжительность пребывания сотрудника в данной зоне, временная отметка, атрибуты зоны и так далее, отображаются в окне свойств.

Модель BandView используется для отображения шаблонов перемещений и сырых данных при исследовании отклонений в маршруте сотрудника. Следует отметить, что модель BandView позволяет легко обнаружить, где и когда произошла аномалия, как долго она длилась, благодаря возможности визуально сравнивать маршрут сотрудника с маршрутами его коллег.

Для работы с исходными данными реализован гибкий механизм фильтрации, позволяющий строить сложные логические выражения, используя всевозможные атрибуты перемещений: идентификатор сотрудника, должность, офис, продолжительность пребывания в зоне, идентификатор зоны, этаж и так далее.

7. Граф контролируемых зон. Методики визуализации на основе графов широко используются для представления пространственных атрибутов траекторий. В предложенном подходе вершины графа представляют контролируемые зоны, смежные зоны соединены ребрами. Зоны, посещенные сотрудником, выделены цветом в соответствии с цветовой схемой, используемой в BandView. Зоны, которые данный сотрудник не посещал, отмечены серым цветом. Граф контролируемых зон позволяет установить связь между посещенными сотрудником зонами и их атрибутами: наличие кафе, конференц-

зала, лестницы или лифта, расположение рабочего места сотрудника и его коллег и так далее. Исследование этих атрибутов позволяет объяснить причины посещения контролируемой зоны сотрудником. Для этого атрибуты контролируемых зон также кодируются цветом, и аналитик может выбирать режимы просмотра графа исходя из атрибутов контролируемых зон или их расположения внутри здания.

В предложенном подходе граф также используется для оценки статистических данных о перемещениях сотрудников: аналитику предоставляется информация о средней продолжительности пребывания в контролируемой зоне в течение дня и ее дисперсии, среднем числе посещений в день и общем количестве посещений на протяжении всего анализируемого интервала времени. Эти данные выводятся в таблице свойств объекта F и могут быть отражены в соответствующих вершинах графа. Кроме того, аналитик имеет возможность устанавливать размер вершины в зависимости от числа посещений или среднего времени пребывания в зоне. Авторы считают, что такая возможность способствует обнаружению зон, где локализованы рабочие места сотрудников, и зон, которые они посещают редко.

Так, например, на рисунке 5 изображен граф посещаемых зон, размер вершин которых зависит от количества посещений в течение всего рассматриваемого периода времени. Можно ясно увидеть, что сотрудник проводит большую часть времени в зоне 1-7, где расположено его рабочее место, а зону 3-4 посетил всего один раз. Поскольку в этой зоне находится лифт третьего этажа, можно предположить, что он посетил ее случайно, нажав не на ту кнопку лифта.

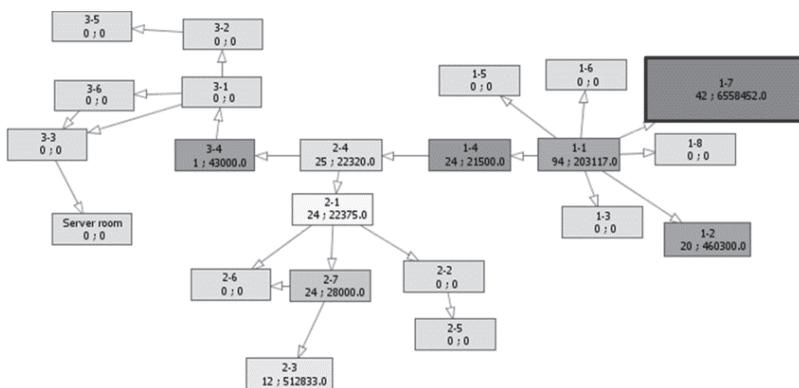


Рис. 5. Граф посещения контролируемых зон

Данные отображаемые на графе контролируемых зон определяются двумя SOM-картами. При выборе элемента на SOM-карте «Сотрудники» на графе цветом выделяются вершины, обозначающие зоны, посе-

ценные выбранной группой сотрудников. SOM-карта «Дни» отфильтровывает данные и отображает пространственный шаблон перемещений для заданной группы сотрудников в течение выбранного множества дней.

8. Тепловая карта аномалий. Цель тепловой карты — отразить наличие потенциально аномальных отклонений в движении персонала. Аномалии обычно проявляются в нерегулярных нечастых изменениях в поведении объектов, таким образом, предлагается искать отклонения в рамках группы сотрудников или группы дней, имеющих схожие шаблоны перемещений. Тепловая карта строится следующим образом: ось Y соответствует сотрудникам одного кластера, ось X представляет атрибуты векторов, сформированные на этапе подготовки исходных данных. Каждый элемент тепловой карты представляет собой расстояние от значения атрибута вектора признаков сотрудника до центроида соответствующего ему кластера.

Однако непосредственное отображение расстояний может привести к возникновению зашумленной картинки в случае, если расстояния от центроида до объектов внутри кластера примерно одинаковы, а также к исчезновению отклонений на карте, если разница между отдельными расстояниями слишком велика. Например, если сотрудник находится приблизительно 2 часа в определенной зоне в течение заданного интервала времени, тогда отклонение в 10 минут может считаться незначительным, однако в ситуации, когда сотрудник проводит на 10 минут больше в зоне, в которой обычно он находится 20 минут, может иметь место аномалия. Кроме того, оценка отклонений может существенно зависеть от специфики конкретной зоны. Например, отклонение в 20 минут в офисе совершенно несопоставимо с ситуацией, когда сотрудник проводит 20 минут на лестничной площадке.

Для решения таких проблем предлагается механизм рейтинговой оценки отклонений, рассматривающий их в контексте некоторого среднего времени пребывания сотрудника в определенной зоне в течение заданного промежутка времени. Основной целью механизма оценки является выбор всех потенциально аномальных отклонений из общего множества и помещение их в фокус внимания аналитика.

Механизм оценки состоит из двух частей — расчет значимости (рейтинга) отклонения и определение пороговых значений для каждой контролируемой зоны. В основе расчета рейтинга отклонения лежит расчет z-показателя (z-score), отражающего расстояние данного значения от среднего значения по набору отклонений. Значения z-показателя распределены на интервале $[-4;4]$ и свидетельствуют, на какую долю стандартного отклонения текущее значение больше или меньше среднего. Значения z-показателя в диапазоне $[-1.65;1.65]$ представляют ожидаемый результат, то есть в такие моменты времени поведение сотрудников почти не отклоняется от их типичного поведения в рамках должностных обязанностей. Значения, выходящие за границы $[-2.58;2.58]$

свидетельствуют о том, что в данные моменты времени сотрудники демонстрируют нестандартную модель поведения, которая не является результатом случайного процесса [27]. Все отклонения, попавшие за границы данного интервала, можно считать потенциально-аномальными, и поэтому рекомендованы для дальнейшего более подробного изучения. На рисунке б продемонстрирован результат применения механизма оценки значимости отклонений в маршрутах сотрудников одного отдела: на рисунке ба представлены исходные расстояния между значениями атрибутов центроида и объектов данного кластера, на рисунке бб — расстояния, преобразованные в z-показатель.

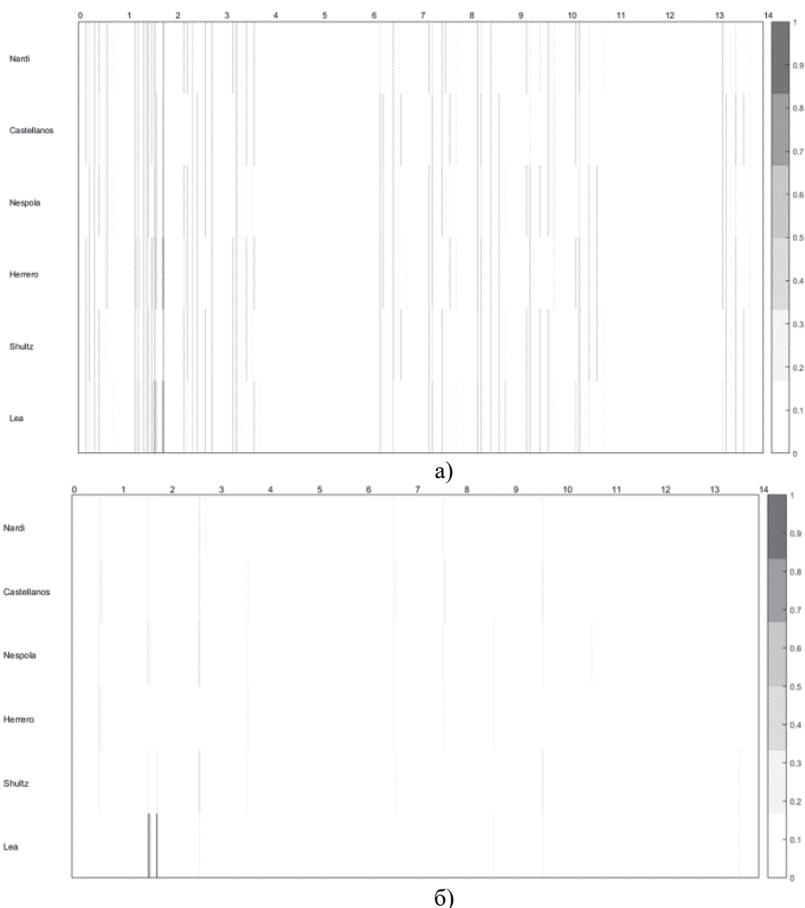


Рис. 6. Результат применения механизмов оценки отклонения: а) тепловая карта расстояний между значениями атрибутов центроида и объектов данного кластера; б) тепловая карта z-показателей отклонений

Очевидно, что уровень шума на тепловой карте значительно снизился, и отчетливо стало видно аномальное отклонение в поведении одного из сотрудников группы на второй изучаемый день. Пороговые значения для каждой зоны непосредственно определяют, какие отклонения будут отображаться на тепловой карте, а какие — нет. Фильтрация отклонений производится по их временным и пространственным атрибутам, таким как продолжительность, зона в которой произошло отклонение и соответствующий временной интервал. Таким образом, аналитик может поместить в фокус внимания только отклонения, происходящие на интервале с 8 до 12 часов утра или в холле первого этажа или совместить оба атрибута.

9. Эксперимент и обсуждение результатов. Для оценки предложенного подхода был использован набор данных, представленный в рамках конкурса TheVASTChallenge 2016: Mini-Challenge 2 [10]. Он содержит журнал логов датчиков, регистрирующих появление пользователя в контролируемых зонах здания организации. Когда сотрудник при помощи электронной карты доступа попадает в новую контролируемую зону, данные о его карте автоматически распознаются датчиком и записываются в журнал. Следует отметить, что большая часть зон доступна для сотрудников, даже если они забыли свои пропуска. Журнал содержит записи датчиков за две недели. Аналитик имеет в своем распоряжении схему планировки здания и расположения офисов, включая карты подконтрольных зон, а также список сотрудников с указанием их должностей и офисов.

Исследование траекторий сотрудников организации показало, что большинство сотрудников одного отдела передвигается одинаково, а имеющиеся отличия в моделях поведения можно объяснить спецификой выполняемых задач и расположения рабочего места.

В качестве примера рассмотрим перемещения сотрудников отдела охраны. На рисунке 2 представлена SOM-карта, отражающая результат кластеризации их маршрутов. Типичные маршруты сотрудников отдела представлены на рисунке 7.

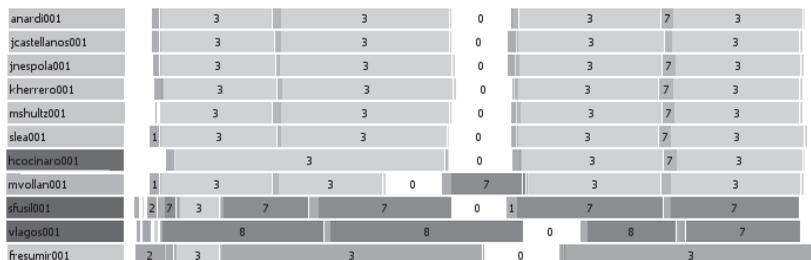


Рис. 7. Типичные маршруты сотрудников отдела «Охрана»

Анализ траекторий членов наиболее многочисленной группы №1, показал, что они имеют два основных шаблона перемещений: один для рабочих дней, другой — для выходных. Их рабочий день обычно начинается около 8:00 и заканчивается в 17:00. Примерно в 12:00 они выходят из здания на час. Большую часть рабочего времени участники этой группы проводят в зоне 2-3 второго этажа, где находятся их рабочие места. Каждые полтора-два часа они покидают свой офис и находятся в смежной зоне 2-7 в течение 5-10 минут. Исходя из того, что охранники посещают только одну зону, расположенную на втором этаже, а также тот факт, что в этой зоне помимо офисов находятся кафе и туалетные комнаты, можно предположить, что они выходят освежиться, не совершая при этом обход здания. В выходные дни они не приходят на работу.

Исследование аномалий в их поведении с помощью тепловой карты позволило обнаружить значительные отклонения в поведении одного из сотрудников на второй день (рисунок 6б). Изучение логов датчиков контроля доступа с помощью модели визуализации Bandview показал отсутствие одного из сотрудников на рабочем месте. Авторы обнаружили еще одну интересную особенность: другой работник группы имеет дублирующиеся записи в журнале для каждого датчика в этот же день, причем вторая временная отметка появляется несколькими секундами позднее первой. Аномалия также хорошо заметна на графе посещаемых зон, отображающего статистику посещений за день для каждой из зон. Факт отсутствия записей в журнале для одного сотрудника и дублирование для другого позволяет предположить, что первый сотрудник использовал электронный пропуск второго сотрудника для прохода в контролируемую зону.

Два ближайших к этой группе кластера (группы №2 и №3) состоят из одного человека. Их маршруты достаточно похожи на маршруты сотрудников группы №1, поскольку их офисы также расположены в зоне 2-3, где они и проводят большую часть времени. Основное отличие для сотрудника группы №2 заключается в том, что он не покидает зону 2-3 в первой половине дня и посещает зону 2-7 только во второй половине дня около 15:00. Исследование его перемещений с помощью модели визуализации BandView и тепловой карты позволило выявить небольшое отклонение, произошедшее на третий день всего контролируемого периода. В этот день во время послеобеденного обхода зоны 2-7 он не возвращается в свою комнату, а в 17:00 идет домой из зоны 2-7. Это отклонение можно объяснить тем, что сотрудник мог забыть приложить электронный пропуск, возвращаясь на рабочее место из зоны 2-7. Цвет ячейки SOM-карты, соответствующей кластеру №3, значительно темнее, и различия в

траекториях выражены сильнее. Данный сотрудник отправляется обедать на час раньше, а с 12:00 до 13:00 находится в зоне 1-7. Каких-либо значимых отклонений у данного сотрудника выявлено не было.

Оставшиеся две группы (№4 и №5) расположены отдельно от остальных. Группа №4 состоит из двух сотрудников, чьи офисы располагаются на первом этаже. Они имеют 3 шаблона маршрутов, зависящих от дня недели, представленные на рисунке 8. Во все рабочие дни, кроме каждого вторника, они проводят большую часть времени на рабочем месте, расположенном в зоне 1-7. Каждый полтора-два часа совершают обход зоны 1-1, а в обеденное время с 12:00 до 13:00 покидают здание. Каждый вторник в первой половине дня они посещают зону 2-3, где располагаются офисы их коллег, что позволяет предположить, что они посещают еженедельные собрания, на которых проходит очередной инструктаж. Авторы обнаружили два интересных отклонения в перемещениях одного из сотрудников этого кластера. Данный сотрудник посетил зону 3-4, расположенную на третьем этаже, только один раз за весь наблюдаемый двухнедельный промежуток времени, причем длительность пребывания в этой зоне не превысила 1 минуты. Вторая аномалия связана с нетипичной длительностью пребывания в зоне 1-2. Проводя обычно в этой зоне около 2-3 минут, он в один из дней провел в ней 15 минут. Поскольку в этой зоне находится кафе, можно предположить, что обычно он совершает обход помещений первого этажа, в то время как во время длительного пребывания в зоне 2-3 он общался с кем-то из сотрудников, также находившихся в этой зоне. Однако для уточнения этого предположения следует изучить, кто еще в это время находился в зоне 2-3.

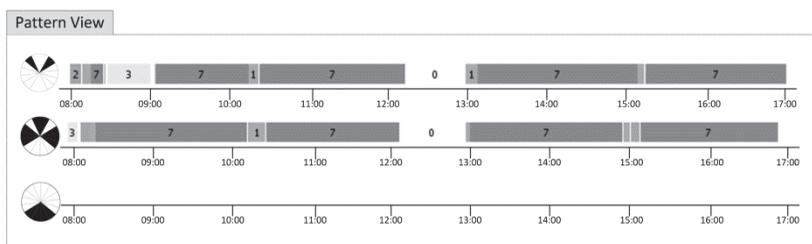


Рис. 8. Шаблоны маршрутов сотрудников группы №4

Группа №5 состоит из одного сотрудника. Большую часть времени он проводит на третьем этаже, где находится его офис. Тот факт, что на третьем этаже располагаются офисы руководства компании, позволяет предположить, что этот сотрудник является начальником отдела безопасности. Его перемещения достаточно разнообразны и сильно зависят от дня недели. Он начинает свой рабочий день с посе-

щения местного кафе, находящегося в зоне 1-2, и затем поднимается в свой офис. Каждый вторник он заходит в зону 2-3 и проводит там около получаса, а затем возвращается в свой офис. Анализ результатов представления исходных данных о перемещениях сотрудников отдела безопасности, полученный с помощью модели визуализации BandView, позволяет заключить, что каждый вторник руководитель отдела проводит совещания со всем отделом. Кроме того, можно установить, что представитель отдела безопасности должен постоянно находиться на первом этаже с восьми часов утра до пяти часов вечера. По этой причине некоторые из сотрудников отдела выходят на обеденный перерыв в другое время; а работники, чьи офисы располагаются на первом этаже, посещают собрания только раз в неделю, чтобы обеспечить присутствие охраны на входе в здание.

Предложенный подход позволил описать основные модели поведения сотрудников организации, определить существующие ограничения в их перемещениях. Использование тепловой карты в сочетании с оценкой отклонений в маршрутах позволяет достаточно легко обнаружить аномалии разного рода, начиная от отсутствия сотрудника на рабочем месте и заканчивая дублированием логов и нетипичной длительностью нахождения в контролируемой зоне.

Следует отметить, что в большинстве случаев анализ контактов персонала способствует лучшему пониманию причин возникновения аномалий. Модель BandView может быть полезна для понимания возможного взаимодействия между работниками. Однако она работает лишь в том случае, когда число отображаемых работников ограничено и не превышает 10-15 человек, иначе отслеживание взаимодействий между сотрудниками сильно затруднено.

Таким образом, одно из основных направлений будущей работы связано с разработкой методики визуального анализа, позволяющей установить паттерны взаимодействия между сотрудниками организации. Возможными способами решения данной проблемы является построение графов контактов, отражающие структурные особенности взаимодействия сотрудников, тепловые карты посещения зон различными группами сотрудников для выявления различных типов взаимодействия. Другое направление будущей работы затрагивает анализ данных, полученных из различных источников. Журналы операционной системы, например события входа-выхода и события клавиатуры, доказывают, что сотрудник находится на рабочем месте. Показания сенсоров инфраструктур здания, например, таких как тепло-вентиляционная система, могут также объяснить нетипичное поведение сотрудников. Корреляция этих данных требует разработки

новых методов визуального анализа с учетом особенностей исходных данных.

10. Заключение. В статье представлен подход к анализу траекторий сотрудников критических инфраструктур, основанный на использовании методик визуализации и автоматического анализа данных. Он позволяет сформировать пространственно-временные шаблоны маршрутов, установить имеющиеся ограничения и обнаружить аномалии в перемещениях сотрудников. Ключевыми элементами подхода являются интерактивные SOM-карты, которые используются для выявления групп сотрудников, имеющих схожие траектории перемещений, и установления возможной периодичности в их маршрутах. Выявление аномалий в перемещениях сотрудников осуществляется с помощью тепловой карты, дополненной механизмом оценки отклонений в контексте пространственно-временного шаблона поведения сотрудника. Авторы представили методики взаимодействия, связывающие все визуальные компоненты и обеспечивающие процесс анализа. Для иллюстрации предложенного подхода был использован набор данных, предоставленный в рамках конкурса VastChallenge 2016. В статье представлено обсуждение полученных результатов и определены основные направления будущей работы.

Учитывая требования к исходным данным — наличие метки времени посещения контролируемой зоны, идентификатор сотрудника, идентификатор контролируемой зоны, данная методика естественным образом может быть добавлена в систему управления и контроля доступом в качестве модуля, выполняющего интеллектуальный анализ данных. В связи с ростом актуальности задачи выявления внутреннего нарушителя она может быть реализована в компоненте визуального анализа поведения сотрудников организации систем управления информационной безопасностью при условии включения датчиков системы доступа в качестве сенсоров данных, обрабатываемой SIEM-системой [28, 29].

Литература

1. Demšar U. et al. Analysis and visualisation of movement: an interdisciplinary review // *Movement Ecology*. 2015. vol. 3. no. 1.
2. Lerman Y., Rofe Y., Omer I. Using Space Syntax to Model Pedestrian Movement in Urban Transportation Planning // *Geographical Analysis*. 2014. vol. 46(4). pp. 392–410.
3. Ferreira N. et al. Visual Exploration of Big Spatio-Temporal Urban Data: A Study of New York City Taxi Trips // *IEEE Transactions on Visualization and Computer Graphics*. 2013. vol. 19. pp. 2149–2158.
4. Tan L., Hu M., Lin H. Agent-based simulation of building evacuation // *International Journal of Information Sciences*. 2015. vol. 295. no. C. pp. 53–66.
5. Hubstuff Employee Monitoring Software. URL: https://hubstuff.com/employee_monitoring_software (дата обращения: 29.07.2017).

6. WaveTrend Access Control. URL: <http://www.wavetrend.net/access-control.php> (дата обращения: 29.07.2017).
7. Employee Monitoring and Productivity Analysis. URL: <https://www.intesecurity.com/employee-monitoring-and-productivity-analysis/> (дата обращения: 29.07.2017).
8. ObserveIT Insider Threat Solution. URL: <https://www.observeit.com/insider-threat-solution> (дата обращения: 29.07.2017).
9. *Bussa T., Litan A., Phillips T.* Market Guide for User and Entity Behavior Analytics. URL: <https://www.gartner.com/doc/3538217/market-guide-user-entity-behavior> (дата обращения: 29.07.2017).
10. Vast Challenge Website. URL: <http://vacommunity.org/> (дата обращения: 05.04.2017).
11. *Kisilevich S., Mansmann F., Nanni M., Rinzivillo S.* Spatio-temporal clustering: a survey // *Data Mining and Knowledge Discovery Handbook*. 2010. pp. 855–874.
12. *Andrienko G., Andrienko N.* Exploration of massive movement data: a visual analytics approach // *Proceedings of the 11th AGILE International Conference on Geographic Information Science*. 2008. URL: https://agile-online.org/conference_paper/cds/agile_2008/pdf/66_doc.pdf (дата обращения: 29.07.2017).
13. *Kotenko I., Novikova E.* VisSecAnalyzer: a Visual Analytics Tool For Network Security Assessment // *International Conference on Availability, Reliability, and Security*. 2013. LNCS 8128. pp. 345–360.
14. *Schreck T., Bernard J., Von Landesberger T., Kohlhammer J.* Visual cluster analysis of trajectory data with interactive Kohonen map // *Information Visualization*. 2009. vol. 8. no. 1. pp.14–29.
15. *Guo D., Chen J., MacEachren A.M., Liao K.* A visualization system for space-time and multivariate patterns (VIS-STAMP) // *IEEE Transactions on Visualization and Computer Graphics*. 2006. vol. 12(6). pp. 1461–1474.
16. *Andrienko N., Andrienko G.* Visual analytics of movement: an overview of methods, tools and procedures // *Information Visualization*. 2013. vol. 12(1). pp. 3–24.
17. *Andrienko G. et al.* Scalable Analysis of Movement Data for Extracting and Exploring Significant Places // *IEEE Transactions on Visualization and Computer Graphics*. 2013. vol. 19. pp. 1078–1094.
18. *Ho Q., Nguyen P.H., Åström T., Jern M.* Implementation of a Flow Map Demonstrator for Analyzing Commuting and Migration Flow Statistics Data // *Procedia – Social and Behavioral Sciences*. 2011. vol. 21. pp. 157–166.
19. *Abel J., Sander N.* Quantifying Global International Migration Flows // *Science*. 2014. vol. 343. Issue 6178. pp. 1520–1522.
20. *Andrienko G., Andrienko N., Schumann H., Tominski C.* Visualization of Trajectory Attributes in Space–Time Cube and Trajectory Wall // *Cartography from Pole to Pole*. 2014. pp. 157–163.
21. *Guo C. et al.* Dodeca-Rings Map: Interactively Finding Patterns and Events in Large Geo-temporal Data // *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST)*. 2014. pp. 353–354.
22. *Choo J. et al.* Exploring Anomalies in GAStech: VAST Mini Challenge 1 and 2 // *Proceedings of IEEE Symposium on Visual Analytics Science and Technology Challenge (VAST)*. 2014. pp. 347–348.
23. *Tominski C., Schumann H., Andrienko G., Andrienko N.* Stacking-Based Visualization of Trajectory Attribute Data // *IEEE Transactions on Visualization and Computer Graphics*. 2012. vol. 18. no. 12. pp. 2565–2574.
24. *Shneiderman B.* Dynamic queries for visual information seeking // *IEEE Software*. 2003. vol. 11. no. 6. pp.70–77.

25. *Kohonen T., Honkela T.* Kohonen network // Scholarpedia. 2007. vol. 2(1). pp. 1568.
26. *Ultsch A.* Self-organizing neural networks for visualization and classification // Information and Classification. pp. 307–313.
27. *Caldas de Castro M., Singer B.* Controlling the False Discovery Rate: A New Application to Account for Multiple and Dependent Test in Local Statistics of Spatial Association // Geographical Analysis. 2006. vol. 38. pp. 180–208.
28. *Котенко И.В., Саенко И.Б.* Построение системы интеллектуальных сервисов для защиты информации в условиях кибернетического противоборства // Труды СПИИРАН. 2012. Вып. 3(22). С. 84–100.
29. *Kotenko I., Chechulin A.* Attack Modeling and Security Evaluation in SIEM Systems // International Transactions on Systems Science and Applications. 2012. vol. 8. pp.129–147.

Новикова Евгения Сергеевна — к-т техн. наук, старший научный сотрудник лаборатории проблем компьютерной безопасности Федерального государственного бюджетного учреждения науки Санкт-Петербургского института информатики и автоматизации Российской академии наук (СПИИРАН), доцент кафедры информационных систем, Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (СПбГЭТУ). Область научных интересов: визуальная аналитика, вредоносное программное обеспечение, двухключевая криптография. Число научных публикаций — 80. novikova.evgenia123@gmail.com, <http://www.comsec.spb.ru/en/staff/novikova>; 14-я линия В.О., 39, Санкт-Петербург, 199178; р.т.: +7(812)328–2642.

Муренин Иван Николаевич — студент магистратуры, Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (СПбГЭТУ). Область научных интересов: анализ траекторий, визуализация многомерных данных. Число научных публикаций — 2. imurenin@gmail.com; ул. Профессора Попова, 5, Санкт-Петербург, 197376; р.т.: +7(812) 234-27-73.

Поддержка исследований. Работа выполнена при финансовой поддержке РФФИ (проект № 16-07-00625) в СПбГЭТУ и при частичной поддержке бюджетных тем № 0073-2015-0004 и 0073-2015-0007 в СПИИРАН.

E.S. NOVIKOVA, I.N. MURENIN
**THE TECHNIQUE OF THE VISUAL ANALYSIS OF THE
 ORGANIZATION EMPLOYEES ROUTES FOR ANOMALY
 DETECTION**

Novikova E. S., Murenin I. N. The Technique of the Visual Analysis of the Organization Employees Routes for Anomaly Detection.

Abstract. The detection of anomalies in the movement of employees is an important task of the cyber-physical security of enterprises, including critical infrastructures. The paper presents a technique to analyze the routes of the organization employees based on combination of the data mining and interactive visualization techniques. It includes two stages – detection of the groups of the employees with similar behavior and anomaly discovery. The self-organizing Kohonen maps are used to group employees on the basis of their behavior. To present spatio-temporal patterns, authors developed special visualization model named BandView. To detect anomalies authors present a rating mechanism assessing spatiotemporal attributes of the movement. The visualization of the anomalies is done using heatmaps that allow an analyst to spot place and time with a possibly suspicious activity. The technique is tested against data set provided within VAST MiniChallenge-2 contest that contains logs from access control sensors describing employees' movement within organization building.

Keywords: anomaly detection in trajectories, visual analytics, behavior patterns, behavior deviation assessment, heatmaps.

Novikova Evgenia Sergeevna — Ph.D., senior researcher of the computer security problems laboratory, St. Petersburg Institute for Informatics and Automation of Russian Academy of Sciences (SPIIRAS), associate professor of the information systems department, Saint-Petersburg State Electrotechnical University “LETI” (ETU). Research interests: security visual analytics, malware, public key cryptography. The number of publications — 80. evgeshka19@mail.ru, <http://www.comsec.spb.ru/en/staff/novikova>; 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone: +7(812)328–2642.

Murenin Ivan Nikolaevich — trajectories analysis, multidimensional data visualization. Research interests: master student Saint-Petersburg State Electrotechnical University “LETI” (ETU). The number of publications — 2. imurenin@gmail.com; 5, Professor Popov str., St.Petersburg, 197376, Russia; office phone: +7(812) 234-27-73.

Acknowledgements. This research is supported by RFBR (projects No. 16-07-00625 in ETU and partly by the budget (projects no. 0073-2015-0004 and 0073-2015-0007) in SPIIRAS.

References

1. Demšar U., Buchin K., Cagnacci F., Safi K., Speckmann B., Van de Weghe N., Weibel R. Analysis and visualisation of movement: an interdisciplinary review. *Movement Ecology*. 2015. vol. 3. no. 1.
2. Lerman Y., Rofe Y., Omer I. Using Space Syntax to Model Pedestrian Movement in Urban Transportation Planning. *Geographical Analysis*. 2014. vol. 46(4), pp. 392–410.
3. Ferreira N. et al. Visual Exploration of Big Spatio-Temporal Urban Data: A Study of New York City Taxi Trips. *IEEE Transactions on Visualization and Computer Graphics*. 2013. vol. 19. pp. 2149–2158.

4. Tan L., Hu M., Lin H. Agent-based simulation of building evacuation. *International Journal of Information Sciences*. 2015. vol. 295. no. C. pp. 53–66.
5. Hubstuff Employee Monitoring Software. Available at: https://hubstuff.com/employee_monitoring_software (accessed: 29.07.2017).
6. WaveTrend Access Control, URL: <http://www.wavetrend.net/access-control.php> (accessed: 29.07.2017).
7. Employee Monitoring and Productivity Analysis. Available at: <https://www.intesecurity.com/employee-monitoring-and-productivity-analysis/> (accessed: 29.07.2017).
8. ObserveIT Insider Threat Solution. Available at: <https://www.observeit.com/insider-threat-solution> (accessed: 29.07.2017).
9. Bussa T., Litan A., Phillips T. Market Guide for User and Entity Behavior Analytics. Available at: <https://www.gartner.com/doc/3538217/market-guide-user-entity-behavior> (accessed: 29.07.2017).
10. Vast Challenge Website. Available at: <http://vacommunity.org/> (accessed: 05.04. 2017).
11. Kisilevich S., Mansmann F., Nanni M., Rinzivillo S. Spatio-temporal clustering: a survey. *Data Mining and Knowledge Discovery Handbook*. 2010. pp. 855–874.
12. Andrienko G., Andrienko N. Exploration of massive movement data: a visual analytics approach. Proceedings of the 11th AGILE International Conference on Geographic Information Science. 2008. Available at: https://agile-online.org/conference_paper/cds/agile_2008/pdf/66_doc.pdf (accessed: 29.07.2017).
13. Kotenko I., Novikova E. VisSecAnalyzer: a Visual Analytics Tool For Network Security Assessment. International Conference on Availability, Reliability, and Security. 2013. LNCS 8128. pp. 345–360.
14. Schreck T., Bernard J., Von Landesberger T., Kohlhammer J. Visual cluster analysis of trajectory data with interactive Kohonen map. *Information Visualization*. 2009. vol. 8. no. 1. pp.14–29.
15. Guo D., Chen J., MacEachren A.M., Liao K. A visualization system for space-time and multivariate patterns (VIS-STAMP). *IEEE Transactions on Visualization and Computer Graphics*. 2006. vol. 12(6). pp. 1461–1474.
16. Andrienko N., Andrienko G. Visual analytics of movement: an overview of methods, tools and procedures. *Information Visualization*. 2013. vol. 12(1). pp. 3–24.
17. Andrienko G. et al. Scalable Analysis of Movement Data for Extracting and Exploring Significant Places. *IEEE Transactions on Visualization and Computer Graphics*. 2013. vol. 19. pp. 1078–1094.
18. Ho Q., Nguyen P.H, Åström T., Jern M. Implementation of a Flow Map Demonstrator for Analyzing Commuting and Migration Flow Statistics Data. *Procedia - Social and Behavioral Sciences*. 2011. vol. 21. pp. 157–166.
19. Abel J., Sander N. Quantifying Global International Migration Flows. *Science*. 2014. vol. 343. Issue 6178. pp. 1520–1522.
20. Andrienko G., Andrienko N., Schumann H., Tominski C. Visualization of Trajectory Attributes in Space–Time Cube and Trajectory Wall. *Cartography from Pole to Pole*. 2014. pp. 157–163.
21. Guo C. et al. Dodeca-Rings Map: Interactively Finding Patterns and Events in Large Geo-temporal Data. Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST). 2014. pp. 353–354.
22. Choo J. et al. Exploring Anomalies in GASTech: VAST Mini Challenge 1 and 2. Proceedings of IEEE Symposium on Visual Analytics Science and Technology Challenge (VAST). 2014. pp. 347–348.
23. Tominski C., Schumann H., Andrienko G., Andrienko N. Stacking-Based Visualization of Trajectory Attribute Data. *IEEE Transactions on Visualization and Computer Graphics*. 2012. vol. 18. no. 12. pp. 2565–2574.

24. Shneiderman B. Dynamic queries for visual information seeking. *IEEE Software*. 2003. vol. 11. no. 6. pp.70–77.
25. Kohonen T., Honkela T. Kohonen network . *Scholarpedia*. 2007. vol. 2(1). pp. 1568.
26. Ultsch A. Self-organizing neural networks for visualization and classification. *Information and Classification*. pp. 307–313.
27. Caldas de Castro M., Singer B. Controlling the False Discovery Rate: A New Application to Account for Multiple and Dependent Test in Local Statistics of Spatial Association. *Geographical Analysis*. 2006. vol. 38. pp. 180–208.
28. Kotenko I.V., Saenko I.B. [Developing the system of intelligent services to protect information in cyber warfare]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2012. vol. 3(22). pp. 84–100. (In Russ.).
29. Kotenko I., Chechulin A. Attack Modeling and Security Evaluation in SIEM Systems. *International Transactions on Systems Science and Applications*. 2012. vol. 8. pp. 129–147.

Т.М. СУХОВ, С.Ю. СТРАХОВ, Л.Б. КОЧИН, Е.А. БЕЛЯЕВ
**ПРИМЕНЕНИЕ И РЕАЛИЗАЦИИ ПРИНЦИПА ПОСТОЯННОЙ
ЦВЕТОВОЙ ЯРКОСТИ В ЦИФРОВЫХ СИСТЕМАХ
КОДИРОВАНИЯ ВИДЕОИНФОРМАЦИИ**

Сухов Т.М., Страхов С.Ю., Кочин Л.Б., Беляев Е.А. Применение и реализации принципа постоянной цветовой яркости в цифровых системах кодирования видеоинформации.

Аннотация. Современные средства обработки потока видеоданных для цветowego кодирования используют принцип постоянной яркости, предложенный при разработке аналоговой системы цветного телевидения NTSC. Этот принцип, как и его реализации, не свободен от недостатков: потеря информации о четкости кодируемых цветных изображений, ухудшение различимости ахроматических деталей по мере увеличения насыщенности цветов и другие. Кроме того, применение прореживания цветоразностных компонентов (4:2:2, 4:2:0, 4:1:1) в цифровых системах обработки видеоданных также искажает декодируемое видеоизображение.

Для кодирования цветовой видеоинформации существует альтернативный подход — применение принципа постоянной цветовой яркости. В работе представлен сравнительный анализ принципов постоянной яркости и постоянной цветовой яркости. Показано, что при использовании принципа постоянной цветовой яркости можно получить выигрыш более 6 дБ.

В работе предлагаются вещественная и целочисленные реализации принципа постоянной цветовой яркости. Выполнен сравнительный анализ разработанных реализаций и целочисленной реализации принципа постоянной яркости, показавший преимущество использования принципа постоянной цветовой яркости для некоторых современных процессоров.

Повышения качества декодируемых изображений в цифровых системах кодирования видеоинформации можно достичь, применив принцип постоянной цветовой яркости.

Ключевые слова: принцип постоянной яркости, принцип постоянной цветовой яркости, целочисленные преобразования цветowego пространства, сравнение преобразований цветowego пространства.

1. Введение. Человеческий зрительный анализатор, в соответствии с трехцветной теорией цвета [1-3], воспринимает результирующий цвет как сочетание трех основных цветов — красного, синего и зеленого. Поэтому все цвета можно выразить комбинацией этих трех «главных» цветов (в теории цвета их называют опорными). Существует информационная модель зрительной системы [4, 5], согласно которой цветовые отклики основных цветов преобразуются в амплитуду равноэнергетического белого цвета и оппонентные отклики зрительной системы человека, которые несут информацию о насыщенности и цветовом тоне.

При разработке цветных аналоговых телевизионных систем были введены в строй такие системы передачи цветовой информации, как PAL, NTSC, SECAM и их модификации. Основой кодирования цве-

товой информации в этих системах является применение принципа постоянной яркости (ПЯ). При кодировании цветовой информации допускается сокращение пространственного разрешения цветоразностных компонентов в 1,5 и более раз.

Термином «черно-белое изображение» в настоящей работе обозначаются изображения, представленные в градациях серого цвета (англоязычный аналог — «*grayscale*»).

Цифровые телевизионные системы унаследовали от аналоговых принцип ПЯ.

В работе рассмотрен метод кодирования цветовой видеoinформации на основе принципа постоянной цветовой яркости (ПЦЯ), реализации которого имеют преимущества перед используемыми на данный момент реализациями ПЯ. Проведен сравнительный анализ, который показал преимущества принципа ПЦЯ при кодировании цветовой видеoinформации. Высокая вычислительная сложность существующей реализации ПЦЯ затрудняет его использование в цифровых системах кодирования видеoinформации. В работе предлагаются быстрые реализации этого принципа, а также их сравнение с существующими реализациями принципа ПЯ.

Статья построена следующим образом. Во втором разделе описан принцип ПЯ, используемый в существующих системах кодирования видеoinформации. В третьем разделе рассмотрены наиболее распространенные форматы прореживания цветоразностных компонентов для сокращения избыточности видеoinформации. Четвертый раздел посвящен анализу недостатков, возникающих в системах, реализованных на основе принципа ПЯ. В пятом разделе описан альтернативный подход, основанный на принципе ПЦЯ. Проведен сравнительный анализ принципов ПЯ и ПЦЯ, на основе которого сделан вывод, что при применении принципа ПЦЯ лучше сохраняется информация о разрешении по яркости и цветности кодируемых видеоизображений. В шестом разделе приведены программные реализации принципов ПЯ и ПЦЯ. Седьмой раздел посвящен сравнительному анализу реализаций указанных принципов, показано, что принцип ПЦЯ имеет преимущества при его реализации на некоторых современных процессорах. В заключении подведены итоги работы.

2. Принцип постоянной яркости. Подробно суть принципа постоянной яркости изложена в [6]. Первичные цветовые сигналы телевизионной камеры E_R , E_G , E_B после введения предсказаний, необходимых для коррекции нелинейных характеристик устройств отображения [6], преобразуются в сигналы $E'_R = E^\gamma_R$, $E'_G = E^\gamma_G$, $E'_B = E^\gamma_B$,

где $\gamma = 0,45$, из которых далее формируются сигналы E'_Y , E'_{R-Y} , E'_{B-Y} . При выборе опорного источника белого C_{6500} (стандартный источник белого цвета, который соответствует полуденному освещению при сплошной облачности) уравнения связи между сигналами E'_R , E'_G , E'_B и сигналами E'_Y , E'_{R-Y} , E'_{B-Y} выглядят следующим образом:

$$\begin{cases} E'_Y = 0,299E'_R + 0,587E'_G + 0,114E'_B, \\ E'_{R-Y} = 0,701E'_R - 0,587E'_G - 0,114E'_B, \\ E'_{B-Y} = -0,299E'_R - 0,587E'_G + 0,886E'_B. \end{cases} \quad (1)$$

Выбор другого опорного источника белого, например D_{6500} (стандартный источник белого цвета, соответствующий излучению абсолютно черного тела при 6504°K), приводит лишь к изменению значений коэффициентов в этих уравнениях.

Сигнал E'_Y содержит информацию о черно-белой составляющей изображения — яркости и называется яркостным сигналом [6]. Этот сигнал обеспечивает совместимость черно-белых и цветных систем телевидения и передается в более широкой полосе частот по сравнению с сигналами E'_{R-Y} , E'_{B-Y} , так как для наблюдения черно-белых изображений необходима максимальная четкость.

Сигналы E'_{R-Y} и E'_{B-Y} несут информацию о цветности изображения, то есть о цветовом тоне и насыщенности.

Реализация принципа ПЯ в цветных вещательных телевизионных системах способствовала их широкому распространению.

3. Прореживание цветоразностных компонентов. Международный комитет по телевидению и радиовещанию (МККР) рекомендовал для цифровых телевизионных студий основной цифровой стандарт с соотношениями полосы для сигналов яркости и цветности 4:2:2.

В цифровых системах обработки видеoinформации также как и в аналоговых применяется прореживание цветоразностных компонентов для сокращения цветовой избыточности, которое приводит к потере видеoinформации.

3.1. Форматы прореживания. Чтобы понять причины искажения изображений, возникающих при усреднении цветоразностных компонентов, рассмотрим наиболее распространенные форматы прореживания цветоразностных сигналов.

Наиболее распространенные форматы прореживания [7] представлены на рисунке 1. Чаще всего при обработке цветовой информации без

потерь применяется формат без прореживания цветоразностных компонентов (4:4:4). Часто встречаются также форматы с прореживанием вдвое по горизонтали (4:2:2) и с прореживанием вдвое как по горизонтали, так и по вертикали (4:2:0). В таких системах цветового кодирования, как PAL, SECAM, NTSC, используются форматы с прореживанием цветоразностных компонентов в четыре раза по горизонтали (4:1:1).

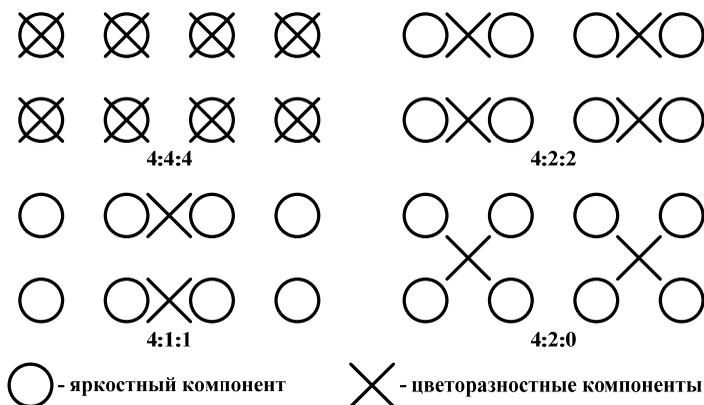


Рис. 1. Форматы прореживания цветоразностных отсчетов

Прореживание цветоразностных сигналов приводит к сокращению потока видеoinформации, что, однако, вызывает потери при передаче цветовой видеoinформации.

4. Недостатки принципа постоянной яркости. Основные недостатки принципа ПЯ: неточное воспроизведение яркости насыщенных цветов, ухудшение четкости в ахроматических деталях и изображениях по мере увеличения насыщенности цветов.

Численно характеризовать потери для различных цветов можно, используя коэффициент постоянной яркости K_L :

$$K_L = L_Y / L, \quad (2)$$

где L_Y — воспроизводимая яркость, L — номинальная яркость.

Рассмотрим случай, когда яркостный сигнал передается по каналу с большей пропускной способностью, чем канал, по которому передаются цветоразностные компоненты. Частотные характеристики системы передачи различных цветов оказываются в этом случае неодинаковыми (рисунок 2). Коэффициент постоянной яркости, заданный формулой (2), зависит от полосы пропускания канала связи.

Анализируя приведенный на рисунке 2 график можно сделать

вывод, что наибольшим потерям подвержены изображения, содержащие красный и синий цвета (29,9% и 11,4% соответственно). Потери для других цветов: пурпурный — 41,3%, зеленый — 58,7%, голубой — 70,1%, желтый — 88,6%.

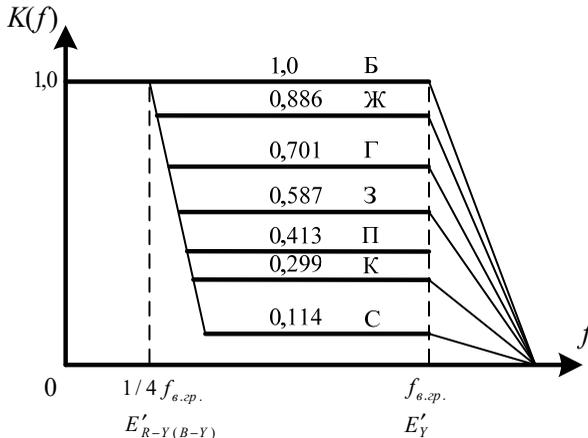


Рис. 2. Частотные характеристики передачи для различных цветов

Потеря видеоинформации из-за преобразований на основе принципа ПЯ недопустима в некоторых системах кодирования: цифровые системы видеонаблюдения [8], цифровые системы улучшенного видения [9], в том числе работающие в условиях низкой освещенности [10], системы компьютерного зрения.

5. Принцип постоянной цветовой яркости. Кодирование цветовой видеоинформации можно осуществить, используя принцип ПЦЯ [6]. Преимущества и недостатки этого принципа рассмотрены в [6, 11].

Прямое цветовое преобразование согласно ПЦЯ можно осуществить следующим образом:

$$\begin{cases} E'_C = \sqrt{(E'_R)^2 + (E'_G)^2 + (E'_B)^2}, \\ E'_1 = E'_R / E'_C, \\ E'_2 = E'_B / E'_C, \end{cases} \quad (3)$$

где E'_C — цветовая амплитуда передаваемого цвета, E'_1 , E'_2 — насыщенность данного цвета красным и синим первичными цветами соответственно.

Обратное преобразование требует получения восстановленных гамма-корректированных цветов — красного и синего:

$$\begin{aligned} E'_R &= E'_C E'_1, \\ E'_B &= E'_C E'_2. \end{aligned} \quad (4)$$

Восстановление зеленого цвета осуществляется по формуле:

$$E'_G = \sqrt{(E'_C)^2 - (E'_R)^2 - (E'_B)^2}. \quad (5)$$

Черно-белые изображения при кодировании с использованием принципа ПЦЯ передаются с полной четкостью, для них $E'_1 = E'_2 = const$, а все изменения передаются по более широкополосному каналу. При кодировании цветных изображений по более узкополосным каналам передается информация о количестве цветных деталей синими и красными цветами, а также изменения пурпурной составляющей цвета.

Использование принципа ПЦЯ позволит повысить эффективность кодирования цветовой информации для лучшего ее восприятия зрительной системой человека [6].

5.1. Сравнение по уровню искажений цветовых компонентов.

Сравнение уровня искажений при потере цветовой информации будем проводить с помощью общепринятой метрики [12] отношения пика сигнала к шуму (PSNR — peak signal-to-noise ratio). Для наглядности построим график зависимости уровня искажения изображения от количества отбрасываемых уровней цветоразностных отсчетов. Такой подход соответствует ранее описанным процедурам прореживания сигналов и дальнейшего кодирования компонентов современными стандартами сжатия видеoinформации [13].

Для сравнительного анализа преобразований предлагается следующий подход. Исходное изображение в формате RGB переводится в яркостные и цветоразностные составляющие с помощью преобразования, согласно принципам ПЯ или ПЦЯ, затем проводится скалярное квантование цветоразностных (цветностных) отсчетов, после чего — обратное преобразование в RGB. Для исходного и полученного RGB-изображений будем использовать метрику PSNR по каждой из составляющих в отдельности (R, G, B).

Стоит отметить, что большинство современных стандартов кодирования видеoinформации после цветового преобразования работают с сокращенным в 1,5–2 раза (в зависимости от формата прореживания) потоком, что равноценно квантованию и восстановлению отсчетов с коэффициентом 2. То есть, применяя прореживание согласно формату 4:2:2, цифровой поток сокращается в 1,5 раза относительно входного, при использовании формата 4:2:0 и 4:1:1 — в 2 раза. Также стоит отметить, что описанный подход характерен для большинства стандартов

сжатия видеоинформации с потерями, где устранение пространственно-временной избыточности достигается путем квантования трансформант с коэффициентами от 2 и более. Далее, после квантования трансформант в большинстве случаев выполняется сжатие полученного потока видеоинформации без потерь.

Для оценки взято изображение «flower_foveon» [14], насыщенное красным и синим цветами. Графики на рисунках 3-5 показывают зависимость уровня искажений восстановленных изображений (относительно оригинала) от коэффициента квантования для красного (3), зеленого (4) и синего (5) цветов.

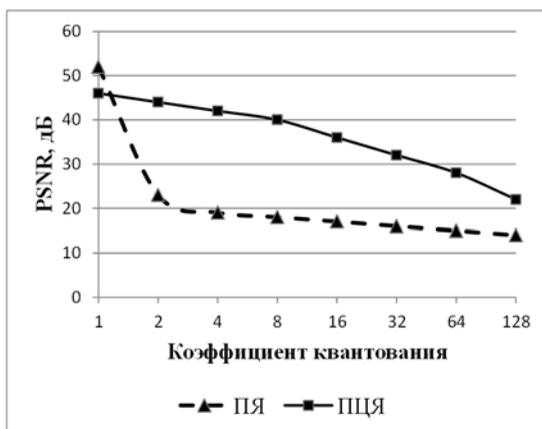


Рис. 3. Зависимость уровня искажения красного цвета восстановленного изображения от коэффициента квантования

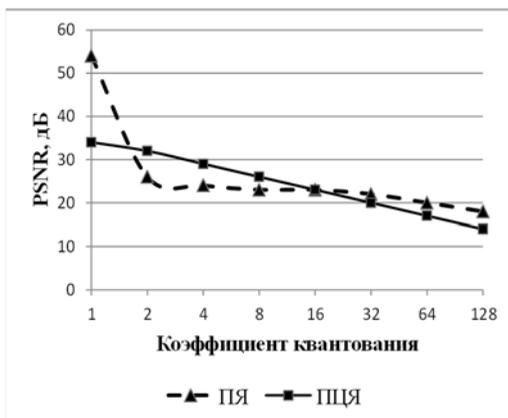


Рис. 4. Зависимость уровня искажения зеленого цвета восстановленного изображения от коэффициента квантования

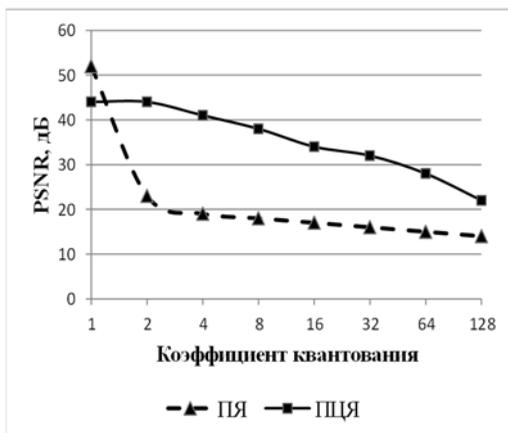


Рис. 5. Зависимость уровня искажения синего цвета восстановленного изображения от коэффициента квантования

Анализируя приведенные графики, можно заметить, что при коэффициентах квантования больше 2 применение цветового преобразования на основе принципа ПЦЯ дает выигрыш по качеству передаваемых цветов, а следовательно, сохраняет и четкость изображений. Значения PSNR показывают, что при обнулении значений составляющих цветоразностных сигналов принцип ПЦЯ меньше искажает обрабатываемое изображение, чем принцип ПЯ.

При использовании принципа ПЦЯ выигрыш при коэффициенте квантования 2 составил: 24,2 дБ на красном цвете, 6,1 дБ на зеленом цвете и 25,0 на синем цвете. Выигрыш применения ПЦЯ при коэффициенте квантования 16 составил: 20,6 дБ на красном цвете, 0,1 дБ на зеленом цвете и 18,2 дБ на синем цвете. При коэффициентах квантования от 32 и более на зеленом цвете проигрыш составил не более 1 дБ.

6. Реализации принципов постоянной яркости и постоянной цветовой яркости. Рассмотрим реализации принципов ПЯ и ПЦЯ.

Пусть R, G, B — оцифрованные первичные основные цвета телевизионной камеры (видеосенсора). Y, C_b, C_r — цифровое представление величин E'_Y, E'_{R-Y}, E'_{B-Y} соответственно. W — целочисленное значение, соответствующее цветовой амплитуде передаваемого цвета (E'_C), P — передает значение насыщенности цвета красным ос-

новным первичным цветом (E'_1), Q — передает значение насыщенности цвета синим основным первичным цветом (E'_2). Разрядность входных и выходных отсчетов — восемь бит, то есть отсчеты представляют собой целочисленные значения в диапазоне от 0 до 255.

6.1. Принцип постоянной яркости. Одна из наиболее быстродействующих целочисленных реализаций согласно формуле (1) предложена в стандарте [15]:

$$\begin{cases} Y = (77 \cdot R + 150 \cdot G + 29 \cdot B) \gg 8; \\ C_b = (-44 \cdot R - 87 \cdot G + 131 \cdot B) \gg 8; \\ C_r = (131 \cdot R - 110 \cdot G - 21 \cdot B) \gg 8, \end{cases} \quad (6)$$

где \gg — операция логического сдвига вправо.

6.2. Принцип постоянной цветовой яркости. Разработаем вещественную реализацию принципа ПЦЯ для цифровых систем обработки видеoinформации. Реализация математических преобразований на языке Си++ приведена в листинге 1.

Первый этап вычисления — нормализация каждого из входных отсчетов к уровню от 0 до 1, которая осуществляется делением значения на константу 255 (см. строки 16-18 листинга 1). Далее с полученными нормализованными значениями вычисления производятся в вещественной арифметике. Вычислим промежуточные значения (строки 19-21 листинга 1):

$$\begin{aligned} w &= \sqrt{r^2 + g^2 + b^2}, \\ q &= r / w, \\ p &= b / w, \end{aligned} \quad (7)$$

где w — вещественное число, определяющее значение цветовой амплитуды, q, p — вещественные числа, определяющие значения насыщенности цвета красным и синим первичными основными цветами соответственно, r, g, b — нормализованные вещественные значения первичных основных цветов.

```

1: void ccl_forward_444_float(unsigned char *RGB24, // Входной поток R,G,B-отчетов
2:   unsigned char *CCL8, // Выходной поток W,U,V-отчетов
3:   unsigned short width, // Ширина кадра
4:   unsigned short height) { // Высота кадра
5:
6:   #define Clip3(MinVal, MaxVal, a) ((a)<(MinVal)) ? (MinVal) : \
7:     ((a)>(MaxVal)) ? (MaxVal) :(a)
8:
9:   unsigned char *line1, *w, *p, *q;
10:  w = CCL8;
11:  p = CCL8 + width * height;
12:  q = p + width * height;
13:  line1 = RGB24;
14:  for ( int c = 0; c < height * width; ++c ) { // Проход по всем пикселям изображения
15:    float r, g, b, W, P, Q;
16:    r = (float) line1[0] / 255.0; // Нормализация красной компоненты
17:    g = (float) line1[1] / 255.0; // Нормализация зеленой компоненты
18:    b = (float) line1[2] / 255.0; // Нормализация синей компоненты
19:    W = sqrt ( r * r + g * g + b * b ); // Вычисление вещественного значения w
20:    P = (W > 0)? (b / W) : (1 / sqrt( 3.0 )); // Вычисление вещественного значения
21:    P
22:    Q = (W > 0)? (r / W) : (1 / sqrt( 3.0 )); // Вычисление вещественного значения
23:    q
24:    // Приведение значений к диапазону [0:255]
25:    w[c] = (unsigned char) Clip3(0,255,255.0 * ( W / sqrt( 3.0 ) ) );
26:    p[c] = (unsigned char) Clip3(0,255,255.0 * P);
27:    q[c] = (unsigned char) Clip3(0,255,255.0 * Q);
28:    line1 += 3;
  }
}

```

Листинг 1. Процедура прямого преобразования согласно ПЦЯ

При $w = 0$ значения $p = q = 1 / \sqrt{3}$ (строки 19-20 листинга 1). Для получения из выражения (7) выходных 8-разрядных значений W , P , требуется привести их к диапазону от 0 до 255 (строки 23-25 листинга 1):

$$\begin{aligned}
 W &= \lfloor w / \sqrt{3} \rfloor, \\
 Q &= \lfloor q \times 255 \rfloor, \\
 P &= \lfloor p \times 255 \rfloor.
 \end{aligned}
 \tag{8}$$

Обратное преобразование выполняется похожим способом (листинг 2). Из отсчетов W , P , Q получают нормализованные (от 0 до 1) промежуточные вещественные отсчеты (строки 16-18 листинга 2):

$$\begin{aligned}
 w &= W \times \sqrt{3} / 255, \\
 q &= Q / 255, \\
 p &= P / 255.
 \end{aligned}
 \tag{9}$$

```

1: void ccl_backward_444_float(unsigned char *CCL8, // Входной поток W,U,V-
2: отчетов
3:     unsigned char *RGB24, // Выходной поток R,G,B-
4: отчетов
5:     unsigned short width, // Ширина кадра
6:     unsigned short height) { // Высота кадра
7:
8: #define Clip3(MinVal, MaxVal, a) ((a)<(MinVal)) ? (MinVal) : \
9: ((a)>(MaxVal)) ? (MaxVal) :(a)
10:
11: unsigned char *line1, *w, *p, *q;
12: w = CCL8;
13: p = CCL8 + width * height;
14: q = p + width * height;
15: line1 = RGB24;
16: for ( int c = 0; c < height * width; ++c ) { // Проход по всем пикселям изображе-
17: ния
18:     float r, g, b, W, P, Q;
19:     // Вычисление значения w, q, p
20:     W = (float)w[c] * sqrt( 3.0 ) / 255.0;
21:     P = p[c] / 255.0;
22:     Q = q[c] / 255.0;
23:     b = P * W;
24:     r = Q * W;
25:     g = W * W - ( r*r + b*b );
26:     g = ( g >= 0 )? sqrt(g) : 0;
27:
28:     line1[0] = (unsigned char) Clip3(0,255,( r * 255.0 ));
29:     line1[1] = (unsigned char) Clip3(0,255,( g * 255.0 ));
30:     line1[2] = (unsigned char) Clip3(0,255,( b * 255.0 ));
31:
32:     line1 += 3;
33: }
}

```

Листинг 2. Процедура обратного преобразования согласно ПЦЯ

Значения, полученные в выражении (9), применяются для вычисления вещественных чисел восстановленных значений (строки 20-22 листинга 2):

$$\begin{aligned}
 r_r &= q \times w, \\
 b_r &= p \times w, \\
 g' &= w^2 - (r_r^2 + b_r^2),
 \end{aligned}
 \tag{10}$$

где r_r , b_r — восстановленные нормализованные вещественные значения выходных отсчетов гамма-корректированных цветов.

Восстановленное нормализованное значение выходного отсчета зеленого цвета вычисляется следующим образом (строка 23 листинга 2):

$$g_r = \begin{cases} 0, & \text{при } g' < 0; \\ \sqrt{g'}, & \text{при } g' \geq 0. \end{cases} \quad (11)$$

Восстановленные компоненты можно получить, используя значения, рассчитанные в (10) и (11), выполнив следующие вычисления (строки 25-27 листинга 2):

$$\begin{aligned} R_r &= \lfloor r_r \times 255 \rfloor \\ G_r &= \lfloor g_r \times 255 \rfloor, \\ B_r &= \lfloor b_r \times 255 \rfloor \end{aligned} \quad (12)$$

где R_r , G_r , B_r — восстановленные 8-разрядные беззнаковые целочисленные значения цветовых компонентов гамма-корректированных цветов.

Вещественная реализация служит для демонстрации работы принципа ПЦЯ, однако является наиболее сложной с точки зрения вычислительных операций.

6.3. Целочисленные реализации принципа постоянной цветовой яркости. К целочисленным системам можно отнести современные процессоры некоторых архитектур (без использования арифметико-логического устройства (АЛУ), выполняющего вещественные операции) и программируемые логические интегральные схемы (ПЛИС).

В данной работе авторами предлагается *табличная реализация* принципа ПЦЯ — наиболее быстродействующая целочисленная реализация при наличии достаточного объема быстродействующей оперативной памяти. Рассмотрим эту реализацию более подробно.

Хранение отсчетов W , P , Q потребует $C=256^3=16777216$ байт (16 МБайт) для каждого из массивов. Однако большинство вычислительных модулей оперируют типами данных, кратными 2^3 — 8, 16, 32 или 64 разрядам, исходя из этого, обращаясь к памяти, наиболее эффективно получать сразу три байта, а не извлекать каждый из отсчетов по отдельности. Поэтому для уменьшения числа операций доступа к па-

мента наиболее целесообразно использовать четырехбайтный тип данных (unsigned integer). Тогда для хранения отсчетов преобразования требуется суммарно 64 МБайта памяти.

Тогда кодирование можно осуществить как получение одного четырехбайтного числа (WQP) из таблицы (wpq) по координатам (R, G, B):

$$WQP = wpq[R][G][B]. \quad (13)$$

Восстановить принятые значения R, G, B можно аналогичным способом, применив таблицу заранее рассчитанных значений (rgb). Извлечение значений R, G, B в зависимости от полученных значений W, V, U проводится следующим образом:

$$RGB = rgb[W][X][Y]. \quad (14)$$

Описанный согласно формулам (13-14) целочисленный метод кодирования и декодирования обеспечивает наибольшее быстродействие для современных процессоров при условии быстрой работы оперативной памяти устройства и ее достаточном объеме.

Требуется предварительная подготовка таблиц кодирования и декодирования при использовании табличной реализации, которая проводится на процессоре с возможностью выполнения вещественных операций. Фрагмент программного кода для формирования значения WQF с целью заполнения таблицы wpq приведен в листинге 3. Значения W, P, Q в листинге 3 вычисляются аналогично тому, как это реализовано в листинге 1. Заполнение таблицы rgb производится похожим образом.

```

1: // R, G, B – целочисленные значения в диапазоне [0;255]
2: unsigned int wpq[256][256][256]; // [R][G][B]
3: int WQP = (unsigned char) Clip3( 0, 255, W / sqrt( 3.0 ) );
4: WQP <<= 8;
5: WQP |= (unsigned char) Clip3( 0, 255, P );
6: WQP <<= 8;
7: WQP |= (unsigned char) Clip3( 0, 255, Q );
8: wpq[R][G][B] = WQP;

```

Листинг 3. Подготовка массива WQP

Разработаем *оптимизированную табличную реализацию принципа ПЦЯ*. Для этого требуется создать таблицу, в которой содержатся значения W и обратные им 24-разрядные значения $invW = 2^{24} / W$ (для сохранения точности вычислений требуется умножение на число 2^{24}). Разместим W и $invW$ в одном четырехбайтном числе:

$$W' = ((W \times 2^{24}) \vee \text{Inv}W), \quad (15)$$

где \vee — операция побитового «или». Умножение на 2^{24} можно заменить логическим сдвигом числа влево на 24 разряда.

Кодирование с применением сокращенной таблицы отсчетов осуществляется следующим образом. Требуется получить число, хранящееся в таблице по координатам $(R^2 + G^2 + B^2)$, и затем получить значения W и $\text{inv}W$:

$$\begin{aligned} W &= (WW[R^2 + G^2 + B^2]) / 2^{24}, \\ \text{inv}W &= (WW[R^2 + G^2 + B^2] \times 2^8) / 2^8, \end{aligned} \quad (16)$$

где WW — массив заранее вычисленных значений W' для всех R, G и B , операции умножения и деления на 2^n реализуются как логические сдвиги вправо и влево на n . Значение W , полученное согласно (16), записывается в поток выходных видеоданных. Для расчета значений P, Q используется полученное в (16) значение $\text{inv}W$:

$$\begin{aligned} Q &= (R \times \text{inv}W) / 2^{24}, \\ P &= (B \times \text{inv}W) / 2^{24}. \end{aligned} \quad (17)$$

Аналогичные действия выполняют при декодировании значений R, G и B .

Для хранения массива WW в таком случае потребуется $256^2 \times 3 \times 4 = 768432$ байта (768 Кбайт), однако для расчетов требуются операции умножения и логического сдвига. Следовательно, описанный метод предпочтителен для систем обработки видеoinформации с ограниченным объемом оперативной памяти и вычислительными ресурсами, достаточными для проведения умножений и логических сдвигов.

7. Сравнительный анализ реализаций принципов постоянной яркости и постоянной цветовой яркости. Сравнение реализаций принципов будем проводить, оценивая число требуемых для вычислений машинных тактов. Основные функциональные блоки вычислительных модулей системы обработки, участвующие при вычислении: операции сложения/вычитания, умножения, сдвиги и обращения к памяти. Для большинства процессоров можно найти задержки, вызываемые выполнением той или иной команды обработки данных. Используем следующие задержки при вычислениях:

ADD — для операции сложения,

SHIFT — для операции логического сдвига,

MULTIPLY — для операции умножения;
и при обращении к памяти:
MEM — при обращении к ОЗУ;
CACHE — при обращении к быстрой памяти (кэш-памяти), расположенной в чипе процессора.

7.1. Сравнение целочисленных реализаций принципа постоянной цветовой яркости. При кодировании согласно табличной реализации для получения отсчетов W , P , Q нужно однократно обратиться к памяти по координатам (R, G, B) , что потребует трех сложений и двух логических сдвигов, то есть $TBL = 2 \times SHIFT + 3 \times ADD + MEM$ операций, что объясняется способом хранения данных в оперативной памяти вычислителя — трехмерный массив раскладывается по строкам, следующим друг за другом.

Проведем оценку числа тактов для оптимизированной реализации, требуемых при вычислении отсчетов W , P , Q . Для получения значения W требует вычислить индекс $i = R^2 + G^2 + B^2$, который используется при обращении к памяти ($w = WW[i]$), и затем сдвинуть полученное значение вправо на 24 разряда ($W = w \gg 24$). Суммарная задержка для выполнения этих операций составит $(3 \times MULTIPLY + 2 \times ADD + MEM + SHIFT)$ тактов.

Для вычисления компонентов P и Q требует выполнить два логических сдвига, чтобы нормализовать значение WW , и затем требуется выполнить одно умножение и один логический сдвиг для каждого из компонентов, что составляет $2 \times SHIFT + 2 \times (MULTIPLY + SHIFT)$ тактов.

Итого, вычисление компонентов W , P , Q потребует $OPT_{MEM} = (5 \times MULTIPLY + 2 \times ADD) + MEM + 5 \times SHIFT$ тактов, что превышает число тактов при реализации табличного метода кодирования. При размещении всей таблицы WW в кэш-памяти процессора потребуется $OPT_{CACHE} = (5 \times MULTIPLY + 2 \times ADD) + CACHE + 5 \times SHIFT$ тактов, что уменьшит задержку при вычислениях.

Как правило, в вычислительных системах величина задержки обращения к внешней памяти в 10 (и более) раз превышает величину задержки обращения к быстрой памяти.

Оптимизированную реализацию целесообразно использовать при наличии требуемого объема быстрой памяти процессора и при выполнении соотношения $TBL \leq OPT_{CACHE}$.

Например, для процессора Intel Atom 330 [16] операция умножения требует 3 такта, сложения — 1 такт, чтения из быстрой памяти —

1 такт, логического сдвига — 1 такт. Значение $OPT_{CACHE}=23$ тактам. Задержка обращения к внешней памяти определяется аппаратной реализацией системы и может составлять от 12 и более тактов ($TBL=11$). При $MEM = 6$ для такой системы предпочтительнее табличная реализация, при $MEM \geq 12$ — оптимизированная реализация.

7.2. Сравнение реализации принципа ПЯ с реализациями принципа ПЦЯ. Оценим число операций, требуемых для кодирования одного пикселя изображения согласно выражению (7). Число тактов для вычисления каждого из значений Y , C_b , C_r равно $L = 3 \times MULTIPLY + 2 \times ADD + SHIFT$. Суммарно потребуется $3 \times L$ тактов для кодирования одного пикселя, что составляет $YCbCr = 9 \times MULTIPLY + 6 \times ADD + 3 \times SHIFT$ тактов.

Сравним производительность, выраженную числом тактов, для следующих моделей процессоров: AMD Kabini, Intel Core i7, Intel Atom 330, Cortex-A57, используя задержки при выполнении операций для процессоров из [16, 17] (таблица 1).

Таблица 1. Число тактов при выполнении операций

Наименование процессора	<i>ADD</i>	<i>SHIFT</i>	<i>MULTIPLY</i>	<i>MEM</i>	<i>CACHE</i>
AMD Kabini	1	1	3	4	1
Intel Core i7	1	1	5	3	1
Intel Atom 330	1	1	6	1	1
Cortex-A57	1	1	3	4	4

Для удобства сведем сравнение числа тактов в таблицу (таблица 2). Сравнивая значения, приведенные в таблице 2, можно сделать вывод, что табличная реализация — наиболее быстрая, оптимизированная реализация принципа ПЦЯ выигрывает у целочисленной реализации принципа ПЯ по числу требуемых для вычислений тактов как при использовании быстрой памяти процессора (*CACHE*), так и при использовании ОЗУ.

Таблица 2. Сравнение числа тактов, требуемых при реализации ПЯ и ПЦЯ

Наименование процессора	ПЦЯ			ПЯ
	<i>TBL</i>	OPT_{MEM}	OPT_{CACHE}	<i>YCbCr</i>
AMD Kabini	9	26	23	36
Intel Core i7	8	35	33	54
Intel Atom 330	6	38	38	63
Cortex-A57	9	26	26	36

8. Заключение. Принцип ПЯ, используемый в современных системах обработки видеoinформации, не свободен от ряда недостатков, например: уменьшение яркости насыщенных цветов и ухудшение четкости цветных изображений. Поэтому применение этого принципа невозможно в некоторых системах, которые не допускают потери информации о первичных цветах.

К потерям в четкости изображений также приводит прореживание цветоразностных сигналов согласно форматам 4:2:2, 4:2:0, 4:1:1.

Улучшить качество систем передачи видеoinформации можно, применив принцип ПЦЯ, использование которого позволяет повысить разрешение декодированных изображений.

В работе проведен сравнительный анализ преобразованных изображений согласно принципам ПЯ и ПЦЯ, который показал преимущество использования принципа ПЦЯ с точки зрения сохранения информации о цветовом разрешении кодируемых данных.

В работе рассмотрены варианты реализации принципа ПЦЯ для вещественных и целочисленных аппаратных средств обработки видеоданных. Описан наиболее быстродействующий метод табличного кодирования и декодирования согласно принципу ПЦЯ. Произведено сокращение количества хранимых данных для табличного кодирования. Оптимизированный метод табличного кодирования согласно принципу ПЦЯ требует в 85,33 раза меньший объем памяти, однако при этом необходимы операции умножения и логические сдвиги.

Проведен сравнительный анализ табличного и оптимизированного методов для выбора реализации принципа ПЦЯ в зависимости от конфигурации аппаратной платформы.

Целочисленная оптимизированная реализация принципа ПЦЯ выигрывает по числу требуемых тактов при вычислении значений у целочисленной реализации принципа ПЯ. В системах, где критична задача четкости цветовой информации и требуется улучшенная цветопередача, наряду с невысокой вычислительной сложностью, предпочтительна оптимизированная целочисленная реализация принципа ПЦЯ. Применение целочисленных реализаций принципа ПЦЯ целесообразно также в системах комплексирования информации [9], где требуется передача улучшенного видеоизображения при сохранении приемлемой вычислительной сложности.

Практическое применение принципа ПЦЯ осуществлено, в частности, при создании системы управления беспроводной передачей энергии [18] к исполнительным устройствам крупногабаритной трансформируемой антенны космического базирования [19], где используется

двухдиапазонная видеосистема в качестве элемента «технического зрения» для наведения лазерного излучения на фотоприемник. В указанном случае принцип постоянной цветовой яркости рационально сочетается с комплексированием видеосигналов от датчиков видимого и инфракрасного диапазонов [9].

Работа выполнена в рамках реализации Федеральной целевой программы «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2014–2020 годы», Министерство образования России, соглашение № 14.577.21.0201 (уникальный идентификатор ПНИЭР RFMEFI 57715X0201): «Создание высокоэффективной системы беспроводной узконаправленной передачи энергии и информации для управления состоянием объектов космического базирования на основе лазерных и волоконнооптических технологий». Работа выполнена на оборудовании ЦКП «Центр радиотехнических и оптических измерений».

Литература

1. *Кривошеев М.И., Кустарев А.К.* Световые измерения в телевидении // М.: «Связь». 1973. 224 с.
2. *Мешков В.В., Матвеев А.Б.* Основы светотехники. Часть 2. Физическая оптика и колориметрия // М.: Энергоатомиздат. 1989. 432 с.
3. *Гонсалес Р., Вудс Р.* Цифровая обработка изображений: издание 3-е, исправленное и дополненное, пер. с англ. // М.: Техносфера. 2012. 1104 с.
4. *Кочин Л.Б.* Методы и средства отображения цветовой видеoinформации // Балт. гос. техн. ун-т. СПб. 2012. 268 с.
5. *Полосин Л.Л.* Представление цветов в метрическом векторном пространстве // Известия вузов России. Радиоэлектроника. 2013. Вып. 2. С. 54–61.
6. *Полосин Л.Л.* Принцип постоянной цветовой яркости в телевидении // Телевизионная техника и связь. Спецвыпуск, посвященный 50-летию НИИТ. 1995. С. 28–37.
7. *Ричардсон Я.* Видеокодирование. H.264 и MPEG-4 — стандарты нового поколения // М.: Техносфера. 2005. 368 с.
8. *Уваров Н.Е.* Цифровая обработка изображений в телевизионных системах наблюдения и охраны // ССТV фокус. 2004. № 3. С. 56–61.
9. *Борейшо А.С. и др.* Комплексирование в оптоэлектронных системах // Инженерно-физический журнал. 2016. Т. 89. № 3. С. 736–741.
10. *Стремоухов М.В., Илюшин М.В., Дворяжкин В.В.* Кодирование изображений с малым динамическим диапазоном на основе нелинейной формирующей системы // Вестник Рязанского государственного радиотехнического университета. 2016. № 58. С. 117–122.
11. *Сухов Т.М., Беляев Е.А.* Использование принципа постоянной цветовой яркости для передачи видеoinформации // Известия вузов. Приборостроение. 2011. Т. 1. С. 55–59.
12. *Куценко С.А., Немцова С.Р., Мкртумов А.С., Цветков О.В.* Теоретико-информационная характеристика качества изображений при их цифровом сжатии // Труды СПИИРАН. 2013. Вып. 8(31). С. 190–202.
13. *Зубакин И.А., Ибатуллин С.М., Фахми Ш.С., Цыцулин А.К.* Измерение качества видеoinформации на выходе цифрового кодера источника и его эффективности // Вопросы радиоэлектроники. Серия: Техника телевидения. 2016. № 4(27). С. 33–39.

14. Image Compression. Benchmark. The New Test Images. URL: http://imagecompression.info/test_images/ (дата обращения: 01.06.2017).
15. RECOMMENDATION ITU-R BT.601-7. Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios. URL: http://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.601-7-201103-I!!PDF-E.pdf (дата обращения: 07.06.2017).
16. Instruction tables. URL: http://www.agner.org/optimize/instruction_tables.pdf (дата обращения: 07.06.2017).
17. Cortex-A57 Software Optimization Guide. URL: http://infocenter.arm.com/help/topic/com.arm.doc.uan0015b/Cortex_A57_Software_Optimization_Guide_external.pdf (дата обращения: 07.06.2017).
18. *Mamseev С.А. и др.* Организация энергоинформационного обмена между устройствами управления формой трансформируемой антенны с применением волоконно-оптических технологий // Оптический журнал. 2016. Т. 83. № 11. С. 73–78.
19. *Бей Н.А., Зимин В.Н.* Трансформируемые антенны больших размеров для геостационарных космических аппаратов. // Антенны. 2005. Вып. 10(101). С. 24–27.

Сухов Тимофей Михайлович — ассистент кафедры радиоэлектронных систем управления, Балтийский государственный технический университет «ВОЕНМЕХ» им. Д.Ф. Устинова (БГТУ «ВОЕНМЕХ» им. Д.Ф. Устинова). Область научных интересов: кодирование цветовой информации, сжатие потока видеоданных, комплексирование и интегрирование данных датчиков различных частотных диапазонов. Число научных публикаций — 25. tsuhov@mail.ru; 1-я Красноармейская ул., 1, Санкт-Петербург, 199005; р.т.: +7(812)495-7753.

Страхов Сергей Юрьевич — д-р техн. наук, доцент, декан факультета информационных и управляющих систем, Балтийский государственный технический университет «ВОЕНМЕХ» им. Д.Ф. Устинова (БГТУ «ВОЕНМЕХ» им. Д.Ф. Устинова). Область научных интересов: радиофизика, лазерные системы, оптико-электронные системы. Число научных публикаций — 79. Strakhov_s@mail.ru; 1-я Красноармейская ул., 1, Санкт-Петербург, 199005; р.т.: +7(812)317-82-49.

Кочин Леонид Борисович — к-т техн. наук, доцент, доцент кафедры радиоэлектронных систем управления, Балтийский государственный технический университет «ВОЕНМЕХ» им. Д.Ф. Устинова (БГТУ «ВОЕНМЕХ» им. Д.Ф. Устинова). Область научных интересов: телевизионные системы, обработка и передача изображений, колориметрия. Число научных публикаций — 35. klen68@mail.ru; 1-я Красноармейская ул., 1, Санкт-Петербург, 199005; р.т.: +7(812)-4957753, Факс: +7(812)-3162409.

Беляев Евгений Александрович — к-т техн. наук, постдоктор лаборатории кодирования и передачи визуальной информации, Датский технический университет. Область научных интересов: сжатие и передача видеoinформации. Число научных публикаций — 64. evbel@fotonik.dtu.dk; Ørstedes Plads, building 343, room 106, Kgs. Lyngby, 2800, Denmark; р.т.: +358449179895.

Поддержка исследований. Работа выполнена в рамках реализации Федеральной целевой программы «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2014–2020 годы», Министерство образования России, соглашение № 14.577.21.0201 (уникальный идентификатор ПНИЭР RFMEFI 57715X0201): «Создание высокоэффективной системы беспроводной узконаправленной передачи энергии и информации для управления состоянием объектов космического базирования на основе лазерных и волоконнооптических технологий». Работа выполнена на оборудовании ЦКП «Центр радиотехнических и оптических измерений».

T.M. SUKHOV, S.YU. STRAKHOV, L.B. KOCHIN, E.A. BELYAEV
**APPLICATION AND IMPLEMENTATIONS OF THE
 PRINCIPLE OF CONSTANT COLOR LUMINANCE IN DIGITAL
 VIDEO CODING SYSTEMS**

Sukhov T.M., Strakhov S.Yu., Kochin L.B., Belyaev E.A. Application and Implementations of the Principle of Constant Color Luminance in Digital Video Coding Systems.

Abstract. Modern hardware systems of processing the video data stream for color coding apply the principle of constant brightness proposed in the development of the NTSC color coding system. This principle, like its implementation, is not free from drawbacks: loss of information on the clarity of the encoded color images, degradation of clarity in achromatic details and images as the color saturation increases, etc. In addition, the use of video data decoding formats in digital video data processing systems, such as 4:2:2, 4:2:0, 4:1:1, distorts the decoded video image.

An alternative approach for encoding a color video stream is to apply the principle of constant color luminance. The work describes the coding according to the principle of constant color luminance. A comparative analysis of the transformed images is carried out with the help of the two principles given. The advantage of applying the principle of constant color brightness in digital video coding systems is shown. It is shown that using the principle of constant color brightness it is possible to obtain a gain of more than 6 dB.

The implementation of the principle of constant color luminance for real and integer modern hardware platforms is described. A comparative analysis of the realizations of the principles of constant brightness and constant color luminance was carried out, showing the advantage of applying the principle of constant color luminance for some modern processors.

The application of the principle of constant color luminance in digital video encoding systems can help improve the quality of recoverable color coded images.

Keywords: principle of constant brightness, the principle of constant color luminance, integer color transformations, comparison of color transformations.

Sukhov Timofey Mikhailovich — assistant of radio-electronic control systems department, Baltic State Technical University "Voenmeh" D.F.Ustinov. Research interests: encoding of color information, compression of video data stream, integration multispectral sensor data. The number of publications — 25. tsuhov@mail.ru; 1, 1-st Krasnoarmeyskaya, St.-Petersburg, 190005, Russia; office phone: +7(812)495-7753.

Strakhov Sergey Yurievich — Ph.D., Dr. Sci., associate professor, dean of information and control systems faculty, Baltic State Technical University "Voenmeh" D.F.Ustinov. Research interests: radiophysics, laser systems, electro-optical systems. The number of publications — 79. Strakhov_s@mail.ru; 1, 1-st Krasnoarmeyskaya, St.-Petersburg, 190005, Russia; office phone: +7(812)317-82-49.

Kochin Leonid Borisovich — Ph.D., associate professor, associate professor of of radio-electronic control systems department, Baltic State Technical University "Voenmeh" D.F.Ustinov. Research interests: television systems, processing and transfer of images, colorimetric. The number of publications — 35. klen68@mail.ru; 1, 1-st Krasnoarmeyskaya, St.-Petersburg, 190005, Russia; office phone: +7(812)-4957753, Fax: +7(812)-3162409.

Belyaev Evgeny Aleksandrovich — Ph.D., postdoctoral researcher of coding and visual communication laboratory, Technical University of Denmark. Research interests: video coding and

transmission. The number of publications — 64. e_beljaev@mail.ru; building 343, room 106, Ørstedes Plads, Kgs. Lyngby, 2800, Denmark; office phone: +358449179895.

Acknowledgements. The work was carried out within the framework of the implementation of the Federal Target Program «Research and Development in Priority Areas for the Development of the Russian Science and Technology Complex for 2014–2020», the Ministry of Education of Russia, agreement No. 14.577.21.0201 (unique identifier of the RFMEFI 57715X0201): «Creating a highly efficient wireless system Focused transmission of energy and information for managing the state of space-based objects based on laser and fiber-optic technologies». The work was performed on the equipment of the Center for Radio-Technical and Optical Measurements.

References

1. Krivosheev M.I., Kustarev A.K. *Svetovyye izmerenija v televidenii* [Light measurements in television]. Moscow: «Svjaz». 1973. 224 p. (In Russ.).
2. Meshkov V.V., Matveev A.B. *Osnovy svetotekhniki, ch. 2. Fizicheskaja optika i kolorimetrija* [Fundamentals of lighting, Part 2. Physical optics and colorimetry]. Moscow: Jenergoatomizdat. 1989. 432 p. (In Russ.).
3. Gonzalez R.C., Woods R.E. *Digital image processing: 3rd ed.* Prentice Hall. 2007. 954 p. (Russ. ed.: Gonsales R., Vuds R. *Tsifrovaia obrabotka izobrazhenii. Izdanie 3-e, ispravlennoe i dopolnennoe.* Moscow: Tekhnosfera Publ. 2012. 1104 p.).
4. Kochin L.B. *Metody i sredstva otobrazhenija cvetovoj videoinformacii* [Methods and means of displaying color video information] Balt. gos. tehn. un-t. SPb. 2012. 268 p. (In Russ.).
5. Polosin L.L. [The representation of colors in a metric vector space]. *Izvestija vuzov Rossii. Radioelektronika — Proceedings of the Russian Universities: Radioelectronics.* 2013. vol. 2. pp. 54–61. (In Russ.).
6. Polosin L.L. [The principle of constant color brightness in television]. *Televizionnaja tehnika i svjaz'. Specvypusk, posvjashhennyj 50-letiju NIIT — Television technology and communication. Special issue dedicated to the 50th anniversary of NIIT.* 1995. pp. 28–37. (In Russ.).
7. Richardson I.E. *H. 264 and MPEG-4 video compression: video coding for next-generation multimedia.* John Wiley & Sons. 2004. 306 p. (Russ. ed.: Richardson Ya. *Videokodirovanie. H.264 i MPEG-4 — standarty novogo pokolenija.* Moscow: Tehnosfera. 2005. 368 p.).
8. Uvarov N.E. [Digital image processing in television surveillance and security systems]. *CCTV focus — CCTV focus.* 2004. vol. 3. pp. 56–61. (In Russ.).
9. Borejsho A.S. et al. [Integration in optoelectronic systems]. *Inzhenerno-fizicheskij zhurnal — Journal of Engineering Physics and Thermophysics.* 2016. vol. 89. no. 3. pp. 736–741. (In Russ.).
10. Stremouhov M.V., Iljushin M.V., Dvorjadkin V.V. [Coding of images with a small dynamic range on the basis of a nonlinear forming system]. *Vestnik Rjazanskogo gosudarstvennogo radiotekhnicheskogo universiteta — Bulletin of the Ryazan State Radio Engineering University.* 2016. vol. 58. pp. 117–122. (In Russ.).
11. Sukhov T.M., Beljaev E.A. [The use of the principle of constant color brightness for the transmission of video information]. *Izvestija vuzov. Priborostroenie — Journal of Instrument Engineering.* 2011. vol. 1. pp. 55–59. (In Russ.).
12. Kucenko S.A., Nemcova S.R., Mkrumov A.S., Cvetkov O.V. [Theoretical and information characteristic of image quality in the course of digital compression]. *Trudy SPIIRAN — SPIIRAS Proceedings.* 2013. vol. 8(31). pp. 190–202. (In Russ.).
13. Zubakin I.A., Ibatullin S.M., Fahmi Sh.S., Cyculin A.K. [Measuring the quality of video information at the output of the digital source coder and its efficiency]. *Voprosy*

- radioelektroniki. Serija: Tehnika televidenija — Questions of radio electronics. Series: Television Technique.* 2016. vol. 4(27). pp. 33–39. (In Russ.).
14. Image Compression. Benchmark. The New Test Images. Available at: http://imagecompression.info/test_images/ (accessed: 01.06.2017).
 15. RECOMMENDATION ITU-R BT.601-7. Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios. Available at: http://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.601-7-201103-I!!PDF-E.pdf (accessed: 07.06.2017).
 16. Instruction tables. Available at: http://www.agner.org/optimize/instruction_tables.pdf (accessed: 07.06.2017).
 17. Cortex-A57 Software Optimization Guide. Available at: http://infocenter.arm.com/help/topic/com.arm.doc.uan0015b/Cortex_A57_Software_Optimization_Guide_external.pdf (accessed: 07.06.2017).
 18. Matveev S.A. et al. [Organization of energy-information exchange between control devices of the form of the transformed antenna using fiber-optic technologies]. *Opticheskij zhurnal — Journal of Optical Technology.* 2016. vol. 83. no. 11. pp. 73–78. (In Russ.).
 19. Bei N.A., Zimin V.N. [Transformable large antenna for geostationary spacecraft]. *Antenny — Antennas.* 2005. vol. 10(101). pp. 24–27. (In Russ.).

Е.П. МИНАКОВ, Б.В. СОКОЛОВ, С.Е. ШАЛДАЕВ
**ИССЛЕДОВАНИЕ ХАРАКТЕРИСТИК И ВАРИАНТОВ
ПРИМЕНЕНИЯ ОКОЛУННОЙ СИСТЕМЫ ПОРАЖЕНИЯ
АСТЕРОИДОВ**

Минаков Е.П., Соколов Б.В., Шалдаев С.Е. Исследование характеристик и вариантов применения окололунной системы поражения астероидов.

Аннотация. Представлен анализ современного состояния исследований по проблеме борьбы с астероидной опасностью, который показал необходимость создания многоуровневой системы поражения астероидов. Приведены разработанные авторами математические модели и исследованы характеристики окололунной системы поражения астероидов. Получены оценки пространственно-временных характеристик и вероятностей поражения астероидов в узловых точках для одного космического аппарата-перехватчика (КАП), на основе которых сделаны выводы о потенциальной эффективности создания такой системы и указаны основные направления дальнейших исследований. Предложена оригинальная динамическая модель планирования применения окололунной системы поражения астероидов группировкой КАП. Проанализированы и обоснованы подходы к решению задач моделирования и планирования ее применения, базирующиеся на комбинированном использовании математического аппарата современной теории оптимального управления и исследования операций. Показано, что разработанное к настоящему времени модельно-алгоритмическое обеспечение можно использовать не только для окололунной, но и на других уровнях системы поражения астероидов.

Ключевые слова: система поражения астероидов, космический аппарат перехватчик, отделяемая часть, узловая точка, окололунное базирование, модели планирования применения.

1. Введение. Степень опасности от падения астероидов на поверхность Земли для жизнедеятельности населения нашей планеты и даже его существования различна и оценивается в зависимости от их размеров и вероятности столкновения с ней. Подобное событие с гигантским астероидом, имеющим более 100 километров в диаметре, может произойти 1 раз в несколько сотен миллионов лет и неминуемо приведет к глобальному уничтожению всего живого на Земле и даже разрушению нашей планеты. Точно также астероиды, диаметр которых составляет более 1 километра, грозят человечеству глобальной катастрофой. От удара о Землю они способны выделить энергию в 10 миллионов раз большую, чем при падении сравнительно небольшого Тунгусского метеорита и, судя по геологическим данным, их столкновения в истории нашей планеты происходили примерно 1 раз в 1-10 миллионов лет. Считается также, что удар астероида диаметром примерно в 1 километр о поверхность Земли приводит к выделению энергии в одну тысячу раз больше, чем Тунгусский метеорит. Его падение в районах Земли с большим количеством населения в наше время может привести к гибели от 1 до 10 миллионов человек. При этом примерная частота столкновения нашей планеты с таким астероидом оценивается как 1 раз в 10000 — 100000 лет.

Хочется надеяться, что наши потомки найдут средства и способы борьбы с указанными выше космическими объектами. Но не все так безнадёжно для случаев столкновения Земли с меньшими по размерам астероидами.

Астероиды диаметром менее 1 километра представляют серьёзную угрозу Земле, поскольку их взрывы вблизи населённых пунктов в результате ударной волны и нагрева могут привести к значительным разрушениям, соизмеримым с поражением от атомного взрыва мощностью до 40 мегатонн. Примером таких астероидов является уже упомянутый Тунгусский метеорит. Считается, что даже меньшие астероиды, как например 325-метровый Апофис, способны вызвать почти такие же серьёзные последствия. Количество астероидов таких размеров, движущихся на опасно близком от Земли расстоянии, составляет около 1 миллиона единиц, а прогнозируемая частота их столкновения с нашей планетой — примерно 1 раз в 100-1000 лет. Интересна оценка характеристик упавшего в результате торможения в атмосфере на поверхности Земли небольшого астероида (размер около 19,8 метра в поперечнике при массе 13 000 тонн при входе в плотные слои атмосферы), так называемого «Челябинского метеорита», 15 февраля 2013 года, то есть немногим более, чем через 100 лет после Тунгусского метеорита. Его приближение к Земле оказалось необнаруженным существующей системой контроля космического пространства. Полёт «Челябинского метеорита» в плотных слоях атмосферы сопровождался разрушением на высоте 15-25 километров с распространением ударных волн и высвобождением энергии по оценкам специалистов РАН порядка 100-200 килотонн в тротиловом эквиваленте. В результате падения этого астероида пострадало 1613 человек, были повреждены здания, а общая сумма ущерба составила около 1 миллиарда рублей.

Проделанный анализ показывает, что в настоящее время астероид и может упасть на Землю практически незамеченным. Однако ущерб от его столкновения с Землей настолько велик, что необходимо незамедлительно готовиться к ликвидации реальной угрозы из космоса. В этой связи уже сейчас должны быть развернуты ширококомасштабные научные исследования и создан научно-технический задел по решению задач как обнаружения, так и уничтожения потенциально опасных астероидов. В частности, глубокой научной проработки требуют подходы к формированию состава и структуры системы поражения астероидов, обоснованию тактико-технических и баллистических характеристик соответствующих средств и способов их применения, оцениванию прогнозируемых эффектов.

С целью понижения негативных последствий от поражения астероидов оно должно происходить как можно дальше от поверхности Земли с вероятностью практически равной единице.

Анализ средств, которые могут быть использованы для поражения астероидов, показывает, что наиболее реализуемыми в ближайшей перспективе являются средства наземного базирования.

В настоящее время у нас в стране и за рубежом предлагается ряд проектов защиты Земли от астероидной опасности [1-2]. Известны следующие основные способы устранения столкновений астероидов с Землей [3-6]:

1) воздействие КАП на астероид для отклонения его траектории от попадания на Землю (например, кинетический ударник, гравитационный «тягач» и т.д.);

2) воздействие по астероиду мощными лазерами;

3) использование «солнечного паруса» или тормозящего «облака» для «увода» астероида от Земли;

4) «мягкая посадка» астероида на поверхность Земли с помощью парашютов;

5) разрушение астероида (например, ракетным ударом, бурением на нем скважины и подрывом заложенного в нее мощного заряда, «испарение» астероида мощным взрывом, применением «камнедробильных» аппаратов, перемалывающих астероиды «в пыль» и т.д.).

В работе [5] рассматриваются возможные пути реализации относительно новой идеи использовать астероид в качестве «снаряда» для изменения траектории опасного астероида. Идея состоит в том, чтобы перевести астероид-снаряд на траекторию, пересекающую траекторию астероида мишени в некоторой точке. Для этого, естественно, необходимо снабдить астероид-снаряд двигателем, пристыковав к нему КАП с необходимым запасом топлива, и/или использовать гравитационный маневр, например, в околоземном космическом пространстве.

Проделанный анализ показывает, что указанные способы либо оказывают слабым воздействием на астероид (на уровне погрешностей определения параметров орбиты) и требуют длительного времени для достижения требуемых эффектов, либо имеют плохо предсказуемые последствия (например, при разрушении астероида велика вероятность образования сравнительно мелких осколков, которые тоже будут представлять опасность для жизнедеятельности людей), либо сопряжены как с экологическими, так и с политическими проблемами при их реализации.

Отмеченные недостатки рассмотренных способов устранения столкновений астероидов с Землей делают целесообразным создание

эшелонированной системы борьбы с астероидами, основные варианты построения которых рассмотрены в работах [6-8].

Так в статье [8] рассматривается вариант системы поражения астероидов, состоящей из трех эшелонов.

Первый эшелон этой системы — подсистемы мониторинга астероидов вообще и в особенности опасных для Земли. Работы по созданию такой подсистемы уже ведутся в настоящее время. В дополнение к ним можно предложить и необходимо исследовать ряд перспективных направлений:

1) расширение сети наземных наблюдателей (достаточно отметить, что ~50% всех околоземных астероидов были обнаружены только одной Лабораторией Линкольна [9]);

2) запуск космического аппарата (КА) с телескопом на орбиту вокруг точки либрации L1 системы Солнце – Земля (SE_L1), который сможет обнаруживать опасные объекты типа «челябинский» астероид, приближающиеся по направлению от Солнца за десятки часов до сближения с Землей [9];

3) реализация идеи создания «космических патрулей» — спутниковых группировок специально для мониторинга астероидов [3, 8];

4) создание глобальной сети оповещения населения об астероидной опасности (достаточно отметить, что до падения астероида в Судане от момента его обнаружения было около 20 часов, но население района об этом так не было предупреждено).

Второй эшелон рассматриваемой системы — подсистема устранения угроз столкновения опасных астероидов с Землей.

Третий эшелон системы — распределенные на различных орбитах КАП астероидов, оснащенных техническими средствами, соответствующими различным способам борьбы с астероидами, дополненные ракетами-перехватчиками, расположенными на поверхности Земли. Варианты размещения ракет-перехватчиков этого эшелона были исследованы и представлены авторами в предыдущей статье [2], в которой впервые были приведены полученные количественные характеристики числа стартовых комплексов моноблочных ракет-перехватчиков и вероятностей поражения ими астероидов, приближающихся к Земле с любого направления. Анализ полученных в этой статье результатов выявил потребность в разработке предложенных и исследовании вариантов размещения КАП астероидов на удаленных от поверхности Земли орбитах.

В соответствии с этим целью предлагаемой статьи является представление новых научных результатов исследования характеристик и вариантов применения КАП, движущихся в окололунном космическом пространстве. Выбранная баллистическая структура размещения КАП объясняется тем, что в случае расположения КАП на поверхности Луны

потребуется их значительное количество. Данный вывод базируется на ранее полученных и представленных в работе [2] результатах. Окололунные траектории движения КАП расширяют «окна» старта КАП за счет использования различных способов фазирования их траектории. В этом случае потребуется значительно меньшее количество КАП по сравнению с вариантом их размещения на поверхности Луны. Для оценивания возможных вариантов применения КАП необходимо, в первую очередь, провести анализ возможных сценариев развертывания соответствующих систем КАП астероидов окололунного базирования, оценить их характеристики. При этом в качестве базового варианта уместно рассмотреть случай, когда на окололунном рубеже поражения астероидов на дежурной орбите развертывается один или несколько КАП, оснащенных ядерными отделяемыми частями (ОЧ) тросового типа [1, 2, 8]. При этом уничтожение астероидов КАП окололунного базирования осуществляется в так называемых узловых точках (УТ) — точках пересечения траекторией движения астероидов плоскости движения КАП и/или ОЧ, задаваемых дальностью от притягивающего центра (центра Луны), углом между направлением на УТ и линией пересечения плоскостей орбит астероида и КАП и временем пролета астероидом плоскости движения КАП. Следует отметить, что предлагаемое авторами перспективное направление исследований, связанное с созданием систем поражения астероидов окололунного базирования и рассматриваемое в данной статье, ориентировано на разработку модельно-алгоритмического обеспечения, позволяющего провести оценивание эффективности возможных способов и технологий уничтожения астероидов КАП. Для этого в статье предлагается оригинальная модель планирования применения окололунной системы поражения астероидов, базирующаяся на математическом аппарате современной теории оптимального управления. К настоящему времени разработаны многочисленные методы, алгоритмы и методики решения задач планирования применения КАП, базирующиеся на математическом аппарате целочисленного программирования, теории расписаний, теории эвристических решений [2, 10-14]. В дополнение к указанным методам и алгоритмам в статье предлагается новый подход к оперативному построению планов применения рассматриваемых КАП, базирующийся на одной из разновидностей реализации принципа максимума Л.С. Понтрягина — методе локальных сечений [15]. В этом случае исходная нестационарная задача теории расписаний большой размерности, в рамках которой описываются различные варианты применения системы КАП, сводится к двухточечной (либо многоточечной) краевой задаче, для решения которой в свое время был разработан комбинированный алгоритм, основанный на вычислительных схемах метода последовательных приближений Крылова-Черноузько и метода ветвей и границ [14, 16].

2. Модель исследования пространственно-временных характеристик применения одиночного КАП системы поражения астероидов окололунного базирования. Пусть исследуемая система поражения астероидов окололунного базирования в своем базовом варианте состоит из одного КАП с ОЧ, развертываемого на окололунных дежурных орбитах [2] в плоскости, перпендикулярной радиусу Земля — Луны (рисунок 1).

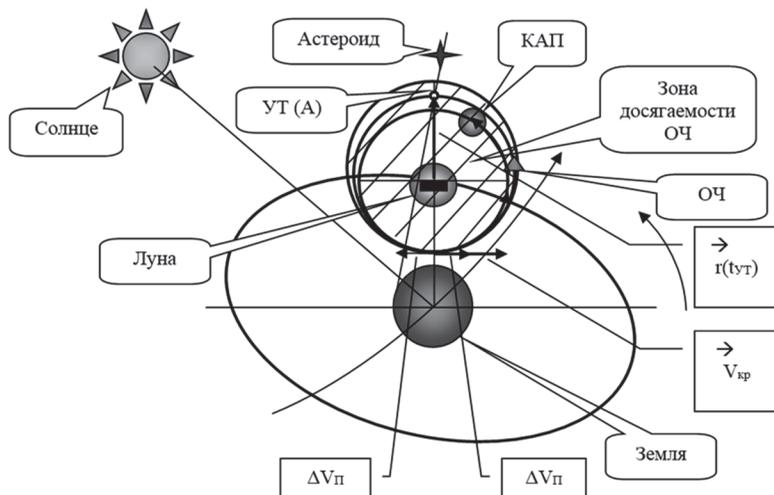


Рис. 1. Зона досягаемости ОЧ КАП окололунного базирования

При разработке математических моделей для исследования пространственно-временных характеристик применения одиночного КАП системы поражения астероидов окололунного базирования были приняты следующие допущения:

- 1) окололунная дежурная орбита КАП — круговая;
- 2) модель Луны — сфера с равномерным распределением масс;
- 3) задействие ОЧ происходит мгновенно в момент пролета астероидом УТ;
- 4) плоскость орбиты ОЧ совпадает с плоскостью орбиты КАП;
- 5) поражение астероида достигается в ходе двухимпульсного маневра ОЧ, когда первый импульс — $\Delta V_{кр / \phi}$ прикладывается на дежурной орбите КАП с целью перевода ОЧ на фазирующую орбиту в точке развернутой на 180^0 по отношению к УТ, второй импульс — $\Delta V_{\phi / УТ}$ прикладывается на фазирующей орбите ОЧ с целью перевода ОЧ на орбиту поражения астероида в той же точке (рисунок 2);

6) сообщаемые ОЧ импульсы скорости — $\Delta V_{кр/\phi}$ и $\Delta V_{\phi/UT}$ направлены по касательным к дежурной орбите КАП и к фазирующей орбите соответственно и не превосходит запаса характеристической скорости, отводимой на доставку ОЧ с КАП к астероиду: $\Delta V_{кр/\phi} - \Delta V_{\phi/UT} \leq \Delta V_3$;

7) перехват астероида осуществляется только в апогее — А или перигее — П орбиты ОЧ.

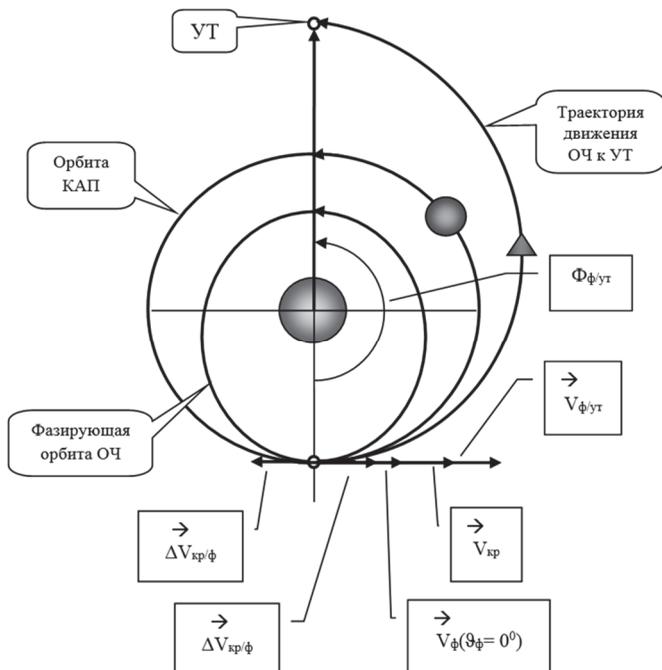


Рис. 2. Схема окололунного маневрирования ОЧ

Указанные допущения позволили авторам на основе кеплеровской теории движения КА и импульсной теории маневров КА впервые разработать модель оценивания пространственного размаха области досягаемости ОЧ КАП с дежурной круговой окололунной орбиты, включающую в себя следующие этапы:

1) вычисление круговой скорости КАП:

$$V_{кр} = \sqrt{K_{лунны} / (R_{лунны} + H_{кр})},$$

где $K_{\text{луны}} = 4890 \text{ км}^3 / \text{с}^2$ — гравитационная постоянная Луны; $R_{\text{луны}} = 1738,14 \text{ км}$ — средний радиус Луны; $H_{\text{кр}}$ — высота круговой орбиты КАП на поверхности Луны;

2) расчет скорости ОЧ в точке отделения от КАП при приложении импульса по направлению его движения:

$$V_+ = V_{\text{кр}} + \Delta V_{\text{кр}/\phi};$$

3) определение радиуса апогея орбиты ОЧ, являющегося радиусом верхней границы области досягаемости ОЧ КАП (при расчете r_A полагаем, что весь запас характеристической скорости ОЧ расходуется на ее движение к указанной границе: $\Delta V_{\text{кр}/\phi} = \Delta V_3$):

$$r_A = (R_{\text{луны}} + H_{\text{кр}}) / (2K_{\text{луны}} / ((R_{\text{луны}} + H_{\text{кр}})V_+^2) - 1);$$

4) расчет скорости ОЧ в точке отделения от КАП при приложении импульса в сторону, противоположную направлению его движения:

$$V_- = V_{\text{кр}} - \Delta V_{\text{кр}/\phi};$$

5) определение радиуса перигея орбиты ОЧ, являющегося радиусом нижней границы области досягаемости ОЧ КАП (при расчете r_{II} полагаем, что весь запас характеристической скорости ОЧ расходуется на ее движение к указанной границе: $\Delta V_{\text{кр}/\phi} = \Delta V_3$):

$$r_{II} = (R_{\text{луны}} + H_{\text{кр}}) / (2K_{\text{луны}} / ((R_{\text{луны}} + H_{\text{кр}})V_-^2) - 1);$$

6) оценивание пригодности окололунного рубежа по критерию пригодности:

$$r_{II} \leq R_{\text{луны}}.$$

В случае удовлетворения указанного критерия при принятых допущениях условия поражения астероида ОЧ имеют вид:

$$\left\{ \begin{array}{l} r_{yT} \leq r_A \\ \left| \vec{r}_{yT} - \vec{r}(t_{yT}) \right| \leq R_{Oч} \end{array} \right.$$

где \vec{r}_{yT} — радиус-вектор УТ астероида в плоскости орбиты КАП; t_{yT} — время пересечения траекторией астероида плоскости орбиты КАП (время

УТ); $\vec{r}(t_T)$ — радиус-вектор центра масс ОЧ в момент времени $t_{УТ}$ (рисунок 2); $R_{ОЧ}$ — радиус зоны захвата астероидов ОЧ ($R_{ОЧ} \geq r_{отн}$).

Временные характеристики применения одиночного КАП системы поражения астероидов окололунного базирования включают в себя время прибытия ОЧ в УТ, время движения ОЧ на фазирующей орбите и время перелета ОЧ с фазирующей орбиты в УТ.

Баланс времени, обеспечивающий одновременное прибытие в УТ астероида и ОЧ, имеет вид:

$$\Delta t_{\phi/УТ} + \Delta t_{\phi} = t_0 - t_{УТ},$$

где t_0 — момент времени начала применения ОЧ (момент времени ее перехода на фазирующую орбиту, момент времени начала маневрирования ОЧ); Δt_{ϕ} — время движения ОЧ на фазирующей орбите; $\Delta t_{\phi/УТ}$ — время перелета ОЧ с фазирующей орбиты в УТ.

Время движения ОЧ на фазирующей орбите определяется выражением:

$$\Delta t_{\phi} = n T_{\phi},$$

где $T_{\phi} = 2\pi\sqrt{a_{\phi}^3 K_{лунны}}$ — период обращения ОЧ на фазирующей орбите; a_{ϕ} — большая полуось фазирующей орбиты ОЧ; n — число полных периодов обращения ОЧ на фазирующей орбите.

Время перелета ОЧ с фазирующей орбиты в УТ может быть вычислено по формуле:

$$\Delta t_{\phi/УТ} = \pi\sqrt{a_{\phi/УТ}^3 K_{лунны}},$$

где $a_{\phi/УТ} = ((R_{лунны} + H_{кр}) + r_{УТ}) / 2$ — большая полуось переходной орбиты ОЧ.

Значение a_{ϕ} позволяет, в свою очередь, оценить потребное значение изменения характеристической скорости — ΔV в точке маневрирования, когда истинная аномалия ОЧ на фазирующей орбите — $\vartheta_{\phi} = 0^\circ$:

$$\Delta V_{кр/\phi} = |V_{кр} - V_{\phi}(\vartheta_{\phi} = 0^\circ)|,$$

где $V_{\phi}(\vartheta_{\phi} = 0^\circ) = \sqrt{K_{лунны} (2 / (R_{лунны} + H_{кр}) - 1 / a_{\phi})}$.

С использованием этой зависимости можно сформулировать критерий пригодности фазирующей орбиты: если $\Delta V_{кр/\phi} \leq \Delta V_{\phi}$, где

ΔV_ϕ — запас характеристической скорости ОЧ (КАП), отводимый на фазирование ОЧ.

3. Модель оценивания эффективности поражения астероида одиночным КАП на окололунном рубеже. В ходе проведенных исследований впервые был оценен минимальный широтный слой на поверхности Земли, «прикрываемый» окололунным рубежом, к северу и югу от экватора определяется углом, рассчитываемым по следующей формуле (рисунок 3):

$$\Psi_{\min} = \alpha - i_{\text{луны}},$$

где $i_{\text{луны}} = 5^\circ$ — наклонение Луны к плоскости экватора Земли; α — угловой размер области досягаемости ОЧ из центра Земли, определяемый по формуле:

$$\operatorname{tg} \alpha = r_A / r_{\text{луны}}.$$

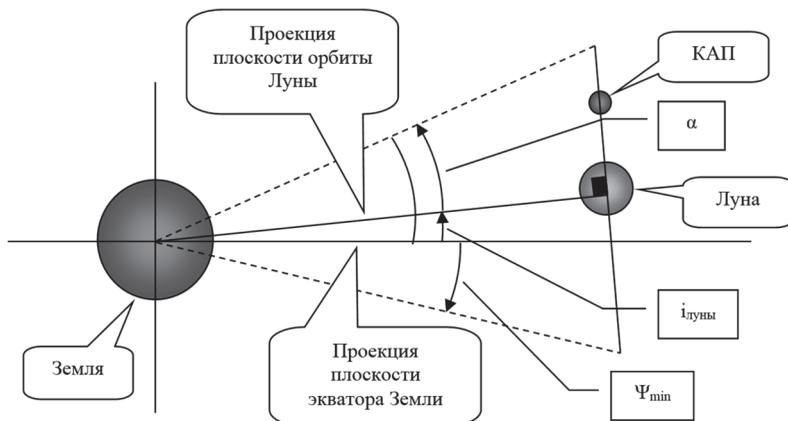


Рис. 3. Угловые элементы системы Земля–Луна–КАП

Особенностью предлагаемого подхода является тот факт, что если астероид обнаружен на расстоянии — $D \geq V_a T_l$ от центра Луны, где V_a — скорость движения астероида, T_l — период обращения Луны, а вероятность выполнения цикла применения по нему ОЧ равна 1, то вероятность поражения астероида на окололунном рубеже будет максимальной, удовлетворяющей неравенству:

$$P_{yl} \geq P_{yl \min} = S_{\Psi \max} / S,$$

где $S_{\Psi_{\max}}$ — площадь поверхности Земли в пределах $\pm\Psi_{\max}$;
 S — площадь поверхности Земли.

При сделанных в пункте 2 допущениях:

$$S_{\Psi_{\max}} = (2\pi - 2\gamma_{\min})(S - 2S_c) / (2\pi),$$

где $S_c = 2\pi R_3 h$ — площадь сферической модели Земли, находящая вне пределов широты $\pm\Psi_{\min}$; $h = R_3 - R_3 \sin \Psi_{\min}$; $S = 4\pi R_3^2$;
 $2\gamma_{\min} = 2\alpha$ — минимально возможное угловое расстояние между положениями Луны, обеспечивающее уничтожение астероида (рисунок 4).

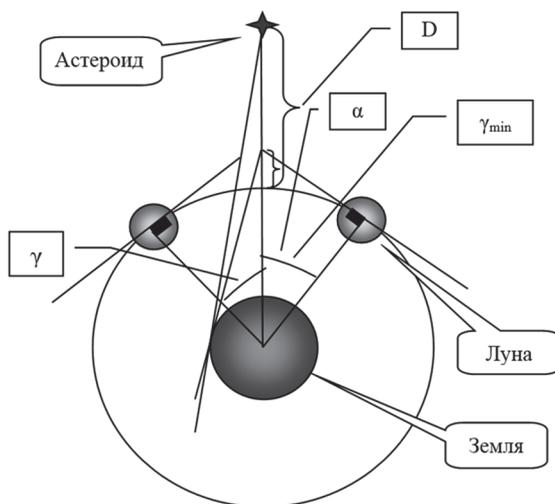


Рис. 4. Углы γ и γ_{\min} при движении Луны вокруг Земли

На рисунке 4 через γ обозначен угол, обеспечивающий поражение астероида, когда он обнаружен на расстоянии D от орбиты Луны независимо от направления его движения. Из рисунка 4 видно, что $D_{\min} = r^* - r_{\text{луны}}$, где $r^* = r_{\text{луны}} / \cos \gamma_{\min}$.

4. Результаты оценивания пространственно-временных характеристик и эффективности применения одиночного КАП системы поражения астероидов окололунного базирования. В ходе проведения вычислительных экспериментов впервые были получены результаты оценивания D_{\min} и $P_{y/l \min}$ при соответствующих исходных данных, которые представлены в таблице 1. Помимо этого в расчетах

принималось, сидерический период обращения Луны вокруг Земли — $T_{\text{луны}} = 27^{\text{CVT}} 07^{\text{U}} 43^{\text{M}}$, круговая скорость — $V_{\text{луны}} = 1,68 \text{ км/с}$, $r_{\text{луны}} = 384400 \text{ км}$.

Таблица 1. Результаты оценивания D_{min} и $P_{yI \text{ min}}$

Ψ_{min} град	s , км ²	h , км	S_c , км ²	γ_{min} , град	$S_{\psi \text{ max}}$, км ²	D_{min} , км	$P_{yI \text{ min}}^n$
39,48144	509805891	1529,607	61199375	44,48144	291671425	154161,60	572123

Если астероид обнаружен на расстоянии $r_{\text{луны}} < D < V_a T_D$ от центра Луны, то время его движения до орбиты Луны составит примерно $t_D = D/V_a$. Тогда $P_{yI}^* = T_{\text{луны}} / (T_{\text{луны}} - t_D)$, что позволяет получить новый научный результат — зависимость вероятности поражения астероида P_{yI}^* от D приведена в таблице 2.

Таблица 2. Зависимость вероятности P_{yI}^* от D

D , км	t_D , сек	P_{yI}^*
1500000	750000	0,836693
1400000	700000	0,8116699
1300000	650000	0,7881001
1200000	600000	0,7658606
1100000	550000	0,7448417
1000000	500000	0,7249458
900000	450000	0,7060851
800000	400000	0,6881809
700000	350000	0,6711623
600000	300000	0,654965
500000	250000	0,6395312
400000	200000	0,6248079
300000	150000	0,6107474
200000	100000	0,5973057
100000	50000	0,5844429
0	0	0,5721225

Соответствующий график представлен на рисунке 5.

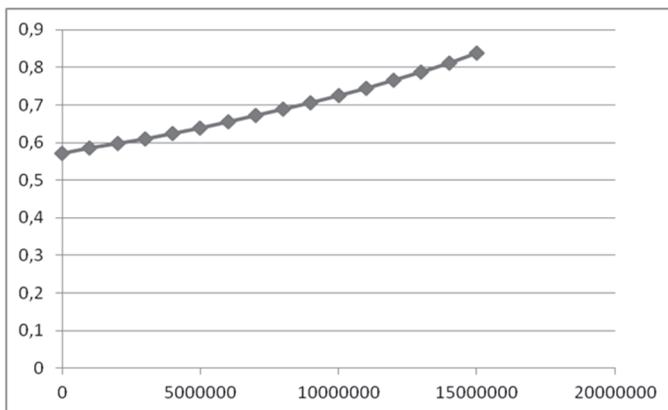


Рис. 5. График зависимости $P_{y,l}^*$ от D

5. Динамическая модель планирования применения системы КАП окололунного базирования.

В основу построения данной модели положены результаты, представленные в предыдущих разделах данной статьи, в соответствии с которыми считается, что пространственные характеристики дежурных орбит КАП в плоскости, перпендикулярной радиусу Земли — Луны определены. Они позволяют рассмотреть самый общий случай ликвидации астероидной опасности, когда в сторону Земли летят несколько астероидов (или их осколков, образовавшихся в результате подрыва основного астероида на дальних подступах к Земле) и для их уничтожения используется несколько КАП, располагаемых на окололунных орбитах. Исходя из предложенных в предыдущих разделах статьи баллистических схем перехвата астероида КАП окололунного базирования, полагается, что для каждого КАП известно соответствующее множество возможных траекторий перехвата. При построении модели планирования применения системы КАП окололунного базирования, предназначенной для перехвата астероидов на дальних рубежах подлета к Земле, будем использовать ранее предложенную в работах [10-11] динамическую интерпретацию процессов выполнения целевых и технологических операций для различных предметных областей. Для этого введем в рассмотрение целевую операцию D_{ρ}^{ij} ($i = 1, \dots, n$; $j = 1, \dots, m$), связанную с перемещением КАП B_j в зону взаимодействия (ЗВ) с астероидом-целью (АЦ) A_i , которая (исходя из особенностей применяемых активных либо пассивных средств воздействия на астероид) формируется в узловой точке, образуемой при пересечении траектории движения АЦ с

плоскостью орбиты соответствующего КАП. При моделировании процесса перехвата астероида КАП весьма существенным параметром указанной операции является время (длительность) перемещения (t_{ρ}^{ij}) КАП. Наряду с t_{ρ}^{ij} рассматриваемая операция D_{ρ}^{ij} характеризуется параметрами относительного движения КАП (относительной дальностью l_{ij}^{omni} , относительной скоростью и ускорением \dot{l}_{ij}^{omni} , \ddot{l}_{ij}^{omni} , затратами ресурсов КАП на перемещение и т.п.). На рисунке 6 изображен график зависимости, связывающий длительность перемещения КАП B_j (из некоторого фиксированного начального положения в ЗВ АЦ A_i) с временем начала выполнения целевой операции D_{ρ}^{ij} . На рисунке 7 изображен график зависимости, связывающей относительную скорость перемещения КАП и АЦ в ЗВ с временем начала выполнения целевой операции D_{ρ}^{ij} . Будем предполагать, что для рассматриваемых целевых операций заданы аналогичные зависимости, связывающие конечные значения параметров операций $x_{i\rho\theta\delta}^{(\theta)}(t)$ с временем начала выполнения указанных операций с учетом условий неразрывности их проведения, ($x_{i\rho\theta\delta}^{(\theta)}(t)$ — конкретное значение параметра $\theta = 1, \dots, \bar{\theta}$ в момент окончания выполнения операции D_{ρ}^{ij} на фиксированном этапе технологического цикла управления КАП $\delta = 1, \dots, \Delta_i$).

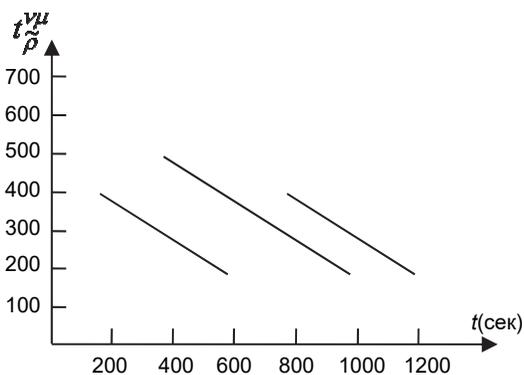


Рис. 6. График зависимости, связывающий длительность перемещения КАП с временем начала выполнения целевой операции

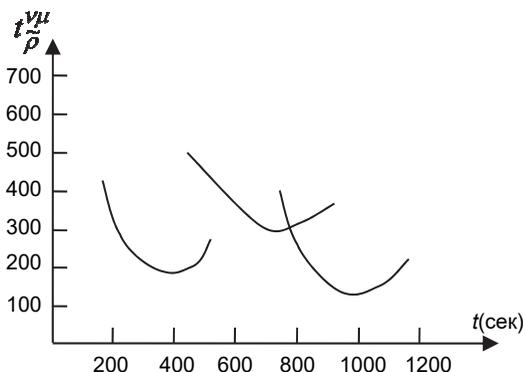


Рис. 7. График зависимости, связывающей относительную скорость перемещения КАП с временем начала выполнения целевой операции

Указанные зависимости можно рассчитать заранее на аналитических и имитационных моделях, решая, например, с заданным шагом во времени соответствующие задачи Эйлера-Ламберта, либо получать их в ходе процесса планирования, обращаясь в каждый момент времени к данным моделям. При этом необходимо отметить, что рассматриваемые зависимости $x_{i\rho\theta\delta}^{(\theta)1}(t)$ имеют периодический характер, связанный с цикличностью процессов управления и движения соответствующих КАП ($\delta = 1, \dots, \Delta_i$).

С учетом вышеизложенного модель планирования применения КАП околулолунного базирования (модель — M_O) можно задать в следующем виде:

$$\dot{x}_{i\rho\theta\delta}^{(\theta)1}(t) = b_{i\rho\theta\delta}^{(\theta)1}(t)\omega_{i\rho\theta\delta}^{(\theta)1}(t), \quad (1)$$

$$\dot{x}_{i\rho\theta\delta}^{(\theta)1}(t) = \sum_{j=1}^{m_\gamma} \sum_{\lambda=1}^{ij} \varepsilon_{ij}(t)u_{i\rho j\lambda}^{(o)1}, \quad (2)$$

$$\omega_{i\rho\theta\delta}^{(\theta)1} \cdot x_{i\rho\theta\delta}^{(\theta)2} = 0, \quad \omega_{i\rho\theta\delta}^{(\theta)1}(t), u_{i\rho\theta\delta}^{(\theta)2}(t) \in \{0, 1\}, \quad (3)$$

где $x_{i\rho\theta\delta}^{(\theta)1}(t)$ — переменная, значение которой в момент времени t численно равно значению параметра θ в момент окончания проведения операции $D_{i\rho}^{ij}$ на фиксированном этапе реализации технологического

цикла управления (ТЦУ) КАП с номером δ при условии, что D_{ρ}^{ij} началась в момент t ; $x_{i\rho\theta\delta}^{(\theta)2}(t)$ — вспомогательная переменная, численно равная времени, которое прошло с момента начала операции D_{ρ}^{ij} ; $u_{i\rho j\lambda}^{(o)1}$ — управляющее воздействие, принимающее значение 1, если принято решение начать выполнение операции D_{ρ}^{ij} (перехват), 0 — в противном случае; λ — номер конкретного средства воздействия на КАП, выбранного для уничтожения астероида A_i ; $\omega_{i\rho\theta\delta}^{(\theta)1}(t)$ — вспомогательное управляющее воздействие, принимающее значение, равное 1, если операция D_{ρ}^{ij} еще не началась, 0 — в противном случае; $b_{i\rho\theta\delta}^{(\theta)1}$ — известная функция времени, с помощью которой при решении дифференциального уравнения (1) задается зависимость, связывающая конечные значения параметров операций $x_{i\rho\theta\delta}^{(\theta)1}(t)$ с временем начала их выполнения при соблюдении условий неразрывности проведения операций [10].

Краевые условия для уравнений (1), (2) могут быть заданы в виде следующих двух вариантов: как в виде варианта К1, так и К2.

Вариант К1:

$$t=t_0 \quad x_{i\rho\theta\delta}^{(\theta)1}(t_0) = x_{i\rho\theta\delta_0}^{(\theta)1}; x_{i\rho\theta\delta}^{(\theta)2}(t_0) = 0, \quad (4)$$

$$t=t_1 \quad x_{i\rho\theta\delta}^{(\theta)1}(t_1) \in \mathfrak{R}^1; x_{i\rho\theta\delta}^{(\theta)2}(t_1) = \mathfrak{R}^1, \quad (5)$$

$$\mathfrak{R}^1 = [0, +\infty).$$

Вариант К2:

$$t=t_0 \quad x_{i\rho\theta\delta}^{(\theta)1}(t_0) = x_{i\rho\theta\delta_0}^{(\theta)1}; x_{i\rho\theta\delta}^{(\theta)2}(t_0) = 0, \quad (6)$$

$$t=t_1 \quad x_{i\rho\theta\delta}^{(\theta)1}(t_1) \in a_{i\rho\theta\delta}^{(\theta)1}; x_{i\rho\theta\delta}^{(\theta)2}(t_1) = \mathfrak{R}^1. \quad (7)$$

где $x_{i\rho\theta\delta_0}^{(\theta)1}$ и $a_{i\rho\theta\delta}^{(\theta)1}$ — заданные величины, определяемые исходя из требуемых условий при взаимодействии средств перехвата КАП $\lambda = 1, \dots, i_j$, направляемых на соответствующий астероид A_i .

Целевые функции, оценивающие качество планирования применения КАП (или по-другому — качество программного управления параметрами целевых операций), имеют следующий вид:

$$J_{i\rho\theta\delta}^{(\theta)} = \sum_{\delta=1}^{\Delta_i} (x_{i\rho\theta\delta}^{(\theta)1} - a_{i\rho\theta\delta}^{(\theta)1})^2 \Big|_{t=t_f}. \quad (8)$$

Данный показатель позволяет оценить полноту выполнения целевых операций, связанных с перехватом астероидов. Данный показатель минимизируется. В наилучшем случае данный показатель должен принимать значение 0 (все АС перехвачены).

В том случае, когда рассматриваются операции, связанные с перемещением КАП в ЗВ АЦ, и в качестве одного из параметров указанной операции выбирается время, затрачиваемое на переход КАП в ЗВ — $t_{\rho}^{V\mu}$, то наряду с уравнениями (1), (2) можно ввести следующие дополнительные соотношения вида:

$$\dot{x}_{i\rho\tilde{\theta}\delta}^{(\theta)3} = \frac{1}{x_{i\rho\tilde{\theta}\delta}^{(\theta)1}} \sum_{j=1}^{m_i} \sum_{\lambda=1}^{i_\mu} u_{i\rho j\lambda}^{(o)1}, \quad (9)$$

$$\left(x_{i\rho\tilde{\theta}\delta}^{(\theta)1}(t_f) - x_{i\rho\tilde{\theta}\delta}^{(\theta)2} \right) z_{i\rho j\lambda}^{(o)1}(t_f) = 0, \quad (10)$$

где $x_{i\rho\tilde{\theta}\delta}^{(\theta)3}$ — вспомогательная переменная, значение которой в момент окончания выполнения операции D_{ρ}^{ij} равно 1, а параметром $\tilde{\theta}$ в данном случае является t_{ρ}^{ij} . Соотношения (10) в рассматриваемой ситуации задают ограничения на неразрывность выполнения операций D_{ρ}^{ij} [10, 12, 13].

Если цель функционирования КАП состоит в оптимизации процесса перемещения КАП относительно ЗВ АЦ, то, используя вспомогательную переменную $x_{i\rho\tilde{\theta}\delta}^{(\theta)3}$, можно задать несколько вариантов показателей качества планирования процесса применения АПО. Среди них можно выделить следующие показатели:

$$J_2^{(\theta)} = \int_{t_0}^{t_f} d\tau, \text{ при ограничении } \sum_{i,\rho,\tilde{\theta},\delta} x_{i\rho\tilde{\theta}\delta}^{(\theta)3}(t_f) = \tilde{K}, \quad (11)$$

$$J_3^{(\theta)} = \sum_{i, \rho, \theta, \delta} x_{i\rho\theta\delta}^{(\theta)3}(t_f) = \tilde{K}, \quad (12)$$

где \tilde{K} — заданное число АЦ, с которыми необходимо осуществить взаимодействие на интервале планирования $\sigma = (t_o, t_f)$. Целевая функция (11) позволяет оценивать время, за которое будет осуществлено взаимодействие с заданным числом объектов A_ν ($\nu = 1, \dots, n_1$). Данный показатель в теории оптимального управления называют показателем максимального быстродействия [15]. Показатель качества (12) количественно оценивает число перехваченных АЦ.

Результаты ранее выполненных исследований показывают, что с помощью метода локальных сечений рассматриваемая задача неклассического вариационного исчисления (1)-(12) сводится к двухточечной (многоточечной) краевой задаче [10, 15, 17]. При этом распределение ресурсов КАП в каждый момент времени осуществляется с учетом величины «динамического» приоритета каждой планируемой целевой операции (операции, связанной с перехватом конкретного АЦ). В состав данного «динамического» приоритета наряду с сопряженными переменными, с помощью которых учитываются логические взаимосвязи между операциями, входят сопряженные переменные, с помощью которых учитываются требования неразрывного выполнения работ, равномерного использования ресурсов КАП. Кроме того, при вычислении приоритета учитывается текущее значение целевой функции. Данный подход позволяет, в отличие от работ [30, 36, 37], при поиске оптимального расписания одновременно учитывать как ресурсные, так и логические ограничения, связанные с процессом планирования и управления перехватом КАП астероидов.

Предварительные исследования показали, что основная трудность решения рассматриваемой двухточечной краевой задачи состоит в том, что в процессе итерационного поиска оптимального плана применения КАП с использованием принципа максимума Л.С. Понтрягина возможны прерывания целевых операции, которые нужно устранять алгоритмически. Существует несколько эвристических правил разрешения конфликтных ситуаций, связанных с ликвидацией разрывов при выполнении работ [11-13]. Применительно к рассматриваемой задаче планирования перехвата астероидов группировкой КАП целесообразно для ее решения использовать ранее разработанный алгоритм устранения разрывов, в основу которого положена идея комбинированного использования метода ветвей и границ и метода последовательных приближений (МПП) Крылова — Черноусько [16]. В этом случае наряду с основной задачей планирования рассматривается релаксированная задача

планирования, в которой сняты ограничения, связанные с требованиями неразрывного выполнения операций перехвата. В ряде выполненных к настоящему моменту времени проектов, результаты которых представлены на сайте <http://litsam.ru> показано, что в этом случае, используя метод локальных сечений, удастся в явном виде получить динамические множители Лагранжа, с помощью которых указанные ограничения учитываются в моделях планирования применения КАП. Кроме того, в рамках предложенного варианта формализации процессов планирования применения КАП можно перейти от исходного класса допустимых управлений к расширенному классу, в котором условия релейности управляющих воздействий заменены на интервальные ограничения. Однако, несмотря на это, из-за линейности дифференциальных уравнений и выпуклости области допустимых управлений сами управляющие воздействия принимают булевы значения. Говоря о практической стороне реализации рассматриваемого модельно-алгоритмического обеспечения, необходимо учитывать особенности технологии управления КАП, которые, в первую очередь, будут определяться составом и структурой его бортовых систем. В работе [10] показано, как, используя предложенный класс логико-динамических моделей, можно непосредственно связать технологию управления целевыми и обеспечивающими операциями при реализации соответствующих программ управления различными классами активных подвижных объектов (в нашем случае КАП) в динамически изменяющейся обстановке. Разработанный язык логико-динамического описания процесса применения группировки КАП позволяет с единых позиций и на одной программно-информационной платформе решать как задачи комплексного моделирования применения КАП, так и задачи планирования и управления реализацией соответствующих планов применения КАП.

Основная особенность предлагаемой модели состоит в том, что при динамической интерпретации нестационарного процесса выполнения целевых операций существенно сокращается размерность задачи планирования применения группировки КАП, которая решается в каждый момент времени. Данная размерность определяется только числом независимых путей в заданном сетевом графике комплекса планируемых операций. При этом степень связности рассматриваемого алгоритма (оцениваемая объемом информации, необходимой для запоминания при переходе от одного шага алгоритма к другому), является минимальной, так как для возобновления счета на ЭВМ при его прерывании нужно в данном случае запомнить только значения компонент векторов состояний основной и сопряженной систем дифференциальных уравне-

ний. Зная указанные величины, можно продолжить интегрирование указанных уравнений, описывающих процесс выполнения комплекса работ, и собственно распределение (планирование) ресурсов КАП. Кроме того, разработанная динамическая модель и алгоритм позволяют при поиске оптимального плана учитывать одновременно как ресурсные, так и логические ограничения, в том числе и связанные с недопустимостью прерываний работ.

6. Заключение. Приведенные в статье новые научные результаты позволяют сделать вывод о том, что размещение одного КАП на окололунном рубеже предоставляет возможность, во-первых, поражать астероиды на большом расстоянии (≈ 384400 км) от Земли; во-вторых, обеспечивать с высокой вероятностью $P_{yI} \approx 1$ поражение астероидов, движущихся в плоскостях близких к плоскости экватора (плоскости орбиты) Земли ($-39^\circ < \psi < 39^\circ$). В тоже время предлагаемая система с недостаточно высокой вероятностью $P_{yI} \approx 0,57$ поражает астероиды, движущиеся по направлению к Земле, с любого другого направления.

Полученные оценки могут быть улучшены путем развертывания на окололунных траекториях уже орбитальной системы, состоящей из нескольких КАП, что требует соответствующих исследований с применением представленного в статье инструментария, который, в свою очередь, предполагает определенное уточнение разработанных моделей. В связи с этим предложена оригинальная динамическая модель планирования применения окололунной системы поражения астероидов группировкой КАП, проанализированы и обоснованы подходы к решению задач моделирования и планирования применения данной группировки, базирующиеся на комбинированном использовании математического аппарата современной теории оптимального управления и исследования операций. Проведенный анализ также показал, что разработанное к настоящему времени модельно-алгоритмическое обеспечение можно использовать не только для окололунной, но и на других уровнях предлагаемой системы защиты Земли от астероидов.

Представленный в статье анализ современного состояния исследований проблем борьбы с астероидной опасностью показал, необходимость создания многоуровневой системы поражения астероидов. Данное обстоятельство связано с основным недостатком предложенной окололунной системы поражения астероидов, которая может обеспечить защиту Земли только с приэкваториальных направлений движения астероидов. Это указывает на необходимость исследования путей борьбы с астероидной опасностью с использованием как КАП, движу-

щихся в околоземном космическом пространстве, так и системы наземного базирования [3], что обеспечит в целом условия существенного уменьшения вероятности падения на Землю астероидов, движущихся к ней с любых направлений.

Литература

1. *Zaitsev A.V.* Preventing catastrophic impact // Room, The Space Journal. 2014.
2. *Минаков Е.П., Соколов Б.В.* Исследование характеристик размещения и применения моноблочных стационарных средств поражения астероидов наземного базирования // Труды СПИИРАН. 2016. Вып. 5(48). С. 182–197.
3. *Башаков А.А., Питьев Н.П., Соколов Л.Л.* Особенности движения астероида 99942 Апофис // Астрономический вестник. 2008. Том 42. № 1. С. 20–29.
4. Угрозы из космоса // Популярная Механика. 2010. № 9(95). С. 31–46.
5. *Назирова Р.Р., Эйсмонт Н.А.* Гравитационные маневры как способ направить малые астероиды на траекторию встречи с опасными околоземными объектами // Космические исследования 2010. Т. 48. № 5. С. 1–6.
6. *Седых Р.* Угрозы из космоса растут // Военное обозрение. 2013. URL: <http://vpk-news.ru/> (дата обращения: 21.06.2017).
7. *Коллин К.К.* Глобальные угрозы развитию цивилизации в XXI веке // Стратегические приоритеты. 2014. №1. С. 12–17.
8. *Бакланов О. и др.* От астероидно-космической опасности Землю защитит «Цитадель» // Воздушно-космическая сфера. 2017. № 1. С. 90–95.
9. *Billings L.* Words matter: A call for responsible communication about asteroid impact hazards and plans for planetary defense // Space Policy. 2015. vol. 33. pp. 8–12.
10. *Kalinin V.N., Sokolov B.V.* A dynamic model and an optimal scheduling algorithm for activities with bans of interrupts // Automation and Remote Control. 1987. vol. 48(1–2). pp. 88–94.
11. *Božek A., Wysłocki M.* Flexible Job Shop with Continuous Material Flow // International Journal of Production Research. 2015. vol. 53(4). pp. 1273–1290.
12. *Dolgui A., Kovalev S.* Min-Max and Min-Max Regret Approaches to Minimum Cost Tools Selection, 4OR: A Quarterly // Journal of Operations Research. 2012. vol. 10(2). pp. 181–192.
13. *Dolgui A., Kovalev S.* Scenario Based Robust Line Balancing: Computational Complexity // Discrete Applied Mathematics. 2012. vol. 160(13–14). pp. 1955–1963.
14. Корбут А.А., Финкельштейн Ю.Ю. Дискретное программирование // М.: Наука. 1969. 368 с.
15. *Понтрягин Л.С., Болтянский В.Г., Гамкрелидзе Р.В., Мищенко Е.Ф.* Математическая теория оптимальных процессов // М.: Наука. 1983. 392 с.
16. *Krylov I.A., Chernousko F.L.* An algorithm for the method of successive approximations in optimal control problems // Zh. Vychisl. Mat. Mat. Fiz. 1972. vol. 12(1). pp. 15–38.
17. *Ye H., Liu R.* A multiphase optimal control method for multi-train control and scheduling on railway lines // Transportation Research Part B: Methodological. 2016. vol. 93. pp. 377–393.
18. *Wang X., Disney S.M., Wang J.* Stability analysis of constrained inventory systems with transportation delay // European Journal of Operational Research. 2012. vol. 223(1). pp. 86–95.

Минаков Евгений Петрович — д-р техн. наук, профессор, профессор кафедры оценивания эффективности, Военно-космическая академия имени А.Ф. Можайского (ВКА им. А.Ф. Можайского). Область научных интересов: системный анализ, баллистическое обеспечение полетов космических аппаратов, эффективность применения космических комплексов и систем. Число научных публикаций — 200. er.minakov12345@mail.ru; ул. Ждановская, 13, Санкт-Петербург, 197198; р.т.: +7(812)552-6341.

Соколов Борис Владимирович — д-р техн. наук, профессор, Заслуженный деятель науки РФ, заместитель директора по научной работе, Федеральное государственное бюджетное учреждение науки Санкт-Петербургского института информатики и автоматизации Российской академии наук (СПИИРАН), профессор, ФГАОУ ВО "Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики" (Университет ИТМО). Область научных интересов: разработка научных основ теории управления структурной динамикой сложных организационно-технических систем. Число научных публикаций — 560. sokol@ias.spb.su; 14-я линия В.О., 39, Санкт-Петербург, 199178; р.т.: +7-812-328-3311.

Шалдаев Сергей Евгеньевич — д-р техн. наук, доцент, начальник управления военного института (научно-исследовательского), Военно-космическая академия имени А.Ф. Можайского (ВКА им. А.Ф. Можайского). Область научных интересов: радиолокация, радионавигация, построение и испытание сложных технических систем. Число научных публикаций — 100. shs99@yandex.ru; ул. Ждановская, 13, Санкт-Петербург, 197198; р.т.: +7(812) 347-95-59.

Поддержка исследований. Работа выполнена при финансовой поддержке Российского научного фонда (проект № 16-19-00199).

M.E. MINAKOV, B.V. SOKOLOV, S.E. SHALDAEV
**INVESTIGATION OF CHARACTERISTICS AND OPTIONS FOR
 THE APPLICATION OF THE NEAR-MOON SYSTEM FOR
 HITTING ASTEROIDS**

Minakov E.P., Sokolov B.V., Shaldaev S.E. Investigation of the Characteristics of the Near-Moon System for Hitting Asteroids.

Abstract. The paper presents an analysis of the current state of research into the problems of the control of the asteroid hazard, which has shown the need to create a multilevel system for hitting asteroids. Mathematical models and characteristics of the near-moon system of damage to asteroids are developed and investigated. For a basic system consisting of a single interceptor spacecraft (ISC), estimates of the probabilities of asteroids are given, on the basis of which conclusions are made about the potential effectiveness of this system, and the main directions for further research are indicated. An original dynamic model for planning the application of the near-moon system of hitting asteroids by the ISC group is proposed. Approaches to the solution of problems of modeling and planning the application of this group based on the combined use of the mathematical apparatus of the modern theory of optimal control and operations research are analyzed and substantiated. The analysis also showed that the model-algorithmic support, developed for the present time, can be used not only for the near-moon but also for other levels of the proposed system for protecting the Earth from asteroids.

Keywords: asteroids hitting system, interceptor spacecraft, detachable part, nodal point, near-moon basing, application planning models.

Minakov Evgeniy Petrovich — Ph.D., Dr. Sci., professor, professor of effectiveness evaluation department, Mozhaisky Military Space Academy. Research interests: system analysis, provision of ballistic spacecraft, effectiveness of space complexes and systems. The number of publications — 200. ep.minakov12345@mail.ru; 13, Zhdanovskaya street, St.-Petersburg, 197198, Russia; office phone: +7(812)552-6341.

Sokolov Boris Vladimirovich — Ph.D., Dr. Sci., professor, Honored scientist of Russian Federation, deputy director for research, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS), professor, ITMO University (Saint Petersburg National Research University of Information Technologies, Mechanics and Optics). Research interests: development of research fundamentals for the control theory by structural dynamics of complex organizational-technical systems. The number of publications — 560. sokol@iias.spb.su; 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone: +7-812-328-3311.

Shaldaev Sergey Evgenjevich — Ph.D., Dr. Sci., assistant professor, head of department of military institute (scientific-research), Mozhaisky Military Space Academy. Research interests: radiolocation, radio navigation, construction and testing of complex technical systems. The number of publications — 100. shs99@yandex.ru; 13, Zhdanovskaya street, St.-Petersburg, 197198, Russia; office phone: +7(812) 347-95-59.

References

1. Zaitsev A.V. Preventing catastrophic impact. *Room, The Space Journal*. 2014.
2. Minakov E.P., Sokolov B.V. [Investigation of the location and application of monoblock stationary ground-based asteroid weapons]. *Trudy SPIIRAN — SPIIRAS Proceedings*. 2016. vol. 5(48). pp. 182–197. (In Russ.).

3. Bashakov A. A., Pit'ev N. P., Sokolov L.L. [Features of the asteroid movement 99942 Apofis]. *Osobennosti dvizheniya asteroida 99942 Apofis — Astronomicheskij vestnik*. 2008. vol. 42. no. 1. pp. 20–29. (In Russ.).
4. [Threats from space]. *Ugrozy' iz kosmosa — Populyarnaya Mexanika*. 2010. vol. 9(95). pp. 31–46. (In Russ.).
5. Nazirov R.R., E'j-smont N.A. [Gravitational maneuvers as a way to direct small asteroids to a meetings trajectory with dangerous underground objects]. *Gravitacionny'e manevry' kak sposob napravit' maly'e asteroidy' na traektoriyu vstrechi s opasny'mi okolozemny'mi ob'ektami — Kosmicheskie issledovaniya*. 2010. vol. 48. no. 5. pp. 1–6. (In Russ.).
6. Sedyh R. [Treats from space are growing]. *Ugrozy' iz kosmosa rastut — Voennoe obozrenie*. 2013. Available at: <http://vpk-news.ru/> (access: 21.08.2017). (In Russ.).
7. Kolin K.K. [Global threats of civilization development in the XXI century] *Global'ny'e ugrozy' razvitiyu civilizacii v XXI veke — Strategicheskie priorityety*. 2014. vol. 1. pp. 12–17. (In Russ.).
8. Baklanov O. et al. [From asteroid and space danger The Earth will be protected by the Citadel] *Ot asteroidno-kosmicheskoy opasnosti Zemlyu zashhiti "Citadel" — Vozdushno-kosmicheskaya sfera*. 2017. vol. 1. pp. 90–95. (In Russ.).
9. Billings L. Words matter: A call for responsible communication about asteroid impact hazards and plans for planetary defense. *Space Policy*. 2015. vol. 33. pp. 8–12.
10. Kalinin V.N., Sokolov B.V. A dynamic model and an optimal scheduling algorithm for activities with bans of interrupts. *Automation and Remote Control*. 1987. vol. 48(1–2). pp. 88–94.
11. Bożek A., Wysocki M. Flexible Job Shop with Continuous Material Flow. *International Journal of Production Research*. 2015. vol. 53(4). pp. 1273–1290.
12. Dolgui A., Kovalev S. Min-Max and Min-Max Regret Approaches to Minimum Cost Tools Selection, 4OR: A Quarterly. *Journal of Operations Research*. 2012. vol. 10(2). pp. 181–192.
13. Dolgui A., Kovalev S. Scenario Based Robust Line Balancing: Computational Complexity. *Discrete Applied Mathematics*. 2012. vol. 160(13–14). pp. 1955–1963.
14. Korbut A.A., Finkel'shtejn Yu.Yu. *Diskretnoe programmirovaniye* [Discrete programming]. M.: Nauka. 1969. 368 p. (In Russ.).
15. Pontryagin L.S., Boltyanskij V.G., Gamkrelidze R.V., Mishhenko E.F. *Matematicheskaya teoriya optimal'ny'x processov* [Mathematical theory of optimum processes]. M.: Nauka. 1983. 392 p. (In Russ.).
16. Krylov I.A., Chernousko F.L. An algorithm for the method of successive approximations in optimal control problems. *Zh. Vychisl. Mat. Mat. Fiz.* 1972. vol. 12(1). pp. 15–38.
17. Ye H., Liu R. A multiphase optimal control method for multi-train control and scheduling on railway lines. *Transportation Research Part B: Methodological*. 2016. vol. 93. pp. 377–393.
18. Wang X., Disney S.M., Wang, J. Stability analysis of constrained inventory systems with transportation delay. *European Journal of Operational Research*. 2012. vol. 223(1). pp. 86–95.

С.А. КАБАНОВ, А.И. КРИВУШОВ, Ф.В. МИТИН
**МОДЕЛИРОВАНИЕ СОВМЕСТНОГО РАСКРЫТИЯ
ЭЛЕМЕНТОВ КРУПНОГАБАРИТНОГО
ТРАНСФОРМИРУЕМОГО РЕФЛЕКТОРА КОСМИЧЕСКОГО
БАЗИРОВАНИЯ**

Кабанов С.А., Кривушов А.И., Митин Ф.В. Моделирование совместного раскрытия элементов крупногабаритного трансформируемого рефлектора космического базирования.

Аннотация. Крупногабаритные трансформируемые конструкции космического базирования доставляются на орбиту в сложном состоянии, в следствии чего встает задача их надежного раскрытия. В данной статье предлагается использовать в качестве исполнительного органа привод в виде электрической машины. Применение данного вида актюатора позволит управлять процессом раскрытия.

В качестве крупногабаритной трансформируемой конструкции рассматривается рефлектор космического базирования. В настоящее время перевод аппаратов из сложного состояния в рабочее осуществляется поэтапно. В работе рассмотрено совместное выполнение двух этапов: разворот корневого звена спицы и выдвигание промежуточного звена. Разработаны математические модели для вращательного и поступательного видов движений, учитывающие такие параметры как изгиб и сжатие спицы. Проведено моделирование и проанализированы результаты различных вариантов совместного раскрытия элементов рефлектора: использование двигателя на каждую из компонент движения и использование центробежной силы для выдвигания спицы.

Рассмотрено применение алгоритма коррекции параметров структуры управления. Одним из важных достоинств алгоритма является возможность построения управления в режиме реального времени. Его можно использовать для вычисления опорного управления в алгоритмах, основанных на принципе двухканальности.

Ключевые слова: математическая модель, вращательное движение, поступательное движение, моделирование, коррекция параметров структуры управления, крупногабаритный трансформируемый рефлектор.

1. Введение. В настоящее время антенны являются неотъемлемой частью современных космических аппаратов. История их создания неразрывно связана с историей развития космической и ракетной техники. Активные исследования и разработки, направленные на создание в космосе конструкций различного класса, имеющих большие размеры, ведутся как в России, так и за рубежом. Разработка раскрывающихся антенн, устанавливаемых на космические аппараты различного назначения, является одним из важных и стремительно развивающихся направлений в области создания крупногабаритных космических конструкций. Несмотря на значительные успехи в области проектирования раскрывающихся крупногабаритных трансформируемых конструкций космического базирования, остается важной задача их плавного и надежного раскрытия [1-5].

Отличительной особенностью крупногабаритных космических аппаратов (крупногабаритные антенны, космические телескопы и т.п.) является наличие в их составе множества подвижных узлов. Особый интерес представляют трансформируемые антенны, профиль которых можно изменять как на стадии подготовки к работе, так и в процессе функционирования. Вследствие больших размеров перевод рефлектора из сложного состояния в рабочее разбивается на этапы, которые можно рассматривать как независимые друг от друга.

Развертывание различных видов крупногабаритных трансформируемых рефлекторов (ферменных, зонтичных, кольцевых и др.) производится за счет энергии пружин или деформированных упругих элементов [6-11].

К недостаткам данного способа раскрытия можно отнести невозможность управлением процессом раскрытия, наличие колебаний и прогибов конструкции после завершения этапа раскрытия, необходимость создания макета и проведения дополнительных расчетов для определения степени сжатия пружин или закручивающих элементов.

Альтернативой классическим системам раскрытия могут являться автоматические системы, в которых энергия для развертывания элементов рефлектора производится за счет приводов, в частности электрических машин. Применение данного вида актуаторов позволит управлять процессом раскрытия, уменьшить колебания и прогибы конструкции крупногабаритного космического аппарата.

Управляемое раскрытие позволит уменьшить влияние возмущающих факторов и привести КА в рабочее состояние за заданное время.

Моделирование процесса создания и поддержания формы крупногабаритного трансформируемого рефлектора (КТР) космического базирования является сложной задачей. При этом возникают проблемы оптимального быстродействия системы, учета возмущений различного характера, действующих на конструкцию, как во время раскрытия, так и после. Для улучшения характеристик возникает необходимость использования различных методов активного управления. Имеет смысл применить принцип двухканальности Б.Н. Петрова, в котором сочетаются опорное управление и управление по возмущению. Опорное управление можно формировать с применением методов оптимального управления [12].

На базе имеющихся в настоящее время научных подходов для решения подобных задач требуется разработать методику и алгоритмы, которые позволят получать надежное решение задачи оптимизации.

Один из возможных подходов к решению задачи синтеза оптимального управления — разработка алгоритма оптимизации движения с коррекцией параметров структуры управления [13]. При этом управление может вычисляться в реальном времени в процессе движения как

функция текущих значений фазовых координат. В этом случае из применения принципа максимума определяется структура управления, параметры которой оптимизируются во вспомогательной задаче. Оптимизацию параметров удобно производить по критерию Красовского [20], известного также как функционал обобщенной работы, или ФОР, с использованием алгоритма с прогнозирующей моделью.

2. Математическая модель совместного движения. Процесс раскрытия крупногабаритного космического аппарата можно разбить на несколько отдельных этапов. Рассмотрим подробнее наиболее часто встречающиеся виды движения, описывающие эти этапы: вращательное и поступательное.

В ходе исследования было предложено осуществлять одновременное раскрытие нескольких звеньев КТР космического базирования для уменьшения времени выведения аппарата в рабочее положение.

Рассмотрим задачу раскрытия корневое звена 1 и одновременное выдвигание промежуточного звена 2 концевых звеньев силового каркаса, представленных на рисунке 1.

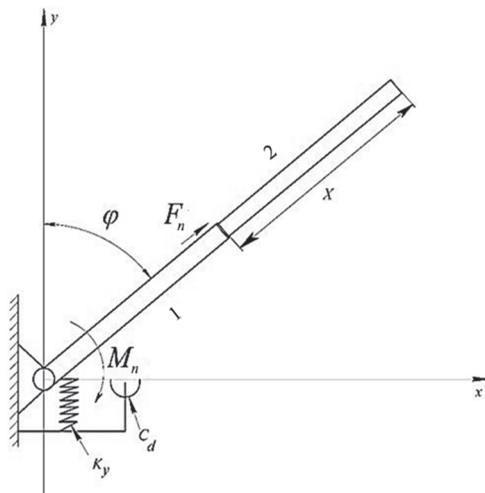


Рис. 1. Схема взаимного расположения звеньев спицы основания силового каркаса КТР

Необходимо за определенное время развернуть корневое звено спицы и одновременно выдвинуть промежуточное звено. На основе работ [14, 15] рассмотрим математические модели угловой и поступательной компоненты движения элементов конструкции.

Поскольку КТР предполагается к использованию на околоземной орбите, а также ввиду отсутствия механизмов с накоплением энергии (таких как пружины) примем, что рефлектор находится на бесконечно большом расстоянии от любых тел. Таким образом, КТР обладает общим нулевым уровнем потенциальной энергии сил всемирного тяготения $W = 0$ [16].

Макроскопическое движение тел сопровождается необратимыми процессами трения, приводящими со временем к прекращению движения. Кинетическая энергия тел при этом переходит в тепло или диссипирует. Для рассматриваемого случая допустим, что имеет место линейное изотропное сопротивление [17].

При вращательном движении тела происходят следующие процессы:

- растяжение материала, причем абсолютное растяжение Δl зависит от угловой скорости вращения и влияет на момент инерции тела I ;
- изгиб тела относительно его средней линии.

Выражение для прогиба h [18] (отклонение спицы от луча с углом φ) в первом приближении имеет вид (1):

$$h = \frac{Ml}{2EI}, \quad (1)$$

где M — момент, действующий на спицу, l — длина колеблющейся части, E — модуль Юнга, I — момент инерции.

Для создания полезного момента вращения M_n выбран бесколлекторный (вентильный) двигатель постоянного тока. Так как они обладают большим сроком службы, обеспечивают высокие моменты, высокую эффективность и низкое тепловыделение.

Управление раскрытием происходит за счет изменения напряжения питания электродвигателя $u = U_1$. Поскольку напряжение питания прямо пропорционально полезному моменту M_n электродвигателя:

$$M_n = \frac{m_1 \cdot p \cdot E_0 \cdot U_1}{\omega \cdot X_c} \cdot \sin \vartheta, \quad (2)$$

где m_1 — число фаз ротора, p — число пар полюсов магнитного поля, E_0 — действующее значение ЭДС на обмотке статора, X_c — синхронное сопротивление ($X_c = X_\sigma + X_a$, где X_σ и X_a — индуктивные сопротивления

ния рассеяния и реакции якоря соответственно), ϑ — угол рассогласования (между U_1 и E_0 , для двигателя находится в пределах $[0, \pi/2]$), то фактически управление происходит за счет изменения полезного момента.

При описании вращательного движения спицы следует учитывать действие механизма упора и фиксации. Реальные физические процессы, происходящие во время установки элементов конструкции на упоры и фиксаторы, довольно сложны. Простейшая модель таких устройств может быть представлена в виде упругого k_y и демпфирующего C_d элементов с соответствующими характеристиками [17].

Система дифференциальных уравнений, описывающая вращательное движение с учетом всех выше перечисленных допущений с использованием уравнения Лагранжа второго рода примет вид (3):

$$\begin{aligned} \dot{\varphi} &= \omega - \dot{h}, \\ \dot{\omega} &= \frac{M_n - M_{\text{тр}} - M_{\text{упор}} - M_{\text{фиксатор}} - \mu \dot{\varphi}}{I}, \\ \dot{h} &= \frac{M}{2EI}. \end{aligned} \quad (3)$$

где ω — угловая скорость; φ — угол поворота элемента конструкции; M_n — полезный момент; $M_{\text{тр}}$ — момент трения; $M_{\text{упор}}$ — момент, создаваемый упором; $M_{\text{фиксатор}}$ — момент, создаваемый фиксатором; μ — коэффициент диссипативности.

Здесь $M_{\text{упор}} = (k_y \cdot \Delta\varphi + C_d \cdot \dot{\varphi}) \cdot H_y(\Delta\varphi)$, где k_y — коэффициент жесткости упругой компоненты модели упора; C_d — коэффициент демпфирования вязкой компоненты модели упора; $\Delta\varphi = (\varphi - \varphi_k)$ — относительный угол, соответствующий моменту установки на упор; φ_k — заданный конечный угол раскрытия спицы; $H_y(\Delta\varphi)$ — функция, которая принимает значение 0 ($\Delta\varphi > 0$) или 1 ($\Delta\varphi < 0$). $M_{\text{фиксатор}} = (k_\phi \cdot \Delta\varphi + C_\phi \cdot \dot{\varphi}) \cdot H_\phi(\Delta\varphi)$, где k_ϕ — коэффициент жесткости упругой компоненты модели фиксатора; C_ϕ — коэффициент демпфирования вязкой компоненты модели фиксатора; $H_\phi(\Delta\varphi)$ — функция,

которая принимает значение 1 ($\Delta\varphi > 0$) или 0 ($\Delta\varphi < 0$) после того, как сработала функция $H_y(\Delta\varphi)$.

Для получения уравнений поступательного движения штанги также воспользуемся уравнением Лагранжа второго рода. Для вывода уравнений диссипативных сил и потенциальной энергии воспользуемся выводами, полученными для вращательного движения. Таким образом, КТР обладает общим нулевым уровнем потенциальной энергии сил всемирного тяготения $W = 0$. В рассматриваемом случае движение происходит вдоль одной оси. Примем, что толкающую силу задает электрический двигатель, аналогичный рассмотренному при вращательном движении.

Для создания полезной силы F_n необходимо учитывать конструкцию передачи, например цепную, зубчатую и так далее, что само по себе является отдельной технической задачей. Для полного описания необходимо учитывать люфты и «мертвые» зоны механизма. Ввиду всего вышесказанного уравнение для F_n будет состоять из полезного момента M_n , умноженного на коэффициент k_f , который отвечает за преобразование вращающего момента в толкающую силу (4):

$$F_n = k_f \cdot M_n. \quad (4)$$

Аналогично вращательному движению необходимо учитывать воздействие механизма фиксации. Продольную деформацию штанги $\Delta x_{сж}$ [19], возникающую при ее выдвигении, можно записать в виде (5):

$$\Delta x_{сж} = \varepsilon \cdot x = \frac{F \cdot x}{E_1 \cdot S}, \quad (5)$$

где ε — относительная продольная деформация, x — длина выдвигающегося участка штанги, E_1 — модуль упругости, S — площадь поперечного сечения. Отметим, что растяжение и сжатие возникают при «жестко» закрепленном конце штанги.

С учетом вышесказанных замечаний система дифференциальных уравнений, описывающих динамику выдвигения штанги, будет иметь вид (6):

$$\begin{aligned}\dot{x} &= V - \Delta\dot{x}_{\text{сж}}, \\ \dot{V} &= \frac{F_{\text{п}} - F_{\text{тр}} - F_{\text{упор}} - F_{\text{фиксатор}} - \mu\dot{x}}{m}, \\ \Delta\dot{x}_{\text{сж}} &= \frac{F}{E_1 \cdot S}.\end{aligned}\tag{6}$$

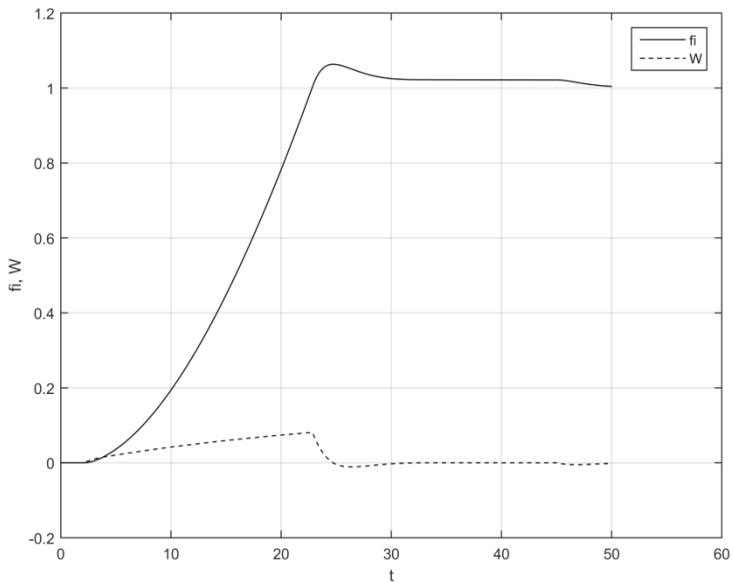
где V — линейная скорость; $F_{\text{тр}}$ — сила трения, $F_{\text{упор}}$ — сила, создаваемая упором; $F_{\text{фиксатор}}$ — сила, создаваемая фиксатором, m — масса выдвигаемой части.

Здесь $F_{\text{упор}}$, $F_{\text{фиксатор}}$ задаются аналогичным образом, что и $M_{\text{упор}}$, $M_{\text{фиксатор}}$.

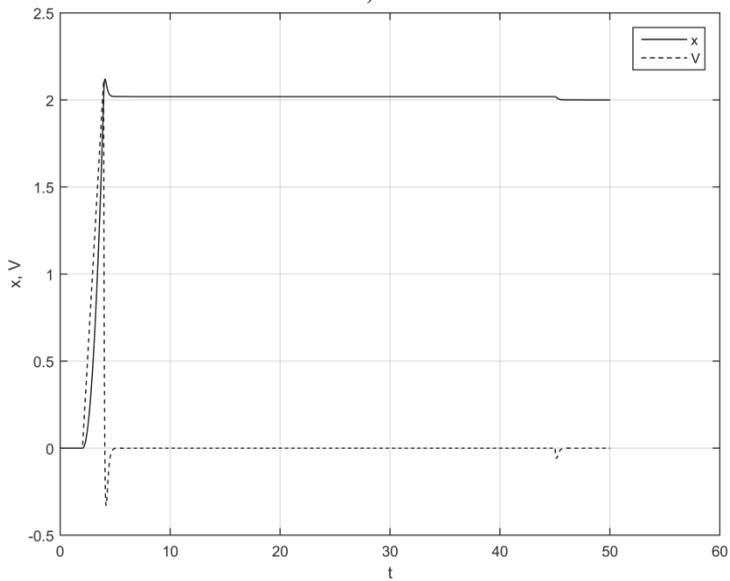
Система, описывающая совместное раскрытие элементов КТР, объединяет системы (3) и (6).

3. Моделирование. Первый способ совместного раскрытия элементов КТР космического базирования — использование двигателя на каждую из компонент движения. При моделировании динамики совместного раскрытия спицы варьировались коэффициенты демпфирования C_{θ} и упругой компоненты упора k_y , исполнительные устройства на оба вида движений имели одинаковые параметры и включались одновременно. Были выбраны следующие параметры системы при моделировании: материал АБС пластик $QH F$ — 0140: $\rho = 1040$ кг·м³, $E = 9900 \cdot 10^6$ Па, напряжение питания, подаваемое на исполнительный орган $U_1 = 5$ В, конечное значение выдвигающейся части $x_k = 2$ м, конечное значение угла раскрытия $\varphi_k = 1$ рад. Поведение системы представлено на рисунках 2 и 3.

Исходя из рисунков 2 и 3, можно сделать вывод, что использование нескольких исполнительных устройств позволяет реализовывать различные программы раскрытия элементов конструкции, позволяющие корректировать управляющие воздействие. К недостаткам можно отнести тот факт, что с ростом количества исполнительных машин будет расти и масса всей конструкции, и инерционность ее звеньев. При этом изменение упругой составляющей упора k_y приводит к изменению характеристики переходного процесса.

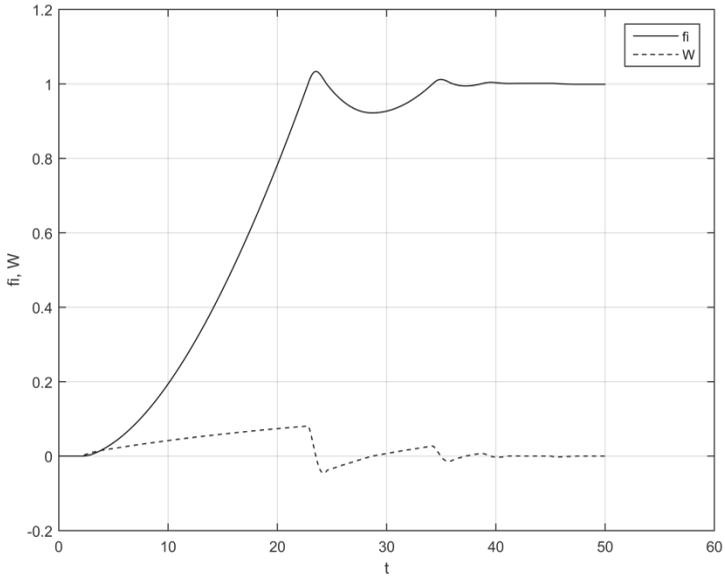


а)

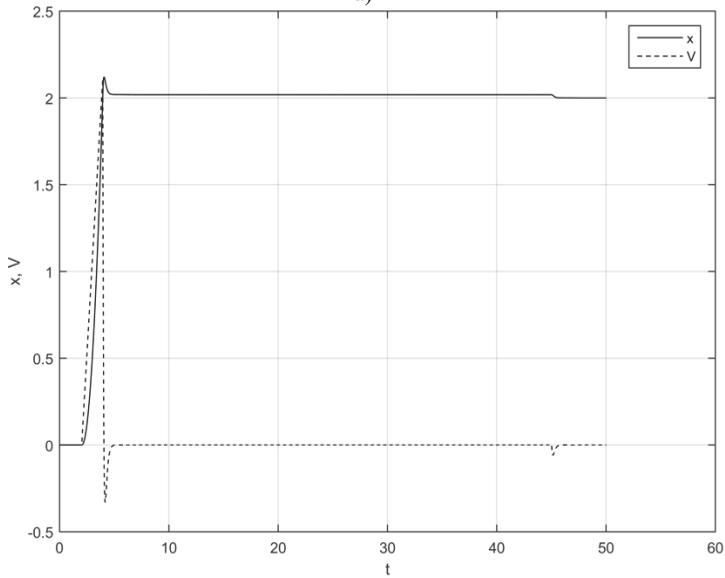


б)

Рис. 2. Характер динамики совместного раскрытия рефлектора: а) график $\varphi(t)$ и $\omega(t)$ при $k_y = 10$, $C_o = 25$; б) график $x(t)$ и $V(t)$ при $k_\phi = 100$, $C_\phi = 25$



а)



б)

Рис. 3. Характер динамики совместного раскрытия рефлектора: а) график $\varphi(t)$ и $\omega(t)$ при $k_y = 100$, $C_o = 25$; б) график $x(t)$ и $V(t)$ при $k_\phi = 100$, $C_\phi = 25$

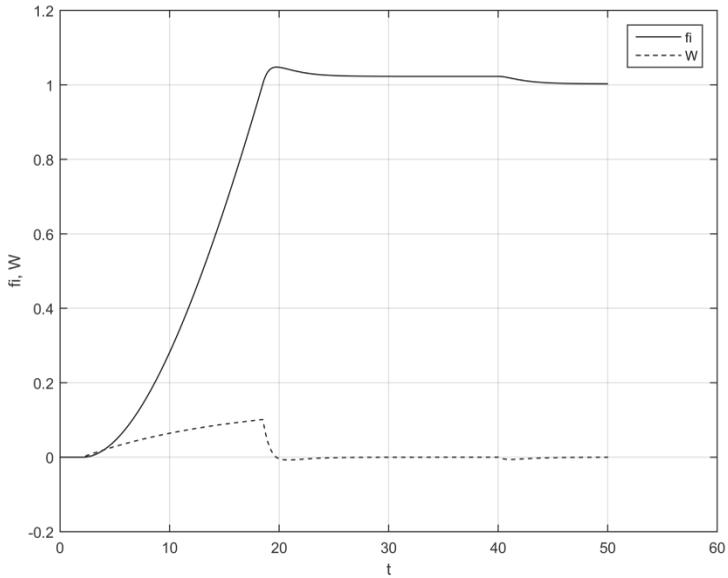
Напряжение питания U_1 действует со второй по 45 секунду, этим объясняется наличие небольшого колебания после выключения электродвигателя. U_1 становится равным нулю, момент или сила (в зависимости от вида движения) упора возвращает спицу к заданному конечному значению при этом срабатывает фиксатор (функция $H_\phi(\Delta\varphi)$) и спица зажимается в заданном положении.

Второй способ — использование центробежной силы для выдвижения промежуточного звена спицы. В данном способе полезная сила F_n создается центробежной силой. Система, описывающая совместное раскрытие элементов конструкции имеет такой же вид, что и системы (3), (6), за исключением того, что (7):

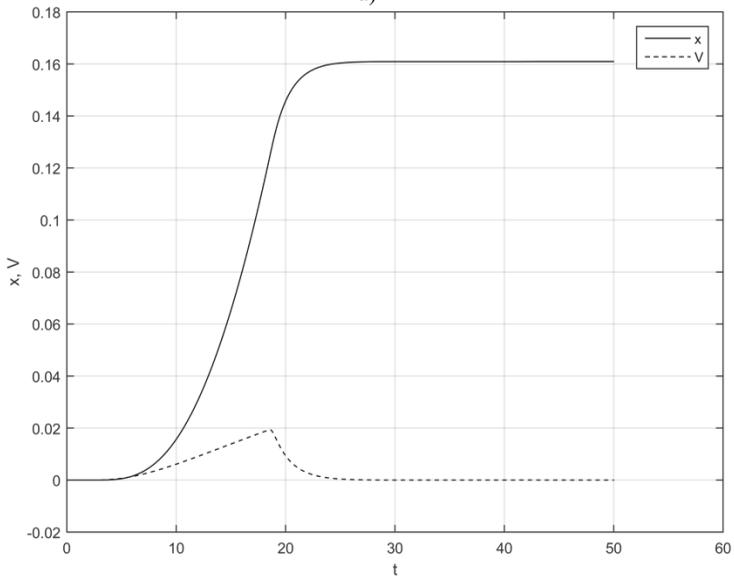
$$F_n = 0,5 \cdot m \cdot \omega^2 \cdot (x + 0,5 \cdot l). \quad (7)$$

В ходе моделирования были использованы те же параметры системы, что и для первого способа. Из рисунка 4 видно, что исполнительное устройство углового движения создает недостаточную центробежную силу для полного выдвижения промежуточного звена при исходных параметрах. Исходя из этого, управляющий сигнал U_1 был увеличен с 5 В до 15 В. Из-за увеличения управляющего сигнала увеличился и крутящий момент. На рисунке 5 показан процесс раскрытия и выдвижения при $U_1 = 15$ В. Из полученной характеристики, можно сделать вывод о том, что удалось добиться выполнения постановки на упоры как корневого звена, так и промежуточного звена рефлектора.

Данный вариант реализации лишен недостатка первого способа в виде дополнительного веса исполнительных устройств. Задача управления напряжением усложняется, поскольку возникает необходимость варьировать U_1 исполнительного устройства исходя из желаемых характеристик не только для углового движения, но и для поступательного.

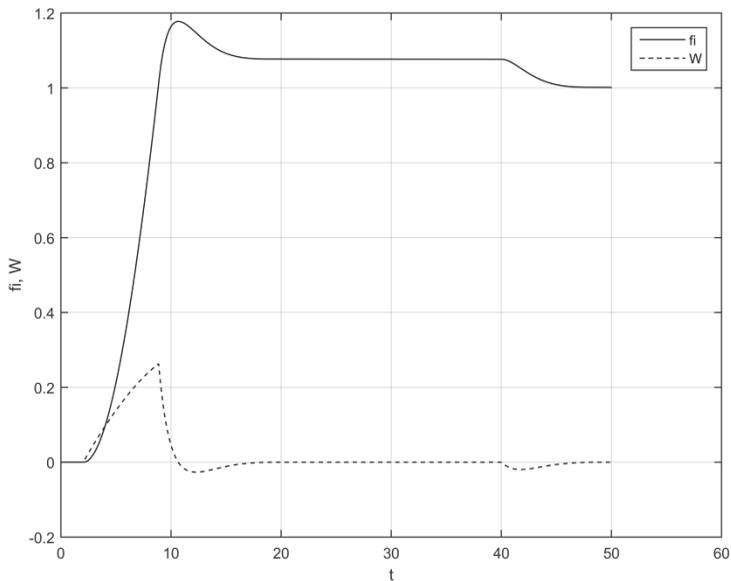


а)

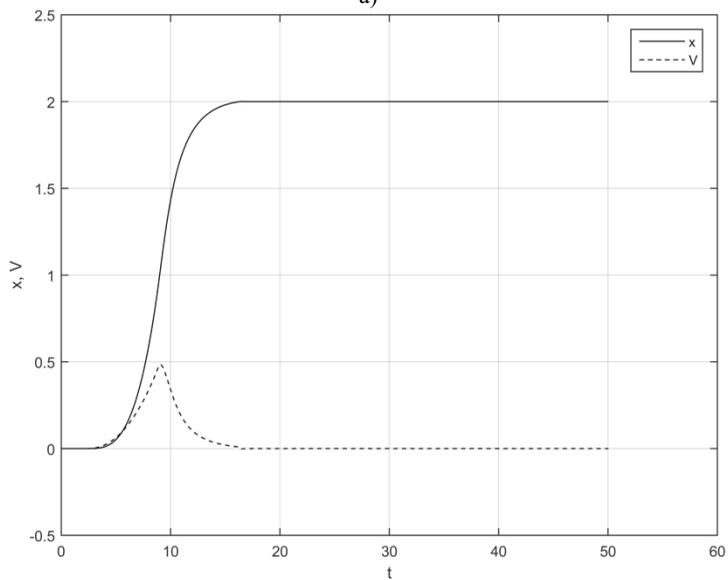


б)

Рис. 4. Характер динамики совместного раскрытия рефлектора при $U_1 = 5$ В:
а) график $\varphi(t)$ и $\omega(t)$; б) график $x(t)$ и $V(t)$



а)



б)

Рис. 5. Характер динамики совместного раскрытия рефлектора при $U_1 = 15$ В:

а) график $\varphi(t)$ и $\omega(t)$; б) график $x(t)$ и $V(t)$

4. Применение алгоритма коррекции параметров структуры управления. Для максимального использования возможностей электродвигателя и практических требований к развертыванию КТР, а также учитывая ограничения на мгновенные переключения, выберем релейную структуру управления с одним переключением для перевода системы из заданного начального положения в конечное.

Процесс динамики, описываемый системами (3), (6), можно в общем виде записать следующим образом (8):

$$\dot{\tilde{x}} = f(\tilde{x}, u, t), \quad (8)$$

где \tilde{x} — n -мерный вектор состояния, u — m -мерный вектор управления. Здесь $f = (\varphi, \omega, h, x, V, \Delta x_{сж})^T$ — вектор переменных состояния. На управление накладывается ограничение $u(t) \in \tilde{U}$. Решается задача перевода системы из начального положения, полагая $\tilde{x}(t_0) = \tilde{x}_0$, в конечное: $\tilde{x}(t_f) = \tilde{x}_f$, при минимуме функционала (9):

$$J = \int_0^T dt. \quad (9)$$

Было выполнено решение задачи с использованием принципа максимума. Численное решение возникающей задачи осуществлялось методом Ньютона [12], результаты вычислений показали, что затруднительно выбрать начальные значения сопряженных переменных, обеспечивающих сходимость метода, при этом даже незначительное изменение этих параметров приводит к тому, что метод расходится.

Ввиду затруднений в сходимости метода Ньютона было рассмотрено применение метода И.А. Крылова — Ф. Л. Черноушко [12]. Данным методом удалось добиться решения поставленной задачи. Однако при учете существующих в реальных системах ограничений на управление применение этого метода значительно усложняется.

Для преодоления трудностей реализации известных численных решений целесообразно обратиться к алгоритму с коррекцией параметров структуры управления [13].

Используя момент времени переключения управления τ в качестве параметра, представим структуру управления в виде (10):

$$u(t) = u_1(t) + \Delta u(t)l(t, \tau), \quad (10)$$

где $i = 2$, i — номер участка структуры управления, $l(t, \tau) = 1(t - \tau)$ — единичная функция Хэвисайда, $\Delta u = -2 \cdot u_1 \cdot l(t, \tau)$, u_1 — максимально допустимое значение управления, полученное исходя из поставленной задачи.

К исходной системе (8) добавляется уравнение $\dot{\tau} = v$, где, v — производная от τ по времени, принимаемое за управление. За критерий оптимальности принимается ФОР (11):

$$J = V_f(x, t_f) + 0,5 \int_{t_0}^{t_f} v^2 k^{-2} dt + 0,5 \int_{t_0}^{t_f} v_0^2 k^{-2} dt, \quad (11)$$

где $V(\tilde{x}, t_f) = 0,5 \Delta \tilde{x}_f^T \rho \Delta \tilde{x}_f$ — терминальная часть ФОР, $\rho = \text{diag}(\rho_\varphi, \rho_\omega, \rho_h, \rho_x, \rho_V, \rho_{\Delta x_{\text{сж}}})$ — диагональная матрица весовых коэффициентов, $\Delta \tilde{x}_f = \tilde{x}(t_f) - \tilde{x}_f$ — вектор разницы значений вектора переменных состояния в конечный момент времени t_f и заданных граничных значений, k — коэффициент, v_0 — оптимальное значение v .

Тогда уравнения для обобщенного вектора состояния примут вид (12):

$$\begin{aligned} \dot{\tilde{x}} &= f(\tilde{x}, u, t) l(t, t_f), \\ \dot{\tau} &= v l(t, t_f). \end{aligned} \quad (12)$$

Время окончания t_f является корректируемым параметром, а u представляется в виде структуры (10). Введением нового управления v исходная задача оптимизации по определению u решается косвенно — через нахождение v .

Отметим, что если количество переключений больше одного, то при вычислениях последовательность переключений τ_i остается неизменной. При сближении соседних значений моментов переключений до минимальной величины $(\tau_{i+1} - \tau_i) < \varepsilon_0$ принимается $\tau_{i+1} = \tau_i$, и количество всех переключений в структуре уменьшается на единицу, где $\varepsilon_0 > 0$ — заданная малая величина.

Рассмотрена задача оптимального управления раскрытием КТР, описываемая уравнениями вида (3), (6). В рассмотренном методе для различных случаев управления меняется не алгоритм, а начальные значения вектора сопряженных переменных и весовые коэффициенты при управлении в целевом функционале (11).

Для уравнения вращательного движения задача состоит в приведении разворачиваемого элемента с начальными условиями $\varphi_0 = 0$, $\omega_0 = 0$, $h_0 = 0$ и для поступательной компоненты $x_0 = 0$, $V_0 = 0$, $\Delta x_{\text{сж}0} = 0$ в заданное конечное состояние $\varphi_f = 1$, $x_f = 2$, $h_f = 0$,

$\omega_f = 0$, $V_f = 0$, $\Delta x_{\text{сж}f} = 0$ при ограничении на полезный момент $M_n = \pm 0,25 \text{ Н}\cdot\text{м}$.

Гамильтониан систем (3) и (6) имеет вид (13):

$$\begin{aligned}
 H = & P_\varphi \cdot \omega - P_\varphi \cdot \dot{h} + P_\omega \cdot \frac{M_n - M_{mp.} - M_{упор} - M_{фиксатор} - \mu \cdot \dot{\varphi}}{I} + P_\omega \cdot \frac{M}{2EI} \\
 + & P_x \cdot V - P_x \cdot \Delta \dot{x}_{\text{сж}} + P_V \cdot \frac{F_n - F_{mp.} - F_{упор} - F_{фиксатор} - \mu \cdot \dot{x}}{m} \\
 + & P_{\Delta \dot{x}_{\text{сж}}} \cdot \frac{F}{ES} + \frac{1}{2} \int_{t_0}^{t_f} \left(\frac{M}{k} \right)^2 dt.
 \end{aligned} \tag{13}$$

Найдем сопряженные переменные (14):

$$\begin{aligned}
 \dot{P}_\varphi = & -\frac{\partial H}{\partial \varphi} = P_\omega \cdot \left(\frac{k_y}{I} \cdot H(\Delta\varphi) - \frac{k_\phi}{I} \cdot H_\phi(\Delta\varphi) \right), \\
 \dot{P}_\omega = & -\frac{\partial H}{\partial \omega} = -P_\varphi + P_\omega \cdot \left(\frac{H(\Delta\varphi)}{I} \cdot C_\pi - \frac{H_\phi(\Delta\varphi)}{I} \cdot C_\phi + \frac{\mu}{I} \right) - P_V \cdot \frac{m \cdot \omega \cdot (x + 0.5 \cdot l)}{E \cdot S}, \\
 \dot{P}_h = & -\frac{\partial H}{\partial h} = P_\varphi - P_\omega \cdot \left(\frac{H(\Delta\varphi)}{I} \cdot C_\pi - \frac{H_\phi(\Delta\varphi)}{I} \cdot C_\phi + \frac{\mu}{I} \right), \\
 \dot{P}_x = & -\frac{\partial H}{\partial x} = P_V \cdot \left(\frac{k_y}{m} \cdot H(\Delta x) - \frac{k_\phi}{m} \cdot H_\phi(\Delta x) - \frac{m \cdot \omega^2}{2 \cdot E \cdot S} \right), \\
 \dot{P}_V = & -\frac{\partial H}{\partial V} = -P_x + P_V \cdot \left(\frac{H(\Delta x)}{m} \cdot C_\pi - \frac{H_\phi(\Delta x)}{m} \cdot C_\phi + \frac{\mu}{m} \right), \\
 \dot{P}_{x_{\text{сж}}} = & -\frac{\partial H}{\partial x_{\text{сж}}} = P_x - P_V \cdot \left(\frac{H(\Delta x)}{m} \cdot C_\pi - \frac{H_\phi(\Delta x)}{m} \cdot C_\phi + \frac{\mu}{m} \right).
 \end{aligned} \tag{14}$$

Система (14) решается при граничных условиях (15):

$$\begin{aligned}
 P_\varphi(t_f) = & \rho_1 \cdot [\varphi(t_f) - \varphi_f], P_\omega(t_f) = \rho_2 \cdot [\omega(t_f) - \omega_f], \\
 P_h(t_f) = & \rho_3 \cdot [h(t_f) - h_f], P_x(t_f) = \rho_4 \cdot [x(t_f) - x_f], \\
 P_V(t_f) = & \rho_5 \cdot [V(t_f) - V_f], P_{x_{\text{сж}}}(t_f) = \rho_6 \cdot [x_{\text{сж}}(t_f) - x_{\text{сж}f}].
 \end{aligned} \tag{15}$$

Алгоритм построения управления из последовательности следующих действий [12, 20]:

1. Интегрирование в прямом времени системы дифференциальных уравнений (12), включающей системы (3) и (6) при $v = 0$ на интервале $[t_0, t_f]$.
2. Вычисление граничных условий по формулам (15).
3. Интегрирование в обратном времени систем (14) и вычисление управления $v(t) = -k^2 \cdot P_t(t)$.
4. Моделирование системы дифференциальных уравнений (3), (6) в прямом времени с найденным управлением.

На рисунке 6 представлены зависимости $\varphi(t), \omega(t), x(t), V(t)$. Из графика видно, что удалось добиться перевода и развертываемого, и выдвигающегося элемента из начального состояния в конечное при ограничении на управление.

Результаты, приведенные на рисунках 6-7, показывают, что алгоритм оптимизации с коррекцией параметров структуры управления позволяет обеспечить успешное выполнение поставленной задачи. Одним из важных достоинств алгоритма является возможность построения управления в режиме реального времени. Его можно использовать для вычисления опорного управления в алгоритмах, основанных на принципе двухканальности [12, 20].

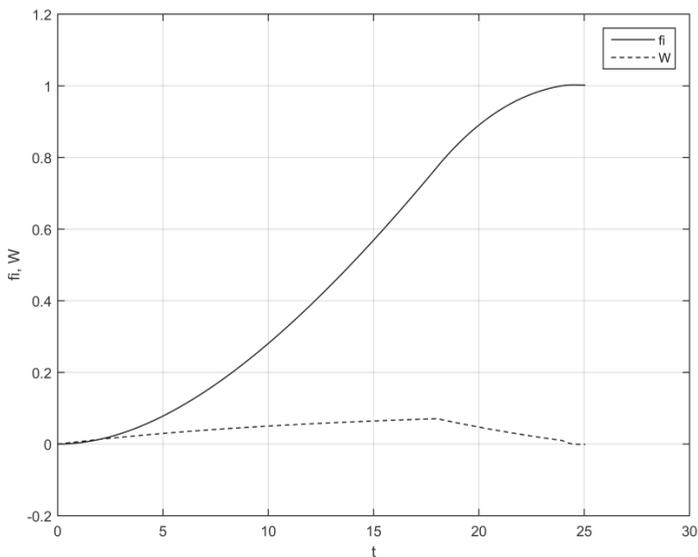


Рис. 6а. Характеристики раскрытия и выдвигения элементов конструкции:
 $\varphi(t)$ и $\omega(t)$

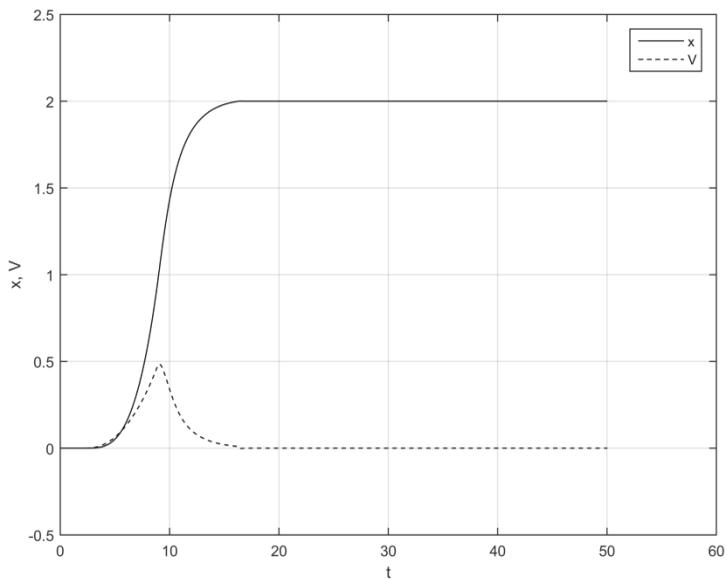


Рис. 6б. Характеристики раскрытия и выдвижения элементов конструкции: $x(t)$ и $V(t)$

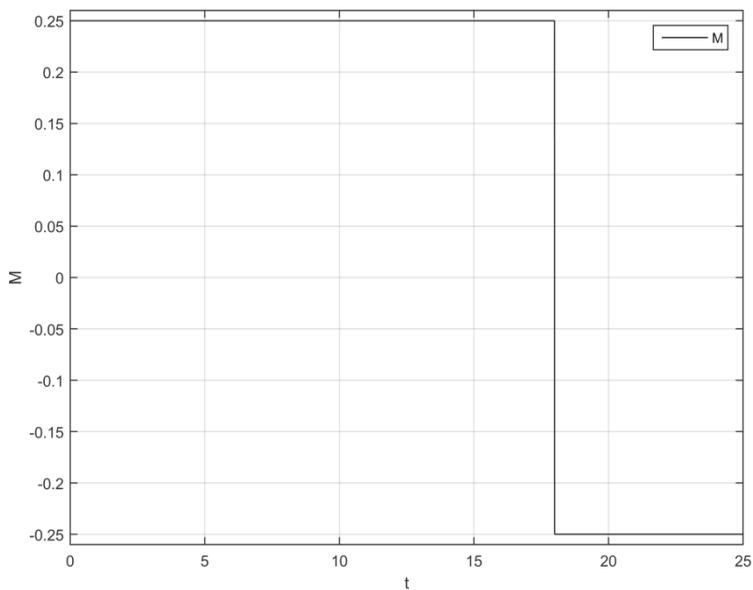


Рис. 7. Характеристика изменения структуры управления

В рассмотренном выше случае переход управления из одного предельного положения в другое осуществляется мгновенно (рисунок 7), что является идеализированным случаем. На самом деле смена управления осуществляется с некоторой задержкой: с ограничением на скорость изменения управления в структуре (10). При этом алгоритм позволяет реализовать различные варианты перехода.

Таким образом, можно рекомендовать предложенный в данной работе алгоритм к реализации совместного (вращательного и поступательного) раскрытия элементов КТР космического базирования в зависимости от времени раскрытия. Если время за которое необходимо произвести раскрытие КТР известно, то целесообразно использовать другую структуру управления, например, в виде экспоненты $z = a \cdot e^{-b \cdot t}$ с управлением параметрами $\dot{a} = v_1$, $\dot{b} = v_2$ или полинома второй степени $Z = a \cdot t^2 + b \cdot t + c$, $\dot{a} = v_1$, $\dot{b} = v_2$, $\dot{c} = v_3$.

5. Заключение. В ходе исследования были получены две универсальные модели одновременного осуществления нескольких этапов раскрытия рефлектора. Универсальность заключается в возможности их применения для конструкций с различными параметрами: от изменения материала, массы, габаритов до выбора управляющего органа (двигателя). Учет таких важных параметров, как величина прогиба элементов конструкции, диссипативность, динамическое удлинение, влияние свойств упора на угловое движение тела при постановке на упор, позволил описывать поведение системы с достаточно высокой точностью.

Два рассмотренных варианта позволяют управлять процессом раскрытия. При этом для реализации первого варианта требуется дополнительное исполнительное устройство, второй вариант лишен этого недостатка и является предпочтительным. При применении обоих методов стоит учитывать ограничения на максимальные нагрузки и моменты конструкции, поскольку создаваемые системой раскрытия усилия могут превысить критические.

Результаты данных исследований можно применять для выбора исполнительного органа, оценки необходимых усилий. Данная математическая модель является приближенной. Для повышения качества системы и точности результатов необходимо рассмотреть варианты решения прямых и обратных задач динамики для физических систем, включающих в себя учет энергетических составляющих движения, представленных в [21, 22].

Рассмотрен вариант управления процессом раскрытия КТР на основе алгоритма коррекции структуры управления. При этом реализация данного варианта позволяет менять компоненты структуры управления.

Использование центробежной силы, создаваемой при разворачивании корневого звена спицы силового каркаса, позволяет снизить предполагаемый вес системы. Важным достоинством алгоритма является возможность построения управления в режиме реального времени. Его можно использовать для вычисления опорного управления в алгоритмах, основанных на принципе двухканальности.

Работа выполнена в рамках реализации Федеральной целевой программы «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2014–2020 годы», Министерство образования России, соглашение № 14.577.21.0201 (уникальный идентификатор ПНИЭР RFMEFI 57715X0201): «Создание высокоэффективной системы беспроводной узконаправленной передачи энергии и информации для управления состоянием объектов космического базирования на основе лазерных и волоконнооптических технологий». Работа выполнена на оборудовании ЦКП «Центр радиотехнических и оптических измерений».

Литература

1. *Бей Н.А., Зимин В.Н.* Трансформируемые антенны больших размеров для геостационарных космических аппаратов. // Антенны. 2005. Вып. 10 (101). С. 24–27.
2. *Лопатин А. В., Рутковская М. А.* Обзор конструкций современных трансформируемых космических антенн (часть 1) // Вестник СибГАУ. 2007. № 2. С. 78–81.
3. *Seefeldt P. et al.* Gossamer-1: Mission concept and technology for a controlled deployment of gossamer spacecraft // *Advances in Space Research*. 2017. vol. 59. Issue 1. pp. 434–456.
4. *Khoroshilov V.S., Zakrzhevskii A.E.*, Dynamics of spacecraft due to elastic ring antenna deployment // *Vibrations in Physical Systems*. 2012. vol. 25. pp. 229–234.
5. *Isaenko S., Sochivko O., Dalyaev I.* Analysis of Vibration Modelling Validity of Space-Borne Robotic System // *Proceedings of the 26th DAAAM International Symposium*. 2016. pp. 0548–0553.
6. *Гриневич Д.В.* Исследование динамики раскрывающихся протяженных конструкций // *Вопросы электромеханики. Труды ВНИИЭМ*. 2013. Т. 134. С. 37–42.
7. *Кабдулин Г.В., Комков В.А., Мельников В.М., Харлов Б.Н.* Динамика управляемого раскрытия центробежными силами космических конструкций с компенсацией кинетического момента // *Космонавтика и ракетостроение*. 2009. №1(54). С. 189–198.
8. *Bagheri G.P., Malaek S.M.* On dynamic stiffness of spacecraft flexible appendages in deployment phase // *Aerospace Science and Technology*. 2015. vol. 47. pp. 1–9.
9. *Van Dyne D., Jennings A., Black J.* Simulation of Locking Space Truss Deployments // *2nd AIAA Spacecraft Structures Conference*. 2015. pp. 0227.
10. *Zakrzhevskii A.E.* Spacecraft dynamics with regard to elastic pantograph deployment // *Journal of Spacecraft and Rockets*. 2013. vol. 50. no. 2. pp. 475–479.
11. *Wang Y. et al.* Design and deployment analysis of modular deployable structure for large antennas // *Journal of Spacecraft and Rockets*. 2015. vol. 52. no. 4. pp. 1101–1111.
12. *Кабанов С.А.* Оптимизация динамики систем при действии возмущений // М.: ФИЗМАТЛИТ. 2008. 200 с.
13. *Кабанов С.А., Емельянов В.Ю., Митин Ф.В.* Оптимизация динамики системы создания формы крупногабаритных трансформируемых антенн космического базирования // *Вопросы радиоэлектроники. Серия ОТ*. 2016. № 8. Вып.6. С. 54–58.

14. *Кривушов А.И.* Создание и исследование математической модели вращательного движения спицы крупногабаритного рефлектора космического базирования // Вопросы радиоэлектроники. 2017. Вып. 7. С. 126–130.
15. *Митин Ф.В.* Разработка и анализ математической модели поступательного движения трансформируемых космических конструкций // Вопросы радиоэлектроники. 2017. Вып. 7. С. 121–125.
16. *Зубов В.Г.* Механика // М.: Наука. 1978. 352 с.
17. *Зимин В.Н.* Механика трансформируемых структурных космических конструкций // Вестник Самарского государственного университета. 2007. № 4(54). С. 105–114.
18. *Работонов Ю.Н.* Сопротивление материалов // М. Физматгиз. 1962. 456 с.
19. *Дарков А.В., Шапиро Г.С.* Сопротивление материалов // М.: Высшая школа. 1975. 654 с.
20. Справочник по теории автоматического управления/ Под ред. А.А. Красовского // М.: Наука. 1987. 712 с.
21. *Костин Г.В., Саурин В.В.* Метод интегриродифференциальных соотношений для анализа собственных колебаний мембран // Прикладная математика и механика. 2009. Том 73. Вып. 3. С. 459–473.
22. *Костин Г.В., Саурин В.В.* Моделирование пространственных движений упругой балки на основе метода интегриродифференциальных соотношений // Труды XIV международной конференции «Современные проблемы механики сплошной среды». 2010. Том 2. С. 165–169.

Кабанов Сергей Александрович — д-р техн. наук, профессор, профессор кафедры систем управления и компьютерных технологий, Балтийский государственный технический университет «ВОЕНМЕХ» им. Д.Ф. Устинова (БГТУ «ВОЕНМЕХ» им. Д.Ф. Устинова). Область научных интересов: процессы управления. Число научных публикаций — 150. kaba-sa@mail.ru; 1-я Красноармейская ул., 1, Санкт-Петербург, 199005; р.т.: +7(812)317-82-49.

Кривушов Алексей Игоревич — инженер НИЛ РИУС, Балтийский государственный технический университет «ВОЕНМЕХ» им. Д.Ф. Устинова (БГТУ «ВОЕНМЕХ» им. Д.Ф. Устинова). Область научных интересов: оптимальное управление. Число научных публикаций — 2. krivushov.alexey@yandex.ru; 1-я Красноармейская ул., 1, г. Санкт-Петербург, 199005; р.т.: +7 (963) 328-41-72.

Митин Фёдор Васильевич — аспирант кафедры систем управления и компьютерных технологий, Балтийский государственный технический университет «ВОЕНМЕХ» им. Д.Ф. Устинова (БГТУ «ВОЕНМЕХ» им. Д.Ф. Устинова). Область научных интересов: системы обработки информации и управления. Число научных публикаций — 6. fedor28@list.ru; 1-я Красноармейская ул., 1, Санкт-Петербург, 199005; р.т.: +7 (905) 254-41-49.

Поддержка исследований. Работа выполнена в рамках реализации Федеральной целевой программы «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2014–2020 годы», Министерство образования России, соглашение № 14.577.21.0201 (уникальный идентификатор ПНИЭР RFMEFI 57715X0201): «Создание высокоэффективной системы беспроводной узконаправленной передачи энергии и информации для управления состоянием объектов космического базирования на основе лазерных и волоконнооптических технологий». Работа выполнена на оборудовании ЦКП «Центр радиотехнических и оптических измерений».

S.A. KABANOV, A.I. KRIVUSHOV, F.V. MITIN
MODELING OF JOINT DEPLOYMENT OF UNITS OF THE LARGE-SIZED TRANSFORMABLE REFLECTOR OF SPACE BASING

Kabanov S.A., Krivushov A.I., Mitin F.V. Modeling of Joint Deployment of Units of the Large-Sized Transformable Reflector of Space Basing.

Abstract. The transformable designs of space basing are delivered into orbit in a folded state, which creates the task of their reliable disclosure. In this paper we propose to use an actuator in the form of electrical machines as the executive body. The use of this type of actuator allows controlling the process of deployment.

As a large-sized transformable structure we consider the space-based reflector. At present, the transfer of the machines from the folded state to the operation state is carried out by stages. The paper considers the joint implementation of two stages: the rotation of the root unit of the spoke and the extension of the intermediate unit. Mathematical models for rotational and translational motions are developed which take into account such parameters as bending and contraction of the spoke. Modeling and analysis of the results of different variants of the joint disclosure of the reflector elements are made: the use of the engine for each of the components of the motion and the use of centrifugal force for the extension of the spoke.

The application of the algorithm for correcting the control parameters is considered. One of the important advantages of the algorithm is the ability to carry control in real time. It can be used to calculate the reference control in algorithms based on the two-channel principle.

Keywords: mathematical model, rotational motion, translational motion, modeling, correction of the parameters of the control structure, large-sized transformable reflector.

Kabanov Sergei Aleksandrovich — Ph.D., Dr. Sci., professor, professor of control systems and computer technologies department, Baltic State Technical University "Voenmeh" D.F.Ustinov. Research interests: control processes. The number of publications — 150. kaba-sa@mail.ru; 1, 1-st Krasnoarmeyskaya, St.-Petersburg, 190005, Russia; office phone: +7(812)317-82-49.

Krivushov Aleksei Igorevich — engineer of NIL RIUS, Baltic State Technical University "Voenmeh" D.F.Ustinov. Research interests: optimal control. The number of publications — 2. krivushov.alexey@yandex.ru; 1, 1-st Krasnoarmeyskaya, St.-Petersburg, 190005, Russia; office phone: +7 (963) 328-41-72.

Mitin Fedor Vasilievich — Ph.D. student of control systems and computer technologies department, Baltic State Technical University "Voenmeh" D.F.Ustinov. Research interests: information processing and control systems. The number of publications — 6. fedor28@list.ru; 1, 1-st Krasnoarmeyskaya, St.-Petersburg, 190005, Russia; office phone: +7 (905) 254-41-49.

Acknowledgements. The work was carried out within the framework of the implementation of the Federal Target Program «Research and Development in Priority Areas for the Development of the Russian Science and Technology Complex for 2014–2020», the Ministry of Education of Russia, agreement No. 14.577.21.0201 (unique identifier of the RFMEFI 57715X0201): «Creating a highly efficient wireless system Focused transmission of energy and information for managing the state of space-based objects based on laser and fiber-optic technologies». The work was performed on the equipment of the Center for Radio-Technical and Optical Measurements.

References

1. Bei N.A., Zimin V.N. [Transformable large antenna for geostationary spacecraft]. *Antenny – Antennas*. 2005. vol. 10(101). pp. 24–27. (In Russ.).

2. Lopatin A. V., Rutkovskaja M.A. [Overview of the designs of modern transformed space antennas (Part 1)]. *Vestnik SibGAU – Vestnik of SibGAU*. 2007. vol 2. pp. 78–81. (In Russ.).
3. Seefeldt P. et al. Gossamer-1: Mission concept and technology for a controlled deployment of gossamer spacecraft. *Advances in Space Research*. 2017. vol. 59. Issue 1. pp. 434–456.
4. Khoroshilov V.S., Zakrzhevskii A.E., Dynamics of spacecraft due to elastic ring antenna deployment. *Vibrations in Physical Systems*. 2012. vol. 25. pp. 229–234.
5. Isaenko S., Sochivko O., Dalyaev I. Analysis of Vibration Modelling Validity of Space-Borne Robotic System. Proceedings of the 26th DAAAM International Symposium. 2016. pp. 0548–0553.
6. Grinevich D.V. [Investigation of the dynamics of unfolding extended structures]. *Voprosy jelectromehaniki. Trudy VNIIEJEM – Electromechanical matters. VNIIEJEM studies*. 2013. Issue 134. pp. 37–42. (In Russ.).
7. Kabdulin G.V., Komkov V.A., Mel'nikov V.M. Harlov B.N. [Dynamics of controlled opening by centrifugal forces of cosmic structures with compensation of the kinetic moment]. *Kosmonavtika i raketostroenie – Cosmonautics and rocket engineering*. 2009. vol. 1(54). pp. 189–198. (In Russ.).
8. Bagheri G.P., Malaek S.M. On dynamic stiffness of spacecraft flexible appendages in deployment phase. *Aerospace Science and Technology*. 2015. vol. 47. pp. 1–9.
9. Van Dyne D., Jennings A., Black J. Simulation of Locking Space Truss Deployments. 2nd AIAA Spacecraft Structures Conference. 2015. pp. 0227.
10. Zakrzhevskii A.E. Spacecraft dynamics with regard to elastic pantograph deployment. *Journal of Spacecraft and Rockets*. 2013. vol. 50. no. 2. pp. 475–479.
11. Wang Y. et al. Design and deployment analysis of modular deployable structure for large antennas. *Journal of Spacecraft and Rockets*. 2015. vol. 52. no. 4. pp. 1101–1111.
12. Kabanov S.A. *Optimizacija dinamiki sistem pri dejstvii vozmushhenij* [Optimization of the dynamics of systems under the influence of perturbations]. M.: FIZMATLIT. 2008. 200 p. (In Russ.).
13. Kabanov S.A., Emel'janov V.Ju., Mitin F.V. [Optimization of the dynamics of the system for creating the shape of large-size transformed space-based antennas]. *Voprosy radioelektroniki. Serija OT – Questions of radio-electronics. General technical series*. 2016. vol. 8. no. 6. pp. 54–58. (In Russ.).
14. Krivushov A.I. [Creation and investigation of the mathematical model of the rotational movement of the spokes of a large-sized space-based reflector]. *Voprosy radioelektroniki – Questions of radio-electronics*. 2017. vol. 7. pp. 126–130. (In Russ.).
15. Mitin F.V. [Development and analysis of the mathematical model of translational motion of transformed space structures]. *Voprosy radioelektroniki – Questions of radio-electronics*. 2017. vol. 7. pp. 121–125. (In Russ.).
16. Zubov V.G. *Mehanika* [Mechanics]. M.: Nauka. 1978. 352 p. (In Russ.).
17. Zimin V.N. [Mechanics of transformable structural space structures]. *Vestnik Samarskogo gosuniversiteta – Vestnik of Samara State University*. 2007. vol. 4(54). pp. 105–114. (In Russ.).
18. Rabotonov Ju.N. *Soprotivlenie materialov* [Strength of materials]. M. Fizmatgiz. 1962. 456 p. (In Russ.).
19. Darkov A.V., Shapiro G.S. *Soprotivlenie materialov* [Strength of materials]. M.: Vysshaja shkola. 1975. 654 p. (In Russ.).
20. *Spravochnik po teorii avtomaticheskogo upravlenija. Pod red. A.A. Krasovskogo* [Handbook on the theory of automatic control. Edited by A.A. Krasovskij]. M.: Nauka. 1987. 712 p. (In Russ.).
21. Kostin G.V., Saurin V.V. [The method of integro differential relations for the analysis of natural vibrations of membranes]. *Prikladnaja matematika i mehanika – Journal of Applied Mathematics and Mechanics*. 2009. Issue 73. vol. 3. pp. 459–473. (In Russ.).
22. Kostin G.V., Saurin V.V. [Modeling of spatial motions of an elastic beam on the basis of the method of integro-differential relations]. *Trudy XIV mezhdunarodnoj konferencii «Sovremennye problemy mehaniki sploshnoj sredy»* [Proceedings of the XIV International Conference "Contemporary Problems of Continuum Mechanics"]. 2010. vol. 2. pp. 165–169. (In Russ.).

А.В. ПОНОМАРЕВ
**МЕТОДЫ ОБЕСПЕЧЕНИЯ КАЧЕСТВА РЕЗУЛЬТАТОВ В
СИСТЕМАХ КРАУД-ВЫЧИСЛЕНИЙ: АНАЛИТИЧЕСКИЙ
ОБЗОР**

Пономарев А.В. Методы обеспечения качества результатов в системах крауд-вычислений: аналитический обзор.

Аннотация. В настоящее время все большее распространение получают крауд-вычисления, то есть привлечение к задачам обработки информации широкого круга людей, взаимодействующих посредством инфокоммуникационных технологий. Тем не менее практическое применение этой технологии в значительной мере сдерживается неопределенностью качества получаемых результатов. В этих условиях задача систематизации сведений об используемых на данный момент методах обеспечения качества и идентификации перспективных направлений их развития является особенно актуальной. В статье обсуждаются результаты систематического обзора журнальных публикаций полнотекстовых баз ScienceDirect и IEEE Xplore, вышедших после 2012 года. Выделены наиболее распространенные на данный момент направления в обеспечении качества, используемые модели и принимаемые допущения, обозначены границы применимости методов. Отмечено, что наибольшее распространение получили методы, основанные на согласовании оценок, полученных от разных участников, и методы, основанные на применении теоретико-игровых моделей.

Ключевые слова: крауд-вычисления, человеко-машинные вычисления, краудсорсинг, социальные вычисления, человеко-машинные системы, человеческие факторы, систематический обзор литературы.

1. Введение. Крауд-вычисления (англ. *crowd computing*), или масштабные человеко-машинные вычисления в настоящее время широко применяются для выполнения различных задач, для которых не существует надежных способов автоматического решения, но которые являются естественными для человека и в которых человек может быть эффективен (например, разметка и аннотирование изображений и аудиозаписей, обработка сложных данных и данных с неопределенностью). Слово «вычисления» в составе этого термина трактуется максимально широко, не ограничиваясь рамками математических вычислений и неся скорее значение «обработка информации». Так, одно из популярных применений крауд-вычислений — создание размеченных эталонных наборов для задач машинного обучения (например, [1]). В этом случае человек-«вычислитель» обрабатывает какой-то объект (как правило, изображение, аудио- или видеофрагмент), и результатом обработки (или «вычислений») является некоторая метка, связанная с содержанием объекта и отвечающая цели создания набора.

Использование крауд-вычислений сопряжено с рядом принципиальных ограничений [2], среди которых проблема качества

результатов, получаемых таким образом, является наиболее острой и во многом именно она ограничивает применение крауд-вычислений.

Причиной этого является как общая склонность к ошибкам человека, включаемого в процесс обработки информации (даже при добросовестном выполнении заданий), так и различные виды сознательных нарушений инструкций участником для получения максимальной выгоды при минимизации затрачиваемых усилий (например, [3]).

Под качеством в широком смысле понимается соответствие результата выполнения некоторой задачи (сбора или обработки информации) с помощью крауд-вычислений и «истинного» результата этой задачи. Роль последнего на практике, как правило, выполняет результат, полученный заинтересованным в выполнении задачи компетентным «экспертом». Конкретные способы оценки качества в значительной мере зависят от области применения. Например, в задачах разметки результатом крауд-вычислений является множество меток, присвоенных объектам участниками, и для оценки качества этого множества часто используются меры точности (precision) и полноты (recall), позаимствованные из информационного поиска, а в роли эталонного множества выступает множество меток, полученных от эксперта.

Цель данного исследования — выявление основных подходов и методов, используемых в настоящее время для управления качеством информации, получаемой с помощью краудсорсинга (управления качеством решения задач с применением краудсорсинга), а также идентификация актуальных направлений исследований в этой области.

Для достижения поставленной цели был проведен отбор современных научных публикаций, касающихся вопросов обеспечения качества, а затем на основе анализа этих публикаций выявлены основные методы и модели, определены их особенности и типовые области применения.

2. Методика исследования. В основу исследования положена методика проведения систематических обзоров литературы, предложенная Б. Китченхем [4, 5]. Одним из важнейших достоинств этой методики является воспроизводимость результатов за счет точной спецификации используемых источников информации и поисковых запросов, а также фиксации всех промежуточных результатов в протоколе исследования. Вместе с тем строгое применение данной методики для анализа обширных областей знаний связано с рядом технических трудностей — в первую очередь, это управление размером выборки исследуемых публикаций и учет их значимости. Для преодоления этих трудностей методика была расширена.

Систематический обзор был применен к относительно небольшому множеству современных публикаций, затем множество исследуемых публикаций было расширено за счет приставейных списков литературы отобранных публикаций.

Исследование, результаты которого изложены в данной статье, состояло из двух этапов:

1) Систематический обзор современной литературы по методике Б. Китченхем, ставящий целью выявление актуальных методов и подходов к обеспечению качества в системах крауд-вычислений.

2) Анализ источников, цитируемых в отобранных статьях, для идентификации наиболее значимых исторических моделей и методов (к которым, в частности, предлагаются усовершенствования в публикациях, отобранных на этапе 1). Эта информация дополняет современную проблематику, выявленную на этапе 1, и позволяет раскрыть историческую перспективу и логику развития данной области исследований.

В качестве источника публикаций для систематического обзора были использованы две полнотекстовые библиографические базы: ScienceDirect и IEEE Xplore. В каждой из баз был осуществлен поиск статей в периодических изданиях (журналах), опубликованных за последние 5 лет (с 2012 года), и удовлетворяющих следующему поисковому запросу: «quality AND ("crowd computing" OR crowdsourcing OR "crowd-sourcing" OR "crowd sourcing" OR "social computing" OR "social computations" OR "human computations")» или аналогичному (с учетом поддерживаемого библиографической базой поискового языка). Таким образом, дальнейшему анализу подвергались статьи, название, аннотация или ключевые слова которых содержали слово quality (качество) и один из терминов, которыми наиболее часто описывается процесс масштабных человеко-машинных вычислений, производимых с помощью сети Интернет. Ограничение поиска статьями в периодических изданиях (журналах) вызвано, во-первых, тем, что результаты, опубликованные в таких изданиях, как правило, прошли более серьезную апробацию и строгий отбор, во-вторых, тем, что это позволяет снизить общий объем первичного материала для ознакомления (сконцентрировавшись на — предположительно — наиболее апробированном его подмножестве). В результате выполнения запросов из базы ScienceDirect было отобрано 124 статьи, а из IEEE Xplore — 79.

Затем в ходе скрининга аннотаций полученных 203 статей были отобраны статьи, удовлетворяющие хотя бы одному из следующих условий: а) статья посвящена исследованию того или иного способа обеспечения качества при использовании краудсорсинга (например [1, 6]); б) статья посвящена описанию способа применения краудсорсинга

для решения практической задачи, но в аннотации авторы подчеркивают важность обеспечения качества и заявляют об использовании того или иного подхода или метода (независимо от степени оригинальности) — например [7, 8]. Вторичные публикации (обзоры) были отнесены в отдельную категорию. В результате для дальнейшего исследования была отобрана 41 статья, из которых 3 обзорных.

При выполнении второго этапа исследуемая литература уже не ограничивалась определенными библиографическими базами и характером публикации (принимались во внимание и не только публикации в периодических изданиях). Решающую роль здесь играла частота (и семантика) ссылок из статей, отобранных в первом этапе.

Таким образом, используемая методика позволяет одновременно получить объективный «срез» публикаций на заданную тему в авторитетных периодических изданиях и выявить базовые методы и наиболее влиятельные публикации.

3. Количественный анализ и типизация. Для определения фазы, в которой находятся исследования в данной области, была проанализирована динамика количества публикаций, удовлетворяющих сформулированному запросу (рисунок 1). Приведенный график показывает, что увеличение количества публикаций происходит нарастающими темпами, то есть проблема переживает фазу активного исследования, сопровождающегося значительным энтузиазмом по поводу потенциала систем, основанных на краудсорсинге и крауд-вычислениях. Заметим также, что учет общего роста количества публикаций в выбранных полнотекстовых базах со временем не вносит существенных изменений в представленную закономерность.

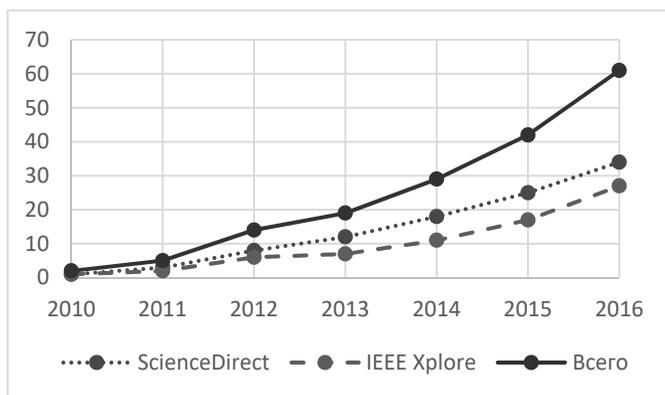


Рис. 1. Количество статей, посвященных исследуемой тематике, по годам (в полнотекстовых библиографических базах ScienceDirect и IEEE Xplore)

3.1. Классификация. В ходе детального анализа отобранных публикаций выделено 6 основных направлений, по которым могут быть сгруппированы методы обеспечения качества: методы согласования (consensus), методы проектирования потока работ, методы централизованного назначения работ, теоретико-игровые методы, методы, основанные на учете свойств заданий, и методы, основанные на анализе действий пользователя и воздействии на него. Помимо этого, в отдельные категории отнесены обзорные статьи и статьи, которые были признаны нерелевантными после анализа полного текста. Количественное распределение работ по категориям приведено в таблице 1.

Таблица 1. Классификация отобранных работ по направлениям

Направление	Ссылки	Количество
Методы согласования	[6, 7, 9-17]	11
Методы проектирования потока работ	[18-21]	4
Методы централизованного назначения	[8, 22-25]	5
Теоретико-игровые методы	[26-32]	7
Методы, основанные на учете свойств заданий	[1, 33, 34]	3
Методы, основанные на анализе действий пользователя и воздействии на него	[35-37]	3
Обзорные статьи	[38-40]	3
Нерелевантные (исключенные)	[41-45]	5

3.2. Обзорные статьи. В ряде обзорных статей неизбежно затрагивается вопрос обеспечения качества. В сформированную выборку вошли следующие обзорные статьи.

Статья [39] содержит большое количество вводной информации по многим аспектам, касающимся организации краудсорсинга, а также богатую библиографию, однако непосредственно вопросам обеспечения качества уделено не так много внимания.

Обзорная статья [40] посвящена широкому кругу проблем, связанных с данными, получаемыми с помощью краудсорсинга. В частности, выделяется три проблемы, связанные с этим: управление качеством, управление стоимостью и управление латентностью (временем). Приводится достаточно подробный обзор различных методов организации краудсорсинга, а также применений его в области обработки данных, однако рассуждения об управлении качеством в основном ограничены методами согласования, что несколько уже, чем в настоящей статье.

Обзорная статья [38] содержит обзор основных методов обеспечения качества. В частности, предлагает лаконичную платформу для классификации различных методов обеспечения качества, однако некоторые методы, набирающие популярность в последние годы, остались неосвещенными (в первую очередь, это касается теоретико-игровых методов и методов централизованного назначения).

Таким образом, настоящая статья, будучи ориентированной на обзор современных подходов к обеспечению качества и представляющая результаты обобщения четко определенного множества первичных работ, отличается от перечисленных статей, дополняет их и, возможно, наряду с ними послужит распространению крауд-вычислений и дальнейшему развитию исследований в этой области.

3.3. Нерелевантные статьи. Ряд статей, отобранных в результате скрининга аннотаций, оказались не вполне соответствующими цели исследования. В первую очередь, это статьи по пространственному краудсорсингу, в которых поднимается вопрос качества информации (Quality of Information, QoI), однако интерпретируется это качество просто как выполнение требований по покрытию, в определенном смысле, полнота информации: [41, 42, 44].

4. Направления в обеспечении качества. При применении крауд-вычислений типовым является следующий сценарий:

- разработка задания;
- распределение заданий между участниками (исполнителями);
- выполнение заданий участниками;
- обработка результатов.

Поскольку методы обеспечения качества имеют первостепенную важность в практическом применении таких систем, существующие (в том числе и найденные в процессе обзора литературы) методы охватывают все этапы этого сценария. Применимость методов различных направлений к этапам типового сценария отображена в таблице 2.

Таблица 2. Охват различных направлений в обеспечении качества

Направление	Разработка задания	Распределение заданий	Выполнение	Обработка результатов
Методы согласования				+
Методы проектирования потока работ		+	+	+
Методы централизованного назначения		+		
Теоретико-игровые методы	+	+	+	+
Методы, основанные на учете свойств заданий			+	+
Методы, основанные на анализе действий пользователя и воздействия на него	+		+	+

Помимо этого, ключевую роль при обсуждении методов обеспечения качества играют: а) набор допущений, которые делаются относительно задач, исполнителей и имеющейся информации об исполнителе; б) дополнительные характеристики (помимо качества) процесса, которые необходимо учесть (скорость решения задачи, стоимость и пр.).

Оставшаяся часть данного раздела построена следующим образом. Отдельный подраздел выделен под каждое из упомянутых направлений; в начале подраздела дается краткая характеристика направления, основные идеи, которые лежат в его основе (для пояснения этих базовых идей используются источники, отобранные на втором этапе обзора литературы — часто основополагающие и особенно цитируемые публикации), затем приводится краткая характеристика всех современных работ по соответствующему направлению, удовлетворяющих условиям систематического обзора литературы (этап 1).

4.1. Методы согласования (результатов, полученных от разных исполнителей). Данные методы получили широкое распространение в тех вариантах организации краудсорсинга, где *недоступна априорная информация об исполнителях* (например, об ожидаемом качестве выполнения заданий исполнителем) и о заданиях (например, их трудность). В этом смысле алгоритмы согласования являются наиболее универсальными, потому что накладывают минимальные ограничения на содержимое профиля исполнителя и могут работать с анонимными исполнителями. Общий принцип работы таких методов заключается в том, что каждое задание выполняется несколькими участниками системы крауд-вычислений, затем полученные результаты обобщаются (согласуются) для получения итогового результата.

Простейшим алгоритмом такого обобщения является мажоритарное голосование (majority voting, MV) [46], которое при определенных допущениях позволяет получать результаты, качество которых является достаточным для многих прикладных задач [6].

Недостаток мажоритарного голосования в том, что при агрегации данных, полученных от разных участников, значимость этих фрагментов считается одинаковой, что полностью игнорирует возможные различия в уровне компетентности и добросовестности участников. Более сложные алгоритмы пытаются преодолеть этот недостаток. Один из основных алгоритмов такого плана — алгоритм DS, предложенный в [47], являющийся одним из наиболее часто упоминаемых в статьях, посвященных методам поиска консенсуса, и

очень часто выступающий в качестве базового при оценке предлагаемых методов (например, в работах [6, 10, 16]).

Методы согласования естественным образом разбиваются на две группы — методы, основанные на EM-алгоритме (expectation minimization), и методы, не основанные на EM-алгоритме. В первую группу попадают, например, DS, GLAD [46], RY [48], KOS [49], Spectral DS [50], ZenCrowd [51], а во вторую — мажоритарное голосование, AWMV [6].

EM-алгоритм — широко распространенный алгоритм математической статистики для нахождения оценок максимального правдоподобия параметров вероятностных моделей [52, 53]. Все алгоритмы согласования (поиска консенсуса), основанные на EM-алгоритме, одновременно оптимизируют параметры модели и вычисляют обобщенные ответы для заданий (в идеале, приближающиеся к «истинным»). Алгоритмы согласования, относящиеся к семейству EM, в сущности, различаются набором параметров модели, которые определяют ответ на задание, получаемый от участника. Так, например, в DS моделью является матрица ошибок каждого участника, GLAD — уровни компетентности участников и уровни сложности заданий, через взаимодействие которых выражаются обобщенные оценки и так далее.

Формально задача поиска консенсуса формулируется следующим образом [6]. Задан набор примеров $E = \{e_i\}_{i=1}^I$, где каждый пример $e_i = (x_i, y_i)$ состоит из набора значений признаков x_i и истинной метки y_i . Множество участников системы будем обозначать $U = \{u_j\}_{j=1}^J$. Каждая метка, полученная от участника, принадлежит множеству классов $C = \{c_k\}_{k=1}^K$. Для удобства индексы элементов этих множеств используются как идентификаторы соответствующих объектов, то есть можно говорить о том, что участник i отнес пример j к классу k . Чаще всего рассматривается (наиболее простая) задача с двумя классами; в этом случае класс c_1 ($k=1$) соответствует положительным («+») примерам, а класс c_2 ($k=2$) — отрицательным («-»). Каждому примеру i соответствует множество меток $I_i = \{I_{ij}\}_{j=1}^J$, элемент I_{ij} которого получен от участника j . Все метки примеров в наборе данных формируют матрицу аннотаций $L = \{I_i\}_{i=1}^I$, $I_{ij} \in \{c_1, 0, c_2\}$, где 0 означает, что участник не предоставил метки для данного примера. Участник j характеризуется также

матрицей $N^{(j)} = \{n_{ik}^j\}, 1 \leq i \leq I, 1 \leq k \leq K$, каждый элемент которой равен количеству раз, когда участник i отнес пример j к классу k . На практике, как правило, $n_{ik}^j \in \{0, 1\}$. Известны также априорные вероятности каждого из классов.

Задача состоит в том, чтобы найти метки \hat{y}_i , минимизирующие функцию ошибки:

$$R = \frac{1}{I} \sum_{i=1}^I \mathbf{I}(\hat{y}_i \neq y_i)$$

при заданных L и $y \in (c_1, c_2)$. Здесь \mathbf{I} — индикаторная функция.

Алгоритм DS решает эту задачу следующим образом. Пусть каждый участник j системы описывается матрицей ошибок классификации (confusion matrix) $\pi_{kl}^{(j)}$, задающей вероятность того, что данный участник укажет класс l для примера, истинным классом которого является k . На E-шаге по известной матрице аннотаций L алгоритм оценивает вероятности того, что пример i принадлежит классу k (для каждого класса) в соответствии со следующим уравнением:

$$P(\hat{y}_i = c_k | L) = \frac{\prod_{j=1}^J \prod_{l=1}^L (\pi_{kl}^{(j)})^{n_{il}^j} P(c_k)}{\sum_{q=1}^K \prod_{j=1}^J \prod_{l=1}^L (\pi_{ql}^{(j)})^{n_{il}^j} P(c_q)}$$

где $P(c_k)$ и $P(c_q)$ — априорные вероятности принадлежности примера классам c_k и c_q соответственно, вычисленные на M-шаге алгоритма. На M-шаге происходит уточнение значений матрицы ошибок классификации и априорных вероятностей классов в соответствии со следующими соотношениями:

$$\hat{\pi}_{kl}^{(j)} = \frac{\sum_{i=1}^I \mathbf{I}(\hat{y}_i = c_k) n_{il}^{(j)}}{\sum_{l=1}^K \sum_{i=1}^I \mathbf{I}(\hat{y}_i = c_k) n_{il}^{(j)}},$$

$$\hat{P}(c_k) = \frac{1}{I} \sum_{i=1}^I \mathbf{I}(\hat{y}_i = c_k).$$

Поскольку истинные классы примеров неизвестны, вместо $\mathbf{I}(\hat{y}_i = c_k)$ используется математическое ожидание этой величины, то есть $P(\hat{y}_i = c_k | L)$.

На вход алгоритма подаются начальные оценки «правильных» меток для примеров. На основе этого выполняется М-шаг, в ходе которого вычисляются матрица ошибок и априорные вероятности классов, далее выполняется Е-шаг, на котором вычисляются новые оценки вероятностей принадлежности примеров каждому из классов, и так до достижения сходимости.

Большинство методов согласования так или иначе обладает следующими ограничениями. Во-первых, тип ответа на задание. В большинстве публикаций рассматриваются задания, предполагающие бинарные ответы, то есть суть задания сводится к бинарной классификации какого-то объекта. Как правило, бинарная классификация может быть расширена на выбор одного из множества ответов, однако это усложняет и без того достаточно сложные модели. Во-вторых, значительная вычислительная сложность моделей (особенно основанных на EM-алгоритме). В-третьих, для работы алгоритмов, основанных на EM, необходимо, чтобы каждое задание было выполнено значительным количеством участников (около 10), что может быть связано с существенными финансовыми затратами, если задания выполняются на коммерческой основе. Наконец, при практической реализации могут возникнуть трудности с определением начальных значений параметров для EM-алгоритма [16].

Тем не менее, несмотря на определенные ограничения, наличие строгого математического базиса привело к тому, что методы обеспечения качества в крауд-вычислениях, основанные на согласовании мнений различных участников, оказываются, пожалуй, наиболее проработанными и получают значительное внимание исследователей. Так, среди 41 публикации, вошедшей в систематический обзор, 11 посвящены различным методам согласования (таблица 1).

Следует также упомянуть попытки стандартизованного сравнения различных алгоритмов согласования, например, SQUARE [54] или BATC [55].

Современные работы, вошедшие в систематический обзор, посвящены развитию методов согласования по следующим направлениям.

Важным направлением развития методов согласования является разработка специализированных методов, учитывающих особенности того или иного класса задач (и множества меток). Так, в ряде работ рассматривается ситуация, когда между метками, которые являются ожидаемым результатом выполнения задания участником, существуют определенные логические связи. Несмотря на то что на практике такая

ситуация встречается достаточно часто, хорошо разработанных и принятых научным сообществом методов агрегации меток для нее пока не существует. Разработке таких методов посвящена работа [10], авторы которой предлагают расширение алгоритма DS и исследуют разные модели представления отношений между метками (включая байесовскую сеть как компактное представление совместного распределения). В статье [13] предлагается вероятностная модель разметки для иерархической классификации, то есть для ситуации, когда метками, которые участники присваивают объектам, являются классы, организованные в иерархию (классификация книг, товаров). Адаптация и разработка методов согласования для заданий, выходящих за рамки простых бинарных схем, рассматривается также в работах [7, 11, 15]. В [11] предлагается метод согласования ответов в непрерывной шкале, в [15] речь идет о методе согласования для ответов в рамках порядковой шкалы, а в [7] — о методах согласования текстовых данных.

Особой темой исследований в области методов поиска консенсуса является разработка способов учета смещения (bias) в определении меток [6, 17], то есть, ситуации, когда несмотря на то, что участники добросовестно выполняют свою задачу, восприятие размечаемых объектов у разных групп пользователей различается. Например, у неспециалистов (коими обычно являются участники систем краудсорсинга) может быть склонность классифицировать объекты не так, как у специалистов.

Еще одним направлением является учет различных моделей поведения пользователей при формировании обобщенной оценки. Так, в работе [11] предлагается метод выявления согласованной оценки для непрерывных измерений при отсутствии золотого стандарта и векторов свойств объектов. Авторы статьи рассматривают различные стили поведения исполнителей (спамер, смещенный, злоумышленник и пр.) и формализуют их в виде байесовских моделей. В статье [15] предлагается способ выявления мусорных (спамерских) результатов. Выявление спамеров базируется на наблюдении, что популярным методом спама является генерация случайных результатов и генерация одних и тех же результатов (для всех заданий исполнителя).

Предлагаются также альтернативные (по отношению к популярному EM-алгоритму) способы поиска консенсуса. Так, в статье [16] предлагается алгоритм получения обобщенных меток на основе кластеризации. В задаче разметки, в которой каждому объекту может быть присвоено несколько меток (несколько классов), вероятность принадлежности каждому из классов, оцененная на основе собранных меток с помощью метода максимального правдоподобия,

рассматривается как вектор признаков объекта. Объекты кластеризуются по сходству этих векторов и итоговая метка для всех объектов, попавших в кластер, определяется как признак с наибольшей общей суммой. В статье [14] предложен оригинальный подход к обеспечению качества краудсорсинга, заключающийся в применении помехоустойчивого кодирования.

В статье [12] решается интересная задача определения оптимального количества эталонных значений (ground truths) и распределения их среди исполнителей для получения максимального качества разметки (в рамках применения краудсорсинга для разметки набора данных). Как и в большинстве алгоритмов, основанных на EM, предполагается, что есть существенное пересечение множеств заданий, выполненных различными участниками. Необычной особенностью предлагаемого авторами подхода является также оценивание доверительного интервала точности исполнителей.

Таким образом, на основе проведенного анализа можно сделать следующие выводы. Ключевыми особенностями методов согласования являются:

- использование вероятностной модели участника процесса крауд-вычислений (это может быть просто вероятность правильного ответа, может быть матрица вероятностей ошибок и пр.);
- возможность применения в условиях минимальной информации об участниках. В то же время многие методы (особенно основанные на EM) позволяют учесть имеющуюся информацию);
- необходимость относительно большого пересечения множеств заданий, обработанных разными участниками.

Текущими направлениями развития методов согласования являются:

- разработка методов согласования для заданий с множеством вариантов ответа, в особенности, когда элементы этого множества взаимосвязаны;
- оценка доверительного интервала параметров модели (например, ожидаемой точности участника);
- разработка методов, обладающих высокой вычислительной эффективностью и обеспечивающих хорошее качество согласования;
- учет субъективных характеристик участников, разработка более точных моделей поведения участников, учитывающих их субъективные характеристики (bias).

4.2. Методы проектирования потока работ. Все рассмотренные ранее модели обеспечения качества, основанные на согласовании ответов, полученных от разных участников, предполагали простейший сценарий взаимодействия человека и системы крауд-

вычислений — разным участникам назначаются однотипные задания, участники дают ответы (выполняют задания), полученные ответы обобщаются с помощью того или иного математического алгоритма, чтобы получить истинные (наиболее вероятные) ответы для заданий. Причем для каждого задания, как правило, осуществляется сбор фиксированного количества ответов от участников. Поток работ в крауд-вычислениях может быть и гораздо более сложным. Он может предполагать разные типы заданий, служащие для одной цели (например, генерация набора меток и проверка релевантности конкретной метки), учитывать разные компетенции участников и, главное, быть адаптивным, динамически подстраиваться под текущую ситуацию — оценивать нехватку информации и имеющееся множество ресурсов (исполнителей), формируя на основе этой оценки необходимые задания, стремясь тем самым сократить общее количество работы, которая должна быть выполнена для получения результата определенного качества.

В ряде вошедших в выборку работ рассматриваются именно такие методы обоснованного формирования (в том числе динамического) сложных потоков работ, учитывающих различные параметры.

Научные исследования здесь можно условно разбить на две ветви. Во-первых, это разработка математических (в первую очередь, аналитических) моделей существующих и широко распространенных механизмов обеспечения качества — мажоритарного голосования, контрольных групп и так далее, и дальнейшее использование этих математических моделей для рационального формирования потоков работ. К этой ветви относится, например, статья [19], в которой производится сравнение мажоритарного голосования и использования контрольной группы (нанимаемой из состава участников). Авторами построена аналитическая модель, позволяющая увязать вероятность правильного выполнения задачи и ее стоимость.

Ко второй ветви относятся работы, в которых предлагаются оригинальные математические модели и методы формирования потока работ. Так, в статье [20] исследуются методы нахождения наилучших альтернатив из заданного набора (top K). Задача формулируется с помощью трех ключевых факторов: латентность (задержка), денежная стоимость и качество ответов. Статья [21] посвящена использованию краудсорсинга для создания набора правил, позволяющих формировать осмысленные правила сочетания характеристик для описания заболеваний в соответствии со схемой кодирования заболеваний ICD 11. В частности, использовались моделирование на основе байесовской сети и обучение с подкреплением. Особенность работы

заключается в том, что авторы предлагают динамическую модель потока работ; на основе байесовской сети определяется вероятность того или иного ответа, и если вероятность оказывается выше пороговой, то задание больше не назначается. В статье [18] предлагается разработка агентов, основанных на обучении байесовских сетей в сочетании с частично наблюдаемыми марковскими процессами принятия решений (POMDPs) для поиска компромисса между стоимостью и качеством потоков работ, выполняемых с помощью краудсорсинга. Основная идея работы также состоит в динамической адаптации потока работ в зависимости от получаемых результатов.

В качестве оценки таких методов помимо практических экспериментов используется и имитационное моделирование [20].

Подводя краткий итог по этой группе методов, следует отметить следующее. Методы проектирования потока работ можно рассматривать как своего рода развитие методов обобщения. Их общей чертой является, например, применение вероятностных моделей для описания поведения и качества результатов, получаемых от отдельных участников. Однако методы динамического проектирования позволяют рассматривать и другие важные характеристики процесса (в том числе стоимость), позволяя получать результаты определенного качества потенциально эффективнее, чем методы согласования.

4.3. Методы централизованного назначения. Одной из достаточно типовых постановок является выбор распределения заданий между исполнителями, максимизирующий ожидаемое качество результата. При этом неизбежно возникает вопрос о том, как это ожидаемое качество оценивать. Популярным ответом на этот вопрос является долговременная характеристика пользователя, обычно называемая «репутацией». При этом считается, что ожидаемое качество результата, получаемого от участника, не убывает с ростом его репутации. Существует множество моделей для вычисления репутации [56-58], а также работ по противодействию попыткам «обмана» механизма уточнения репутации (например, [59, 60]).

Наибольшее распространение методы централизованного назначения получили в контексте некоторых прикладных разновидностей краудсорсинга, в частности пространственного краудсорсинга и краудсенсинга, то есть получения информации о состоянии физической среды с использованием мобильных устройств участников ([22, 23]). Вероятно, это связано с тем, что в подобных приложениях при распределении заданий традиционно учитывается целый ряд факторов (покрытие необходимой области пространства, расходование энергии, затраты на передачу трафика), качество

получаемой информации становится еще одним из критериев, вписываемых в структуру оптимизационной задачи.

Базовая модель такого типа записывается следующим образом. Пусть r_{ij} — ожидаемое качество выполнения работы i исполнителем j (репутация), c_{ij} — стоимость выполнения работы i исполнителем j , B — бюджет рекрутера, L_j — количество работ, которое может выполнить исполнитель j , а x_{ij} — бинарная переменная, принимающая значение 1 в случае назначения работы i участнику j и 0 в противном случае. Тогда:

$$\begin{aligned} & \max \sum_{i=1}^N \sum_{j=1}^M x_{ij} r_{ij} \\ & \sum_{i=1}^N x_{ij} \leq L_j, \quad j \in \{1, \dots, M\} \\ & \sum_{i=1}^N \sum_{j=1}^M x_{ij} c_{ij} \leq B. \end{aligned}$$

То есть, задача заключается в том, чтобы найти такое распределение работ по исполнителям (участникам системы крауд-вычислений), чтобы максимизировать ожидаемое качество (считая его для простоты линейной функцией), при условии ограничений на загрузку каждого исполнителя и бюджет рекрутера.

При рассмотрении централизованного назначения математические модели и, соответственно, методы обеспечения качества, основанные на использовании этих моделей, можно разделить на две группы. В первую группу входят методы, в которых процесс обновления репутации рассматривается независимо от процесса распределения задач (например, [24]), то есть при распределении задач ожидаемое качество результатов, получаемого от каждого из участников, считается известным. Во вторую группу входят более общие методы, в которых предполагаемое качество каждого участника считается неизвестным, и процесс организуется таким образом, чтобы сначала оценить ожидаемое качество, а затем распределить задания между участниками на основании этих оценок (например, [8, 25]).

К первой группе относится работа [24], в которой обеспечение качества основывается на учете репутации, то есть характеристики исполнителя, отражающей его «надежность» при выполнении заданий. Причем, поскольку данная статья ориентирована на разработку методов, используемых в системах пространственного краудсорсинга, то в ней решается задача оптимизации распределения заданий с учетом

как репутации, так и пространственного положения пользователей. В частности, задача моделируется как задача о рюкзаке (вариант с несколькими рюкзаками — Multiple Choice Knapsack Problem). В силу сложности данной задачи для ее решения применяют различные эвристики. Кроме того, при такой постановке отдельной проблемой является способ увязывания репутации с ожидаемым качеством ответа.

В [22] рассматривается задача назначения заданий, а именно — выбора сервисного узла (service node) в задачах мобильного краудсенсинга, однако здесь рассматривается уже многокритериальная задача оптимизации по ряду качественных критериев, актуальных для систем краудсенсинга (покрытие, задержки, достоверность и пр.). При этом предполагается, что известен вклад каждого из участников в значение каждой из мер качества (которые предполагаются аддитивными). В статье [23] предлагается оптимизационная модель для задач мобильного краудсенсинга, учитывающая вопросы приватности исполнителей, то есть рассматривается поиск компромисса между тремя метриками — приватность, полезность и эффективность.

Ко второй группе относятся работы [8] и [25]. Поскольку в них предполагается, что изначально информации о предполагаемой эффективности пользователей нет, в обеих работах используются элементы обучения с подкреплением, в частности модель многоруких бандитов (multi-armed bandits) [61]. Многорукые бандиты — это класс оптимизационных задач, в которых действия агента связаны с изначально неизвестными наградами, которые необходимо выявить посредством (зашумленных) экспериментов. Общей же целью оптимизации является максимизация полезности за счет выбора различных действий. Это соответствует ситуации выбора исполнителя задания — действиями агента (платформы краудсорсинга) являются назначение задач различным участникам, а результат выполнения (в частности его полезность) зависит от неизвестных заранее характеристик участника. Особенностью применения модели многоруких бандитов при распределении задач в системах краудсорсинга является то, что большинство классических вариантов этой модели (например, [62]) не рассматривают стоимость, которая может быть связана с каждым действием агента, а также возможные ограничения на количество заданий, которые могут быть назначены одному участнику (загрузка участника). С связи с этим в статьях, рассматривающих применение этой модели к распределению заданий в краудсорсинге, предлагаются собственные модификации.

В статье [8] рассматривается ситуация мобильного, пространственного краудсорсинга. Решается двоякая задача распределения заданий (с учетом ограниченного бюджета) и выявления

качества участников с использованием предлагаемого подхода, основанного на динамическом обучении (online learning). Эффективность демонстрируется с помощью имитационных моделей.

В работе [25] рассматривается ситуация выполнения работы с помощью участников, начальная эффективность которых неизвестна. Авторы подчеркивают, что ориентируются, в первую очередь, на сложные задания (краудсорсинг с помощью экспертов), мотивируя это тем, что такие условия применения краудсорсинга обладают следующими особенностями: существенная зависимость от уровня умений (отсюда особенная важность выбора лучших), открытый и динамичный рынок (нет базы всех исполнителей), сильный разброс цен на выполнение задания. Предлагаемое решение работает следующим образом: бюджет задачи разделяется на две части (в соответствии с задаваемым параметром ϵ), а исполнение разбито на две фазы. На первой фазе происходит выяснение эффективности участников (с использованием первой части бюджета), а на второй фазе задачи назначаются только наилучшим участникам (с помощью решения задачи о рюкзаке).

Важной характеристикой методов централизованного распределения является перечень факторов, учитываемых при распределении: ограничения бюджета, ограничения на загрузку исполнителей, временные показатели, а также набор критериев: качество, время и другое.

Таким образом:

- с прикладной точки зрения централизованное распределение заданий между участниками, осуществляемое посредством решения оптимизационной задачи, представляет собой удобную и расширяемую платформу для постановки и достижения различных целей (по качеству, оперативности, стоимости и пр.);

- решаемая при этом оптимизационная задача представляет собой вариант задачи о рюкзаке и (или множественной задачи о рюкзаке), является вычислительно трудной, что вынуждает использовать различные эвристики для ее решения. Эвристический подход может быть основан как на применении метаэвристик (генетическое программирование, имитация отжига и многих других), так и на создании специальных эвристик и в этом видится существенное поле для исследований;

- перспективным представляется использование методов обучения с подкреплением и активного обучения (в частности модель многоруких бандитов) для выяснения качеств участника, используемых в дальнейшем при формировании оптимального назначения.

4.4. Теоретико-игровые методы. Теория игр дает понятийный и математический аппарат, который, как показывают современные работы, может быть успешно использован для проектирования систем человеко-машинных вычислений. Использование теоретико-игровых методов позволяет рассматривать проблему обеспечения качества еще шире. Если методы, основанные на поиске согласия, предлагают только способ вывода истинных значений из накопленного набора данных, методы, связанные с проектированием потока работ, предлагают способы формирования потока работ, оптимального по ряду параметров, на основе известных (или оцениваемых в процессе) характеристик точности и прилежности исполнителей, то применение теоретико-игровых методов включает человека-исполнителя в рамки рассматриваемой системы, предлагая способ увязывания целей исполнителя (как активного элемента, действующего в своих интересах) с целями инициатора процесса вычислений (преследующего свои цели). И здесь проблема обеспечения качества находится в тесной связи с используемым механизмом стимулирования (*incentivization*), назначением которого как раз и является такое увязывание. В общем случае перечень задач, стоящих при разработке стратегии стимулирования, оказывается шире, чем обеспечение качества. Сюда входит, в первую очередь, привлечение участников к функционированию в системе, в том числе долговременному (если это предусмотрено — StackOverflow, Wikipedia и пр.).

Идея применения теоретико-игровых методов заключается в том, что (потенциальные) участники системы рассматриваются как рациональные агенты, основной целью которых является максимизация полезности от участия в системе. Функция полезности для участника определяется как разница между размером вознаграждения и стоимостью выполнения задания. Связь с качеством здесь в том, что модель может предусматривать разные стратегии участника, соответствующие различным уровням усилий, которые он может прикладывать, а от уровня усилий зависит качество получающихся результатов. Таким образом, участник осуществляет выбор уровня усилий (учитывая, что в системе есть и другие участники, действующие в своих интересах), чтобы максимизировать полезность. Задачей же разработчика системы является создание такого механизма распределения вознаграждения, чтобы приложение участниками максимальных усилий было равновесной по Нэшу стратегией.

В работе [63] таксономия методов стимулирования разбита на две части в зависимости от цели соответствующего механизма: предметно-ориентированные механизмы и общие (многоцелевые) механизмы.

Немного особняком стоит применение теоретико-игровых методов в краудсенсинге, поскольку участие в такого рода системах связано с расходованием определенных ресурсов — заряда батареи, сетевого трафика.

Применение теоретико-игровых методов связано и с определенными ограничениями. Например, необходимо считать известными функции зависимости качества от уровня усилий для участников, субъективную полезность для участников тех или иных видов вознаграждения, а также тот факт, что участники не всегда поступают рационально. Собственно, примерно по тем же причинам, по которым наряду с математической теорией игр появилась поведенческая теория игр [64]. Однако, несмотря на это, теоретико-игровые модели являются хорошим базисом для изучения и проектирования систем человеко-машинных вычислений.

В частности, теория игр и теоретико-игровые модели позволяют решать две важные задачи: во-первых, анализ существующих механизмов, лежащих в основе систем крауд-вычислений; во-вторых, разработка новых механизмов, удовлетворяющих требуемым свойствам. В список отобранных публикаций попали работы, посвященные решению каждой из перечисленных задач. Важные свойства, которым должны удовлетворять «хорошие» механизмы:

- Правдивость (Truthfulness). Ни один участник не может непосредственно увеличить свою полезность, делая ставку, отличающуюся от его стоимости выполнения задания.

- Индивидуальная рациональность (Individual rationality). Неотрицательность полезности для каждого участника.

- Вычислительная эффективность (Computational efficiency). Возможность вычислить результат (размер вознаграждения) за полиномиальное время. В [31] вычислительная эффективность также трактуется как возможность работы в реальном времени.

- Сбалансированность бюджета (Budget-balance) или прибыльность для организатора (Platform profitable). Неотрицательность полезности для организатора (в данном случае, платформы крауд-вычислений).

Одной из первых публикаций по применению теории игр для анализа устройства систем крауд-вычислений является [28], в которой предлагается обобщенная аналитическая модель игры ESP и производится анализ параметров этой аналитической модели на качество получаемых результатов. Игра ESP предложена в [65] и предназначена для аннотирования изображений. Суть игры в том, что двум участникам предъявляется одно и то же изображение и каждый из них должен

описать это изображение с помощью ключевых слов. Количество набранных участниками очков зависит от того, совпали ли ключевые слова. Таким образом, каждый из участников, преследуя цель набрать наибольшее количество очков, пытается выбрать те слова, которые, как он считает, вероятнее всего выберет другой участник.

В статье [27] производится анализ распространенных в области краудсорсинга микрозаданий механизмов (случайного назначения, дополненного мажоритарным голосованием, и централизованного назначения с последующей проверкой) с позиций теории игр и показываются их недостатки. Предлагаемый же авторами механизм основан на том, что участникам назначаются тренировочные задания, способные повышать качество их результатов за счет эффекта обучения. Участник моделируется вероятностью того, что результаты его решений будут приемлемого качества, предполагается также, что множество решений бесконечно и вероятность того, что два участника получают одинаковое ложное решение, равна 0. Стратегии участника связаны с уровнем усилий, которые он может приложить, известна эта функция, удовлетворяющая ряду условий. Выигрыш (полезность) участника при наличии механизма M зависит от выбранного качества решения и качества решений, полученных от других участников. Механизмы сравниваются и оцениваются по их симметрическим равновесным по Нэшу стратегиям (symmetric Nash equilibria), среди которых наибольший интерес представляют те, в которых участники имеют стимул для того, чтобы генерировать ответы наилучшего возможного качества, прилагая максимум усилий. С другой стороны, механизм также оценивается по стоимости для организатора (рекрутера). В частности, по способности гарантированно «укладываться» в любой неотрицательный заранее известный бюджет (подробнее про эту постановку можно почитать в [27]).

В работе [29] предлагается механизм стимулирования для мобильного краудсенсинга, основанный на аукционе по качеству. Основной областью применения являются такие приложения краудсенсинга, в которых вознаграждение участника зависит от качества полученных от него данных, а не от времени работы и других параметров. В работе теоретически доказан ряд свойств предложенного механизма: правдивость, индивидуальная рациональность, прибыльность для организатора. Показано также применение предложенного механизма при разработке приложения для позиционирования внутри помещений, основанного на методе радиоотпечатков Wi-Fi.

В статье [30] предлагается механизм стимулирования для мобильного краудсенсинга. В отличие от множества других механизмов, ориентирующихся, в первую очередь, на измерение

качества вклада определенного пользователя и подборе адекватного стимула для него, в данной статье предлагается механизм социального стимулирования, направленный на стимулирование социального окружения участника, выполняющего задание.

Отдельную и относительно хорошо изученную группу механизмов составляют аукционы.

В статье [32] предлагается механизм стимулирования для участников систем краудсенсинга, комбинирующий элементы обратных аукционов (reverse auctions) и аукционов Викри (Vikrey auctions). Показано, что механизм является вычислительно эффективным, индивидуально рациональным, сбалансированным по бюджету и правдивым. Имитационные эксперименты показали, что применение механизма позволяет улучшить качество данных, получаемых от системы. Основной сценарий, рассматриваемый в статье, заключается в том, что участники могут делать ставки на выполнение заданий (указывая, за какую компенсацию они согласны выполнить задание), по результатам аукциона задание назначается участникам, а факт получения вознаграждения зависит от качества данных, предоставленных участником. При этом основной фокус статьи сосредоточен на том, чтобы разработанный механизм не допускал злонамеренного искажения ставок.

В работе [26] предложен концептуальный каркас (framework) стимулирования для краудсенсинга, интегрирующий три популярных метода стимулирования: обратные аукционы, геймификацию и систему репутации, и затрагивающая все три стороны, заинтересованные в процессе краудсенсинга: участников, рекрутеров и саму платформу. Особенностью этой работы является включение рекрутера в общую схему. В частности, решается проблема правдивости «приема» заданий рекрутером, для чего задача по приему заданий назначается участникам, используя геймификацию в качестве механизма, и в зависимости от решения участников рекрутер либо получает, либо не получает возмещение.

В статье [31] рассматривается задача создания механизма, удовлетворяющего шести необходимым свойствам: вычислительная эффективность, индивидуальная рациональность, прибыльность для организатора, правдивость (участники сообщают истинную стоимость решения задачи и истинные сроки ее выполнения), возможность каждого участника выиграть аукцион, близка к оптимальному, при полностью известной информации об участниках. Общая модель взаимодействия такая — участник характеризуется временем прихода, временем ухода, стоимостью выполнения задания и набором заданий,

которые он может выполнить (это составляет тип участника). Участник делает ставку, и среда должна решить, покупает ли она услуги участника за эту ставку до момента ухода участника. При этом рекрутер преследует цель получения максимальной полезности в рамках своих денежных ограничений. Однако качество здесь сводится только лишь к требуемому количеству повторений задания.

Таким образом:

– теория игр и в особенности ее раздел, посвященный дизайну механизмов (mechanism design), дают очень хорошую концептуальную базу, позволяющую производить анализ систем крауд-вычислений, а также проектировать системы, рациональные участники которых заинтересованы в качественном выполнении заданий. И упоминание рациональности здесь очень важно, поскольку при ее отсутствии и/или при целенаправленных попытках нарушения работы системы теоретико-игровые модели перестают адекватно описывать поведение участников. Вместе с тем качество определяется не только заинтересованностью участника в выполнении задания, поэтому теоретико-игровые методы, по всей видимости, должны сочетаться с методами других групп;

– к перспективным направлениям развития этой группы методов относится поведенческая теория игр, объединение теоретико-игровых моделей систем крауд-вычислений с практическими исследованиями демографии площадок крауд-вычислений, качественными и количественными моделями поведения участников таких систем.

4.5. Методы, основанные на учете свойств заданий.

Исходными данными для большинства методов и подходов к обеспечению качества является информация об участниках и множество накопленных результатов выполнения заданий. Характеристики самих заданий практически не учитываются, и это объяснимо, поскольку крауд-вычисления традиционно применяются в тех областях, где машинная обработка информации оказывается недостаточно надежной. Однако среди отобранных источников оказалось и несколько статей, учитывающих свойства заданий. Эти статьи позволяют наметить область применимости подобного подхода.

В статье [34] краудсорсинг используется для разметки научных данных (астрономических снимков). Авторы предлагают оригинальную модель, в которой предлагается оценивать непротиворечивость результатов, полученных от самого участника, на основе того факта, что классификаторы хорошо работают с незашумленными данными. Для этого авторы используют распространенную модель автоматической классификации астрономических объектов по визуальным признакам и

обучают ее на качественных данных, чтобы оценить границы точности предсказаний. Затем модель обучается на подмножестве данных, размеченных участником, и проверяется ее предиктивная «сила». Если оказывается, что она существенно меньше, чем предиктивная «сила» модели, обученной на реальных данных, то делается вывод, что разметка, полученная от пользователя, является зашумленной. Ограничение метода в том, что для тренировки модели требуется большое количество аннотаций (хорошие показатели у авторов получаются, например, при 100 аннотаций).

Есть также попытки адаптации методов очистки данных к идентификации противоречивых результатов исполнения заданий участниками системы крауд-вычислений. Так, в статье [33] предлагается несколько методов очистки — основанный на известном методе «полировки» [66], основанный на самообучающихся классификаторах и основанный на кластерном анализе. Интересно, что предлагаемая авторами разновидность метода «полировки» и очистка, основанная на самообучающихся классификаторах, по принципу работы напоминают различные методы согласования, с тем отличием, что источниками меток являются не люди, а модели классификации, обученные на разных подмножествах. Очистка же с помощью кластерного анализа требует, чтобы в рамках одного кластера объектов (заданий) результаты, полученные от людей, были одинаковы. Схожие подходы рассмотрены и проанализированы в [1].

Обобщая эти работы можно отметить, что свойства самих заданий (или результат машинной обработки информации, связанной с заданиями) актуально использовать по крайней мере в двух случаях:

1) Если для них существуют качественные математические модели, позволяющие получить точность обработки, сопоставимую с точностью обработки человеком.

2) Когда часть заданий (вызывающую наибольшие противоречия) можно не учитывать в дальнейшей обработке (например, при подготовке данных для машинного обучения устранение лишних данных — фильтрация — ведет часто лишь к большей устойчивости моделей).

4.6. Методы, основанные на анализе действий пользователя и воздействии на него. Отдельное направление методов обеспечения качества связано с анализом действий пользователя в ходе выполнения заданий и влиянием на процесс выполнения заданий.

Основной идеей здесь является построение классификатора, позволяющего связать поведение пользователя (движение указателя, активизация тех или иных управляющих элементов формы, временные задержки, сложность ввода, следование «мягким» правилам при

выполнении задания) с его «добросовестностью» при выполнении заданий. Для обучения такой модели, как правило, используются задания с заранее известными результатами. Сами признаки, определяющие вид модели классификации, в значительной мере определяются видом заданий. Примером такой работы является [36].

В работе [35] для обеспечения качества применяется подход, связанный с мониторингом психических характеристик участника (усталости) с помощью эталонных заданий. Авторы внедряют задания с известным результатом для того, чтобы обнаружить момент, когда из-за усталости качество результатов участника снижается, и прекращают назначать ему задания.

Помимо работ, в которых происходит мониторинг действий пользователя, в эту же группу можно отнести и работы, в которых подчеркиваются различные аспекты проектирования заданий, которые могут повлиять на качество получаемых результатов.

Статья [37] посвящена оригинальному, скорее, организационному способу повышения качества результатов в краудсорсинге микрозаданий — применению техники прайминга (психологический и маркетинговый термин, обозначающий фиксирование установки [67]). На практике это проявлялось в том, что заданиям сопутствовали изображения и аудио-фрагменты, призванные вызвать определенное эмоциональное состояние исполнителя (сформировать установку на качественное выполнение работы). Приведенные в статье результаты показывают, что использование техник прайминга действительно позволяет добиться краткосрочного улучшения результатов.

В рассмотренной уже ранее статье [21] вспомогательным методом повышения качества являлось помещение на форму выполнения задания ссылок, поясняющих специальные термины. И хотя это решение представляется разумным, в статье нет экспериментальных свидетельств о влиянии решения на качество результатов.

5. Заключение. Проведен аналитический обзор современных (вышедших после 2012 года) статей библиографических баз ScienceDirect и IEEE Xplore, посвященных методам обеспечения качества результатов при использовании крауд-вычислений. Выявлены основные направления, по которым происходит развитие таких методов. Показано, что наибольшее внимание научного сообщества сосредоточено в двух во многом противоположных направлениях: обеспечение качества, основанное на согласовании данных, полученных от разных участников, и обеспечение качества за счет использования теоретико-игровых методов. Противоположность же этих направлений связана с

тем, что методы, основанные на согласовании, рассматривают задачу в максимально узкой постановке — как наиболее правильно интерпретировать собранные данные, а теоретико-игровые методы, напротив, предлагают концептуальную платформу для проектирования и разработки таких систем.

Выявленные подходы сопоставлены с основными этапами процесса применения крауд-вычислений к решению задачи (таблица 2). Понимание места того или иного подхода в рамках процесса реализации крауд-вычислений имеет существенное значение, потому что различные платформы, с помощью которых реализуются подобные системы, обладают разными особенностями работы и возможностями, что может накладывать ограничения на допустимые методы. Например, популярная платформа краудсорсинга микрозаданий Amazon Mechanical Turk осуществляет распределение заданий в так называемом «пулл»-режиме, то есть участники сами выбирают задания. Следовательно, методы, основанные на централизованном распределении оказываются в этом случае неприменимы. Наименьшие ограничения в этом смысле у методов согласования, поскольку для их применения достаточно лишь набора данных, полученного от участников.

В работах [68, 69] делается попытка выделения причин получения некачественных результатов в крауд-вычислениях. Среди основных факторов, влияющих на качество, названы компетентность, заинтересованность в выполнении заданий (мотивация), ясность представления заданий, наличие или отсутствие «злого умысла» (целенаправленного подрыва работы системы). Различные классы методов в основном учитывают воздействие лишь какого-то подмножества факторов. Так, учет мотивационных факторов происходит с помощью теоретико-игровых методов, выявление и использование сведений о компетентности участников происходит в большинстве методов разных групп, наконец, свойства задания частично рассматриваются в последней группе. Таким образом, комплексный подход к проблеме обеспечения качества должен базироваться на сочетании различных методов.

Следует также отметить, что в настоящий момент отсутствуют публикации, в которых сравнивались бы принципиально различные подходы к обеспечению качества результатов, получаемых с помощью крауд-вычислений. В то же время только такие комплексные работы способны заложить основы рационального и научно обоснованного подхода к принятию проектных решений при создании человеко-машинных систем, использующих данную технологию.

Литература

1. *Li C. et al.* Noise filtering to improve data and model quality for crowdsourcing // Knowledge-Based Syst. 2016. vol. 107. pp. 96–103.
2. *Bernstein A., Klein M., Malone T.W.* Programming the global brain // Commun. ACM. 2012. vol. 55. no. 5. pp. 41–43.
3. *Kazai G., Kamps J., Milic-Frayling N.* Worker types and personality traits in crowdsourcing relevance labels // Proc. 20th ACM Int. Conf. Inf. Knowl. Manag. (CIKM '11). USA: ACM Press. 2011. pp. 1941–1944.
4. *Kitchenham B.* Procedures for Performing Systematic Reviews // Keele. 2004. vol. 33. no. 2004. pp. 1–26.
5. *Kitchenham B., Charters S.* Guidelines for performing Systematic Literature Reviews in Software Engineering. Technical Report EBSE-2007-01. 2007. 65 p.
6. *Zhang J., Sheng V, Li Q., Wu J., Wu X.* Consensus algorithms for biased labeling in crowdsourcing // Information Sciences. 2017. vol. 382. pp. 254–273.
7. *Matsunaga A., Mast A., Fortes J.A.B.* Workforce-efficient consensus in crowdsourced transcription of biocollections information // Future Generation Computer Systems. 2016. vol. 56. pp. 526–536.
8. *Han K., Zhang C., Luo J.* Taming the Uncertainty: Budget Limited Robust Crowdsensing Through Online Learning // Biological Cybernetics. 2016. vol. 24. no. 3. pp. 1462–1475.
9. *Dang D., Liu Y., Zhang X., Huang S.* Crowdsourcing Worker Quality Evaluation Algorithm on MapReduce for Big Data Applications // IEEE Transactions on Parallel and Distributed Systems. 2016. vol. 27. no. 7. pp. 1879–1888.
10. *Duan L., Oyama S., Sato H., Kurihara M.* Separate or joint? Estimation of multiple labels from crowdsourced annotations // Expert Systems with Applications. 2014. vol. 41. no. 13. pp. 5723–5732.
11. *Kara Y.E., Genc G., Aran O., Akarun L.* Modeling annotator behaviors for crowd labeling // Neurocomputing. 2015. vol. 160. pp. 141–156.
12. *Kubota T., Aritsugi M.* Assignment strategies for ground truths in the crowdsourcing of labeling tasks // The Journal of Systems and Software. 2017. vol. 126. pp. 113–126.
13. *Otani N., Baba Y., Kashima H.* Quality control of crowdsourced classification using hierarchical class structures // Expert Systems with Applications. 2016. vol. 58. pp. 155–163.
14. *Vempaty A., Varshney L.R., Varshney P.K.* Reliable Crowdsourcing for Multi-Class Labeling // IEEE Journal of Selected Topics in Signal Process. 2014. vol. 8. no. 4. pp. 667–679.
15. *Vuurens J.B.P., de Vries A.P.* Obtaining High-Quality Relevance Judgments Using Crowdsourcing // IEEE Internet Comput. 2012. vol. 16. no. 5. pp. 20–27.
16. *Zhang J., Sheng V., Wu J., Wu X.* Multi-Class Ground Truth Inference in Crowdsourcing with Clustering // IEEE Transactions on Knowledge and Data Engineering. 2016. vol. 28. no. 4. pp. 1080–1085.
17. *Zhang J., Wu X., Sheng V.S.* Imbalanced Multiple Noisy Labeling // IEEE Transactions on Knowledge and Data Engineering. 2015. vol. 27. no. 2. pp. 489–503.
18. *Dai P., Lin C., Weld D.* POMDP-based control of workflows for crowdsourcing // Artificial Intelligence. 2013. vol. 202. pp. 52–85.
19. *Hirth M., Hofffeld T., Tran-Gia P.* Analyzing costs and accuracy of validation mechanisms for crowdsourcing platforms // Mathematical and Computer Modeling. 2013. vol. 57. no. 11–12. pp. 2918–2932.
20. *Lee J., Lee D., Hwang S.* CrowdK: Answering top- k queries with crowdsourcing // Information Sciences. 2017. vol. 399. pp. 98–120.
21. *Lou Y. et al.* Use of Ontology Structure and Bayesian Models to Aid the Crowdsourcing of ICD-11 Sanctioning Rules // Journal of Biomedical Informatics. 2017. vol. 68. no. C. pp. 20–34.

22. *Ding S., He X., Wang J.* Multiobjective Optimization Model for Service Node Selection Based on a Tradeoff Between Quality of Service and Resource Consumption in Mobile Crowd Sensing // IEEE Internet of Things Journal. 2017. vol. 4. no. 1. pp. 258–268.
23. *Gong Y. et al.* Optimal Task Recommendation for Mobile Crowdsourcing With Privacy Control // IEEE Internet of Things Journal. 2016. vol. 3. no. 5. pp. 745–756.
24. *Miao C., Yu H., Shen Z., Leung C.* Balancing quality and budget considerations in mobile crowdsourcing // Decision Support Systems. 2016. vol. 90. pp. 56–64.
25. *Tran-Thanh L., Stein S., Rogers A., Jennings N.* Efficient crowdsourcing of unknown experts using bounded multi-armed bandits // Artificial Intelligence. 2014. vol. 214. pp. 89–111.
26. *Dai W., Wang Y., Jin Q., Ma J.* An Integrated Incentive Framework for Mobile Crowdsourced Sensing // Tsinghua Science and Technology. 2016. vol. 21. no. 2. pp. 146–156.
27. *Gao Y., Chen Y., Liu K.J.R.* On Cost-Effective Incentive Mechanisms in Microtask Crowdsourcing // IEEE Transactions on Computational Intelligence and AI in Games. 2015. vol. 7. no. 1. pp. 3–15.
28. *Wang B.C., Lin C.W., Chen K.T., Chen L.J.* An analytical model for generalized ESP games // Knowledge-Based Systems. 2012. vol. 34. pp. 114–127.
29. *Wen Y. et al.* Quality-Driven Auction-Based Incentive Mechanism for Mobile Crowd Sensing // IEEE Transactions on Vehicular Technology. 2015. vol. 64. no. 9. pp. 4203–4214.
30. *Yang G., He S., Shi Z., Chen J.* Promoting Cooperation by the Social Incentive Mechanism in Mobile Crowdsensing // IEEE Communications Magazine. 2017. vol. 55. no. 3. pp. 86–92.
31. *Zhao D., Li X., Ma H.* Budget-Feasible Online Incentive Mechanisms for Crowdsourcing Tasks Truthfully // IEEE/ACM Transactions on Networking. 2016. vol. 24. no. 2. pp. 647–661.
32. *Zhu X. et al.* A Fair Incentive Mechanism for Crowdsourcing in Crowd Sensing // IEEE Internet of Things Journal. 2016. vol. 3. no. 6. pp. 1364–1372.
33. *Nicholson B., Sheng V.S., Zhang J.* Label noise correction and application in crowdsourcing // Expert Systems with Applications. 2016. vol. 66. pp. 149–162.
34. *Shamir L., Diamond D., Wallin J.* Leveraging Pattern Recognition Consistency Estimation for Crowdsourcing Data Analysis // IEEE Transactions on Human-Machine Systems. 2016. vol. 46. no. 3. pp. 474–480.
35. *Burmania A., Parthasarathy S., Busso C.* Increasing the Reliability of Crowdsourcing Evaluations Using Online Quality Assessment // IEEE Transactions on Affective Computing. 2016. vol. 7. no. 4. pp. 374–388.
36. *Mok R.K.P., Chang R.K.C., Li W.* Detecting Low-Quality Workers in QoE Crowdttesting: A Worker Behavior-Based Approach // IEEE Transactions on Multimedia. 2017. vol. 19. no. 3. pp. 530–543.
37. *Morris R.R., Dontcheva M., Gerber E.M.* Priming for Better Performance in Microtask Crowdsourcing Environments // IEEE Internet Computing. 2012. vol. 16. no. 5. pp. 13–19.
38. *Allahbakhsh M. et al.* Quality Control in Crowdsourcing Systems: Issues and Directions // IEEE Internet Computing. 2013. vol. 17. no. 2. pp. 76–81.
39. *Chittilappilly A.I., Chen L., Amer-Yahia S.* A Survey of General-Purpose Crowdsourcing Techniques // IEEE Transactions on Knowledge and Data Engineering. 2016. vol. 28. no. 9. pp. 2246–2266.
40. *Li G., Wang J., Zheng Y., Franklin M.J.* Crowdsourced Data Management: A Survey // IEEE Transactions on Knowledge and Data Engineering. 2016. vol. 28. no. 9. pp. 2296–2319.
41. *Liu C.H. et al.* QoI-Aware Energy-Efficient Participatory Crowdsourcing // IEEE Sensors Journal. 2013. vol. 13. no. 10. pp. 3742–3753.
42. *Liu C.H. et al.* Toward QoI and Energy Efficiency in Participatory Crowdsourcing // IEEE Transactions on Vehicular Technology. 2015. vol. 64. no. 10. pp. 4684–4700.

43. *Yung D., Li M.-L., Chang S.* Evolutionary approach for crowdsourcing quality control // *Journal of Visual Languages and Computing*. 2014. vol. 25. no. 6. pp. 879–890.
44. *Zhang B. et al.* Privacy-preserving QoI-aware participant coordination for mobile crowdsourcing // *Computing Networks*. 2016. vol. 101. pp. 29–41.
45. *Zhang X. et al.* Robust Trajectory Estimation for Crowdsourcing-Based Mobile Applications // *IEEE Transactions on Parallel and Distributed Systems*. 2014. vol. 25. no. 7. pp. 1876–1885.
46. *Sheng V.S., Provost F., Ipeirotis P.G.* Get another label? improving data quality and data mining using multiple, noisy labelers // *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD 08)*. 2008. pp. 614–622.
47. *Dawid A.P., Skene A.M.* Maximum Likelihood Estimation of Observer Error-Rates Using the EM Algorithm // *Applied Statistics*. 1979. vol. 28. no. 1. pp. 20–28.
48. *Raykar V.C. et al.* Learning from crowds // *Journal of Machine Learning Research*. 2010. vol. 11. pp. 1297–1322.
49. *Karger D.R., Oh S., Shah D.* Iterative learning for reliable crowdsourcing systems // *Proceedings of the 24th International Conference on Neural Information Processing Systems*. 2011. pp. 1953–1961.
50. *Zhang Y., Chen X., Zhou D., Jordan M.I.* Spectral methods meet EM: A provably optimal algorithm for crowdsourcing // *Proceedings of the 27th International Conference on Neural Information Processing Systems, NIPS'14, Advances in Neural Information Processing Systems*. 2014. vol. 2. pp. 1260–1268.
51. *Demartini G., Dfallah D.E., Cudré-Mauroux P.* ZenCrowd: leveraging probabilistic reasoning and crowdsourcing techniques for large-scale entity linking // *Proceedings of the 21st international conference on World Wide Web (WWW '12)*. 2012. pp. 469–478.
52. *Gupta M.R., Chen Y.* Theory and Use of the EM Algorithm // *Foundations and Trends in Signal Processing*. 2010. vol. 4. no. 3. pp. 223–296.
53. *McLachlan G.J., Krishnan T.* *The EM algorithm and extensions*: 2nd ed. // *Wiley-Interscience*, 2008. 359 p.
54. *Sheshadri A., Lease M.* Square: A benchmark for research on computing crowd consensus // *Proceedings of the 1st AAAI Conference on Human Computations and Crowdsourcing*. 2013. pp. 156–164.
55. *Quoc Viet Hung N., Tam N.T., Tran L.N., Aberer K.* An Evaluation of Aggregation Techniques in Crowdsourcing // *Web Information Systems Engineering – WISE 2013*. 2013. LNCS 8181. pp. 1–15.
56. *Khosravifar B., Bentahar J., Gomrokchi M., Alam R.* CRM: An efficient trust and reputation model for agent computing // *Knowledge-Based Systems*. 2012. vol. 30. pp. 1–16.
57. *Fang H., Guo G., Zhang J.* Multi-faceted trust and distrust prediction for recommender systems // *Decision Support Systems*. 2015. vol. 71. pp. 37–47.
58. *Wahab O.A., Bentahar J., Otrok H., Mourad A.* A survey on trust and reputation models for Web services: Single, composite, and communities // *Decision Support Systems*. 2015. vol. 74. pp. 121–134.
59. *Yu H. et al.* Filtering trust opinions through reinforcement learning // *Decision Support Systems*. 2014. vol. 66. p. 102–113.
60. *Liu S., Yu H., Miao C., Kot A.C.* A fuzzy logic based reputation model against unfair ratings // *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*. 2013. pp. 821–828.
61. *Auer P., Cesa-Bianchi N., Fischer P.* Finite-time Analysis of the Multiarmed Bandit Problem // *Machine Learning*. 2002. vol. 47. no. 2/3. pp. 235–256.
62. *Bubeck S., Cesa-Bianchi N.* Regret analysis of stochastic and nonstochastic multi-armed bandit problems. 2012. 138 p.

63. *Restuccia F., Das S.K., Payton J.* Incentive Mechanisms for Participatory Sensing: Survey and Research Challenges // *ACM Transactions on Sensor Networks*. 2016. vol. 12. no. 2. Article no. 13.
64. *Camerer C.F.* Behavioral game theory: Predicting human behavior in strategic situations // *Advances in Behavioral Economics*. 2011. pp. 374–392.
65. *Von Ahn L., Dabbish L.* Labeling images with a computer game // *Proceedings of Conference on Human Factors in Computing Systems*. 2004. pp. 319–326.
66. *Teng C.M.* Correcting noisy data // *Proceedings of the Sixteenth International Conference on Machine Learning*. 1999. pp. 239–248.
67. *Фаликман М.В., Койфман А.Я.* Виды прайминга в исследованиях восприятия и перцептивного внимания. Часть 1 // *Вестник Московского университета. Серия 14. Психология*. 2005. № 3. С. 86–97.
68. *Frenay B., Verleysen M.* Classification in the Presence of Label Noise: A Survey // *IEEE Transactions on Neural Networks and Learning Systems*. 2014. vol. 25. no. 5. pp. 845–869.
69. *Eickhoff C., de Vries A.P.* Increasing cheat robustness of crowdsourcing tasks // *Information Retrieval*. 2013. vol. 16. no. 2. pp. 121–137.

Пономарев Андрей Васильевич — к-т техн. наук, старший научный сотрудник лаборатории интегрированных систем автоматизации, Федеральное государственное бюджетное учреждение науки Санкт-Петербургского института информатики и автоматизации Российской академии наук (СПИИРАН). Область научных интересов: системы коллективного интеллекта, социальные вычисления, краудсорсинг, рекомендующие системы, технологии поддержки принятия решений, дискретная оптимизация. Число научных публикаций — 46. ponomarev@iiias.spb.su; 14-я линия В.О., 39, Санкт-Петербург, 199178; р.т.: +7(812)328-8071, Факс: +7(812)328-4450.

Поддержка исследований. Работа выполнена при финансовой поддержке РФФИ (проекты № 16-37-60107 и № 16-07-00466).

A.V. PONOMAREV
**QUALITY CONTROL METHODS IN CROWD COMPUTING:
 LITERATURE REVIEW**

Ponomarev A.V. Quality Control Methods in Crowd Computing: Literature Review.

Abstract. Currently, crowd computing is gaining popularity. However, the quality of results obtained by means of crowd computing is often unpredictable, and that fact puts limits on the practical applicability of this technology. Therefore, systematization of information about modern methods for quality control in crowd computing is an important task that can pave the way to new research efforts in this area and therefore widen the scope of its applicability. The paper discusses the results of systematic literature review of journal articles from ScienceDirect and IEEE Xplore bibliographic databases, published after 2012. The paper also identifies main directions in crowd computing quality control, corresponding models and assumptions. In particular, it shows that most scientific attention is concentrated around consensus methods and game theoretic methods.

Keywords: crowd computing, human-machine computing, crowdsourcing, social computing, human-machine systems, human factors, systematic literature review.

Ponomarev Andrew Vasil'evich — Ph.D., senior researcher of computer aided integrated systems laboratory, St.Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS). Research interests: collective intelligence systems, social computing, crowdsourcing, machine learning, recommender systems, decision support technologies, discrete optimization. The number of publications — 46. ponomarev@iias.spb.su; 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone: +7(812)328-8071, Fax: +7(812)328-4450.

Acknowledgements. The research is financially supported by RFBR (projects no. 16-37-60107 and 16-07-00466).

References

1. Li C. et al. Noise filtering to improve data and model quality for crowdsourcing // *Knowledge-Based Syst.* 2016. vol. 107. pp. 96–103.
2. Bernstein A., Klein M., Malone T.W. Programming the global brain // *Commun. ACM.* 2012. vol. 55. no. 5. pp. 41–43.
3. Kazai G., Kamps J., Milic-Frayling N. Worker types and personality traits in crowdsourcing relevance labels // Proc. 20th ACM Int. Conf. Inf. Knowl. Manag. (CIKM '11). USA: ACM Press. 2011. pp. 1941–1944.
4. Kitchenham B. Procedures for Performing Systematic Reviews // Keele. 2004. vol. 33. no. 2004. pp. 1–26.
5. Kitchenham B., Charters S. Guidelines for performing Systematic Literature Reviews in Software Engineering. Technical Report EBSE-2007-01. 2007. 65 p.
6. Zhang J. et al. Consensus algorithms for biased labeling in crowdsourcing. *Information Sciences.* 2017. vol. 382. pp. 254–273.
7. Matsunaga A., Mast A., Fortes J.A.B. Workforce-efficient consensus in crowdsourced transcription of biocollections information. *Future Generation Computer Systems.* 2016. vol. 56. pp. 526–536.
8. Han K., Zhang C., Luo J. Taming the Uncertainty: Budget Limited Robust Crowdsensing Through Online Learning. *Biological Cybernetics.* 2016. vol. 24. no. 3.

- pp. 1462–1475.
9. Dang D., Liu Y., Zhang X., Huang S. Crowdsourcing Worker Quality Evaluation Algorithm on MapReduce for Big Data Applications. *IEEE Transactions on Parallel and Distributed Systems*. 2016. vol. 27. no. 7. pp. 1879–1888.
 10. Duan L., Oyama S., Sato H., Kurihara M. Separate or joint? Estimation of multiple labels from crowdsourced annotations. *Expert Systems with Applications*. 2014. vol. 41. no. 13. pp. 5723–5732.
 11. Kara Y.E., Genc G., Aran O., Akarun L. Modeling annotator behaviors for crowd labeling. *Neurocomputing*. 2015. vol. 160. pp. 141–156.
 12. Kubota T., Aritsugi M. Assignment strategies for ground truths in the crowdsourcing of labeling tasks. *The Journal of Systems and Software*. 2017. vol. 126. pp. 113–126.
 13. Otani N., Baba Y., Kashima H. Quality control of crowdsourced classification using hierarchical class structures. *Expert Systems with Applications*. 2016. vol. 58. pp. 155–163.
 14. Vempaty A., Varshney L.R., Varshney P.K. Reliable Crowdsourcing for Multi-Class Labeling. *IEEE Journal of Selected Topics in Signal Process.* 2014. vol. 8. no. 4. pp. 667–679.
 15. Vuurens J.B.P., de Vries A.P. Obtaining High-Quality Relevance Judgments Using Crowdsourcing. *IEEE Internet Computing*. 2012. vol. 16. No. 5. pp. 20–27.
 16. Zhang J., Sheng V., Wu J., Wu X. Multi-Class Ground Truth Inference in Crowdsourcing with Clustering. *IEEE Transactions on Knowledge and Data Engineering*. 2016. vol. 28. no. 4. pp. 1080–1085.
 17. Zhang J., Wu X., Sheng V.S. Imbalanced Multiple Noisy Labeling. *IEEE Transactions on Knowledge and Data Engineering*. 2015. vol. 27. no. 2. pp. 489–503.
 18. Dai P., Lin C., Weld D. POMDP-based control of workflows for crowdsourcing. *Artificial Intelligence*. 2013. vol. 202. pp. 52–85.
 19. Hirth M., Hoßfeld T., Tran-Gia P. Analyzing costs and accuracy of validation mechanisms for crowdsourcing platforms. *Mathematical and Computer Modeling*. 2013. vol. 57. no. 11–12. pp. 2918–2932.
 20. Lee J., Lee D., Hwang S. CrowdK: Answering top- k queries with crowdsourcing. *Information Sciences*. 2017. vol. 399. pp. 98–120.
 21. Lou Y. et al. Use of Ontology Structure and Bayesian Models to Aid the Crowdsourcing of ICD-11 Sanctioning Rules. *Journal of Biomedical Informatics*. 2017. vol. 68. no. C. pp. 20–34.
 22. Ding S., He X., Wang J. Multiobjective Optimization Model for Service Node Selection Based on a Tradeoff Between Quality of Service and Resource Consumption in Mobile Crowd Sensing. *IEEE Internet of Things Journal*. 2017. vol. 4. no. 1. pp. 258–268.
 23. Gong Y. et al. Optimal Task Recommendation for Mobile Crowdsourcing With Privacy Control. *IEEE Internet of Things Journal*. 2016. vol. 3. no. 5. pp. 745–756.
 24. Miao C., Yu H., Shen Z., Leung C. Balancing quality and budget considerations in mobile crowdsourcing. *Decision Support Systems*. 2016. vol. 90. pp. 56–64.
 25. Tran-Thanh L., Stein S., Rogers A., Jennings N. Efficient crowdsourcing of unknown experts using bounded multi-armed bandits. *Artificial Intelligence*. 2014. vol. 214. pp. 89–111.
 26. Dai W., Wang Y., Jin Q., Ma J. An Integrated Incentive Framework for Mobile Crowdsourced Sensing. *Tsinghua Science and Technology*. 2016. vol. 21. no. 2. pp. 146–156.
 27. Gao Y., Chen Y., Liu K.J.R. On Cost-Effective Incentive Mechanisms in Microtask Crowdsourcing. *IEEE Transactions on Computational Intelligence and AI in Games*. 2015. vol. 7. no. 1. pp. 3–15.
 28. Wang B.C., Lin C.W., Chen K.T., Chen L.J. An analytical model for generalized ESP games. *Knowledge-Based Systems*. 2012. vol. 34. pp. 114–127.
 29. Wen Y. et al. Quality-Driven Auction-Based Incentive Mechanism for Mobile Crowd

- Sensing. *IEEE Transactions on Vehicular Technology*. 2015. vol. 64. no. 9. pp. 4203–4214.
30. Yang G., He S., Shi Z., Chen J. Promoting Cooperation by the Social Incentive Mechanism in Mobile Crowdsensing. *IEEE Communications Magazine*. 2017. vol. 55. no. 3. pp. 86–92.
 31. Zhao D., Li X., Ma H. Budget-Feasible Online Incentive Mechanisms for Crowdsourcing Tasks Truthfully. *IEEE/ACM Transactions on Networking*. 2016. vol. 24. no. 2. pp. 647–661.
 32. Zhu X. et al. A Fair Incentive Mechanism for Crowdsourcing in Crowd Sensing. *IEEE Internet of Things Journal*. 2016. vol. 3. no. 6. pp. 1364–1372.
 33. Nicholson B., Sheng V.S., Zhang J. Label noise correction and application in crowdsourcing. *Expert Systems with Applications*. 2016. vol. 66. pp. 149–162.
 34. Shamir L., Diamond D., Wallin J. Leveraging Pattern Recognition Consistency Estimation for Crowdsourcing Data Analysis. *IEEE Transactions on Human-Machine Systems*. 2016. vol. 46. no. 3. pp. 474–480.
 35. Burmania A., Parthasarathy S., Busso C. Increasing the Reliability of Crowdsourcing Evaluations Using Online Quality Assessment. *IEEE Transactions on Affective Computing*. 2016. vol. 7. no. 4. pp. 374–388.
 36. Mok R.K.P., Chang R.K.C., Li W. Detecting Low-Quality Workers in QoE Crowdtesting: A Worker Behavior-Based Approach. *IEEE Transactions on Multimedia*. 2017. vol. 19. no. 3. pp. 530–543.
 37. Morris R.R., Dontcheva M., Gerber E.M. Priming for Better Performance in Microtask Crowdsourcing Environments. *IEEE Internet Computing*. 2012. vol. 16. no. 5. pp. 13–19.
 38. Allahbakhsh M. et al. Quality Control in Crowdsourcing Systems: Issues and Directions. *IEEE Internet Computing*. 2013. vol. 17. no. 2. pp. 76–81.
 39. Chittilappilly A.I., Chen L., Amer-Yahia S. A Survey of General-Purpose Crowdsourcing Techniques. *IEEE Transactions on Knowledge and Data Engineering*. 2016. vol. 28. no. 9. pp. 2246–2266.
 40. Li G., Wang J., Zheng Y., Franklin M.J. Crowdsourced Data Management: A Survey. *IEEE Transactions on Knowledge and Data Engineering*. 2016. vol. 28. no. 9. pp. 2296–2319.
 41. Liu C.H. et al. QoI-Aware Energy-Efficient Participatory Crowdsourcing. *IEEE Sensors Journal*. 2013. vol. 13. no. 10. pp. 3742–3753.
 42. Liu C.H. et al. Toward QoI and Energy Efficiency in Participatory Crowdsourcing. *IEEE Transactions on Vehicular Technology*. 2015. vol. 64. no. 10. pp. 4684–4700.
 43. Yung D., Li M.-L., Chang S. Evolutionary approach for crowdsourcing quality control. *Journal of Visual Languages and Computing*. 2014. vol. 25. no. 6. pp. 879–890.
 44. Zhang B. et al. Privacy-preserving QoI-aware participant coordination for mobile crowdsourcing. *Computing Networks*. 2016. vol. 101. pp. 29–41.
 45. Zhang X. et al. Robust Trajectory Estimation for Crowdsourcing-Based Mobile Applications. *IEEE Transactions on Parallel and Distributed Systems*. 2014. vol. 25. no. 7. pp. 1876–1885.
 46. Sheng V.S., Provost F., Ipeirotis P.G. Get another label? improving data quality and data mining using multiple, noisy labelers. Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD 08). 2008. pp. 614–622.
 47. Dawid A.P., Skene A.M. Maximum Likelihood Estimation of Observer Error-Rates Using the EM Algorithm. *Applied Statistics*. 1979. vol. 28. no. 1. pp. 20–28.
 48. Raykar V.C. et al. Learning from crowds. *Journal of Machine Learning Research*. 2010. vol. 11. pp. 1297–1322.
 49. Karger D.R., Oh S., Shah D. Iterative learning for reliable crowdsourcing systems. Proceedings of the 24th International Conference on Neural Information Processing Systems. 2011. pp. 1953–1961.

50. Zhang Y., Chen X., Zhou D., Jordan M.I. Spectral methods meet EM: A provably optimal algorithm for crowdsourcing. Proceedings of the 27th International Conference on Neural Information Processing Systems, NIPS'14, Advances in Neural Information Processing Systems. 2014. vol. 2. pp. 1260–1268.
51. Demartini G., Difallah D.E., Cudré-Mauroux P. ZenCrowd: leveraging probabilistic reasoning and crowdsourcing techniques for large-scale entity linking. Proceedings of the 21st international conference on World Wide Web (WWW '12). 2012. pp. 469–478.
52. Gupta M.R., Chen Y. Theory and Use of the EM Algorithm. Foundations and Trends in Signal Processing. 2010. vol. 4. no. 3. pp. 223–296.
53. McLachlan G.J., Krishnan T. The EM algorithm and extensions. 2nd ed. 2008. 359 p.
54. Sheshadri A., Lease M. Square: A benchmark for research on computing crowd consensus. Proceedings of the 1st AAAI Conference on Human Computations and Crowdsourcing. 2013. pp. 156–164.
55. Quoc Viet Hung N., Tam N.T., Tran L.N., Aberer K. An Evaluation of Aggregation Techniques in Crowdsourcing. Web Information Systems Engineering – WISE 2013, 2013. LNCS 8181. pp. 1–15.
56. Khosravifar B., Bentahar J., Gomrokchi M., Alam R. CRM: An efficient trust and reputation model for agent computing. *Knowledge-Based Systems*. 2012. vol. 30. pp. 1–16.
57. Fang H., Guo G., Zhang J. Multi-faceted trust and distrust prediction for recommender systems. *Decision Support Systems*. 2015. vol. 71. pp. 37–47.
58. Wahab O.A., Bentahar J., Otok H., Mourad A. A survey on trust and reputation models for Web services: Single, composite, and communities. *Decision Support Systems*. 2015. vol. 74. pp. 121–134.
59. Yu H. et al. Filtering trust opinions through reinforcement learning. *Decision Support Systems*. 2014. vol. 66. p. 102–113.
60. Liu S., Yu H., Miao C., Kot A.C. A fuzzy logic based reputation model against unfair ratings. Proceedings of the 2013 International Conference on Autonomous Agents and Multi-Agent Systems. 2013. pp. 821–828.
61. Auer P., Cesa-Bianchi N., Fischer P. Finite-time Analysis of the Multiarmed Bandit Problem. *Machine Learning*. 2002. vol. 47, no. 2/3. pp. 235–256.
62. Bubeck S., Cesa-Bianchi N. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. 2012. 138 p.
63. Restuccia F., Das S.K., Payton J. Incentive Mechanisms for Participatory Sensing: Survey and Research Challenges. *ACM Transactions on Sensor Networks*. 2016. vol. 12. no. 2. Article no. 13.
64. Camerer C.F. Behavioral game theory: Predicting human behavior in strategic situations. *Advances in Behavioral Economics*. 2011. pp. 374–392.
65. Von Ahn L., Dabbish L. Labeling images with a computer game. Proceedings of Conference on Human Factors in Computing Systems. 2004. pp. 319–326.
66. Teng C.M. Correcting noisy data. Proceedings of the Sixteenth International Conference on Machine Learning. 1999. pp. 239–248.
67. Falikman M.V., Koifman A.Ya. [Types of priming in perception and perceptive attention research. Part 1]. *Vestnik Moskovskogo Universiteta. Seriya 14. Psihologiya – Herald of Moscow State University. Series 14. Psychology*. 2005. vol. 3. pp. 86–97. (In Russ.).
68. Frenay B., Verleysen M. Classification in the Presence of Label Noise: A Survey. *IEEE Transactions on Neural Networks and Learning Systems*. 2014. vol. 25. no. 5. pp. 845–869.
69. Eickhoff C., de Vries A.P. Increasing cheat robustness of crowdsourcing tasks. *Information Retrieval*. 2013. vol. 16. no. 2. pp. 121–137.

С.А. КРАСНОВ, А.С. ИЛАТОВСКИЙ, А.Д. ХОМОНЕНКО, В.Н. АРСЕНЬЕВ
**ОЦЕНКА СЕМАНТИЧЕСКОЙ БЛИЗОСТИ ДОКУМЕНТОВ НА
ОСНОВЕ ЛАТЕНТНО-СЕМАНТИЧЕСКОГО АНАЛИЗА С
АВТОМАТИЧЕСКИМ ВЫБОРОМ РАНГОВЫХ ЗНАЧЕНИЙ**

Краснов С.А., Илатовский А.С., Хомоненко А.Д., Арсеньев В.Н. Оценка семантической близости документов на основе латентно-семантического анализа с автоматическим выбором ранговых значений.

Аннотация. Предлагается метод оценки семантической близости документов на основе латентно-семантического анализа, учета динамики изменения сингулярных значений матрицы «терм-документ» и автоматического определения диапазона ранговых значений. Оценка семантической близости документов рассматривается применительно к решению задач выявления дублирования и противоречий в базах данных.

Приводится краткий обзор подходов, используемых при оценке семантической близости документов, выявлении дублирования и противоречий в базах данных и хранилищах данных. Приводятся результаты численных примеров оценки семантических зависимостей между терминами документов в интересах выявления дублирования и противоречий в базах данных. При этом в качестве результирующей характеристики рассчитывается степень соответствия λ сравниваемых документов.

Приведены сравнительные оценки расчета степени соответствия λ документов с помощью основных методов (косинусной меры близости, векторной модели, коэффициента ранговой корреляции Спирмена, статической меры tf-idf — частота термина — обратная документная частота).

Показано, что использование предложенного метода анализа динамики изменения сингулярных чисел матрицы «терм-документ» с автоматическим выбором диапазона используемых ранговых значений позволяет устранить зависимость метода латентно-семантического анализа от выбора оптимального ранга.

Ключевые слова: оценка семантической близости документов, выявление дублирования и противоречий, базы данных, латентно-семантический анализ, статистический анализ, косинусная мера близости, векторная модель.

1. Введение. В настоящее время наблюдается большой рост объемов информации, хранимой и обрабатываемой в базах данных (БД) и информационных системах, сетях Интернет и Интранет. При этом достаточно важной и распространенной является задача выявления и устранения дублированной и противоречивой информации. В основе ее решения лежит теория интеллектуального анализа данных [1, 2].

При решении указанной задачи и других задач интеллектуальной обработки рассматриваются различные варианты прикладных областей и разные подходы. В частности, в [3] предлагаются методы выявления дублированной и противоречивой информации в Википедии. В [4, 5] для устранения конфликтов на

уровне семантики используется онтологический подход. В [6, 7] предлагаются решения задач разрешения противоречий в электронной библиотеке eLIBRARY.ru и автоматизированной классификации документов с использованием нечеткого логического вывода по алгоритму Мамдани. В [8] решается задача автоматического устранения конфликтов в БД реляционного типа. Предлагаемые подходы служат для обнаружения конфликтов именования и структурных конфликтов или для предотвращения конфликтов путем введения более детализированных описаний связей между объектами, запрещения определенных действий. В [9] рассматривается задача стилевой и жанровой классификации на основе глубокого парсинга (разбора) текста.

В статье [10] представлено разрешение конфликтов и управление репликацией при интеграции данных из нескольких источников с помощью объектной модели. Разрешение конфликтов выполняется с помощью операций переключения и вспомогательной алгебры объекта. В работе [11] выявляются противоречия в тексте путем отыскания конфликтных операторов как основы информационного анализа текста. Вводится определение противоречия и предлагается их типология. Отмечается достижение хороших характеристик производительности анализа текста на основе предложенного подхода с использованием отрицания и антонимии.

В числе современных работ в области интеллектуального анализа данных отметим следующие. В [12] затрагиваются вопросы применения библиотеки BigARTM с открытым кодом для тематического моделирования больших текстовых коллекций. В [13] дается математическое обоснование метода категориальных векторов, обеспечивающего повышение эффективности поисковых информационных запросов. В [14] систематически изложены вопросы информационного поиска с подробными запросами.

Решение задачи выявления и устранения дублирования и противоречий в БД и многих других задач, например информационного поиска, кластеризации, классификации (рубрикация) и фильтрации, часто требует оценки семантической близости документов. При этом широко используется метод латентно-семантического анализа (ЛСА) [15-19].

Целью настоящей статьи является повышение *точности* оценки СБД с помощью метода ЛСА для выявления дублирования и противоречий в БД и информационных системах. Дублирования и противоречия в БД могут возникать из-за ошибок ввода и

редактирования данных, выполняемых человеком; ошибок, возникающих в процессе передачи данных между БД в распределенных информационных системах и так далее.

Современные системы управления базами данных (СУБД) реляционного типа имеют механизмы защиты от возникновения избыточности данных и соответствующих аномалий с помощью нормализации отношений [17]. Средства СУБД также позволяют обеспечить целостность БД.

Противоречия (конфликты) в БД разделяют в основном на три группы:

1. Конфликты наименований — из-за использования различной терминологии в процессе описания некоторой предметной области, приводят к омонимии и синонимии.

2. Структурные конфликты — связаны с различными представлениями одного и того же объекта в различных источниках.

3. Семантические конфликты — возникают, когда данные противоречивы.

Выявляемые в статье противоречия охватывают все три группы конфликтов в той или иной мере. Под такими противоречиями понимаем наличие близких, но не совпадающих значений одних и тех же полей БД или документов, что может приводить к неоднозначности их интерпретации.

Подход к выявлению и устранению дублированной и противоречивой информации описан в работе [20], где показана необходимость провести оценку смысловой близости документов с помощью метода ЛСА. Отметим, что в статье [21] приводится обзор работ по оценке близости текстов. В ней выделяются три подхода: основанные на строках, основанные на теле документа и основанные на знаниях. В статье [22] рассматривается метод сравнения текстов на русском языке для определения семантической близости. Дается обзор существующих методов сравнения. Предложен метод определения степени подобия между текстовыми пассажами в пределах семантического класса.

Для достижения поставленной нами цели требуется комплексное применение исходных и разработанных методов в области семантического анализа информации [23-26].

Метод ЛСА позволяет получить более точные результаты по отношению к простой векторной модели, однако без предварительной обработки данных метод ЛСА обладает

недостаточной точностью для выявления дублирования и противоречий в БД. Кроме того, у метода ЛСА отсутствует универсальный способ автоматической аппроксимации исходной матрицы матрицей, содержащей только k значимых сингулярных значений, отражающих основную структуру ассоциативных зависимостей исходной матрицы. Отсюда следует необходимость решения задачи выбора оптимального ранга аппроксимирующей матрицы [7, 17, 19, 20].

2. Этапы выявления дублирования и противоречий в БД.

Как отмечалось, для выявления и устранения дублирования и противоречий в текстовых данных требуется провести оценку их смысловой близости с помощью метода ЛСА. Кроме того, ЛСА измеряет корреляционные зависимости типа «терм-терм», «терм-вектор» и «вектор-вектор». Результативность данного метода зависит не только от частот использования слов (термов) в документах, но и от выявления более глубоких (скрытых) связей [7, 23].

Для устранения факторов, затрудняющих выявления дублирования и противоречий в БД с высокой точностью, предлагается комплексное решение по подбору параметров метода, опирающееся на результаты [20] и основанное на комбинации:

- предложенных методов предварительной обработки данных, повышающих качество формирования векторной модели данных;
- метода ЛСА с анализом динамики изменения сингулярных значений матрицы «терм-документ», дающего возможность автоматически выделить диапазон ранговых значений сингулярных чисел;
- косинусной меры близости (КМБ) на основе скалярного произведения векторов, позволяющей провести оценку смысловой близости сравниваемых данных.

В результате анализа лингвистических и статистических методов предварительной обработки данных определены факторы, влияющие на точность выявления дублирования и противоречий в БД с помощью метода ЛСА:

- оценка важности слова в контексте документов (статистические меры);
- нормирование векторов (НВ);

Представлением документа (в векторном пространстве термов) будет вектор вещественных чисел, $\vec{d}_i = (d_{i1}, \dots, d_{i|m|})^T$, в котором каждое вещественное число будет координатой вектора, соответствовать

конкретному термину и иметь значение, равное весу термина в соответствующем документе. При оценке СБД с помощью метода ЛСА исследовано влияние статической меры *tf-idf* и процедуры НВ:

$$d_{ij} = \frac{w_{ij}}{|w_{ij}|}, \quad w_{ij} = tf_{ij} \times \log \frac{|D|}{df_j}, \quad tf_{ij} = \frac{n_i}{\sum_k n_k}, \quad (1)$$

где d_{ij} — нормированные веса по *tf-idf*, (частота встречаемости термина — обратная документная частота), $0 \leq d_{ij} \leq 1$;

$|D|$ — число анализируемых документов;

n_i — количество употреблений слова в документе;

$\sum_k n_k$ — общее количество слов, содержащихся в документе;

tf_{ij} — частота встречаемости слова в документе (число раз, которое j -е слово встретилось в i -м документе);

df_j — документная частота (число документов, в которых встретилось j -ое слово);

$|w_{ij}|$ — нормированный вектор w_{ij} в евклидовом пространстве.

В *tf-idf* наибольший вес получают слова с высокой частотой в пределах документа и с небольшой частотой встречаемости в других документах. При НВ матрицы «терм-документ» скалярное произведение не зависит от нормы векторов. Это позволяет упростить сравнение результатов скалярных произведений. Операция нормирования производится перед расчетом λ — степени соответствия документов.

Существенным фактором, влияющим на точность выявления дублирования и противоречий в БД при использовании метода ЛСА, является определение оптимального ранга аппроксимирующей матрицы [7, 12, 20]. Поэтому анализируется влияние сингулярных значений на точность выявления дублирования и противоречий в БД и предлагается метод анализа динамики изменения сингулярных значений матрицы «терм-документ» с автоматическим определением диапазона используемых ранговых значений. Этапы процесса выявления дублирования и противоречий представлены в таблице 1.

Таблица 1. Этапы выявления дублирования и противоречий в БД

Этап	Описание
1. Лингвистический анализ	
Лексический анализ; морфологический анализ.	Разделение текста документа на лексемы и выделение термов (для проведения экспериментов использовался алгоритм стемминга М. Портера).
2. Статистический анализ	
Частотный анализ и вычисление весовых коэффициентов термов; нормализация векторов.	Подсчет числа повторений термов; получение весовых характеристик; формирование матрицы «терм-документ».
3. ЛСА	
Сингулярное разложение матрицы «терм-документ»; определение значимого диапазона значений рангов рассматриваемой матрицы; получение аппроксимированных матриц с меньшими рангами.	Выбор оптимального диапазона ненулевых сингулярных значений на основе метода анализа динамики изменения сингулярных чисел анализируемой матрицы с автоматическим выбором диапазона ранговых значений.
3. Анализ соответствия векторов	
Оценка косинусной меры близости (λ векторов); вычисление результирующего λ полученных значений; задание порогового значения λ ; сравнение полученного значения λ с пороговым значением.	Вычисление скалярных произведений векторов для каждого значимого ранга; вычисление среднего арифметического для каждого значимого ранга; выявление дублирования и противоречий в БД.

Для экспериментальной части работы использован язык программирования Python 3.4 и его библиотеки с открытым исходным кодом: NumPy и SciPy. Такой выбор обусловлен тем, что в них имеется поддержка больших многомерных массивов, матриц и высокоуровневых математических функций для операций с ними.

3. Метод анализа динамики изменения сингулярных чисел матрицы «терм-документ» с автоматическим выбором диапазона используемых ранговых значений. На первом шаге необходимо сформировать матрицу A «терм-документ», которая опишет анализируемые документы и будет содержать исходные данные для метода ЛСА. Ее элементы будут содержать веса термов, полученные

после применения статистической меры $tf-idf$ (отношение частоты слов к обратной частоте документа) $0 \leq d_{ij} \leq 1$ и НВ матрицы. Далее необходимо выполнить сингулярное разложение матрицы A в произведение трех матриц:

$$A = U W V^T, \quad (2)$$

где U и V — унитарные матрицы, которые состоят из левых и правых сингулярных векторов, а W — матрица с неотрицательными элементами на диагонали, которые называются сингулярными значениями матрицы A .

Согласно теореме Экарта-Янга, если в матрице W оставить только наибольшие сингулярные значения σ , а в матрицах U и V — соответствующие этим значениям столбцы, то матрицы U_σ и V_σ будут лучшими их приближениями, отражающими ассоциативные зависимости представления термов и документов в пространстве размерности σ [17, 20].

Следующим шагом является оптимальный выбор ненулевых сингулярных значений $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n > 0$ матрицы A , которые влияют на результат выявления дублирования и противоречий в БД. Результатом приведения матрицы A к рангам, имеющим близкие к нулю сингулярные значения, являются равные матрицы, учет которых ведет к увеличению вычислительной сложности метода ЛСА и тем самым снижается оперативность выявления и устранения дублирования и противоречий в БД. Чтобы решить задачу оптимального выбора сингулярных значений, предлагается описанный ниже эвристический метод выбора значимых рангов, который является сущностью метода анализа динамики изменения сингулярных чисел матрицы «терм-документ» с автоматическим выбором диапазона используемых ранговых значений.

Определим функцию $f(i) = \sigma_i$, $i \in N$, $i < P$, где P — количество документов. Значимыми рангами являются только ранги $r_p, r_{p+1}, \dots, r_{m-1}, r_m$, $p \leq m$, заключенные между соответствующими сингулярными значениями $\sigma_p \geq \sigma_m \geq 0$, претерпевающими резкое изменение $\Delta\sigma_i = \sigma_i - \sigma_{i-1}$, $i \in \{p; m\}$ относительно предыдущих сингулярных значений $\sigma_i \geq \sigma_p$, $i \leq p$; $\sigma_i \geq \sigma_m$, $i \leq m$.

Определение границ значимых рангов осуществляется с помощью понятия производной функции сингулярных значений $f'(i) = \sigma_i - \sigma_{i-1}$, $i \in N$, $i < n$.

Далее осуществляется поиск максимального значения производной функции f'_{\max} , достижимого при σ_{\max} , $1 < \max \leq \frac{n}{2}$.

Затем определяется первый локальный минимум σ_p , следующий за σ_{\max} . Ранги $r_1, r_2, \dots, r_i, \dots, r_{p-1}, r_p, i \leq p$, соответствующие сингулярным значениям, большим σ_p , признаются незначимыми.

В качестве правой границы значимых рангов выбирается ранг r_n , соответствующий последнему ненулевому сингулярному значению σ_n .

На заключительном этапе необходимо рассчитать λ документов, используя КМБ:

$$\cos(\bar{X}_j, \bar{X}_i) = \sum_{i=1}^M x_j^{(i)} x_i^{(i)}, \quad (3)$$

где $x_j^{(i)} x_i^{(i)}$ — элементы разных векторов, между которыми вычисляется мера близости; M — размерность пространства векторов.

Значения КМБ ограничены промежутком $[-1; 1]$ при использовании операции НВ. Степень соответствия $\lambda_{j,i}^l$ векторов $\bar{X}_j, \bar{X}_i (i < j \leq P)$ вычисляется для каждого значимого ранга $r_l, p \leq l \leq m$. Далее необходимо вычислить результирующую $\bar{\lambda}_{j,l}$ векторов \bar{X}_j, \bar{X}_i :

$$\bar{\lambda}_{j,l} = \frac{\sum_{l=p}^m \lambda_{j,i}^l}{m - p + 1}. \quad (4)$$

Таким образом, получаем результирующую степень соответствия для конкретной пары документов по всему значимому диапазону ранговых значений. Полученное значение необходимо сравнить, например, с пороговым значением для автоматического или автоматизированного принятия решения по устранению дублирования или противоречия в БД.

Отметим, что метод автоматического определения диапазона используемых ранговых значений позволяет с большей точностью гарантировать, что данные действительно дублированные или противоречивые, потому что значения λ всех пар векторов

оцениваются на каждом значимом ранге. При этом случайные всплески полученных значений λ при неправильных ранговых значениях сглаживаются, а значения λ явно дублированных или противоречивых данных стремятся к единице, так как находится среднее арифметическое по всем значениям λ одной пары векторов матриц, полученных из диапазона используемых ранговых значений.

Реализуемость рассмотренных нами этапов выявления дублирования и противоречий в БД подтверждается представленными результатами экспериментов, а также результатами, изложенными в других источниках, например [20, 27].

4. Сравнительная оценка точности выявления дублирования и противоречий в БД. Оценка точности выявления дублирования и противоречий в БД производится путем сравнения полученных результатов с эталоном по единой выборке. В качестве эталона используется выборка с заранее известной дублированной и противоречивой информацией, размеченной вручную. Для примера рассмотрим выборку D записей электронной библиотеки eLIBRARY.ru. Фамилии авторов изменены (таблицы 1, 2). Полная выборка содержит 60 документов и включает 7 пар дублированных и 3 пары схожих документов. Схожие документы, как и дублированные, имеют минимальные расхождения, поэтому в рамках этой модели рассматриваются как противоречивые.

Таблица 1. Образцы дублированных документов выборки D

1	<i>Модель функционирования системы автоматической рубрикации документов в нестационарном режиме. Иванов А.Д., Кубнов В.П., Петров С.А., Юремин А.С. Проблемы информационной безопасности. Компьютерные системы. 2011. № 4. С. 16.</i>
2	<i>Модель функционирования системы автоматической рубрикации документов в нестационарном режиме. Петров С.А., Юремин А.С., Иванов А.Д., Кубнов В.П. Проблемы информационной безопасности. Компьютерные системы. 2011. № 4. С. 16-23.</i>
3	<i>Иванов А.Д., Петров С.А. Применение метода латентно-семантического анализа для автоматической рубрикации текстов в системах электронного документооборота. Сборник материалов первой международной научно-практической конференции «Интеллектуальные системы на транспорте» (ИнтеллектТранс-2011)/СПб.: ПГУПС, 2011. С. 291-294.</i>
4	<i>Применение метода латентно-семантического анализа для автоматической рубрикации текстов в системах электронного документооборота: доклад Петров С.А., Иванов А.Д. В сборнике: Интеллектуальные системы на транспорте. Сборник материалов Первой международной научно-практической конференции. Редактор А. А. Порниенко. 2011. С. 291-294.</i>

Таблица 2. Образцы схожих документов выборки D

1.	Информационная система геодезического мониторинга транспортных сооружений. Дрынь М.Я., Пикитчин А.А., Ледянин В.В., Иванов А.Д. <i>В сборнике: Интеллектуальные системы на транспорте. ИнтеллектТранс-2012 сборник материалов II Международной научно-практической конференции. Федеральное агентство железно-дорожного транспорта, ФГБОУ ВПО "Петербургский государственный университет путей сообщения"; Редактор: Порниенко А.А.. 2012. С. 150-154.</i>
2.	Информационная система геодезического мониторинга транспортных сооружений. Дрынь М.Я., Пикитчин А.А., Ледянин В.В., Иванов А.Д. <i>В книге: Интеллектуальные системы на транспорте. ИнтеллектТранс-2012 сборник материалов II Международной научно-практической конференции. Федеральное агентство железно-дорожного транспорта, ФГБОУ ВПО "Петербургский государственный университет путей сообщения"; Редактор: Порниенко А.А. 2012. С. 36.</i>
3.	Анализ и интеграция диагностической информации АСУ в условиях катастрофических отказов: доклад. Салиниченко С.В., Иванов А.Д. <i>В сборнике: Проблемы математической и естественно-научной подготовки в инженерном образовании сборник трудов II Международной научно-методической конференции. Редактор: В. А. Родаковский. 2013. С. 99-103.</i>
4.	Анализ и интеграция диагностической информации АСУ в условиях катастрофических отказов. Иванов А.Д., Салиниченко С.В. <i>В книге: Проблемы математической и естественно-научной подготовки в инженерном образовании Тезисы докладов 2-ой международной научно-методической конференции. 2012. С. 168-169.</i>

В качестве меры точности выявления дублирований и противоречий в БД нами используется оценка семантической близости сравниваемых пар документов.

4.1. О применении метода ЛСА. Результаты исследований показали низкую точность выявления дублирования и противоречий в БД (рисунок 1) с помощью простой векторной модели (ПВМ) без использования метода ЛСА.

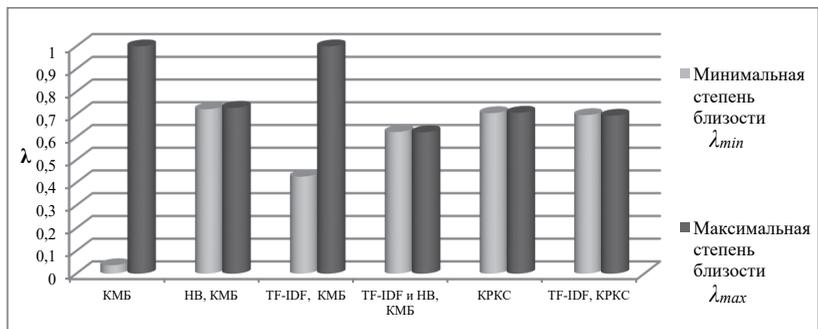


Рис. 1. Оценка семантической близости пар документов без метода ЛСА

То есть необходимо, чтобы левый цилиндр в каждой подгруппе сравниваемых методов был значительно выше правого (рисунок 1). Простая векторная модель не позволяет получить значение Δ , требуемое для выявления дублирования и противоречий в БД. Для наглядности результатов применения ПВМ без операции НВ в матрицы A проведено нормирование полученных результатов: каждое значение ряда заменено отношением к наибольшему по модулю значению ряда.

Операция нормирования при использовании коэффициента ранговой корреляции Спирмена (КРКС) [27] не меняет результатов применения метода, так как порядок расположения отсортированных элементов векторов матрицы A не изменяется.

4.2. Влияние предварительной обработки данных на точность метода ЛСА. На основе выборки D проведен анализ эффективности метода ЛСА в зависимости от различных комбинаций методов предварительной обработки данных (рисунок 2), из которого следует, что применение статической меры $tf-idf$ в совокупности с НВ дает наилучший результат, так как удается достичь устойчивых значений Δ на достаточно продолжительном участке ранговых значений. На рисунке 3 представлены значения Δ для каждого ранга. В качестве меры близости выбрано скалярное произведение векторов. Для наглядности проведено нормирование результатов.

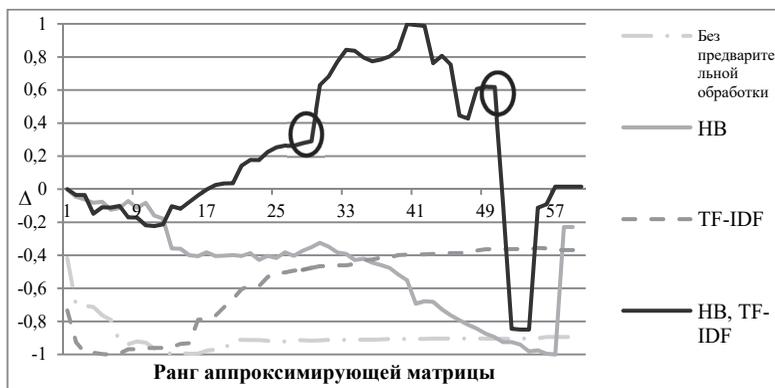


Рис. 2. Точность комбинаций методов предварительной обработки данных

Предположительно, точки изменения характера монотонности графика функции, отмеченные кругами черного цвета, ограничивают промежутки значений рангов, лучше всего подходящих для выявления дублированной и противоречивой информации. Далее выполняется обоснование метода выявления таких точек.

4.3. Влияние мер близости на точность метода ЛСА. Проведен анализ применения методов расчета λ документов при применении метода ЛСА (рисунок 3): с помощью КМБ и с помощью КРКС [27].

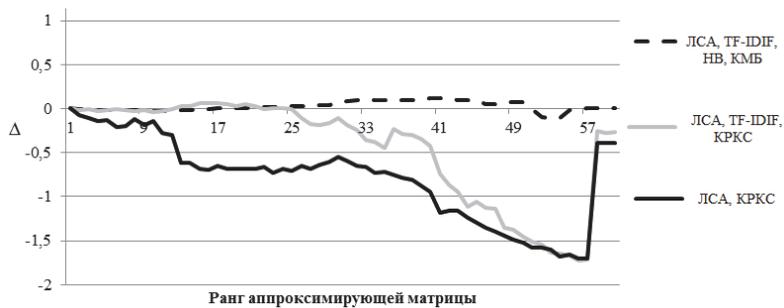


Рис. 3. Точность различных методов расчета λ документов

На рисунке 3 для каждого ранга приведено найденное значение Δ . Анализ показал, что применение КМБ дает более устойчивые к изменению ранга аппроксимирующей матрицы результаты. Устойчивость к изменению ранга позволяет выбрать более широкий диапазон сингулярных значений без критического падения точности метода.

4.4. Оценка точности применения ЛСА с анализом динамики изменения сингулярных чисел матрицы «терм-документ» и автоматическим выбором диапазона ранговых значений. Рассмотрим алгоритм выбора значимых рангов на примере выборки D (рисунок 4). В качестве правой границы значимых рангов примем ранг r_m , соответствующий последнему ненулевому значению σ_m функции $f(i)$.

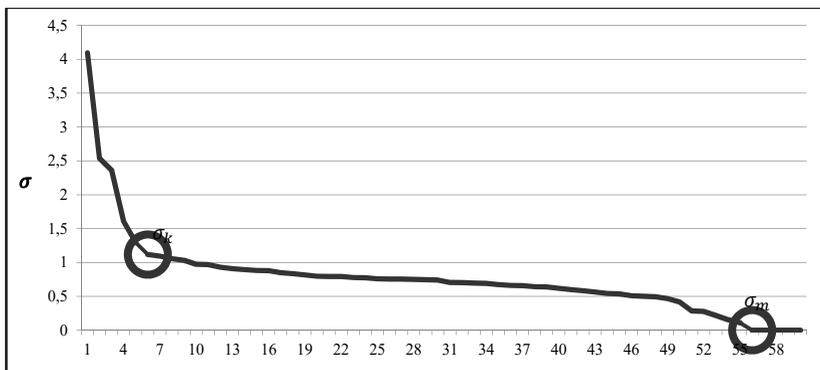


Рис. 4. Сингулярные значения

Для определения левой границы r_k значимых рангов рассмотрим график производной функции сингулярных значений $f'(i)$ (рисунок 5). Резкая смена характера убывания функции сингулярных значений $f(i)$ при $i = k$ (рисунок 5) соответствует первому локальному минимуму $\Delta\sigma_k$ (рисунок 5), следующему за максимальным значением $\Delta\sigma_{max}$ функции $f'(i)$ на интервале $(1; \frac{n}{2})$. Ранг r_k , соответствующий сингулярному значению σ_k , примем за левую границу значимых рангов.

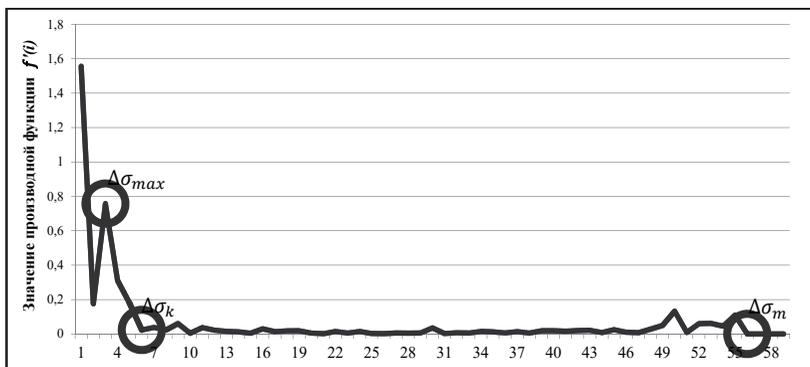


Рис. 5. Производная функция сингулярных значений

Оценим качество подобранных параметров метода ЛСА для выявления дублирования и противоречий в БД (рисунок 6) на основе оценки семантической близости документов. Применение предварительной обработки данных (*tf-idf*, НВ), ранга аппроксимирующей матрицы (ранг 40), методов определения λ документов (КМБ) позволяет получить значение $\Delta_{lsa,40} = 0.120$, которое незначительно превышает значение $\Delta_{mlsa} = 0.103$, полученное с помощью разработанного метода анализа динамики изменения сингулярных чисел матрицы «терм-документ» с автоматическим выбором диапазона используемых ранговых значений.

Оптимальный ранг аппроксимирующей матрицы найден с помощью графика значений Δ (рисунок 2). Значение Δ возможно получить только при заведомо известных дублированных и противоречивых парах документов. Кроме того, значение $\Delta_{lsa,1-60} = 0,086$, полученное при выборе всего рангового диапазона, менее показательны. Таким образом, разработанный метод позволяет

автоматически выбрать диапазон значимых рангов, уменьшить объем вычислений и сохранить точность метода ЛСА.

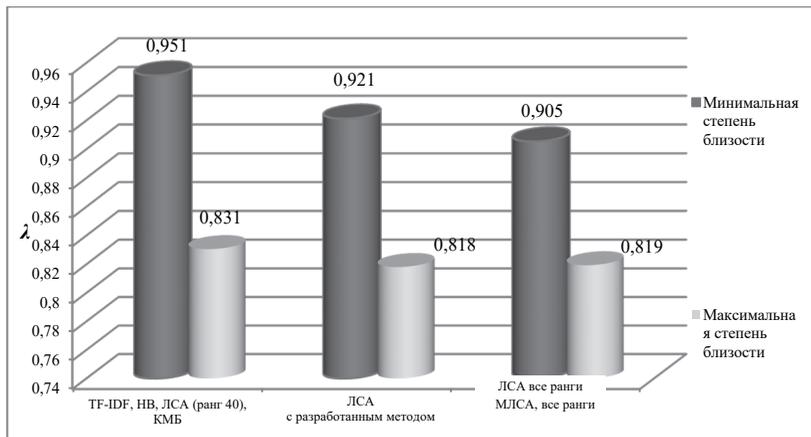


Рис. 6. Точность метода ЛСА с разработанным методом

Рассмотрим кривую распределения количества пар документов, превышающих определенное значение λ (рисунок 7).

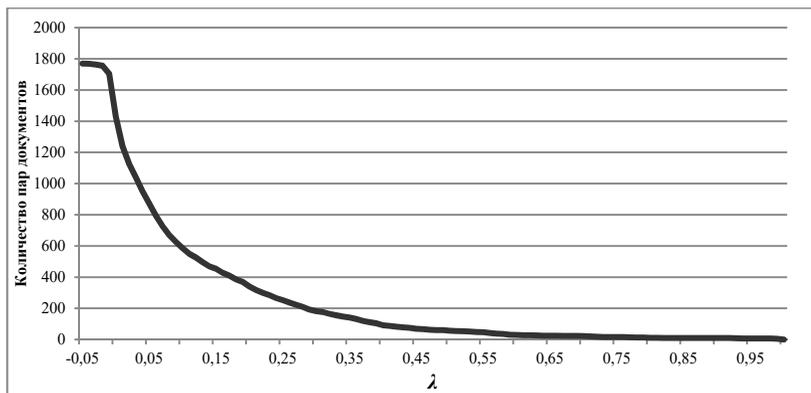


Рис. 7. Распределение пар документов по степени соответствия

Приведенная на рисунке 7 кривая имеет экспоненциальный вид, что дает возможность в автоматическом режиме значительно уменьшить область поиска дублирования и противоречий в БД: с 1770 пар документов до 10 пар документов при пороговом значении $\lambda_{\text{пор.}} = 0,85$. Найденные 10 пар документов являются дублированными или противоречивыми.

Большое значение величины $\Delta_{mlsa} = 0,103$ и экспоненциальный характер распределения позволяют выявить дублирования и противоречия в БД с меньшими временными затратами, так как обеспечивается максимальное значение разности между минимальным значением λ_{\min} пары документов, содержащих дублирования и противоречия, и максимальным значением λ_{\max} пары документов, не содержащих дублирования и противоречий.

5. Заключение. Проведенные эксперименты подтверждают, что на точность выявления дублирования и противоречий в БД на основе оценки семантической близости документов с помощью метода ЛСА существенное влияние оказывают следующие факторы: применение статистической меры tf-idf, нормирование векторов при формировании матрицы «терм-документ», определение оптимального ранга аппроксимирующей матрицы, применение косинусной меры близости.

Использование метода анализа динамики изменения сингулярных чисел матрицы «терм-документ» с автоматическим выбором диапазона используемых ранговых значений при латентно-семантическом анализе совместно с методами предварительной обработки данных позволяет обеспечить высокую точность поиска дублирования и противоречий в БД, а так же снизить временные затраты на их поиск и устранение. При проведении экспериментов использовалась небольшая выборка из 1770 документов. Вполне вероятно, при существенном изменении объема тестовой БД изменятся и рекомендуемые значения параметров. Решение этого вопроса требует отдельного исследования.

Литература

1. *Witten I.H., Frank E., Hall M.A.* Data Mining: Practical Machine Learning Tools and Techniques: 3rd edition // Morgan Kaufmann. 2011. 664 p.
2. *Паклин Н. Б., Орешков В. И.* Бизнес-аналитика: от данных к знаниям (+ CD) // СПб.: Изд. Питер. 2009. 624 с.
3. *Weissman S., Ayhan S., Bradley J., Lin J.* Identifying Duplicate and Contradictory Information in Wikipedia // Proceedings of the 15th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL '15). 2015. pp. 57–60.
4. *Йоцов В.С., Сузрев В.С., Юсупов Р.М., Хомоненко А.Д.* Онтологии для разрешения семантических конфликтов // Труды СПИИРАН. 2008. Вып. 7. С. 26–40.
5. *Ram S., Park J.* Semantic Conflict Resolution Ontology (SCROL): An Ontology for Detecting and Resolving Data- and Schema-Level Semantic Conflicts // IEEE Transactions on Knowledge and Data engineering. 2004. vol. 16. no. 2. pp. 189–202.
6. *Хомоненко А.Д., Войцеховский С.В., Логашев С.В., Дашонок В.Л.* Устранение семантических противоречий в elibrary.ru на основе нечеткого вывода // Проблемы информационной безопасности. Компьютерные системы. 2015. № 1. С. 24–33.

7. *Хомоненко А.Д., Логашев С.В., Краснов С.А.* Автоматическая рубрикация документов с помощью латентно-семантического анализа и алгоритма нечеткого вывода Мамдани // Труды СПИИРАН. 2016. № 1(44) С. 5–19.
8. *Lawrence R.* Automatic Conflict Resolution to Integrate Relational Schema // Ph.D. Thesis. 2001. 165 p.
9. *Galitsky B., Ilvovsky D., Kuznetsov S.O.* Style and Genre Classification by Means of Deep Textual Parsing // Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialogue 2016". 2016. pp. 171–181.
10. *Peng Z., Kambayashi Y.* Resolving Conflicts and Handling Replication during Integration of Multiple Databases by Object Deputy Model // Proceedings of the 20th International Conference on Conceptual Modeling: Conceptual Modeling. 2001. LNCS 2224. pp. 285–298.
11. *De Marneffe M.-C., Rafferty A. N., Manning C.D.* Finding Contradictions in Text // Proceedings of ACL-08: HLT. 2008. pp. 1039–1047.
12. *Воронцов К., Фрей А., Ромов П.* BigARTM: библиотека с открытым кодом для тематического моделирования больших текстовых коллекций // Аналитика и управление данными в областях с интенсивным использованием данных. 2015 С. 28–36.
13. *Бондарчук Д.В., Тимофеева Г.А.* Математические основы метода категориальных векторов в интеллектуальном анализе данных // Вестник Уральского государственного университета путей сообщения. 2015. № 4(28). С. 4–8.
14. *Gupta M., Bendersky M.* Information Retrieval with Verbose Queries // Foundations and Trends in Information Retrieval. 2015. vol. 9. no. 3–4. pp. 209–354.
15. *Landauer T., Foltz P., Laham D.* An introduction to Latent Semantic Analysis // Discourse processes. 1998. vol. 25. no. 2–3. pp. 259–284.
16. *Foltz P.W.* Using latent semantic indexing for information filtering // ACM Conference on Office Information Systems (COIS). 1990. pp. 40–47.
17. *Бубнов В.П. и др.* Модели информационных систем: учеб. пособие // М.: ФГБОУ «Учебно-методический центр по образованию на железнодорожном транспорте». 2015. 188 с.
18. *Dumais S.* Latent semantic indexing: TREC-3 report // Proc. of the Third Text REtrieval Conference. 1995. pp. 219–230.
19. *Соловьев А.Н.* Моделирование процессов понимания речи с использованием латентно-семантического анализа: диссертация на соискание степени к.ф.-м.н. // С.-Петербург. гос. ун-т. Санкт-Петербург. 2008.
20. *Хомоненко А.Д., Дашонок В.Л., Краснов С.А.* Выявление противоречий в семантически близкой информации на основе латентно-семантического анализа // Проблемы информационной безопасности. Компьютерные системы. 2014. № 2. С. 73–84.
21. *Goma W.H., Fahmy A.A.* A Survey of Text Similarity Approaches. International Journal of Computer Applications 2013. vol. 68. no. 13. pp. 13–18.
22. *Бермудес С.Х.Г., Керимова С.У.* О методе определения текстовой близости, основанном на семантических классах // Инженерный вестник Дона. 2016. № 4(43). URL: ivdon.ru/ru/magazine/archive/n4y2016/3832 (дата обращения: 01.08.2017).
23. *Kuznetsov S., Poelmans J.* Knowledge representation and processing with formal concept analysis // Wiley interdisciplinary views: Data mining and knowledge discovery. 2013. vol. 3. pp. 200–215.
24. *Jones K.S.* A statistical interpretation of term specificity and its application in retrieval // Journal of Documentation. 2004. vol. 60. no. 5. pp. 493–502.
25. *Manning C.D., Raghavan P., Schütze H.* An Introduction to Information Retrieval Draft // Online edition. Cambridge University Press. 2009. 544 p.

26. *Бондарчук Д.В.* Использование латентно-семантического анализа в задачах классификации текстов по эмоциональной окраске // Бюллетень результатов научных исследований. 2012. № 2(3). С. 146–152.
27. *Краснов С.А., Хомоненко А.Д., Яковлев Я.В.* Оценка эффективности применения алгоритма вычисления коэффициента ранговой корреляции Спирмена в методе латентно-семантического анализа при автоматической рубрикации документов // Бюллетень результатов научных исследований. 2012. № 2(3). С. 153–162.

Краснов Сергей Александрович — к-т техн. наук, старший преподаватель, Военно-космическая академия имени А.Ф. Можайского (ВКА им. А.Ф. Можайского). Область научных интересов: информационные технологии, защита информации, системы искусственного интеллекта. Число научных публикаций — 30. kras25@rambler.ru; ул. Ждановская 13, Санкт-Петербург, 197198; р.т.: +7(911)7346550.

Илатовский Анатолий Сергеевич — курсант, Военно-космическая академия имени А.Ф. Можайского (ВКА им. А.Ф. Можайского). Область научных интересов: информационные технологии, защита информации, системы искусственного интеллекта. Число научных публикаций — 2. letsbrainup@gmail.com; ул. Ждановская 13, Санкт-Петербург, 197198; р.т.: +7(967)968-20-63.

Хомоненко Анатолий Дмитриевич — д-р техн. наук, профессор, заведующий кафедрой информационных и вычислительных систем, ФГБОУ ВО Петербургский государственный университет путей сообщения Императора Александра I. Область научных интересов: численная теория массового обслуживания, программирование, операционные и информационные системы. Число научных публикаций — 150. khomon@mail.ru, <http://www.pgups.ru>; Московский пр., 9, Санкт-Петербург, 190031; р.т.: 8(812)457-80-23, Факс: 8(812)310-75-25.

Арсеньев Владимир Николаевич — д-р техн. наук, профессор, профессор кафедры бортовых информационных и измерительных комплексов, Военно-космическая академия имени А.Ф. Можайского (ВКА им. А.Ф. Можайского). Область научных интересов: методы анализа и синтеза сложных систем. Число научных публикаций — 100. vladar56@mail.ru; ул. Ждановская 13, Санкт-Петербург, 197198; р.т.: +79112620222.

S.A. KRASNOV, A.S. ILATOVSKY, A.D. KHOMONENKO, V.N. ARSENIYEV
**ASSESSMENT OF SEMANTIC SIMILARITY OF DOCUMENTS ON
 THE BASIS OF THE LATENT SEMANTIC ANALYSIS WITH THE
 AUTOMATIC CHOICE OF RANK VALUES**

Krasnov S.A., Ilatovsky A.S., Khomonenko A.D., Arseniev V.N. Assessment of Semantic Smilarity of Documents on the basis of the Latent Semantic Analysis with the Automatic Choice of Rank Values.

Annotation. The method of assessment of semantic similarity of documents, which is based on the use of the latent and semantic analysis, dynamics of change of singular values of a term-document matrix and automatic determination of a range of rank values, is offered. Assessment of semantic similarity of documents is considered in relation to the solution of problems of identification of duplication and contradictions in databases and storages of data.

A short review of the approaches used at assessment of semantic similarity of documents, identification of duplication and contradictions in databases is provided. Results of numerical examples of assessment of semantic dependences between terms of documents for the benefit of identification of duplication and contradictions in databases and storages of data are given. In this case, the degree of correspondence between the compared documents as the resultant characteristic is calculated.

Comparative estimates of the accuracy of the calculation of the degree of correspondence of λ documents with the help of the main methods (cosine proximity measure, vector model, Spearman rank correlation coefficient, static measure tf-idf — frequency of the term — reverse document frequency) are given.

It is shown that application of the offered method of the latent and semantic analysis with automatic detection of a range of rank values allows eliminating dependence of results of application of a method of the latent semantic analysis on the chosen rank.

Keywords: assessment of semantic similarity of documents, identification of duplications and contradictions, databases, latent semantic analysis, statistical analysis, cosine measure of proximity, vector model.

Krasnov Sergey Aleksandrovich — senior lecturer, Mozhaisky Military Space Academy. Research interests: information technology, information security, artificial intelligence systems. The number of publications — 30. kras25@rambler.ru; 13, Zhdanovskaya street, St.-Petersburg, 197198, Russia; office phone: +7(911)7346550.

Ilatovsky Anatoly Sergeyevich — cadet, Mozhaiskii Military Space Academy. Research interests: information technology, information security, system of artificial intelligence. The number of publications — 2. letsbrainup@gmail.com; 13, Zhdanovskaya street, St.-Petersburg, 197198, Russia; office phone: +7(967)968-20-63.

Khomonenko Anatoly Dmitrievich — Ph.D., Dr. Sci., professor, head of information and computing systems department, Emperor Alexander I St. Petersburg State Transport University. Research interests: queuing systems, artificial intelligence, databases. The number of publications — 150. khomon@mail.ru, <http://www.pgups.ru>; 9, Moskovsky pr., Saint Petersburg, 190031; office phone: 8(812)457-80-23, Fax: 8(812)310-75-25.

Arseniev Vladimir Nikolaevich — Ph.D., Dr. Sci., professor, professor of the pulpit on-board (side) information and measuring complex, Mozhaisky Military Space Academy. Research interests: modeling and testing of complex systems, control of aircraft. The number of publications — 100. vladar56@mail.ru; 13, Zhdanovskaya street, St.-Petersburg, 197198, Russia; office phone: +79112620222.

References

1. Witten I.H., Frank E., Hall M.A. Data Mining: Practical Machine Learning Tools and Techniques: 3rd edition. Morgan Kaufmann. 2011. 664 p.
2. Paklin N.B., Oreshkov V.I. *Biznes-analitika: ot dannyh k znaniyam (+ CD)* [Business analysis: from data to knowledge (+ CD)]. SPb: Izd. Peter. 2009. 624 p. (In Russ.).
3. Weissman S., Ayhan S., Bradley J., Lin J. Identifying Duplicate and Contradictory Information in Wikipedia. Proceedings of the 15th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL '15). 2015. pp. 57–60.
4. Jotsov V.S., Sgurev V.S., Yusupov R.M., Khomonenko A.D. [The Ontology for the Semantic Conflicts Resolution]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2008. vol. 7. pp. 26–40. (In Russ.).
5. Ram S., Park J. Semantic Conflict Resolution Ontology (SCROL): An Ontology for Detecting and Resolving Data- and Schema-Level Semantic Conflicts. *IEEE Transactions on Knowledge and Data engineering*. 2004. vol. 16. no. 2. pp. 189–202.
6. Khomonenko A.D., Voitsekhovskii S.V., Logashev S.V., Dashonok V.L. [Resolving Semantic Inconsistencies in ELIBRARY.RU Based on Fuzzy Inference]. *Problemy informacionnoy bezopasnosti. Komp'yuternye sistemy – Automatic Control and Computer Sciences*. 2015. vol. 1. pp. 24–33. (In Russ.).
7. Khomonenko A.D., Logashev S.V., Krasnov S.A. [Automatic categorization of documents using latent semantic analysis and the Mamdani fuzzy inference algorithm]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2016. vol. 1(44). pp. 5–19. (In Russ.).
8. Lawrence R. Automatic Conflict Resolution to Integrate Relational Schema. Ph.D. Thesis. 2001. 165 p.
9. Galitsky B., Ilvovsky D., Kuznetsov S.O. Style and Genre Classification by Means of Deep Textual Parsing, Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialogue 2016". 2016. pp. 171–181.
10. Peng Z., Kambayashi Y. Resolving Conflicts and Handling Replication during Integration of Multiple Databases by Object Deputy Model. Proceedings of the 20th International Conference on Conceptual Modeling: Conceptual Modeling. 2001. LNCS 2224. pp. 285–298.
11. De Marneffe M.C., Rafferty A. N., Manning C.D. Finding Contradictions in Text. Proceedings of ACL-08: HLT. Proceedings of ACL-08: HLT. 2008. pp. 1039–1047.
12. Vorontsov K., Frei A., Romov P. [BigARTM: open source library for thematic modeling of large text collections]. *Analitika i upravlenie dannyimi v oblastyah s intensivnym ispol'zovaniem dannyh – Analytics and data management in areas with intensive data use*. 2015. pp. 28–36. (In Russ.).
13. Bondarchuk D.V., Timofeeva G.A. [Mathematical foundations of the method of categorical vectors in the intellectual analysis of data]. *Vestnik Ural'skogo gosudarstvennogo universiteta putej soobshheniya – Newsletter Ural State University of Communications*. 2015. vol. 4(28). pp. 4–8. (In Russ.).
14. Gupta M., Bendersky M. Information Retrieval with Verbose Queries. Foundations and Trends in Information Retrieval. vol. 9. no. 3–4. 2015. pp. 209–354.
15. Landauer T., Foltz P., Laham D. An Introduction to Latent Semantic Analysis. *Discourse processes*. 1998. vol. 25. no. 2–3. pp. 259–284.
16. Foltz P.W. Using the latent semantic indexing for information filtering. ACM Conference on Office Information Systems (COIS). 1990. pp. 40–47.
17. Bubnov V.P. et al. *Modeli informacionnyh sistem* [Models of information systems]. Moscow: Izd.FGBOU «The Educational-methodical center on formation on a railway transportation». 2015. 188 p. (In Russ.).

18. Dumais S. Latent semantic indexing: TREC-3 report. Proc. Of the Third Text REtrieval Conference. 1995. pp. 219–230.
19. Soloviev A.N. *Modelirovanie processov ponimaniya rechi s ispol'zovaniem latentno-semanticeskogo analiza: dissertaciya na soiskanie stepeni k.f.-m.n.* [Modeling the processes of understanding speech using latent-semantic analysis. Ph.D. thesis. S.-Peterb. gos. un-t. Sankt-Peterburg. 2008. (In Russ.).
20. Khomonenko A.D., Dashonok V.L., Krasnov S.A. [Revealing of contradictions in semantically close information on the basis of latent-semantic analysis]. *Problemy informacionnoj bezopasnosti. Kompjuternye sistemy – Automatic Control and Computer Sciences*. 2014. vol. 2. pp. 73–84. (In Russ.).
21. Goma W.H., Fahmy A.A. A Survey of Text Similarity Approaches. *International Journal of Computer Applications*. 2013. vol. 68. no. 13. pp. 13–18.
22. Bermudes S.H.G., Kerimova S.U. [About the method of determination of text closeness based on semantic classes]. *Inzhenernyj vestnik Dona – Engineering journal of Don*. 2016. vol. 4(43). Available at: ivdon.ru/ru/magazine/archive/n4y2016/3832 (accessed: 01.08.2017). (In Russ.).
23. Kuznetsov S., Poelmans J. Knowledge representation and processing with formal concept analysis. *Wiley inter disciplinary views: Data mining and knowledge discovery*. 2013. vol. 3. pp. 200–215.
24. Jones K.S. A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*. 2004. vol. 60. no. 5. pp. 493–502.
25. Manning C.D., Raghavan P., Schütze H. *An Introduction to Information Retrieval Draft*. Cambridge University Press. 2009. 544 p.
26. Bondarchuk D.V. [The use of latent-semantic analysis in problems of classification of texts on emotional coloring]. *Bjulleten' rezul'tatov nauchnyh issledovanij – Bulletin of the results of scientific research*. 2012. vol. 2(3). pp. 146–152. (In Russ.).
27. Krasnov S.A., Khomonenko A.D., Yakovlev Ya.V. [Estimation of the effectiveness of the algorithm for calculating the Spearman's rank correlation coefficient in the method of latent-semantic analysis in the automatic classification of documents]. *Bjulleten' rezul'tatov nauchnyh issledovanij – Bulletin of the results of scientific research*. 2012. vol. 2(3). pp. 153–162. (In Russ.).

А.А. ВОЕВОДА, Д.О. РОМАННИКОВ
**СИНТЕЗ НЕЙРОННОЙ СЕТИ ДЛЯ РЕШЕНИЯ ЛОГИКО-
АРИФМЕТИЧЕСКИХ ЗАДАЧ**

Воевода А.А., Романников Д.О. Синтез нейронной сети для решения логико-арифметических задач.

Аннотация. Одна из основных проблем, стоящих перед разработчиком системы с нейронной сетью — выбор структуры нейронной сети, которая могла бы решать поставленные задачи. В настоящее время нет однозначных рекомендаций по выбору такой структуры и таких параметров, как: количество слоев, количество нейронов в слое, тип нелинейности нейрона, метод обучения, параметры метода обучения и другие.

В статье рассматривается подход к синтезу нейронной сети для класса логико-арифметических задач, основанный на формировании сети из предпостроенных элементарных функций. Новизна предлагаемого подхода заключается в формировании нейронной сети по известному алгоритму с использованием предварительно построенных функций. Таким образом, в статье изначально построены элементарные логико-арифметические функции, такие как «и», «или», «исключающее или», «и-не», «или-не», «≥», «≤», «>», «<», которые можно использовать для решения более сложных задач. Также приведен пример решения задачи построения функции по выбору максимального числа из четырех чисел, представленных в бинарном виде тремя разрядами. Синтез нейронной сети вышеприведенным способом выполняется с дальнейшей целью получения обобщенной структуры нейронной сети.

Ключевые слова: нейронные сети, машинное обучение, логико-арифметические задачи, синтез нейронной сети.

1. Введение. В настоящее время нейронные сети широко распространены в решении различных задач: распознавание изображений [1] и речи [2, 3], задачах классификации [4-7], кластеризации [4-6, 8], задачах управления [9] и других.

При решении вышеприведенных задач при помощи нейронных сетей возникает ряд проблем, среди которых можно отметить отсутствие однозначных рекомендаций по выбору структуры нейронной сети, избыточность (*redundancy*) [4-7], недообучение (*underfitting*) и переобучение (*overfitting*) моделей [5, 6], выбор начальных условий [6] и другие.

В частности, при рассмотрении классической задачи машинного обучения — XOR [4, 5] — видно, что избыточность нейронных сетей проявляется в том, что задача решается при различных значениях весовых коэффициентов и значениях смещения нейронов. В [10] приводится решение XOR задачи при помощи нейронной сети с одним скрытым нейронным слоем (рисунок 1).

В практически решаемых задачах, где число параметров может превышать десятки миллионов обучаемых параметров [1-3, 9, 11] и нет однозначных рекомендаций для выбора числа слоев, числа «фильтров», значений ядра свертки, избыточность может привести к тому, что система

«попадает» в локальный минимум или в лучшем случае значительно увеличивается время и затрачиваемые ресурсы на обучение сети.

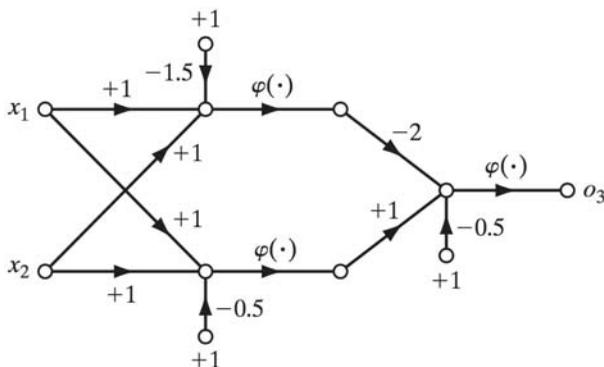


Рис. 1. Схема нейронной сети с тремя нейронами для решения XOR [17]

Также стоит отметить неоднозначность при выборе метода обучения, которая выражается в выборе самого метода обучения, значения параметра обучения (*learning rate*), количества эпох обучений (*epoch*), начальных условий параметров обучения (от их выбора зависит к какому из многих минимумов «сойдется» нейронная сеть при обучении).

Все вышеприведенные недостатки, заключающиеся в неоднозначности выбора параметров системы, на практике решаются при помощи проведения множества экспериментов и сравнении полученных результатов. Среди самых распространенных техник и методик можно выделить обнуление случайно выбранных весов при обучении (*dropout u dropConnect*) [12], ранний останов (*early stopping*) [13, 14], регуляризация [4, 6], кадрирование (*cropping*) и другие.

В отличие от используемого на практике подхода, основанного на выборе одного из известных вариантов структуры нейронной сети и дальнейшего обучения, в данной работе предлагается построение структуры нейронной сети исходя из анализа алгоритма решения рассматриваемой задачи. Иллюстрация предлагаемого подхода выполнена на примере логико-арифметической задачи, а именно рассмотрены примеры построения функций логических операций «и», «или», «не», «исключающего или» и функции выбора максимального числа из массива чисел. При решении данной задачи все входные и выходные данные представлены в виде чисел, где под «единицей» подразумевается

число из диапазона 0.9-1.1, под «нулем» — число из диапазона -0.1-0.1. Это условие необходимо для того, чтобы при прохождении через цепочку нейронов выходное значение выходных нейронов не смещалось.

В разделе 2 приводится реализация основных логико-арифметических функций при помощи нейронных сетей, которые затем могут быть использованы при синтезе нейронной сети для решения более сложных логико-арифметических задач. В частности, в разделе 3 приведен пример решения задачи выбора максимального числа из массива из четырех трехразрядных чисел, представленных в бинарном виде. В разделе 4 приведен сравнительный анализ полученной нейронной сети с нейронными сетями прямого распространения с четырьмя скрытыми слоями для различного числа нейронов в слое. Статья заканчивается разделом с выводами.

2. Реализация основных логико-арифметических функций.

Для решения логико-арифметических задач существует необходимость использовать такие базовые функции, как «и», «или», «исключающее или», «и-не», «или-не», « \geq », « \leq », « $>$ », « $<$ » (например, $x_1 > x_2$, если $x_1 : 1, x_2 : 0$).

Для того чтобы нейронные сети по вычислительным свойствам не вырождались (обусловленность — *condition number* — неудовлетворительная), весовые коэффициенты и смещения в среднем должны не слишком отличаться от единицы по модулю. Это следует из того, что при возрастании размеров системы с вычислительной точки зрения ее свойства существенно ухудшаются. Одним из возможных способов ограничения значений коэффициентов является изменение наклона (сжатие) нелинейной функции (измененный наклон нелинейной функции нейронной сети может негативно сказаться при обучении сети, но так как предлагаемый синтез нейронной сети основан на использовании уже пред-рассчитанных функций, то данным недостатком можно пренебречь). В статье в качестве базовой функции используется сигмоидальная (*sigmoid*) функция $f(x) = 1 / (1 + e^{(-x)})$ как наиболее распространенная в нейронных сетях, график которой приведен на рисунке 2. Далее для различных примеров она будет изменяться в соответствии со спецификой задачи. Также можно использовать и другие монотонно изменяющиеся функции.

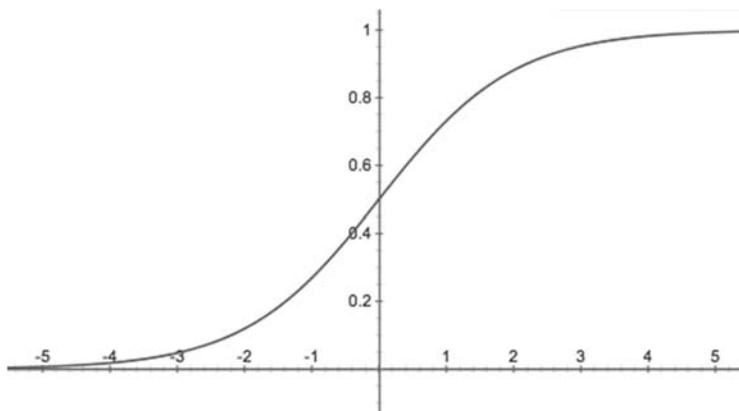


Рис. 2. График сигмоидальной нелинейной функции $f(x) = 1 / (1 + e^{-x})$

Рассмотрим процесс синтеза элементарных функций с использованием нейронных сетей на примере синтеза логической функции «или». Первым шагом необходимо составить таблицу значений входов и выходов, расширенную дополнительной колонкой с суммой значений x_1, x_2 , для рассматриваемой функции (таблица 1).

Таблица 1. Значения входов и выходов для функции «или»

x_1	x_2	y	Сумма входных значений
0	0	0	0
0	1	1	1
1	0	1	1
1	1	1	2

Основной идеей задачи реализации функции на основе нейронов является подбор такой нелинейности, которая подходит под заданный критерий точности. Например, при заявленной ошибке в 10% (далее будет использоваться абсолютное значение 0.1) для соответствия требуемому выходу, функция должна проходить через точки $x = 0.2, y = 0.1$; $x = 0.8, y = 0.9$. Эти значения обуславливаются тем, что в худшем случае значение «нуля» будет равно 0.2, а «единицы» — 0.8. Функция, проходящая через эти точки, представлена на рисунке 3.

Далее необходимо убедиться, что выбранная функция действительно удовлетворяет заданным критериям. Значения функции «или» представлены в таблице 2. В таблице приведена только часть значений входных сигналов с погрешностями, потому что остальные заведомо укладываются в значение заявленной ошибки.

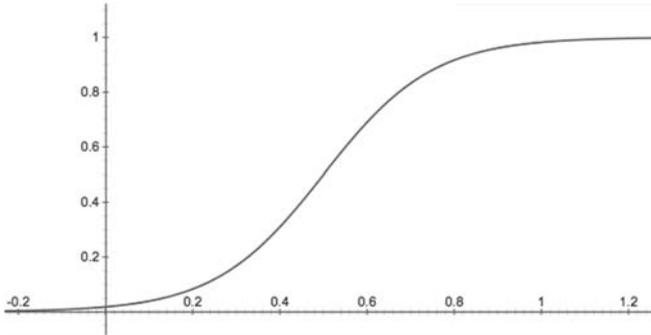


Рис. 3. График сигмоидальной нелинейной функции $f(x) = 1 / (1 + e^{-8*x+4})$

Таблица 2. Значения функции «или» при различных значениях входных сигналов

x1	x2	Сумма входных значений	y
0	0	0	0.01
0	1	1	0.98
1	0	1	0.98
1	1	2	0.99
0.1	0.1	0.2	0.08
-0.1	0.9	0.8	0.92
0.9	-0.1	0.8	0.92
0.9	0.9	1.8	0.99

Нейронная схема полученной функции приведена на рисунке 4.

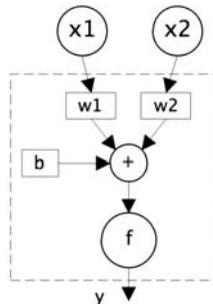


Рис. 4. Операция логического «или», реализованная при помощи одного нейрона

В полученной функции значение смещения будет b равно нулю, значение w_1 и w_2 равняются единицам.

Реализация функции «и» выполняется схожим образом. Изначально необходимо составить таблицу значений входов и выходов, расширенную дополнительной колонкой с суммой значений x_1 , x_2 для функции «и» (таблица 3).

Таблица 3. Значения входов и выходов для функции «и»

x1	x2	y	Сумма входных значений
0	0	0	0
0	1	0	1
1	0	0	1
1	1	1	2

Следующим шагом является получение точек, через которые должна проходить функция. Такими точками будут: $x=1.2, y=0.1$; $x=1.8, y=0.9$. Функция, проходящая через эти точки представлена на рисунке 5.

Результаты значений функции «и» представлены в таблице 4.

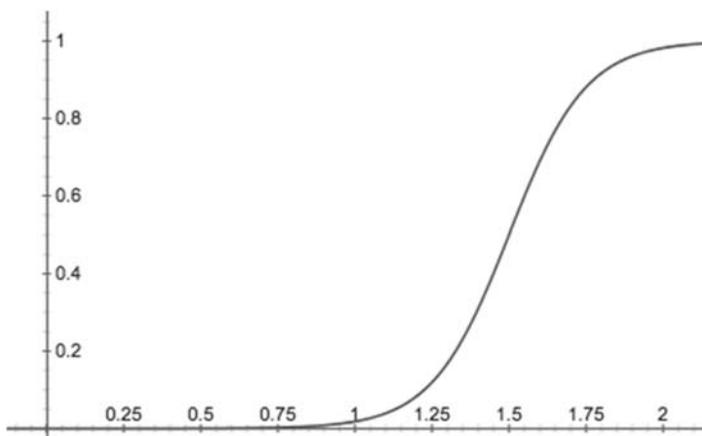


Рис. 5. График сигмоидальной нелинейной функции $f(x) = 1 / (1 + e^{-8*x+12})$

Таблица 4. Значения функции «и» при различных значениях входных сигналов

x1	x2	Сумма входных значений	y
0	0	0	0.01
0	1	1	0.02
1	0	1	0.02
1	1	2	0.98
0.1	0.1	0.2	0.01
0.1	1.1	1.2	0.08
1.1	0.1	1.2	0.08
0.9	0.9	1.8	0.91

Нейронная схема полученной функции также, как и структуры нейронных сетей для функций «и-не», «или-не» и функций сравнения, совпадает со схемой на рисунке 4. Все вышеприведенные функции могут быть реализованы при помощи нейронной сети, состоящей из одного нейрона. Это обусловливается тем, что разделение выходного множества результатов нейронной сети может быть достигнуто при помощи прямой.

Рассмотрим реализацию функции XOR. Согласно порядку синтеза вышеприведенных функций, необходимо составить таблицу значений входов и выходов, расширенную дополнительной колонкой с суммой значений x_1 , x_2 (таблица 5).

Таблица 5. Значения входов и выходов для функции «исключающее или»

x_1	x_2	y	Сумма входных значений
0	0	0	0
0	1	1	1
1	0	1	1
1	1	0	2

Исходя из таблицы 5 получим точки, через которые должна проходить функция нелинейности: $x=0.2$, $y=0.1$; $x=0.8$, $y=0.9$; $x=1.2$, $y=0.9$; $x=1.8$, $y=0.1$. Очевидно, что сигмоидальная функция не может пройти через все указанные точки. Тогда можно либо отказаться от сигмоидальной функции, либо составить результирующую функцию из двух сигмоидальных функций, то есть использовать схему из нескольких нейронов. Стоит отметить, что результирующая функция может быть представлена как сумма двух функций: 1) перевернутой функции на рисунке 3 ($f(x) = -1 / (1 + e^{-8*x+12})$); 2) функции на рисунке 3. Тогда итоговое уравнение полученной функции будет иметь вид: $f(x) = -1 / (1 + e^{-8*x+12}) + 1 / (1 + e^{-8*x+4})$. График ее представлен на рисунке 6.

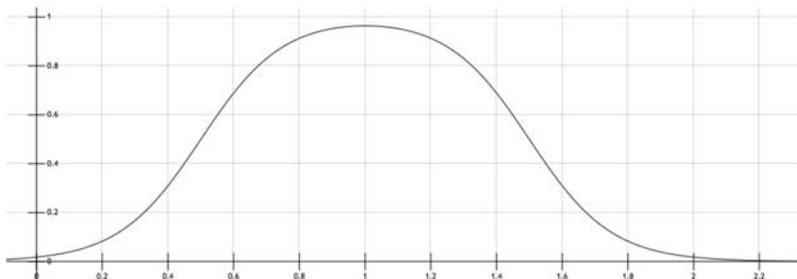


Рис. 6. График сигмоидальной нелинейной функции

$$f(x) = -1 / (1 + e^{-8*x+12}) + 1 / (1 + e^{-8*x+4})$$

Реализация вышеприведенной функции будет выполняться путем разделения данной функции на слагаемые, где каждая из частей будет включена в состав нейрона. После исполнения каждого из нейронов их результат будет суммироваться на третьем нейроне, где в качестве функции нелинейности будет константное значение единицы. Итоговая схема приведена на рисунке 7.

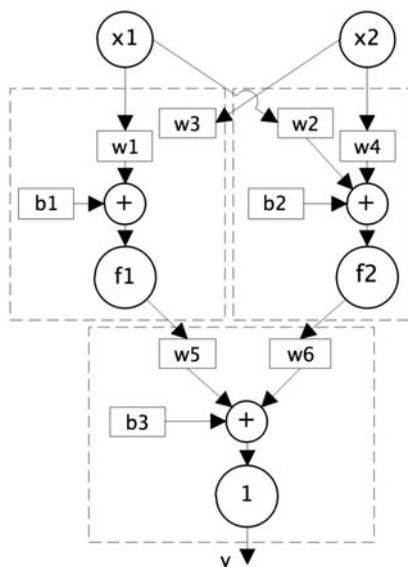


Рис. 7. Операция «исключающего или» (XOR), реализованная при помощи нейронной сети

Получившееся решение для функции XOR совпадает по структуре с решениями, приведенными в [4, 5], но различается по выбранной нелинейности и значениям весов.

Построение других функций выполняется аналогичным образом. Стоит также отметить, что изменение количества входов для разных функций приводит к разным результатам. Например, функции «или» с любым количеством входов может быть реализована с помощью одного нейрона. А для реализации функции XOR с тремя входами требуется большее количество нейронов. В вышеприведенных таблицах помимо колонки с суммой может быть выбрана другая функция, например, разность для реализации функции сравнения «>».

3. Пример. Рассмотрим решение логико-арифметических задач на примере реализации функции выбора максимального числа из трех чисел, представленных в бинарном виде четырьмя разрядами.

Идея, на которой базируется структура нейронной сети (рисунок 8) схожа с идеей, представленной в [15-17], и заключается в том, что числа сравниваются попарно, начиная со старших разрядов и далее к младшим. При сравнении разрядов возможны несколько вариантов: 1) значения разрядов равны; 2) значение разрядов не равны; 3) значения разрядов не сравниваются, если при сравнении более старших разрядов было выявлено, что они не равны. При реализации сравнения может возникнуть ситуация, когда по старшим разрядам можно закончить сравнение, но так как в нейронных сетях, в отличие от сетей Петри, отсутствует возможность задать порядок управления, то необходимо продолжать сравнение с учетом предыдущих результатов.

На рисунке 8 представлена схема сравнения старших разрядов (x_1 , x_2) двух чисел. Если значения старших разрядов x_1 и x_2 являются единицами, то на выходе *and* также будет единица (на рисунках 8, 9, 10, 11 прямоугольниками представлены функции, составленные при помощи нейронов), если оба значения равняются нулю, то на выходе функции *and_not* будет единица. Если значения разрядов не равны, то на выходе двух вышеприведенных функций будут нули. Выявление какой из сравниваемых разрядов больший выполняется функциями *cmp1*, *cmp2* (*compare*), которые на выходе выдают единицу, если $x_1 = 1$, $x_2 = 0$ и $x_1 = 0$, $x_2 = 1$ соответственно. В итоге в результате работы нейронной сети на рисунке 8 может быть единица либо на выходе eq_0 , либо на $x_1 > x_2_0$, либо на $x_1 < x_2_0$.

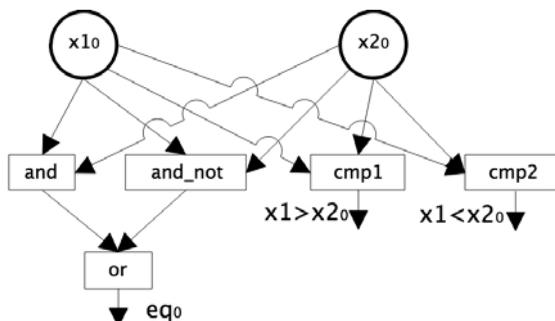


Рис. 8. Сравнение старших разрядов двух чисел

На рисунке 9 приведена схема сравнения последующих разрядов. В отличие от предыдущей схемы (рисунок 8), на этой схеме пунктирными линиями представлены связи нейронов с результатами сравнения предыдущих слоев.

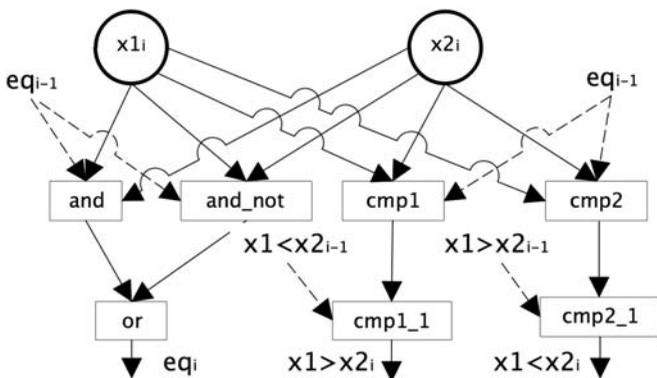


Рис. 9. Сравнение i -тых разрядов двух чисел

Функция *and* выдает единицу на выходе при двух единицах в x_{1i} и x_{2i} , а также если в результате прошлого сравнения $i-1$ -ые разряды были равны. Если eq_{i-1} – ноль (т.е. значения чисел не равны), то выходом функции *and* будет ноль, то есть числа не равны. Аналогично и для функции *and_not*. Функции *cmp1* и *cmp2* работают также, как и в предыдущем слое при условии, что в предыдущем сравнении разряды были равны (значение eq_{i-1} — единица). Выходное значение $x_1 > x_{2i}$ (функция *cmp1_1*) формируется с учетом значения выхода $x_1 < x_{2_{i-1}}$: если его значением является единица, то на выходе $x_1 > x_{2i}$ будет ноль, так как сравнение более старших разрядов приоритетно, если $x_1 < x_{2_{i-1}}$ является нулем, то выход функции *cmp1_1* определяется значением выхода функции *cmp1*. Аналогично работает и функция *cmp2_1*.

Итоговая схема сравнения всех разрядов двух чисел представлена на рисунке 10. Данная схема может быть легко расширена на произвольное количество разрядов путем добавления повторяющихся участков.

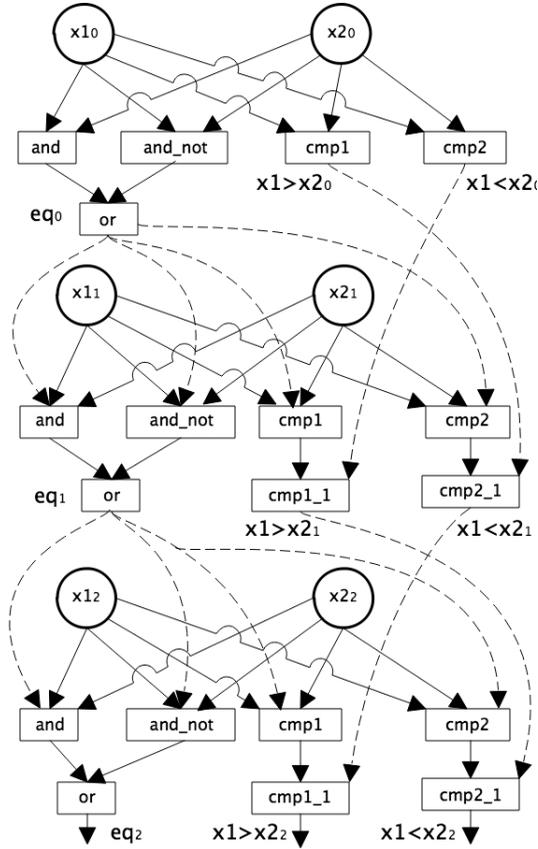


Рис. 10. Сравнение разрядов двух трехразрядных чисел

На рисунке 11 представлена часть нейронной сети, в которой выполняется формирование выходных разрядов двух сравниваемых чисел. Изначально формируются признаки сравнения двух чисел: eq — будет равной единице, если результат сравнения младших разрядов eq_2 будет равен единице (в противном случае — ноль). Выходы $x_1 > x_2$, $x_1 < x_2$ формируются в результате операций «или» для всех $x_1 > x_{2i}$ и $x_1 < x_{2i}$ соответственно. Только один из функций eq , $x_1 > x_2$, $x_1 < x_2$ может содержать значение единицы. Далее значения разрядов и признаков, что соответствующее число больше, поступают на функции «и» (отмечены как and). Для первого числа таким признаком является или выход eq ,

или $x_1 > x_2$, для второго только $x_1 < x_2$. После исполнения функций «и» значение разряда меньшего числа будет содержать ноль, а значение разряда старшего числа само значение разряда. Далее из выходов функций «и» формируется итоговые значения разрядов максимального числа.

Пример получения значений разрядов может быть расширен для произвольного количества разрядов путем добавления повторяющихся участков схемы выбора соответствующего разряда.

Результатом работы нейронной сети на рисунке 11 является максимальное число из двух сравниваемых чисел. Таким образом, для выбора максимального числа из массива чисел необходимо продолжить сравнение текущего максимального числа (т.е. результата нейронной сети на рисунке 11) с оставшимися числами массива. В итоге максимальное число будет выбрано в результате сравнения последнего числа и текущего максимального значения.

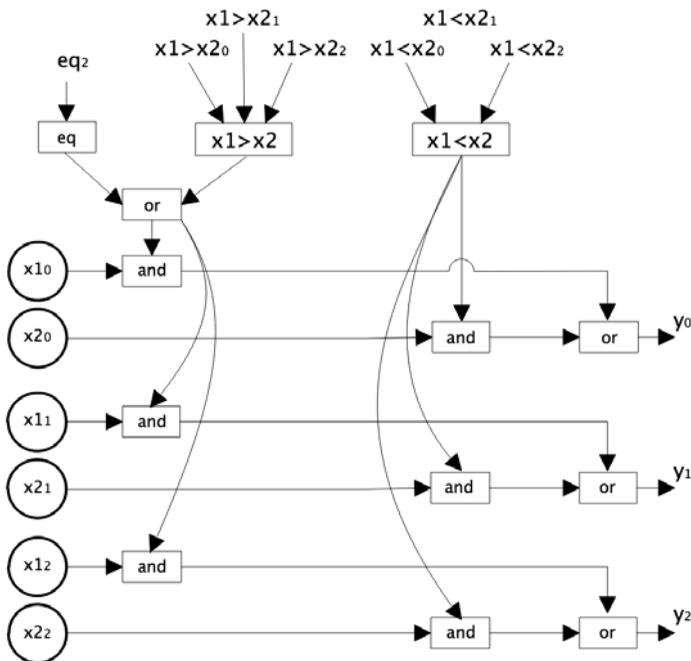


Рис. 11. Формирование разрядов числа

Общая итоговая схема представлена на рисунке 12. Каждая из секций сравнения, представленная на рисунке прямоугольником

comparison, включает в себя сравнение двух поступающих на вход схемы чисел, представленных в бинарном виде тремя разрядами. Результатом работы каждой секции сравнения будет максимальное из двух число, также представленное в бинарном виде (для чисел i_1 и i_2 максимальным будет max_{1-2}). Затем максимальное выбранное число сравнивается со следующим и далее происходит до тех пор, пока не будут сравнены все числа. Итоговое выбранное максимальное число на рисунке 12 обозначено как *max*.

Также стоит отметить, что приведенная схема похожа по своей работе на однопоточный выбор максимального числа в массиве чисел, который можно реализовать при помощи любого языка программирования. Но потенциально работу представленной схемы можно ускорить за счет распараллеливания выполняемых операций. При этом синхронизация работы параллельно выполняющихся операций будет выполняться на любом нейроне, который использует результаты вычислений предыдущих слоев.

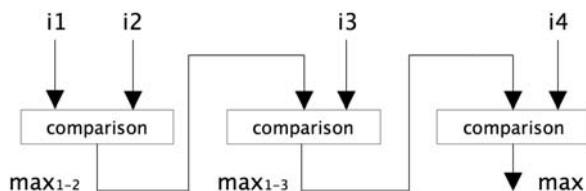


Рис. 12. Формирование i -ого разряда числа

4. Проанализируем рассмотренный в предыдущем разделе пример. Получившаяся система содержит 314 обучаемых параметров в отличие от нейронной сети, например, прямого распространения с четырьмя скрытыми слоями по 12 нейронов в каждом, которая содержит 676 обучаемых параметров. Нейроны обучаемой сети также используют сигмоидальную функцию активации. Обучение выполняется при помощи библиотек для работы с нейронными сетями *keras v2.0.2* с использованием *tensorflow v1.0.0* в качестве вычислительной среды. Начальные условия задаются равномерным распределением из диапазона $\{-1, 1\}$, непосредственно само обучение выполняется за 500 эпох с использованием метода среднеквадратического распространения ошибки (*RMSprop*) с шагами обучения 0.1, 0.01 и бинарной перекрестной энтропией (*binary crossentropy*) в качестве стоимостной функции.

Кроме того, рассматривался пример выбора максимального числа из пяти чисел, представленных в бинарном виде восьмью разрядами (вместо данного примера в работе приведен более компактный, но схожий по логике работы). Данная система была синтезирована аналогичным образом и содержит 1300 обучаемых параметров. В нейронной сети прямого распространения из четырех скрытых слоев по 40 нейронов содержится 6888 обучаемых параметров (остальные параметры обучения такие же, как и у сети выше).

Стоит отметить, что предлагаемый в статье способ построения нейронной сети не потребовал обучения, что связано с простотой схемы и незначительным числом входов каждого из элементов (таблица 4). Достоинством предлагаемого подхода является то, что начальные значения для обучаемых параметров, по крайней мере для простых схем, можно рассчитать, а для более сложных систем — они будут близки к оптимальным.

Отдельно стоит сравнить скорость обучения нейронной сети прямого распространения с 12 нейронами в скрытых слоях и нейронной сети, представленной на рисунках 8-12. Графики стоимостных функций (*loss function*) приведены на рисунках 13-16.

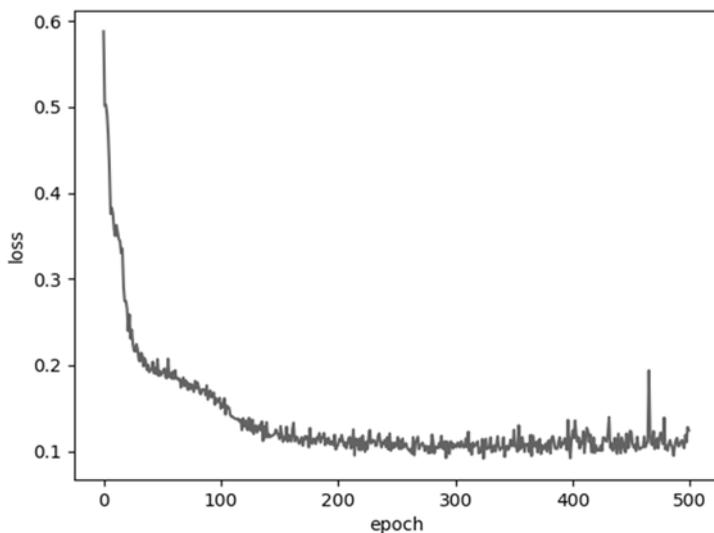


Рис. 13. График стоимостной функции нейронной сети прямого распространения с параметром обучения равным 0.1

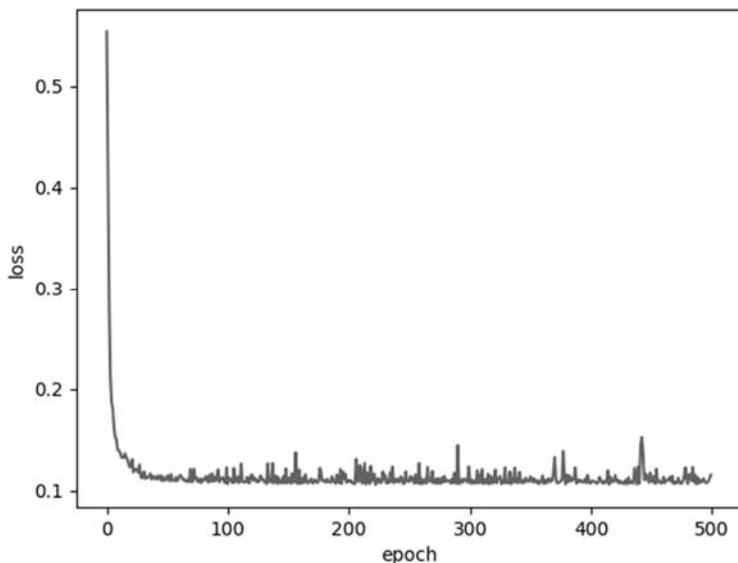


Рис. 14. График стоимостной функции нейронной сети (рисунки 8-12) с параметром обучения равным 0.1

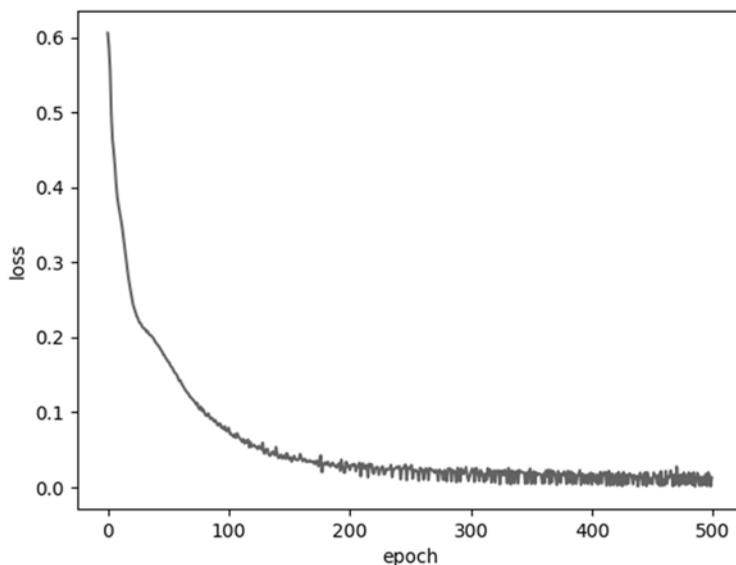


Рис. 15. График стоимостной функции нейронной сети прямого распространения с параметром обучения равным 0.01

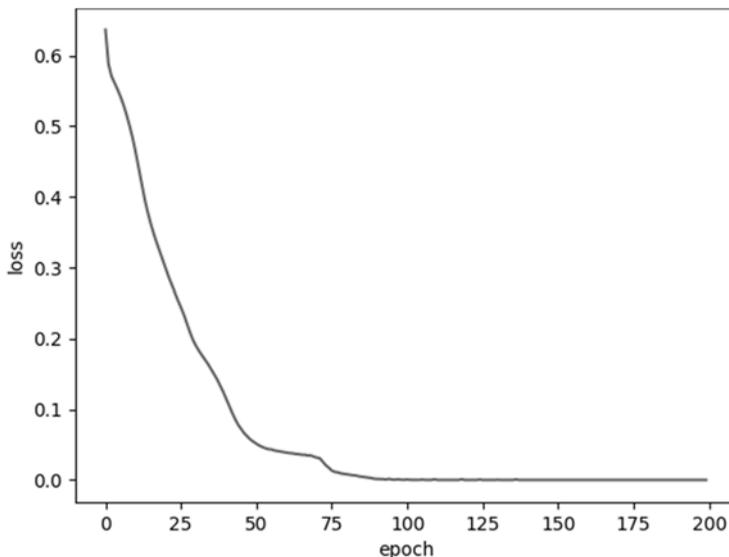


Рис. 16. График стоимостной функции нейронной сети (рисунки 8-12) с параметром обучения равным 0.01

Из приведенных графиков (рисунки 13-16) видно, что обучение выполняется быстрее в синтезированной нейронной сети, что может быть объяснено меньшим количеством параметров обучения.

На рисунке 14 стоимостная функция сходится быстрее и достигает асимптотического значения, но само значение больше, чем у сравниваемой функции на рисунке 13. Это может быть объяснено большим значением параметра обучения (0.1), так как при уменьшенном до 0.01 значение стоимостной функции практически достигает нуля за первые 100 эпох в отличие от рисунка 15, где значение нуля достигается после 400 эпох.

5. Заключение. В статье предложен способ синтеза нейронных сетей для логико-арифметических задач, который заключается в предварительном построении элементарных функций таких как: «и», «или», «и-не», «или-не», «≥», «≤», «>», «<» и других. На основе вышеприведенных функций может конструироваться нейронная сеть, что иллюстрируется на примере построения функции выбора максимального числа из трех чисел, представленных в бинарном виде четырьмя разрядами. Синтез нейронной сети вышеприведенным способом выполняется с дальнейшей целью получения обобщенной структуры нейронной сети. Кроме того, полученная нейронная сеть не потребовала обучения, так как она основывается на предварительно построенных функциях, которые уже содержат нужные значения коэффициентов. А для более сложных систем, начальные значения будут близки к оптимальным, что также положительно скажется на скорости обучения.

С другой стороны, если рассматривать только структуру нейронной сети (без значений весов), то ее обучение выполняется значительно быстрее, чем у сравниваемой нейронной сети прямого распространения.

Литература

1. *Krizhevsky A., Sutskever I., Hinton G.E.* ImageNet Classification with Deep Convolutional Neural Networks // Proceedings of Neural Information Processing Systems (NIPS). 2012. pp. 1090–1098.
2. *Graves A., Mohamed A., Hinton G.* Speech recognition with deep recurrent neural networks // Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2013. pp. 6645–6649.
3. *Deng L., Hinton G.E., Kingsbury B.* New types of deep neural network learning for speech recognition and related applications: An overview // Proceedings of IEEE International Conference on Acoustic Speech and Signal (ICASSP 2013). 2013. pp. 8599–8603.
4. *Haykin S.* Neural networks and learning machines: 3rd edition // Pearson Education. 2009. 938 p.
5. *Bishop C.* Pattern Recognition and Machine Learning // Springer. 2006. 738 p.
6. *Goodfellow I., Bengio Y., Courville A.* Deep Learning // MIT Press. 2016. 781 p.
7. *Haykin S., Deng C.* Classification of radar clutter using neural networks // IEEE Transactions on Neural Networks. 1991. vol. 2. pp. 589–600.
8. *Hastie T., Tibshirani R., Friedman J.* The Elements of Statistical Learning: Data Mining, Inference, and Prediction // Springer. 2001. 552 p.
9. *Hagan M., Demuth H., Jesús O.* A neural network predictive control system for paper mill wastewater treatment // Engineering Applications of Artificial Intelligence. 2003. vol. 16. no. 2. pp. 121–129.
10. *Touretzky D.S., Pomerleau D.A.* What is hidden in the hidden layers? // Byte. 1989. vol. 14(8). pp. 227–233.
11. *LeCun Y., Bengio Y., Hinton G.* Deep learning // Nature. 2015. vol. 521. no. 7553. pp. 436–444.
12. *Srivastava N. et al.* Dropout: A Simple Way to Prevent Neural Networks from Overfitting // Journal of Machine Learning Research. 2014. vol. 15(1). pp. 1929–1958.
13. *Mahsereci M., Balles L., Lassner C., Hennig P.* Early Stopping without a Validation Set // ArXiv e-prints. no.1703.09580. 2017.
14. *Prechelt L.* Early Stopping — But When? // Neural Networks: Tricks of the Trade. 2012. pp. 53–67.
15. *Воевода А.А., Романников Д.О.* Асинхронный алгоритм сортировки массива чисел с использованием ингибиторных сетей Петри // Труды СПИИРАН. 2016. Вып. 48. С. 198–213.
16. *Воевода А.А., Полубинский В.Л., Романников Д.О.* Сортировка массива целых чисел с использованием нейронной сети // Научный Вестник НГТУ. 2016. № 2(63). С. 151–157.
17. *Voevoda A.A., Romannikov D.O.* A Binary Array Asynchronous Sorting Algorithm with Using Petri Nets // Journal of Physics: Conference Series. 2017. vol. 803. no. 1. pp. 012178.

Воевода Александр Александрович — д-р техн. наук, профессор, профессор кафедры автоматизации, Новосибирский государственный технический университет (НГТУ). Область научных интересов: полиномиальный синтез, сети Петри, UML диаграммы. Число научных публикаций — 200. voevoda@ucit.ru; пр. Карла Маркса 20, Новосибирск, 630073; р.т.: +79139223092.

Романников Дмитрий Олегович — к-т техн. наук, доцент кафедры автоматизации, Новосибирский государственный технический университет (НГТУ). Область научных интересов: машинное обучение, нейронные сети, сети Петри. Число научных публикаций — 51. dmitry.romannikov@gmail.com; пр. Карла Маркса 20, Новосибирск, 630073; р.т.: +7 961 223 8567.

A.A. VOEVODA, D.O. ROMANNIKOV
**SYNTHESIS OF NEURAL NETWORK FOR SOLVING
 LOGIC-ARITHMETIC PROBLEMS**

Voevoda A.A., Romannikov D.O. Synthesis of Neural Network for Solving Logical-Arithmetic Problems.

Abstract. One of the main problems facing the developer of a system with a neural network is the choice of the structure of a neural network that could solve the tasks. At present there are no unambiguous recommendations for choosing such a structure and such parameters as: the number of layers, the number of neurons in the layer, the type of neuron nonlinearity, the training method, the parameters of the training method, and others.

The article considers an approach to the synthesis of a neural network for a class of logical-arithmetic problems, based on the formation of a network of pre-constructed elementary functions. The novelty of the proposed approach is the formation of a neural network using a well-known algorithm using pre-built functions. Thus, in the article elementary logical-arithmetic functions such as "and", "or", "exclusive or", "and-not", "or-not", " \geq ", " \leq ", " $>$ ", " $<$ " are built, which can be used to solve more complex problems. An example of a solution to the problem of constructing a function for selecting the maximum number of four numbers represented in binary form in three digits is given. Synthesis of a neural network in the manner described above is performed with the further goal of obtaining a generalized structure of the neural network.

Keywords: neural networks, machine learning, logical-arithmetic problems, synthesis of a neural network.

Voevoda Alexandr Aleksandrovich — Ph.D., Dr. Sci., professor, professor of automation department, Novosibirsk State Technical University (NSTU). Research interests: polynomial synthesis, UML diagrams, Petri nets. The number of publications — 200. voevoda@ucit.ru; 20, Karl Marx Avenue, Novosibirsk, 630073; office phone: +79139223092.

Romannikov Dmitry Olegovich — Ph.D., associate professor of automation department, Novosibirsk State Technical University (NSTU). Research interests: machine learning, neural networks, Petri nets. The number of publications — 51. dmitry.romannikov@gmail.com; 20, Karl Marx Avenue, Novosibirsk, 630073; office phone: +7 961 223 8567.

References

1. Krizhevsky A., Sutskever I., Hinton G.E. ImageNet Classification with Deep Convolutional Neural Networks. Proceedings of Neural Information Processing Systems (NIPS). 2012. pp. 1090–1098.
2. Graves A., Mohamed A., Hinton G. Speech recognition with deep recurrent neural networks. Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2013. pp. 6645–6649.
3. Deng L., Hinton G.E., Kingsbury B. New types of deep neural network learning for speech recognition and related applications: An overview. Proceedings of IEEE International Conference on Acoustic Speech and Signal (ICASSP 2013). 2013. pp. 8599–8603.
4. Haykin S. Neural networks and learning machines: 3rd edition. Pearson Education. 2009. 938 p.
5. Bishop C. Pattern Recognition and Machine Learning. Springer. 2006. 738 p.
6. Goodfellow I., Bengio Y., Courville A. Deep Learning. MIT Press. 2016. 781 p.

7. Haykin S., Deng C. Classification of radar clutter using neural networks. *IEEE Transactions on Neural Networks*. 1991. vol. 2. pp. 589–600.
8. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer. 2001. 552 p.
9. Hagan M., Demuth H., Jesús O. A neural network predictive control system for paper mill wastewater treatment. *Engineering Applications of Artificial Intelligence*. 2003. vol. 16. no. 2. pp. 121–129.
10. Touretzky D.S., Pomerleau D.A. What is hidden in the hidden layers? *Byte*. 1989. vol. 14(8). pp. 227–233.
11. LeCun Y., Bengio Y., Hinton G. Deep learning. *Nature*. 2015. vol. 521. no. 7553. pp. 436–444.
12. Srivastava N. et al. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*. 2014. vol. 15(1). pp. 1929–1958.
13. Mahseeci M., Balles L., Lassner C., Hennig P. Early Stopping without a Validation Set. ArXiv e-prints. no. 1703.09580. 2017.
14. Prechelt L. Early Stopping — But When? Neural Networks: Tricks of the Trade. 2012. pp. 53–67.
15. Voevoda A.A., Romannikov D.O. [Asynchronous Sorting Algorithm for Array of Numbers With the use of Inhibitory Petri Nets]. *Trudy SPIIRAN — SPIIRAS Proceedings*. 2014. vol. 3(34). pp. 218–232. (In Russ.).
16. Voevoda A.A., Polubinskij V. L., Romannikov D.O. [Array of integers sorting with a using of a neural network]. *Nauchnyj Vestnik NGTU — Science Bulletin of NSTU*. 2016. vol. 2(63). pp. 151–157. (In Russ.).
17. Voevoda A.A., Romannikov D.O. A Binary Array Asynchronous Sorting Algorithm with Using Petri Nets. *Journal of Physics: Conference Series*. 2017. vol. 803. no. 1. pp. 012178.

Н.А. БАЛОНИН, М.Б. СЕРГЕЕВ, В.С. СУЗДАЛЬ
**ДИНАМИЧЕСКИЕ ГЕНЕРАТОРЫ КВАЗИОРТОГОНАЛЬНЫХ
МАТРИЦ СЕМЕЙСТВА АДАМАРА**

Балонин Н.А., Сергеев М.Б., Суздаль В.С. **Динамические генераторы квазиортогональных матриц семейства Адамара.**

Аннотация. Исследуется задача построения нелинейных и линейных определенных в конечном поле генераторов квазиортогональных матриц семейства Адамара с малым количеством отличных между собой значений их элементов, не превосходящих по абсолютной величине 1, и глобальным или локальным значением детерминанта. Проанализированы свойства таких динамических систем, приведена классификация полученных с их помощью семейств матриц и их орнаментов, показан путь доказательства существования вещественных и целочисленных матриц, отличный от средств комбинаторного подхода. Значения, которым равны элементы матрицы, названы ее уровнями. Введены понятия адамаровой нормы и определителя квазиортогональной матрицы. Уровни, адамарова норма и определитель играют фундаментальную роль в определениях классов обобщенных матриц семейства Адамара. Выделены классы матриц Адамара, Белевича (конференц-матриц), Себерри (взвешенных матриц), Мерсенна, Эйлера, Одина (Зейделя), Ферма. Приведены формулы для значений их уровней. Орнаменты матриц Эйлера отвечают на вопрос максимальной сложности структуры матриц Адамара — бицикл с двойной каймой.

Ключевые слова: динамические генераторы, квазиортогональные матрицы, детерминированный хаос, конечные поля.

1. Введение Квазиортогональные матрицы — квадратные матрицы, ортогональные по строкам и столбцам с относительно простыми значениями элементов. К простейшим среди них относятся матрицы Адамара с элементами 1 и -1, применяемые в помехоустойчивом кодировании информации [1, 2]. Интерес к ним не проходит, за минувшее столетие вышло не поддающееся обзору количество научных работ.

Тем не менее и сейчас публикуются таблицы матриц [3], пробелы таблиц изучаются специальными исследованиями [4, 5, 6]. Обуславливается это тем, что орнамент (узор) бинарных по знакам элементов матриц усложняется с ростом порядка.

Поясним, почему это важно. Матрицы Адамара — это матрицы максимума детерминанта, чей размер кратен 4 и отвечает числу оснований генетического кода. На первой международной конференции (AIMEE2017), организованной ИМАШ РАН совместно с ассоциацией RAMECS (Китай), отмечалась [7] связь структурной организации алфавитов ДНК с формализмами теории помехоустойчивого кодирования информации (функциями Уолша и Радемахера).

Эта точка зрения, развиваемая в книгах и статьях С.В. Петухова по матрично-алгебраическому анализу системы генетического кодирования живых организмов [8], оказывается привлекательной для

написания совместных работ международным коллективом исследователей, например [9]. Мун Хо Ли, давно работающий с обобщениями матриц Адамара, предложил использовать жакетные матрицы — не ортогональные матрицы, инвертируемые посредством всего лишь инверсии их элементов и масштабирования. Эта черта роднит их с симметричными матрицами Адамара.

На конференции прозвучал также доклад Ю.И. Манина, ныне автора классических работ по теории чисел, искусственному интеллекту и кодированию (публикация готовится). В задачах про орнаменты целочисленное решение в виде матриц Адамара зависит от иррациональных по своей природе бициклов Эйлера. Занимаясь в шестидесятые годы знаменитой гипотезой Морделла, Ю.И. Манин так сформулировал такого рода соответствия [10]: «Взаимодействие алгебраической геометрии с теорией чисел привело к пониманию удивительного и фундаментального принципа: ответы на Диофантовы вопросы о системе уравнений критически зависят от геометрической формы пространства всех комплексных решений этой системы. Например, пространство всех комплексных решений может выглядеть (топологически) как сфера или тор, или сфера с несколькими ручками. Количество ручек называется родом, это очень устойчивый инвариант системы уравнений и, кажется, что он имеет мало общего с арифметическими тонкостями и дискретными точками решетки целочисленных векторов (в проективном пространстве различие между целыми и рациональными точками стирается).

Тем не менее род определяет, когда множество всех рациональных решений может быть бесконечным: только если ручек не больше одной».

В работах сотрудника института сквинтилляционных монокристаллов НАН Украины В.С. Суздаля исследуется связь квазикристаллов Шехтмана [11] с квазиортогональными матрицами золотого сечения [12]. Такие матричные многомерные модели, дополняющие плоские мозаики Пенроуза, рассматриваются впервые. Динамические алгоритмы их получения — новые, неизвестные ранее в кристаллографии модели кристаллизации [13], описывающие иррациональности, присутствующие квазикристаллам. Строение химической таблицы Менделеева, кристаллы, структурные особенности ДНК — все это перекликается с темой построения матриц Адамара и их обобщений.

Вместе с тем динамические системы финитного времени [14] отличаются собственные функции теплицева и ганкелева операторов, порождающих в ряде практически важных случаев ортогональные массивы, в том числе и интересующие нас матрицы. Эта точка зрения нова

и недостаточно освещена в научной литературе в силу долгого превалирования чисто комбинаторных методов исследования матриц Адамара. Итерационные процессы, ведущие к ортогональным матрицам, можно реализовать в конечных полях Галуа, но в этом случае приходится считаться с фактом существования поля. Поля не создать для составных размеров матриц, но, как показывает опыт, сохраняются присущие решениям задач в конечном поле явная или скрытая симметрии кодовых последовательностей [15-17].

Это наблюдение является предпосылкой для нахождения соответствующих орнаментов матриц на всех порядках матриц Адамара, выявляемых алгоритмами оптимизации детерминанта. По отношению к комбинаторным методам, оперирующим целочисленными матрицами, оптимизационный подход демонстрирует некоторые преимущества. Например, еще в античные времена было обнаружено, что диагональ прямоугольного с равными катетами треугольника не приблизить отношениями целых величин. Эта задача разрешима на более широком множестве чисел, порожденном изучением итерационных процессов и рядов. Квазиортогональные матрицы с их иррациональными элементами выступают как посредники в решении известной проблемы Адамара относительно существования матриц на всех порядках, кратных 4.

2. Предварительные сведения, определения и свойства. *Определение 1.* Квазиортогональной матрицей [15] будем называть квадратную матрицу A порядка n с приведенным к единице максимумом абсолютных значений ее элементов, удовлетворяющую квадратичному условию связи $A^T A = \omega(n)I$, где n — любое натуральное число, $\omega(n)$ — некоторая весовая функция, определяющая тип матрицы, а I — единичная матрица.

Квазиортогональными в широком смысле будем называть любые ортогональные по столбцам (и строкам) матрицы. В этом случае они включают в себя также ортогональные матрицы с весом $\omega = 1$ и максимальным по модулю элементом $t < 1$.

Определение 2. Значения, которым равны элементы матрицы, будем называть ее уровнями.

Например, матрица Адамара с элементами $\{1, -1\}$ имеет два уровня (двухуровневая), а матрица Белевича (конференц-матрица, взвешенная матрица) с элементами $\{0, 1, -1\}$ — трехуровневая [1]. Квазиортогональные матрицы семейства Адамара с элементами, не превышающими по модулю единицы, лежат на пересечении M двух классов (рисунок 1): D -матриц абсолютного максимума детерминанта (не ортогональных по столбцам) и квазиортогональных W -матриц (безотносительных к оптимуму детерминанта) с некоторым (желательно небольшим) количеством уровней.

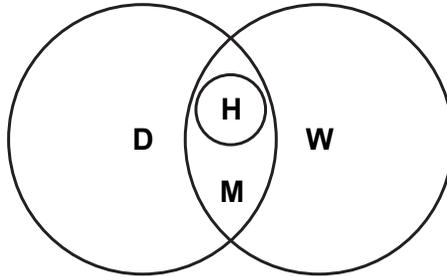


Рис. 1. Диаграмма Венна пересечения M множеств D -матриц и W -матриц, отражающая подмножество H матриц Адамара

Теорема. Утверждается, что $|\det(A)| \leq n^{n/2}$ для всех матриц A с элементами, не превышающими по модулю единицы [1].

Теорема вводит в рассмотрение нестрогое *неравенство Адамара*, равенство достижимо только на матрицах Адамара H с их экстремально малым количеством уровней (два) при экстремально большом детерминанте. Классические матрицы Адамара образуют на пересечении D -матриц и W -матриц подмножество H всех матриц с единичными по модулю элементами. Возможность обобщения матриц Адамара базируется на том, что матрицы класса H существуют не для всех значений порядков n . Соответственно, можно сформулировать задачу поиска квазиортогональных матриц с максимально достижимым значением детерминанта, то есть матриц M , к которым принадлежит, например, часть матриц Белевича.

Вес $\omega = 1$ характерен для ортогональных матриц, к которым квазиортогональные матрицы, в частности матрицы Адамара, помимо тривиальной матрицы первого порядка, не относятся. Вместе с тем это матрицы весьма близкие к ортогональным, получаемым из A элементарным нормированием их столбцов, после чего максимальный по модулю элемент (m -норма) уменьшается до $m < 1$ для порядков, больших 1.

Определение 3. Минимаксными квазиортогональными M -матрицами [15, 17] в строгом смысле будем называть матрицы, обладающие минимумом m -нормы на классе квазиортогональных матриц порядка n .

Несложно заметить, что $|\det(A)| = \omega^{n/2}$, причем $\omega = 1/m^2$.

Матрица Адамара H , обладающая максимумом модуля детерминанта, имеет минимальное значение $m = 1/\sqrt{n}$, то есть является частным случаем M -матриц с весом $\omega = n$. Отсутствие оптимального решения на границе неравенства Адамара искусственное, поскольку возможен поиск достижимого максимума модуля детерминанта.

Определение 4. Минимаксными квазиортогональными М-матрицами [15, 17] в общем смысле будем называть матрицы, обладающие глобальным или локальным минимумом m -нормы на классе квазиортогональных матриц порядка n .

Такие М-матрицы более широко представлены, они обладают глобальным или локальным максимумами модуля детерминанта.

Определение 5. Адамаровой нормой (h -нормой, взвешенной m -нормой) матрицы будем называть показатель $h = m\sqrt{n}$.

Адамарова норма отражает близость матрицы к матрице Адамара, у которой она имеет минимально возможное значение $h = 1$. Модуль детерминанта квазиортогональной матрицы $|\det(A)| = 1/m^n = n^{n/2}/h^n$ включает в себя правую часть неравенства Адамара $n^{n/2}$, пониженную до достижимой величины делением на n -ю степень h -нормы.

Определение 6. Приведенным определителем или адамаровым определителем матрицы будем называть величину $D = 1/h^2 = 1/(nm^2)$.

Вычислить адамаров определитель несложно, он удобен тем, что выступает множителем в выражении $|\det(A)| = (Dn)^{n/2}$, $D \leq 1$. Уровни, адамаровы норма и определитель играют фундаментальную роль в определениях классов обобщенных матриц семейства Адамара [3, 16]. Для всех матриц Адамара $D = 1$. Для остальных малоуровневых матриц семейство определяется не константой, а значением функций уровней и вытекающих из них зависимостей $D = D(n)$ или $h = h(n)$.

3. Орнаменты ортогональных матриц. Циклические и бициклические матрицы с малым количеством уровней объединяют в себе черты орнаментов, обладающих фиксированным узором, и матриц, обладающих детерминантом. Семейство орнаментов с *двумя значениями* элементов $a, -b$ описывается тремя инвариантами $\{n, k, \lambda\}$, где n — порядок матрицы, характеризующий величину узора, k — количество одинаковых элементов каждой строки и столбца, λ — количество одинаковых элементов, имеющих одну и ту же позицию в каждой паре строк или столбцов. Для квазиортогональных матриц с элементом $a = 1$ модуль второго элемента $0 \leq b \leq 1$. В том случае, когда ортогональность строк и столбцов недостижима назначением двух элементов, она достигается введением особого уровня для диагонали $0 \leq d \leq b$ или каймы из элементов первой строки и столбца матрицы с уровнем $b \leq s \leq 1$. В обоих случаях это требуется для матриц, превосходящих размер матриц Адамара на 1. Выбор наименований орнаментов в работе [17] согласован

с порядками, образующими числовые последовательности, на которых они наблюдаются. На порядках, равных числам Мерсенна $2^k - 1$ (здесь k — натуральное число), существуют двухуровневые циклические матрицы, называемые матрицами Мерсенна. Трехуровневые матрицы Ферма существуют на порядках, равных числам Ферма $2^{2^k} + 1$.

Принцип вложения. Адамар первым сформулировал принцип вложения, согласно которому уровневые (не орнаментальные) свойства матриц базовой последовательности, в его случае — последовательности Сильвестра 2^k , распространяются на последовательность $4t$, в которую они вложены. Соответственно, матрицы Мерсенна существуют вне базовой последовательности на порядках $4t - 1$, в которую числа $2^k - 1$ вложены. Что касается чисел Ферма, то они вложены не в одну, а в несколько числовых последовательностей, среди них $2^k + 1$ (здесь k — четное целое число), поэтому на $4t + 1$ уровневые свойства этих матриц не соблюдаются.

На одном и том же порядке могут сосуществовать несколько видов орнамента, описываемых одним и тем же набором орнаментальных инвариантов. На рисунке 2 приведены две матрицы Мерсенна [15, 17] порядка 15 с инвариантами $\{15, 7, 3\}$, каждая имеет по 7 одинаковых клеток в каждой строке и столбце и по 3 — в каждой паре строк или столбцов. Два цвета клеток соответствуют двум уровням a и $-b$.

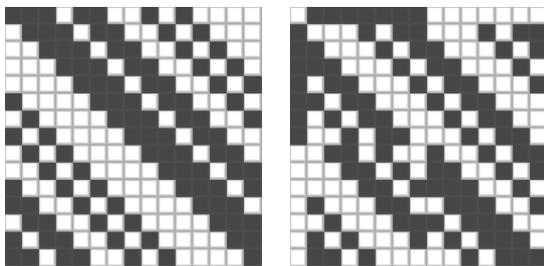


Рис. 2. Два $\{15, 7, 3\}$ -орнамента матриц Мерсенна

Первый циклический орнамент характерен для порядков, равных числам Мерсенна, второй более универсальный бициклический орнамент разрешим для всех порядков $4t - 1$, в которые числа Мерсенна вложены. Таким образом, орнаменты сопровождают числовые последовательности и имеют с ними взаимно-однозначное соответствие. Это положение звучит более широко, чем гипотеза Адамара, поскольку последняя связывает порядки и не говорит о структурах. Как будет показано

далее, орнамент матриц Мерсенна с иррациональными уровнями с точностью до знака соответствует орнаментам целочисленных матриц Адамара в их нормализованной форме с каймой из 1.

Поскольку на поиск вещественных матриц Мерсенна нет ограничений, выдвигаемых комбинаторными методами, это путь доказательства гипотезы Адамара косвенно через рассматриваемый далее оптимизационный подход.

Возможные сочетания инвариантов $\{n, k, \lambda\}$ для матриц ограничивает их порядок. Реализуемые параметры связывает выделенное еще исследованиями матриц Адамара [1] квадратичное *диофантово уравнение* I вида $k(k-1) = \lambda(n-1)$. Второе столь же общее матричное квадратичное уравнение $A^T A = \omega I$ отражает условие ортогональности столбцов и строк квадратной матрицы.

Для матриц с двумя элементами $a, -b$ (матрица с положительными элементами не может быть ортогональной) скалярное произведение любых двух отличающихся индексами строк содержит λ произведений вида a^2 , $2(k-\lambda)$ произведений ab ($k-\lambda$ элементов a каждой из строк умножено на b) и $n-2k+\lambda$ произведений b^2 . Отсюда следует квадратичное *характеристическое уравнение* II ортогонального дизайна,

$$(n-2k+\lambda)b^2 - 2(k-\lambda)ab + \lambda a^2 = 0, \quad (1)$$

записанное в виде, удобном для поиска корней при превалировании количества положительных элементов над отрицательными. Для матриц с целыми элементами $a, -b$ оно дает квадратичное *диофантово уравнение* II.

4. Оптимизационный подход к поиску квазиортогональных матриц. Итерационная схема реализации оптимизационного подхода к поиску оптимальных и субоптимальных по детерминанту квазиортогональных матриц [17] показана на рисунке 3 в виде динамического генератора.

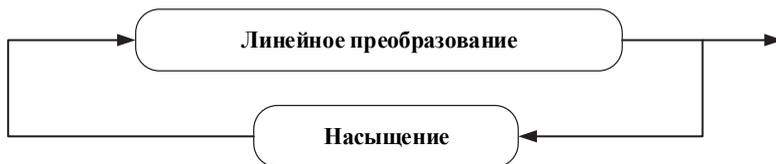


Рис. 3. Динамический генератор малоуровневых матриц

Эта схема напоминает классическую схему Айзермана, служащую для анализа аттракторов, возникающих в замкнутой нелинейной

элементом линейной системе посредством составления параметрически зависимых от амплитуды осцилляций передаточных функций. Состояние и выход дискретной линейной динамической системы в нашем случае описывается квадратной матрицей $Y = [y_1, y_2, \dots, y_n]$, получаемой из матрицы входа $U = [u_1, u_2, \dots, u_n]$ цепочкой ортогональных преобразований $y_1 = Q_1 u_1, y_2 = Q_2 u_2, \dots, y_n = Q_n u_n$, для поворота векторов входа до ближайшей к ним ортогональной конфигурации матрицы выхода. Содержание этих операций может быть разнообразным, взятым из практики вычислительных методов линейной алгебры: QR-преобразование, преобразование Грама-Шмидта, сингулярное разложение матрицы и обратный синтез с округленными до целых значений 1 и -1 собственными числами.

Цель динамической прямой связи — получить после нормирования амплитуд Y к 1 квазиортогональную матрицу A . Нормирование выполняет нелинейный элемент цепи обратной связи, дополняя ее насыщением $U = f(A)$ так, чтобы элементы U не превосходили по своей абсолютной величине некоторого порога $p \leq 1$: $u_{ij} = a_{ij}$ для $|a_{ij}| \leq p$, $u_{ij} = \text{sign}(a_{ij}) p$, для $a_{ij} > p, i, j \in \{1, n\}$.

Поскольку детерминант матрицы $|\det(A)| = 1/m^n$ обратно пропорционален значению степени от m -нормы, которую мы уменьшаем насыщением максимальных элементов, аттрактором динамического процесса является минимаксная матрица, принимающая характерную для нее малоуровневую конфигурацию. Этот процесс отражен на рисунке 4 при помощи трех гистограмм абсолютных значений элементов матрицы пятого порядка A_5 . Модули элементов в этом случае представлены в виде амплитуд.

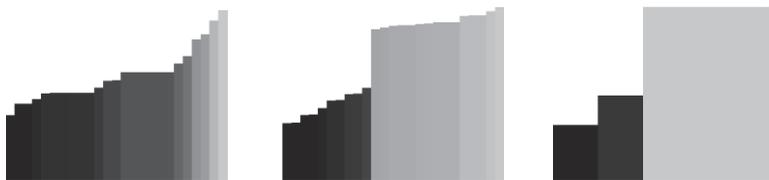


Рис. 4. Гистограммы амплитуд элементов матрицы A_5 трех стадий процесса

На рисунке 5 приведены орнаменты и гистограммы значений амплитуд их элементов квазиортогональных матриц абсолютного максимума детерминанта нечетных порядков, значения элементов матриц (уровни) отображены оттенками серого.

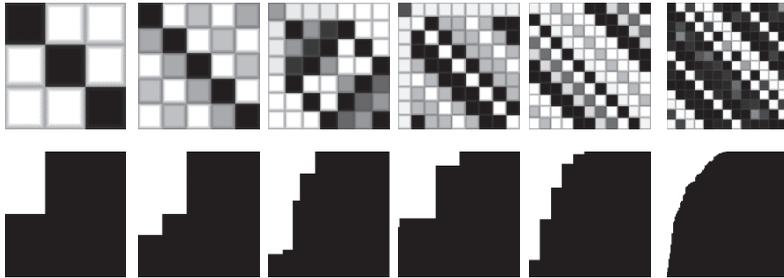


Рис. 5. Портреты матриц $A_3, A_5, A_7, A_9, A_{11}, A_{13}$ с гистограммами

Для повышения эффективности схождения итераций к матрицам с заданным орнаментом помимо применяемых в практике вычислительных методов перестановок столбцов итерируемой матрицы на вход динамической системы на старте подается импульс в виде циклической, бициклической и других подобных матриц, после чего порог p плавно повышается от стартового значения к единице.

5. Критские матрицы. Качественный анализ аттракторов модифицированной схемы Айзермана (рисунок 3) позволяет выделить детали, роднящие ее (ввиду квадратичного характера уравнения связи $A^T A = \omega(n)I$ с математическими моделями детерминированного хаоса. В задачах с матрицами четных порядков $4t$ или $4t - 2$ наблюдается постепенный выход динамического процесса на аттрактор с элементами 1 и -1, в последнем случае с 0 на диагонали после сортировки элементов.

Цель комбинаторных методов [1] — достичь равенства $A^T A = \omega(n)I$ последовательными изменениями знаков элементов с использованием переборных процедур. Оптимум детерминанта гарантирован свойствами матриц соответствующей размерности. В рассматриваемой схеме, наоборот, оптимизация детерминанта выступает в качестве основной мотивации ее построения, результатом ее работы являются матрицы с вещественными элементами.

В том случае, когда структура разрешима целочисленным решением, округление дает матрицы Адамара и Белевича, ранее получаемые совершенно иным путем. Количество уровней оптимальных по детерминанту матриц нечетного порядка с ростом размера растет почти ли-

нейно (с отклонением ± 1). Есть и критическая точка, порядок 13, на котором экстремуму детерминанта, как видно из рисунка 5, соответствует дисперсное состояние уровней. В ранних работах этот эффект был отнесен к влиянию ограничений разрядной сетки компьютера, не позволяющей вычислить точный результат. Позже [17] было высказано подтвердившееся предположение, что как и у аттракторов Лоренца, дисперсное состояние уровней аттрактора обобщенной схемы Айзермана — свойство самой задачи, а не ошибка вычислительного инструмента. Строго оптимальной матрицы A_{13} с семью уровнями (± 1) попросту нет, а если и есть, то она не оптимальна по детерминанту. Однако субоптимальная матрица была найдена.

Показатель D удобен для контроля качества вычислений и выработки критерия останова итераций в схеме 2. Значение адамарова детерминанта $D_{13} = 1/h^2$ равно примерно 0,8 (так же, как и константу Фейгенбаума, это критическое значение можно уточнять). Аттракторы оптимальных матриц порядков $n > 13$ с $D \geq 0,8$ отличаются дисперсным характером элементов. Преодоление указанной границы служит важным критерием выхода итерационного процесса на малоуровневую структуру.

Побочным, а для нас главным результатом работы оптимизатора детерминанта являются матрицы локального максимума детерминанта, отличающиеся от прочих малым количеством уровней. Детерминант D этих упорядоченных структур с ростом порядка стремится к 1. Такие малоуровневые структуры в [17] для краткости названы *критскими матрицами*, поскольку анонсированы были на Крите (2014 г.). В настоящее время это название широко используется как в отечественных, так и зарубежных публикациях.

6. Систематизация критских матриц. Малоуровневая матрица локального максимума детерминанта отличается от прочих матриц тем, что любая ограниченная вариация ее параметров, не выводящая их за пределы области разрешенных значений (амплитуды элементов ≤ 1), понижает значение модуля ее детерминанта. Кроме того, на порядках $4t - 3$, представимых суммой квадратов двух целых чисел 1, 5, 9, 13, ... и т. п. (в первой сотне целых чисел исключениями являются 21, 33, 57, 69, 77, 93) встречаются трехуровневые матрицы Одина с выделенной уровнем диагональю, которые отвечают седловым точкам этой задачи. Со сделанной оговоркой мы отнесем их к критским матрицам ввиду того, что они дополняют двухуровневые матрицы Мерсенна порядков $4t - 1$, теория которых без них будет не полна. Матрицы Одина не только прерываются на некоторых порядках, но и являются более сложными, чем матрицы Мер-

сенна, поскольку не имеют ограничений на структуру. Для них нет понятия универсальной структуры, в рамках которой они могут быть найдены. Для порядка 45 орнамент включает неординарные блоки, для порядков 65 и 85 простых орнаментов заведомо нет, сложные орнаменты не найдены. Теория этих матриц в стадии развития.

Заметно проще бициклические матрицы Эйлера порядков $4t - 2$, которые, в отличие от матриц Белевича, существуют для любого отведенного им порядка. В таблице 1 приведены семейства выделенных нами критских матриц с двумя и тремя уровнями.

Таблица 1. Значения уровней семейств критских матриц

Символ	Порядок n	Матрица	Значения элементов
H	$4t$	Адамара	1, -1
C	$2t, 4t$	Белевича	1, -1, 0
W	$t, 2t, 3t, 4t$	Себерри (взвешенная)	1, -1, 0
M	$4t - 1$	Мерсенна	$b = \frac{t}{t + \sqrt{t}}$ 1, -b, где
E	$4t - 2$	Эйлера	$b = \frac{t}{t + \sqrt{2t}}$ 1, -b, где
S	$4t - 3$ (если разложимо на сумму 2 квадратов)	Одина (Зейделя)	1, -b, d, где $b = 1 - 2d$, $d = \frac{1}{1 + \sqrt{n}}$
F	$4t + 1$ (если представимо числами Ферма или последовательностям и, в которые они вложены)	Ферма	1, -b, s, где $q = n - 1 = 4u^2$, $p = q + \sqrt{q}$, $b = \frac{2n - p}{p} = 1 - \frac{2u - 1}{2u + 1} \times \frac{1}{u}$, $s = \frac{\sqrt{nq} - 2\sqrt{q}}{p} = \frac{\sqrt{nu} - 1}{2u + 1} \times \frac{1}{\sqrt{u}}$

Как отмечалось ранее, для уровней диагональных элементов и элементов каймы в виде первых строки и столбца (за исключением первого, как правило, единичного элемента) сделаем исключение, помечая их персональными буквами d и s .

7. Динамические системы в конечных полях Галуа. Эффективным средством расчета критских матриц высоких порядков явля-

ется моделирование их структуры с помощью анализа поведения линейной динамической системы первого порядка в конечном поле $\text{GF}(p^m)$. Роль ограничения в виде обратной связи расширенной схемы Айзермана, придающей необходимые свойства аттрактору, играет ограниченный размер поля.

Реакция линейной динамической системы в конечном поле качественно отличается от импульсной весовой характеристики апериодического звена в виде экспоненты. Число различных между собой значений показательной функции g^k , где g — ненулевой элемент поля Галуа, ограничено $p^m - 1$. В отличие от сдвигового регистра, которым ищут иногда оригинальные матрицы порядков чисел Мерсенна в поле $\text{GF}(2)$ [17], диапазон возможных значений элементов g не столь узок.

В поле $\text{GF}(p^m)$ можно рассматривать g^k как адреса негативных рациональных или иррациональных элементов в последовательности чисел, формирующих первую строку циклической матрицы. В поле $\text{GF}(p^m)$ значения показательной функции векторные, однако смысловую нагрузку адресов они не теряют, если вместо g^k рассматривать порядковый номер элемента. Кроме того, всегда имеется возможность рассматривать не одну, а две динамические системы первого порядка и фиксировать моменты пересечения значений этих векторных последовательностей, трактуя скалярные значения моментов пересечений как адреса.

В качестве примера поля Галуа $\text{GF}(p)$ возьмем набор целых чисел $0, 1, \dots, p-1$. Мультипликативная группа поля Галуа, образуемая операцией умножения по модулю p , обозначим ее как $\text{GF}^*(p)$, состоит из всех элементов поля, кроме 0. Поскольку количество адресов негативных элементов заведомо меньше p (около половины), для построения циклической матрицы нам нужна не столько вся группа, сколько подгруппа, существование которой гарантируется теорией групп.

Пример. Рассмотрим группу Галуа размера $p = 11$. Размеры ее подгрупп являются делителями числа $p-1 = 1 \times 2 \times 5 = 10$. Они отвечают длинам цепочек последовательностей g^k с неповторяющимися элементами — циклическим подгруппам группы $\text{GF}^*(11)$, перечисленным в таблице 2.

Элементы 2, 6, 7, 8 первого столбца таблице 2 называются *примитивными*, они порождают циклическую подгруппу максимальной длины

$GF^*(11)$. Такая орбита, как ее еще иногда называют, порождает ортогональную циклическую матрицу I (единичную) порядка 12 с элементами $a = 1$, $b = 0$, где элементы орбиты — адреса элемента b в первой строке матрицы I , состоящей из 0, за исключением первой 1.

Таблица 2. Циклические подгруппы $GF^*(11)$

g	g^k
1	1
2	1,2,4,8,5,10,9,7,3,6
3	1,3,9,5,4
4	1,4,5,9,3
5	1,5,3,4,9
6	1,6,3,7,9,10,5,8,4,2
7	1,7,5,2,3,10,4,6,9,8
8	1,8,9,6,4,10,3,2,5,7
9	1,9,4,3,5
10	1,10

Орбита минимальной длины, отвечающая $g = 1$, порождает ту же самую матрицу, если произвести изменение значения ее элементов на противоположные и добавить циклический сдвиг строк на один шаг назад.

Более содержательна орбита средней длины 5, образованная элементами 3, 4, 5, 9 первого столбца таблицы. Отметим, что эти элементы являются также значениями g^k , помимо первого единичного элемента $g^0 = 1$. Она порождает квазиортогональную циклическую матрицу размера 11 с элементами $a = 1$, $-b$, изображенную на рисунке 6. Если индексировать элементы с 0, то числа 1, 3, 4, 5, 9 отвечают положениям элемента $-b$.

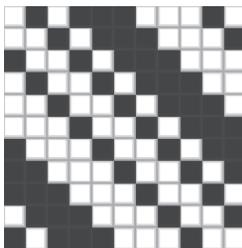


Рис. 6. Циклическая квазиортогональная матрица

Диофантово уравнение I вида $k(k-1) = \lambda(n-1)$ с параметрами $\lambda = t$, $k = 2t$ заведомо разрешимо для порядков в виде простых чисел $n = 4t - 1$. После подстановки параметров в уравнение связи (1), оно упрощается до $(t-1)b^2 - 2tba + ta^2 = 0$, положительный корень этого полинома дает уровень вещественных ортогональных моноциклов Мерсенна $b = \frac{t}{t + \sqrt{t}}$ при $a = 1$, отображенных в таблице 1.

Анализ таблицы 2 показывает, что есть еще бинарная матрица с двумя элементами. Такие матрицы отличаются сравнительно низким значением детерминанта (единица у единичной матрицы), они есть для многих порядков и мало интересны для приложений как слишком простые матрицы.

8. Бициклические матрицы. Моноциклы (циклические матрицы) существуют далеко не всегда. В связи с этим результативные поиски часто ограничивают бициклическими матрицами (бициклами)

$\begin{pmatrix} A & B \\ B^T & -A^T \end{pmatrix}$ фиксированной сложности, ограниченной двумя циклическими блоками A, B с параметрами $\{n, k_1, \lambda_1\}$, $\{n, k_2, \lambda_2\}$.

Нижняя часть этого орнамента зависима, ортогональность вложенных блоков, как правило, не интересует (ортогональны строки и столбцы матрицы в целом), поэтому для описания орнамента достаточно четырех параметров $\{n, k_1, k_2, \lambda\}$, где $\lambda = \lambda_1 + \lambda_2$. Диофантово уравнение I бицикла имеет вид $k_1(k_1-1) + k_2(k_2-1) = \lambda(n-1)$. Заменой переменных $x = p - k_1$, $y = p - k_2$, при $p = k_1 + k_2 - \lambda$, оно сводится к уравнению окружности:

$$x^2 + y^2 = p + \lambda(v - 2p), \quad (2)$$

решаемому в целых числах. Характеристическое уравнение:

$$\lambda b^2 - 2(k_1 + k_2 - \lambda)ab + (n - 2(k_1 + k_2) + \lambda)a^2 = 0, \quad (3)$$

в котором $k_1, k_2 < p$, $p = k_1 + k_2 - \lambda = (n-2)/4$, отражают условие ортогональности строк бицикла с элементами $a, -b$. Оно записано в виде, удобном для поиска корней при превалировании количества положительных элементов над отрицательными.

В теории бициклов наиболее интересны те из них, которые разрешимы для $n = 4t - 2$, не зависимо от разложимости числа $n-1$ на сумму

квадратов двух чисел. Напомним, что этим условием ограничены квазиортогональные матрицы с целочисленными элементами [1], в частности матрицы Белевича. После нормирования в 1 элементов каймы (помним диагонального элемента 0) и отделения каймы, матрицы Белевича переходят в матрицы Одина нечетных порядков той же структуры. Ввиду пропусков и сложности орнаментов матрицы Одина имеют узко теоретическое значение. Так как критериями разложимости чисел на суммы квадратов целых чисел занимался Эйлер, бициклы, восполняющие пропуски матриц Белевича, названы матрицами Эйлера.

Для всех матриц Эйлера порядков $n = 4t - 2$ (размер плеча $v = n/2 = 2p - 1$, $p = t$) диофантово уравнение (2) сводится к уравнению равнобедренного прямоугольного треугольника $x^2 + y^2 = 2$, разрешимое для точки $x = y = 1$ окружности с квадратом радиуса $p + \lambda(v - 2p) = 2$. Поскольку $k_1 = p - x$ и $k_2 = p - y$, количества элементов одного знака в плечах бицикла равны $k_1 = k_2 = p - 1 = (v - 1)/2$, $\lambda = p - 2 = (v - 3)/2$. Дизайн вида $\{n = 2v; k_1; k_2; \lambda\} = \{n = 2v; (v - 1)/2; (v - 1)/2; (v - 3)/2\}$ назовем *эйлеровым*.

Наиболее проста реализация бицикла Эйлера с равными плечами $A = B$, $\lambda_1 = \lambda_2$, $\lambda_1 = \lambda_1 + \lambda_2$, которые представляют собой, в частности, рассмотренные ранее циклические матрицы Мерсенна вдвое меньшего порядка $v \equiv 3 \pmod{4}$. Пусть разложение на базе одной и той же матрицы в обоих плечах невозможно, тогда раскрывается скрытый ресурс бицикла: плечи его утрачивают равновесие $A \neq B$, отклонение $\lambda_1 = \lambda_2 + 1$ берет в расчет блоки, которым не обязательно быть ортогональными, ведь ортогональна конструкция в целом. Характеристическое уравнение (3) отражает условие ортогональности строк с элементами $a, -b$, причем $(n - 2(k_1 + k_2) + \lambda) = (n - 2p + \lambda) = p$, при $\lambda = p - 2$, $2(k_1 + k_2 - \lambda) = 2p$, и для $n = 4t - 2$, $p = t$. Положительный корень этого полинома дает уровень матриц Эйлера $b = \frac{t}{t + \sqrt{2t}}$ при $a = 1$.

Возможно построение блоков матрицы Эйлера из двух циклических матриц Мерсенна, находимых итерациями в поле Галуа. Циклические матрицы Одина также порождают матрицу Эйлера, при росте размера матрицы вдвое уровень диагонального элемента перестает отставать от 1 (можно задать 1). В отличие от матриц Одина вдвое большая по размеру бициклическая матрица Эйлера имеет резерв на существование даже тогда, когда первых не существует. Ввиду экстремальности свойств эти критские матрицы локального максимума детерминанта

независимо от порядка находит динамическая система оптимизации, рассмотренная ранее.

9. Переход от матриц Эйлера к матрицам Адамара. Формальный переход от бициклической матрицы Эйлера к матрице Адамара опосредованно через промежуточную матрицу Эйлера венчает построение теории, показывающей конечность возможных структур матриц Адамара. Подсчет количества элементов матриц Эйлера показывает, что эти бициклы переходят в матрицы Мерсенна элементарным добавлением каймы из одинакового количества негативных и позитивных элементов (см. рисунок 7) с пересчетом уровней элементов, согласно таблице 1. Матрицы Мерсенна, в свою очередь, при помощи монотонной каймы и инверсии знаков элементов переходят в матрицы Адамара с приведением уровней элементов до 1 и -1.

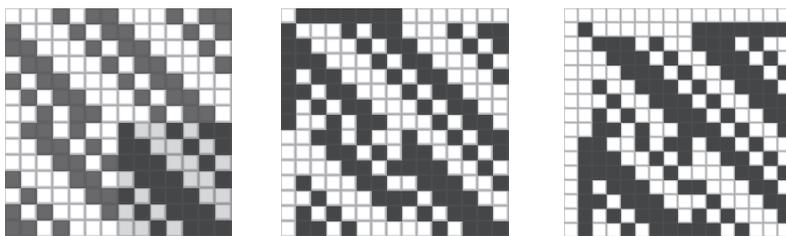


Рис. 7. Переход от бицикла Эйлера к матрице Мерсенна и Адамара

10. Получение квазиортогональных матриц в теории грамианов. В теории линейных динамических систем ортогональные матрицы возникают при диагонализации грамианов системных свойств управляемости, наблюдаемости и идентифицируемости. Впервые на это обратил внимание Гловер в работе [18], предложив использовать ненулевые элементы диагональных матриц грамианов управляемости и наблюдаемости в качестве аналогов сингулярных чисел матриц для задач редукции моделей динамических систем.

При реакции на импульсное воздействие вектор состояния модели динамической системы в сбалансированной форме Гловера содержит взаимно-ортогональные между собой функции. Обобщение этого метода состоит в рассмотрении широкого класса ассоциированных с линейной динамической системой линейных операторов, симметрия которых гарантирует ортогональность их сингулярных функций. Флип-метод определения сингулярных функций ганкелева оператора и оператора свертки позволяет выделить ортогональные последовательности в процессе математического или натурального моделирования [14].

Для нахождения малоуровневых матриц эти методы возбуждения динамических систем и создания условий, когда они выдают ортогональные последовательности, недостаточны, в связи с чем возникают аналитические методы поиска с помощью полей Галуа. Они очень эффективны для поиска матриц высокой размерности [17], но не отвечают на все вопросы теории, поскольку относятся к комбинаторным методам. В их основе лежит перебор вариантов знаков элементов, поиск подгруппы и т. п. Комбинаторная теория широко оперирует орнаментами целочисленных матриц Белевича, также являющихся блоками матриц Адамара конструкции Пэли [1]. Проблематика матриц Белевича (заведомо существуют не всегда, орнамент не всегда ясен) сформировала точку зрения на орнаменты матрицы Адамара как неясные и, в общем, непредсказуемые. То же самое касается вопроса о существовании матриц на всех порядках $4t$.

11. Заключение. Существующие на порядках с шагом 4 бициклы Эйлера составляют антитезу известной гипотезе Райзера [17] о существовании матрицы Адамара в форме моноцикла не выше 4-го порядка. Орнаменты матриц Эйлера отвечают на вопрос максимальной сложности структуры квазиортогональных матриц, к которым матрица Адамара относится, сводя основную трудность решения задачи к поиску блоков бицикла. Добавление одинарной или двойной каймы сложности не привносит. Переход к динамическим моделям оптимизации детерминанта, разумеется, не делает трудные задачи поиска матриц Адамара высоких порядков легким занятием. Эти трудности сохраняются в виде сложности указания областей устойчивости матриц локального максимума детерминанта. Но перспектива доказательства существования квазиортогональных матриц с возможно иррациональными элементами выглядит иначе. Сам по себе факт существования области сходимости к локальному максимуму безотносителен к порядку.

Таким образом, намечается продуктивный и независимый от прежних воззрений путь рассмотрения одной из наиболее интригующих гипотез прошлого века (гипотезы Адамара).

Литература

1. *Awyzio G., Seberry J.* On good matrices and skew Hadamard matrices // Algebraic Design Theory and Hadamard Matrices. 2015. pp. 13–28.
2. *Kim J., Susilo W., Au M.H., Seberry J.* Efficient semi-static secure broadcast encryption scheme // 6th International Conference Pairing-Based Cryptography (Pairing 2013). 2014. LNCS 2738. pp. 62–76.
3. *Holzmann W.H., Kharaghani H., Tayfeh-Rezaie B.* Williamson Matrices up to Order 59 // Designs, Codes and Cryptography. 2008. vol. 46. pp. 343–352.
4. *Doković D.Ž.* Williamson Matrices of Order $4n$ for $n=33;35;39$ // Discrete Math. 1993. vol. 115. pp. 267–271.
5. *Matteo O.Di, Djokovic D.Z., Kotsireas I.S.* Symmetric Hadamard matrices of order 116 and 172 exist // Special matrices. 2015. vol. 3. pp. 227–234.
6. *Doković D.Ž.* Generalization of Scarpi's theorem on Hadamard matrices // Linear and Multilinear Algebra. pp. 1–3. URL:

<http://www.tandfonline.com/doi/abs/10.1080/03081087.2016.1265062?journalCode=glma20>. (дата обращения: 21.11.2016).

7. *Petoukhov S.V.* The Genetic Coding, United-Hypercomplex Numbers and Artificial Intelligence // *Advances in Artificial Systems for Medicine and Education*. 2017. pp. 2–13.
8. *Петухов С. В.* Матричная генетика, алгебры генетического кода, помехоустойчивость // Москва: РХД. 2008. 316 с.
9. *Lee M.H., Hai H., Lee S.K., Petoukhov S.V.* A Mathematical Proof of Double Helix DNA to Reverse Transcription RNA for Bioinformatics // *Advances in Artificial Systems for Medicine and Education*. 2017. pp. 23–38.
10. *Манин Ю.И.* Математика как метафора // Москва: МЦМНО. 2008. 400 с.
11. *Shechtman D., Blech I., Grattias D., Cahn J.W.* Metallic Phase with LongRange Orientational Order and No Translational Symmetry // *Physical Review Letters*. 1984. vol. 53. pp. 1951–1953.
12. *Balotin N.A., Suzdal V.S.* Symmetry of Life in Crystals // *Functional materials*. 2016. vol. 23(4). pp. 1–7.
13. *Балонин Н.А., Сергеев М.Б., Суздаль В.С.* Матричные модели обобщенной кристаллографии // *Информационно-управляющие системы*. 2016. № 4(83). С. 27–33.
14. *Балонин Н.А.* Дискретные частотные характеристики элементарных динамических звеньев // *Информационно-управляющие системы*. 2015. № 4 (77). С. 17–24.
15. *Sergeev A.M.* Generalized Mersenne Matrices and Balonin's Conjecture // *Automatic Control and Computer Sciences*. 2014. vol. 48. no. 4. pp. 214–220.
16. *Балонин Ю.Н., Востриков А.А., Сергеев А.М., Егорова И.С.* О взаимосвязях квазиортогональных матриц, построенных на известных последовательностях чисел // *Труды СПИИРАН*. 2017. №. 1(50). С. 209–223.
17. *Балонин Н.А., Сергеев М.Б.* Расширение гипотезы Райзера на двучисленные структуры и разрешимость матриц Адамара орнаментом в виде бицикла с двойной каймой // *Информационно-управляющие системы*. 2017. № 1 (86). С. 2–10.
18. *Glover K.* All Optimal Hankel-norm Approximations of Linear Multivariable Systems // *Intern. J. Control*. 1984. vol. 39. no. 6. pp. 1115–1193.

Балонин Николай Алексеевич — д-р техн. наук, доцент, профессор кафедры вычислительных систем и сетей, Санкт-Петербургский государственный университет аэрокосмического приборостроения (СПбГУАП). Область научных интересов: вычислительные методы, теория управления, теория чисел, идентификация динамических систем, образовательные интернет-технологии, исполняемые алгоритмы. Число научных публикаций — 94. korbendfs@mail.ru; ул. Большая Морская, 67, Санкт-Петербург, 190000; р.т.: +7(911)213-59-07.

Сергеев Михаил Борисович — д-р техн. наук, профессор, заведующий кафедрой вычислительных систем и сетей, института вычислительных систем и программирования, Санкт-Петербургский государственный университет аэрокосмического приборостроения (СПбГУАП), директор института вычислительных систем и программирования, Санкт-Петербургский государственный университет аэрокосмического приборостроения (СПбГУАП). Область научных интересов: численные методы, цифровая обработка данных, оптико-информационные системы. Число научных публикаций — 106. mbsergeev@gmail.com; ул. Большая Морская, 67, Санкт-Петербург, 190000; р.т.: +7(905)223-24-89.

Суздаль Виктор Семенович — д-р техн. наук, ведущий научный сотрудник лаборатории систем управления, Институт сцинтилляционных материалов Национальной академии наук Украины (НАНУ). Область научных интересов: теория управления, идентификация динамических систем, теория матриц, кристаллография. Число научных публикаций — 141. suzdal@isma.kharkov.ua; пр. Науки, 60, Харьков, 61072, Украина; р.т.: +380 97 914 6537.

Поддержка исследований. Работа выполнена при поддержке Минобрнауки РФ при проведении научно-исследовательской работы в рамках проектной части государственного задания в сфере научной деятельности по заданию № 2.2200.2017/4.6.

N.A. BALONIN, M.B. SERGEEV, V.S. SUZDAL
**DYNAMIC GENERATORS OF THE QUASIORTOGONAL
 HADAMARD MATRIX FAMILY**

Balonin N.A., Sergeev M.B., Suzdal V.S. **Dynamic Generators of the Quasiorthogonal Hadamard Matrix Family.**

Abstract. The problem of constructing non-linear and linear finite-field generators of quasi-orthogonal matrices of the Hadamard family with a small number of distinct values of their elements not exceeding by absolute value 1 and a global or local maximum of determinant is investigated. The properties of such dynamical systems are analyzed; the classification of the matrix families and their ornaments, obtained with their help, is described; the way of proving the existence of real and integer matrices different from the combinatorial approach is shown. The values to which the elements of the matrix are equal are called its levels. The concepts of the Hadamard norm and the determinant of a quasiorthogonal matrix are introduced. Levels, Hadamard norm and determinant play a fundamental role in the definitions of classes of generalized matrices of the Hadamard family. The classes of the Hadamard, Belevich (conference matrices), Seberri (weighing matrices), Mersenne, Euler, Odin (Seidel), Fermat matrices are described. The formulas for the values of their levels are given. Ornaments of Euler matrices answer to the question of the maximum complexity of Hadamard matrices — two border two circulant structure.

Keywords: dynamical systems, quasiorthogonal matrices, deterministic chaos, finite fields.

Balonin Nikolay Alekseevich — Ph.D., Dr. Sci., associate professor, professor of computer systems and networks department, Saint Petersburg State University of Aerospace Instrumentation (SUAI). Research interests: numerical methods, number theory, computer simulation. The number of publications — 94. korbendfs@mail.ru; 67, B. Morskaja St., 190000, Saint-Petersburg; office phone: +7(911)213-59-07.

Sergeev Mikhail Borisovich — Ph.D., Dr. Sci., professor, head of computer systems and networks department of Institute of computing systems and programming, Saint Petersburg State University of Aerospace Instrumentation (SUAI), director of Institute of computing systems and programming, Saint Petersburg State University of Aerospace Instrumentation (SUAI). Research interests: numerical methods, digital data performing, optical information systems. The number of publications — 106. mbsergeev@gmail.com; 67, B. Morskaja St., 190000, Saint-Petersburg; office phone: +7(905)223-24-89.

Suzdal Viktor Semenovich — Ph.D., Dr. Sci., leading researcher of control systems laboratory, Institute for Scintillation Materials of National Academy of Sciences of Ukraine. Research interests: control theory, identification of dynamic systems, crystallography. The number of publications — 141. suzdal@isma.kharkov.ua; 60, Nauky ave. Kharkiv, 61072, Ukraine; office phone: +380 97 914 6537.

Acknowledgements. The research leading to these results has received funding from the Ministry of Education and Science of the Russian Federation according to the project part of the state funding assignment No 2.2200.2017/4.6.

References

1. Awyzio G., Seberry J. On good matrices and skew Hadamard matrices. Algebraic Design Theory and Hadamard Matrices. 2015. pp. 13–28.

2. Kim J., Susilo W., Au M.H., Seberry J. Efficient semi-static secure broadcast encryption scheme. 6th International Conference Pairing-Based Cryptography (Pairing 2013). 2014. LNCS 2738. pp. 62–76.
3. Holzmann W.H., Kharaghani H., Tayfeh-Rezaie B. Williamson Matrices up to Order 59. *Designs, Codes and Cryptography*. 2008. vol. 46. pp. 343–352.
4. Doković D.Ž. Williamson Matrices of Order $4n$ for $n=33;35;39$. *Discrete Math*. 1993. vol. 115. pp. 267–271.
5. Matteo O.Di, Djokovic D.Z., Kotsireas I.S. Symmetric Hadamard matrices of order 116 and 172 exist. *Special matrices*. 2015. vol. 3. pp. 227–234.
6. Doković D.Ž. Generalization of Scarpi's theorem on Hadamard matrices. *Linear and Multilinear Algebra*. pp. 1–3. Available at: <http://www.tandfonline.com/doi/abs/10.1080/03081087.2016.1265062?journalCode=glma20>. Published online: 07 Dec 2016. (accepted: 21.11.2016).
7. Petoukhov S.V. The Genetic Coding, United-Hypercomplex Numbers and Artificial Intelligence. *Advances in Artificial Systems for Medicine and Education*. 2017. pp. 2–13.
8. Petoukhov S.V. *Matrichnaia genetika, algebrы geneticheskogo koda, pomekhoustoichivost'* [Matrix genetics, algebra of genetic code, noise immunity]. Moscow: RHD. 2008. 316 p. (In Russ.).
9. Lee M.H., Hai H., Lee S.K., Petoukhov S.V. A Mathematical Proof of Double Helix DNA to Reverse Transcription RNA for Bioinformatics. *Advances in Artificial Systems for Medicine and Education*. 2017. pp. 23–38.
10. Manin Yu.I. *Matematika kak metafora* [Mathematics as a Metaphor]. Moscow: MTsMNO. 2008. 400 p. (In Russ.).
11. Shechtman D., Blech I., Gratias D., Cahn J.W. Metallic Phase with LongRange Orientational Order and No Translational Symmetry. *Physical Review Letters*. 1984. vol. 53. pp. 1951–1953.
12. Balonin N.A., Suzdal V.S. Symmetry of Life in Crystals. *Functional materials*. 2016. vol. 23(4). pp. 1–7.
13. Balonin N.A., Sergeev M.B., Suzdal' V.S. [Matrix Models of Generalized Crystallography]. *Informatsionno-upravliaiushchie sistemy — Information and Control Systems*. 2016. vol. 4(83). pp. 27–33. (In Russ.).
14. Balonin N.A. [Discrete Frequency Characteristics of Elementary Dynamic Units]. *Informatsionno-upravliaiushchie sistemy — Information and Control Systems*. 2015. vol. 4(77). pp. 17–24. (In Russ.).
15. Sergeev A.M. Generalized Mersenne Matrices and Balonin's Conjecture. *Automatic Control and Computer Sciences*. 2014. vol. 48. no. 4. pp. 214–220.
16. Balonin Yu.N., Vostrikov A.A., Sergeev A.M., Egorova I.S. [On Relationships among Quasi-orthogonal Matrices Constructed on the Known Sequences of Prime Numbers]. *Trudy SPIIRAN — SPIIRAS Proceedings*. 2017. vol. 1(50). pp. 209–223. (In Russ.).
17. Balonin N.A., Sergeev M.B. [Ryser's Conjecture Expansion for Bicirculant Structures and Hadamard Matrix Resolvability by Double-Border Bicycle Ornament]. *Informatsionno-upravliaiushchie sistemy — Information and Control Systems*. 2017. vol. 1(86). pp. 2–10. (In Russ.).
18. Glover K. All Optimal Hankel-norm Approximations of Linear Multivariable Systems. *Intern. J. Control*. 1984. vol. 39. no. 6. pp. 1115–1193.

И.С. ПОЛЯНСКИЙ, Ю.С. ПЕХОВ
**БАРИЦЕНТРИЧЕСКИЙ МЕТОД В РЕШЕНИИ
СИНГУЛЯРНЫХ ИНТЕГРАЛЬНЫХ УРАВНЕНИЙ
ЭЛЕКТРОДИНАМИЧЕСКОЙ ТЕОРИИ ЗЕРКАЛЬНЫХ АНТЕНН**

Полянский И.С., Пехов Ю.С. Бариецентрический метод в решении сингулярных интегральных уравнений электродинамической теории зеркальных антенн.

Аннотация. В статье предложено использование бариецентрического метода в решении задач анализа характеристик излучения зеркальных антенн. Формализация задачи выполнена в приближении электродинамической теории зеркальных и полосковых антенн при сведении к решению системы сингулярных интегральных уравнений. В постановке бариецентрического метода решение формируемой системы предполагается выполнять численно с использованием вариационных методов Галёркина и Ритца. Для повышения эффективности решения в сравнении с известными методами аппроксимация Ритца для областей анализа (раскрыты рефлекторы и излучателей) задается без их разбиения на элементарные подобласти (конечные элементы). Для заданной аппроксимации исходная задача сведена к системе линейных уравнений. С целью определения предпочтительности применения бариецентрического метода рассмотрен пример решения задачи анализа зеркальной антенны при сравнении с существующими методами, которые формируют аппроксимацию базисными функциями путем разбиения проводящей поверхности антенны на подобласти простой формы.

Ключевые слова: бариецентрический метод, метод моментов, интегральное сингулярное уравнение, зеркальная антенна.

1. Введение. Одна из основных задач математической теории зеркальных антенн (ЗА) [1] при проектировании, оптимизации и управлении связана с формированием адекватных математических моделей, позволяющих при минимальных затратах вычислительного и емкостного ресурсов ЭВМ с заданной точностью оценить характеристики излучения [2]. Для математического моделирования ЗА в ряде работ [2-10 и др.] предлагается применять методы вычислительной электродинамики, которые составляют основу широко используемых во всем мире программ моделирования высокочастотных электромагнитных полей (ЭМП) [11]: Ansis-Ansoft HFSS, CST Microwave Studio Suite, QWAVE QuickWave3D, Remcom XFDTD, EMSS-SA FEKO, Russian R&D Center LG Electronics RFS и другие. Эти методы условно разделяются относительно вида решаемой внешней или внутренней задачи [12]. Первая группа не ограничивает расчетную область анализа и связана с решением неоднородного волнового уравнения, удовлетворяющим условию излучения Зоммерфельда [13]. К ней относятся метод моментов (ММ) [4] (метод сингулярных интегральных уравнений [6, 14, 15]) и ряд асимптотических методов [2] (геометрической оптики [16], геометрической теории дифракции [17], физической оптики [13], физической теории дифракции [18] и др.). Вторая группа связана с

поиском решения уравнений Максвелла или соответствующих им волновых уравнений в ограниченной расчетной области с заданными граничными условиями (методы конечных разностей [19], конечных элементов [20], матрицы линий передачи [21]). Все перечисленные методы имеют ряд ограничений. Асимптотические методы расчета не обеспечивают требуемого уровня адекватности, поскольку они не способны в требуемой мере учесть эффекты дифракции и переотражения излучения. Остальные методы при практической реализации требуют значительных вычислительных ресурсов, что с точки зрения пользователя затрудняет решение задач анализа зеркальных антенн за обозримый отрезок машинного времени на типовой ЭВМ.

Для устранения ограничения асимптотических методов в [2, 9] предложена модификация токового метода. Однако при решении ряда задач анализа многоэлементных ЗА в условиях взаимного влияния облучающей и отражающей систем разработанное решение [9] приводит к существенному росту вычислительных затрат. Снижения вычислительных затрат можно добиться за счет решения задач анализа многоэлементных зеркальных антенн ММ [4] в приближении электродинамической теории зеркальных и полосковых антенн [10]. Эффективность метода моментов существенным образом зависит от удачного выбора базисных функций, которые при минимальном порядке аппроксимации обеспечивают требуемую адекватность решения формируемого сингулярного интегрального уравнения (СИУ). Известные методы [4, 14, 15, 22-24, 31] задают базисные функции путем разбиения проводящей поверхности зеркальной антенны на подобласти простой формы, размер которых должен быть $\leq 0,1\lambda$, где λ — длина волны. С учетом решений [8, 25] интерес вызывает задание в приближении барицентрического метода (БМ) базисных функций, позволяющих сформировать аппроксимацию векторного и скалярного потенциалов для всей проводящей поверхности ЗА в целом без ее разбиения на подобласти простой формы.

Цель статьи заключается в разработке эффективного решения сингулярных интегральных уравнений электродинамической теории зеркальных антенн с применением БМ [8, 25].

2. Содержательная постановка задачи. Зеркальную антенну представим системой, состоящей из M излучателей с раскрывами $\Omega_1, \Omega_2, \dots, \Omega_M$ и M' рефлекторов с произвольной формой отражающих поверхностей $S_1, S_2, \dots, S_{M'}$ (рисунок 1). В общей системе координат ЗА $OXYZ$ положение раскрыва Ω_m m -го излучателя ($m = \overline{1, M}$) задано

координатами x_m, y_m, z_m , а ориентация — орт-вектором нормали $\vec{s}_m^{(o)}$ плоскости Ω_m . Рефлектор в $OXYZ$ m' -й отражающей поверхности $S_{m'}$ ($m' = \overline{1, M'}$) характеризуется раскрывом $\Omega'_{m'}$ апертуры с центром в $x'_{m'}, y'_{m'}, z'_{m'}$ и функцией $Z_{m'}(x, y)$ формы, из которой по известным правилам [26] определяется орт вектора внешней нормали $\vec{n}_{m'}^0$. Раскрывы апертуры рефлектора $\Omega'_{m'}$ и излучателя Ω_m задаются произвольными многоугольниками с соответствующим числом $N'_{m'}$ и N_m вершин $P_1^{m'}, P_2^{m'}, \dots, P_{N'_{m'}}^{m'}$ и $P_1^m, P_2^m, \dots, P_{N_m}^m$. Относительно $\Omega'_{m'}$ и Ω_m в $OXYZ$ вводятся системы координат $O'_{m'}X'_{m'}Y'_{m'}Z'_{m'}$ и $O_mX_mY_mZ_m$ соответственно.

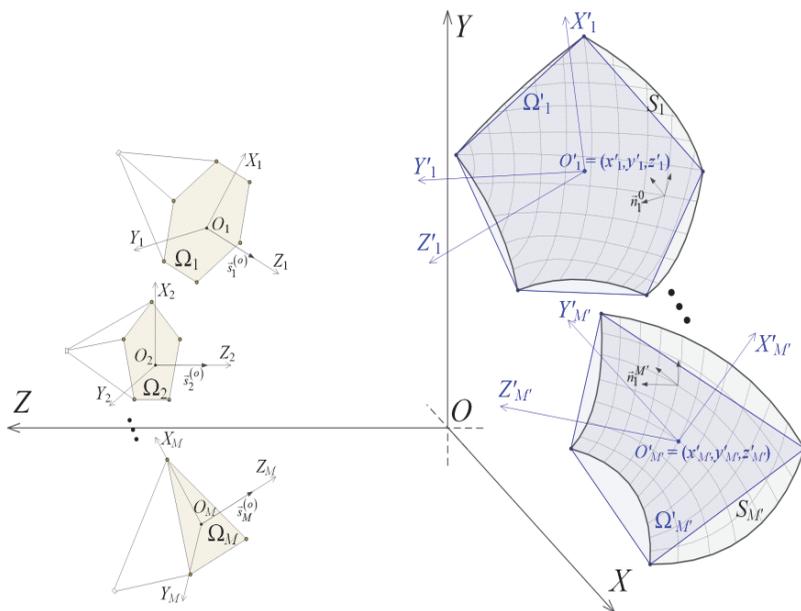


Рис. 1. Пример геометрического представления зеркальной антенны

В соответствии с правилами, приведенными в [27], в плоскостях $O'_{m'}X'_{m'}Y'_{m'}$ и $O_mX_mY_m$ раскрывов $\Omega'_{m'}$ и Ω_m задаются соответствующие барицентрические системы координат (БСК) $\zeta_{n_{m'}}^{m'}$ и $\zeta_{n_m}^m$, где $n'_{m'} = \overline{1, N'_{m'}}$, $n_m = \overline{1, N_m}$. При этом барицентрический

координаты $\zeta_{n_m}'(P')$ и $\zeta_{n_m}^m(P)$ являются относительно $P' \in \Omega_{m'}$ и $P \in \Omega_m$ гармоническими функциями [27], удовлетворяющими свойствам [28]: аффинной инвариантности, положительной определенности, равенства суммы единице.

Для принятых представлений с использованием решений из [10, 30], определим формализованную постановку задачи анализа ЗА. С учетом применяемой симметризации [29] системы уравнений Максвелла для сокращения математической записи далее будем рассматривать решение относительно напряженности \vec{E} электрического поля.

3. Формализованная постановка задачи. Постановка задачи анализа ЗА следует из неоднородного уравнения Гельмгольца [2]:

$$\nabla^2 \vec{E} + \kappa^2 \vec{E} = -(i\omega\epsilon)^{-1} \nabla(\nabla \cdot \vec{J}) + i\omega\mu \vec{J} + \nabla \times \vec{J}^M, \quad (1)$$

где \vec{J} и \vec{J}^M — плотности эклектического и магнитного токов соответственно; $\kappa = \omega\sqrt{\epsilon\mu}$ — волновое число; ϵ и μ — диэлектрическая и магнитная проницаемости среды соответственно; $\omega = 2\pi/\lambda$ — круговая частота.

С учетом известных из [3, 13, 22] граничных условий между рассеянным и наводимым полями на поверхностях раздела (диэлектрика и идеального проводника, двух диэлектриков) и фундаментального решения уравнения Гельмгольца [30] с помощью

функции Грина $\psi(\vec{r}, \vec{r}') = \frac{e^{-\beta R}}{4\pi R}$ ($R = |\vec{r} - \vec{r}'|$; \vec{r} и \vec{r}' — радиус-векторы точек наблюдения и источника) задача (1) сводится к решению СИУ для рефлектора $S_{m'}$ [22]:

$$-\vec{E}_{\tan}^{\text{ind}} = \left[\iint_{S_{m'}} \psi \left(i\omega\mu \vec{J}_{m'} - \frac{1}{i\omega\epsilon} \nabla(\nabla \cdot \vec{J}_{m'}) \right) dS \right]_{\tan}, \quad (2)$$

и излучателя Ω_m [22]:

внешняя область:

$$\vec{E}_{\tan}^{\text{ind}} = - \left[\iint_{\Omega_m} \psi \left(i\omega\mu \vec{J}_m - \frac{\nabla(\nabla \cdot \vec{J}_m)}{i\omega\epsilon} + \nabla \times \vec{J}_m^M \right) d\Omega \right]_{\tan}; \quad (3)$$

внутренняя область:

$$0 = \left[\iint_{\Omega_m} \Psi \left(i\omega\mu\vec{J}_m - \frac{1}{i\omega\epsilon} \nabla(\nabla \cdot \vec{J}_m) + \nabla \times \vec{J}_m^M \right) d\Omega \right]_{\tan},$$

где индекс \tan обозначает тангенциальную составляющую поля; \vec{E}^{ind} — напряженность электрического поля источника, наводимого на соответствующую поверхность.

Объединяя (2), (3) для $S_{m'}$ и Ω_m в единое соотношение, с учетом содержательной постановки получим систему из $M' + 2M$ СИУ, решение которой формализует задачу анализа функционирования ЗА:

$$\begin{aligned} & -\vec{n}_{m'}^0 \times \left[\sum_{k=1}^{M'} \iint_{S_k} \Psi \left(i\omega\mu\vec{J}_k - \frac{1}{i\omega\epsilon} \nabla(\nabla \cdot \vec{J}_k) \right) dS + \right. \\ & \left. + \sum_{k=1}^M \iint_{\Omega_k} \Psi \left(i\omega\mu\vec{J}_k - \frac{1}{i\omega\epsilon} \nabla(\nabla \cdot \vec{J}_k) + \nabla \times \vec{J}_k^M \right) d\Omega \right] = \vec{n}_{m'}^0 \times \sum_{k=1}^M \vec{E}_{k,m'}^{\text{ind}}; \\ & -\vec{s}_m^{(o)} \times \left[\sum_{k=1}^{M'} \iint_{S_k} \Psi \left(i\omega\mu\vec{J}_k - \frac{1}{i\omega\epsilon} \nabla(\nabla \cdot \vec{J}_k) \right) dS + \right. \\ & \left. + \sum_{k=1}^M \iint_{\Omega_k} \Psi \left(i\omega\mu\vec{J}_k - \frac{1}{i\omega\epsilon} \nabla(\nabla \cdot \vec{J}_k) + \nabla \times \vec{J}_k^M \right) d\Omega \right] = \vec{s}_m^{(o)} \times \sum_{k=1}^M \vec{E}_{k,m}^{\text{ind}}; \\ & \vec{s}_m^{(o)} \times \sum_{k=1}^M \iint_{\Omega_k} \Psi \left(i\omega\mu\vec{J}_k - \frac{1}{i\omega\epsilon} \nabla(\nabla \cdot \vec{J}_k) + \nabla \times \vec{J}_k^M \right) d\Omega = 0, \end{aligned} \quad (4)$$

где $\vec{E}_{k,m'}^{\text{ind}}$ — \vec{E}^{ind} (k -й излучатель, где $k = \overline{1, M}$), наведенное на $S_{m'}$; $\vec{E}_{k,m}^{\text{ind}}$ — \vec{E}^{ind} , наведенное на Ω_m .

4. Анализ функционирования зеркальных антенн в приближении барицентрического метода. Решение задачи (4) выполняется численно по аналогии с [14, 15, 22-24, 31 и др.] при сведении в приближении вариационных методов В. Ритца, И.Г. Бубнова и Б.Г. Галёркина системы сингулярных интегральных уравнений к системе линейных уравнений. Для этого задаются

аппроксимации функций $\vec{J}_k(P)$ и $\vec{J}_k^M(P)$ для соответствующей k -й поверхности $P \in \mathcal{S}_k$ (или S_k , или Ω_k) в N_k -мерной барицентрической системе координат $\zeta_1^k, \zeta_2^k, \dots, \zeta_{N_k}^k$ следующим образом.

Теорема 1. Функции $\vec{J}_k(P) = \sum_{j \in \mathbb{M}_p} \hat{J}_{k,j} \vec{\gamma}_j^k(P)$ и $\vec{J}_k^M(P) = \sum_{j \in \mathbb{M}_p} \hat{J}_{k,j}^M \vec{\chi}_j^k(P)$, задающие аппроксимацию порядка p ($p \in \mathbb{N}$)

плотности поверхностных электрического \vec{J}_k и магнитного \vec{J}_k^M токов в интегральных уравнениях вида (4), при:

$$\vec{\gamma}_j^k(P) = \vec{n}_k^0(P) \times \sum_{\substack{n=1 \\ j_n > 0}}^{N_k} [\hat{\alpha}_{j,n}^k(P) \vec{\Theta}_n^k(P)]; \quad (5)$$

$$\vec{\chi}_j^k(P) = \vec{n}_k^0(P) \times \vec{\gamma}_j^k(P),$$

являются аппроксимацией Рунта.

В заданных аппроксимациях $\hat{J}_{k,j}$ и $\hat{J}_{k,j}^M$ — коэффициенты в j -х узловых точках электрического \vec{J}_k и магнитного \vec{J}_k^M поверхностных токов на \mathcal{S}_k .

В выражении (5) $\vec{n}_k^0(P)$ — орт-вектора нормали в $P \in \mathcal{S}_k$; $\hat{\alpha}_{j,n}^k(P)$ — вспомогательный интерполяционный сдвинутый полином Сильвестра–Лагранжа [32] относительно n -й вершины многоугольника, задающего раскрыв \mathcal{S}_k :

$$\hat{\alpha}_{j,n}^k(P) = \hat{R}_{j_n}^k(P) \prod_{\substack{n'=1 \\ n' \neq n}}^{N_k} R_{j_n'}^k(P); \quad (6)$$

$$\hat{R}_{j_n}^k(P) = \begin{cases} \frac{1}{(j_n - 1)!} \prod_{h=1}^{j_n-1} (p \zeta_n^k(P) - h), & \text{if } 2 \leq j_n \leq p; \\ 1, & \text{if } j_n = 1; \end{cases}$$

$\bar{\Theta}_n^k(P) = (\zeta_n^k(P) \nabla \zeta_{n+1}^k(P) - \zeta_{n+1}^k(P) \nabla \zeta_n^k(P)) |P_{n+1}^k - P_n^k|$ — краевые базисные векторные функции Неделека относительно n -й вершины; j — мультииндекс из множества \mathbb{M}_p [33]:

$$\mathbb{M}_p = \left\{ j = (j_1, j_2, \dots, j_n, \dots, j_{N_k}) : j_n \in \mathbb{Z}, j_n \geq 0, \sum_{n \in [1; N_k]} j_n = p \right\}. \quad (7)$$

В выражении (6) $R_{j_n'}^k(P) = \prod_{n=1}^{j_n'} [(p \zeta_{n'}^k(P) - h + 1) / h]$ при $j_n' > 0$,

$R_0^k(P) = 1$ — полином Лагранжа [8].

Доказательство. Из [25] известно, что функции

$$\bar{\beta}_j^k(P) = \sum_{\substack{n=1 \\ j_n > 0}}^{N_k} [\hat{\alpha}_{j,n}^k(P) \bar{\Theta}_n^k(P)]$$
 с учетом свойств гармонических

барицентрических координат ζ_n^k ($n = \overline{1, N_k}$) [27], интерполяционных

форм $\alpha_{j,n}^k(P) = \prod_{n=1}^{N_k} R_{j_n}^k(P)$ и $\hat{\alpha}_{j,n}^k(P)$ являются расширением краевых

базисных функций $\bar{\Theta}_n^k(P)$ и для n -го ребра $e_n^k = (P_n^k, P_{n+1}^k)$ раскрыва \mathcal{S}_k удовлетворяют условиям:

$$\frac{P_{n+1}^k - P_n^k}{|P_{n+1}^k - P_n^k|} \cdot \sum_{\substack{j \in \mathbb{M}_p \\ j_n \neq 0}} \bar{\beta}_j^k(P) = 1; \quad \nabla \cdot \sum_{j \in \mathbb{M}_p} \bar{\beta}_j^k(P) / G_j^k = 0; \quad (8)$$

$$\nabla \times \sum_{j \in \mathbb{M}_p} \bar{\beta}_j^k(P) / G_j = \sum_{n=1}^{N_k} (\nabla \zeta_n^k(P)) \times (\nabla \zeta_{n+1}^k(P)) = \text{const},$$

определяющих, что $\bar{\beta}_j^k(P)$ дивергентно-свободные и имеют постоянную тангенциальную компоненту на e_n^k , что обеспечивает отсутствие ложных решений [3] дифференциальных волновых уравнений вида (1).

В (8) приняты следующие обозначения:

$$\bar{\beta}_j^k(P) = \sum_{\substack{n=1 \\ j_n > 0}}^{N_k} \hat{\alpha}_{j,n}^k(P) \frac{\bar{\Theta}_n^k(P)}{|P_{n+1}^k - P_n^k|}; \quad G_j^k = \begin{cases} 1, & \text{if } \sum_{n=1}^{N_k} \frac{j_n}{p} P_n^k \in \Gamma_k; \\ N_k, & \text{otherwise,} \end{cases} \quad (9)$$

где Γ_k — кусочно-линейная граница раскрыва \mathcal{S}_k .

Из (8) следует, что векторная базисная функция $\bar{\gamma}_j^k(P) = \bar{n}_k^0(P) \times \bar{\beta}_j^k(P)$ на ребре e_n^k при $|\bar{n}_k^0(P)| = 1$ удовлетворяет условиям:

$$\begin{aligned} \frac{P_{n+1}^k - P_n^k}{|P_{n+1}^k - P_n^k|} \cdot \sum_{\substack{j \in \mathbb{M}_p \\ j_n \neq 0}} \bar{\gamma}_j^k(P) &= 1; \quad \nabla \times \sum_{j \in \mathbb{M}_p} \bar{\gamma}_j^k(P) / G_j^k = 0; \\ \nabla \cdot \sum_{j \in \mathbb{M}_p} \bar{\gamma}_j^k(P) / G_j^k &= \text{const}, \end{aligned} \quad (10)$$

где:

$$\bar{\gamma}_j^k(P) = \sum_{\substack{n=1 \\ j_n > 0}}^{N_k} \left[\hat{\alpha}_{j,n}^k(P) \bar{n}_k^0(P) \times \bar{\Theta}_n^k(P) / |P_{n+1}^k - P_n^k| \right]. \quad (11)$$

Свойства (10) определяют то, что при аппроксимации $\bar{J}_k(P) = \sum_{j \in \mathbb{M}_p} \hat{J}_{k,j} \bar{\gamma}_j^k(P)$ на Γ_k нормальная составляющая плотности тока равна нулю, а функция $\bar{\gamma}_j^k(P)$ дивергентно-постоянная, что удовлетворяет уравнению непрерывности [29]. В целом из (10) следует, что $\bar{J}_k(P) = \sum_{j \in \mathbb{M}_p} \hat{J}_{k,j} \bar{\gamma}_j^k(P)$ является для (4) аппроксимацией

Ритца \bar{J}_k на \mathcal{S}_k и удовлетворяет уравнениям Максвелла [13], условию не протекания тока на Γ_k и условию сохранения полного тока в цепи.

Из выделенных свойств векторной базисной функции $\bar{\gamma}_j^k(P)$, аппроксимации $\bar{J}_k(P) = \sum_{j \in \mathbb{M}_p} \hat{J}_{k,j} \bar{\gamma}_j^k(P)$, уравнений Максвелла [13], формализующих зависимость между векторами напряженностей электрического \bar{E}_k и магнитного \bar{H}_k полей, и определением \bar{J}_k и \bar{J}_k^M

токов через векторное произведение \vec{H}_k и \vec{E}_k на $\vec{n}_k^0(P)$ соответственно следует, что $\vec{J}_k^M(P) = \sum_{j \in \mathbb{M}_p} \hat{J}_{k,j}^M \vec{\chi}_j^k(P)$ является для (4)

аппроксимацией Ритца \vec{J}_k^M на k -й поверхности. ■

Из доказательства теоремы 1 с учетом выделенных свойств векторных базисных функций $\vec{\gamma}_j^k(P)$, $\vec{\chi}_j^k(P)$, уравнений Максвелла [13] и принципа симметризации следует следующее утверждение.

Следствие 1. Вариационное решение системы сингулярных интегральных уравнений вида (4) по \vec{H} задается аппроксимацией $\vec{J}_k(P)$ и $\vec{J}_k^M(P)$ через $\vec{\chi}_j^k(P)$ и $\vec{\gamma}_j^k(P)$ соответственно.

Аппроксимации $\vec{J}_k(P) = \sum_{j \in \mathbb{M}_p} \hat{J}_{k,j} \vec{\gamma}_j^k(P)$ и $\vec{J}_k^M(P) = \sum_{j \in \mathbb{M}_p} \hat{J}_{k,j}^M \vec{\chi}_j^k(P)$ позволяют свести исходную систему сингулярных

интегральных уравнений (4) к системе линейных уравнений вида:

$$\mathbf{B}\vec{J} = \vec{C}, \quad (12)$$

где \vec{J} и \vec{C} — блочные векторы:

$$\vec{J} = \left(\left[\vec{J}_1^S \quad \dots \quad \vec{J}_{M'}^S \right] \left[\vec{J}_1^\Omega \quad \dots \quad \vec{J}_M^\Omega \right] \left[\vec{J}_1^{\Omega M} \quad \dots \quad \vec{J}_M^{\Omega M} \right] \right)^T;$$

$$\vec{C} = \left(\left[\vec{C}_1^S \quad \dots \quad \vec{C}_{M'}^S \right] \left[\vec{C}_1^\Omega \quad \dots \quad \vec{C}_M^\Omega \right] \left[\vec{C}_1^0 \quad \dots \quad \vec{C}_M^0 \right] \right)^T,$$

а \mathbf{B} — блочная матрица, составленная из следующих компонент:

$$\mathbf{B} = \begin{pmatrix} \begin{bmatrix} \mathbf{B}_{1,1}^S & \dots & \mathbf{B}_{1,M'}^S \\ \dots & \dots & \dots \\ \mathbf{B}_{M',1}^S & \dots & \mathbf{B}_{M',M'}^S \end{bmatrix} & \begin{bmatrix} \mathbf{B}_{1,1}^\Omega & \dots & \mathbf{B}_{1,M}^\Omega \\ \dots & \dots & \dots \\ \mathbf{B}_{M',1}^\Omega & \dots & \mathbf{B}_{M',M}^\Omega \end{bmatrix} & \begin{bmatrix} \mathbf{B}_{1,1}^{\Omega M} & \dots & \mathbf{B}_{1,M}^{\Omega M} \\ \dots & \dots & \dots \\ \mathbf{B}_{M',1}^{\Omega M} & \dots & \mathbf{B}_{M',M}^{\Omega M} \end{bmatrix} \\ \begin{bmatrix} \tilde{\mathbf{B}}_{1,1}^S & \dots & \tilde{\mathbf{B}}_{1,M'}^S \\ \dots & \dots & \dots \\ \tilde{\mathbf{B}}_{M,1}^S & \dots & \tilde{\mathbf{B}}_{M,M'}^S \end{bmatrix} & \begin{bmatrix} \tilde{\mathbf{B}}_{1,1}^\Omega & \dots & \tilde{\mathbf{B}}_{1,M}^\Omega \\ \dots & \dots & \dots \\ \tilde{\mathbf{B}}_{M,1}^\Omega & \dots & \tilde{\mathbf{B}}_{M,M}^\Omega \end{bmatrix} & \begin{bmatrix} \tilde{\mathbf{B}}_{1,1}^{\Omega M} & \dots & \tilde{\mathbf{B}}_{1,M}^{\Omega M} \\ \dots & \dots & \dots \\ \tilde{\mathbf{B}}_{M,1}^{\Omega M} & \dots & \tilde{\mathbf{B}}_{M,M}^{\Omega M} \end{bmatrix} \\ \begin{bmatrix} \tilde{\mathbf{B}}_{1,1}^0 & \dots & \tilde{\mathbf{B}}_{1,M'}^0 \\ \dots & \dots & \dots \\ \tilde{\mathbf{B}}_{M,1}^0 & \dots & \tilde{\mathbf{B}}_{M,M'}^0 \end{bmatrix} & \begin{bmatrix} \tilde{\mathbf{B}}_{1,1}^\Omega & \dots & \tilde{\mathbf{B}}_{1,M}^\Omega \\ \dots & \dots & \dots \\ \tilde{\mathbf{B}}_{M,1}^\Omega & \dots & \tilde{\mathbf{B}}_{M,M}^\Omega \end{bmatrix} & \begin{bmatrix} \tilde{\mathbf{B}}_{1,1}^{\Omega M} & \dots & \tilde{\mathbf{B}}_{1,M}^{\Omega M} \\ \dots & \dots & \dots \\ \tilde{\mathbf{B}}_{M,1}^{\Omega M} & \dots & \tilde{\mathbf{B}}_{M,M}^{\Omega M} \end{bmatrix} \end{pmatrix}.$$

Блочная матрица \mathbf{B} включает подматрицы:

$$\begin{aligned} \mathbf{B}_{m'_1, m'_2}^S &= \left(b_{j_1, j_2}^{S, m'_1, m'_2} \right)_{|\mathbb{M}_p| \times |\mathbb{M}_p|}; & \mathbf{B}_{m', m}^\Omega &= \left(b_{j_1, j_2}^{\Omega, m', m} \right)_{|\mathbb{M}_p| \times |\mathbb{M}_p|}; \\ \mathbf{B}_{m', m}^{\Omega M} &= \left(b_{j_1, j_2}^{\Omega M, m', m} \right)_{|\mathbb{M}_p| \times |\mathbb{M}_p|}; & \tilde{\mathbf{B}}_{m, m'}^S &= \left(\tilde{b}_{j_1, j_2}^{S, m, m'} \right)_{|\mathbb{M}_p| \times |\mathbb{M}_p|}; \\ \tilde{\mathbf{B}}_{m_1, m_2}^\Omega &= \left(\tilde{b}_{j_1, j_2}^{\Omega, m_1, m_2} \right)_{|\mathbb{M}_p| \times |\mathbb{M}_p|}; & \tilde{\mathbf{B}}_{m_1, m_2}^{\Omega M} &= \left(\tilde{b}_{j_1, j_2}^{\Omega M, m_1, m_2} \right)_{|\mathbb{M}_p| \times |\mathbb{M}_p|}; \\ \tilde{\mathbf{B}}_{m, m'}^0 &= (0)_{|\mathbb{M}_p| \times |\mathbb{M}_p|}, \text{ где } m', m'_1, m'_2 \in \{ \overline{1, M'} \}; m, m_1, m_2 \in \{ \overline{1, M} \}; \end{aligned}$$

$$b_{j_1, j_2}^{S, m'_1, m'_2} = \iint_{S_{m'_1}} \iint_{S_{m'_2}} \Psi \left(i\omega\mu \bar{\gamma}_{j_1}^{m'_1} \bar{\gamma}_{j_2}^{m'_2} - \frac{1}{i\omega\epsilon} [\nabla \cdot \bar{\gamma}_{j_1}^{m'_1}] [\nabla \cdot \bar{\gamma}_{j_2}^{m'_2}] \right) dS_2 dS_1; \quad (13)$$

$$b_{j_1, j_2}^{\Omega, m', m} = \iint_{S_{m'}} \iint_{\Omega_m} \Psi \left(i\omega\mu \bar{\gamma}_{j_1}^{m'} \bar{\gamma}_{j_2}^m - \frac{1}{i\omega\epsilon} [\nabla \cdot \bar{\gamma}_{j_1}^{m'}] [\nabla \cdot \bar{\gamma}_{j_2}^m] \right) d\Omega dS; \quad (14)$$

$$b_{j_1, j_2}^{\Omega M, m', m} = \iint_{S_{m'}} \bar{\gamma}_{j_1}^{m'} \iint_{\Omega_m} \bar{\chi}_{j_2}^m \times \nabla \Psi d\Omega dS; \quad (15)$$

$$\tilde{b}_{j_1, j_2}^{S, m, m'} = \iint_{\Omega_m} \iint_{S_{m'}} \Psi \left(i\omega\mu \bar{\gamma}_{j_1}^m \bar{\gamma}_{j_2}^{m'} - \frac{1}{i\omega\epsilon} [\nabla \cdot \bar{\gamma}_{j_1}^m] [\nabla \cdot \bar{\gamma}_{j_2}^{m'}] \right) dS d\Omega; \quad (16)$$

$$\tilde{b}_{j_1, j_2}^{\Omega, m_1, m_2} = \iint_{\Omega_{m_1}} \iint_{\Omega_{m_2}} \Psi \left(i\omega\mu \bar{\gamma}_{j_1}^{m_1} \bar{\gamma}_{j_2}^{m_2} - (i\omega\epsilon)^{-1} \times \left[\nabla \cdot \bar{\gamma}_{j_1}^{m_1} \right] \left[\nabla \cdot \bar{\gamma}_{j_2}^{m_2} \right] \right) d\Omega_2 d\Omega_1; \quad (17)$$

$$\tilde{b}_{j_1, j_2}^{\Omega_M, m_1, m_2} = \iint_{\Omega_{m_1}} \tilde{\gamma}_{j_1}^{m_1} \iint_{\Omega_{m_2}} \tilde{\chi}_{j_2}^{m_2} \times \nabla \Psi d\Omega_2 d\Omega_1. \quad (18)$$

Блочный вектор \tilde{J} состоит из подвекторов $\tilde{J}_{m'}^S = \left(\tilde{J}_{m', j}^S \right)_{|M_P|}$ в $\tilde{J}_{m'}(P) = \sum_{j \in M_P} \tilde{J}_{m', j}^S \tilde{\gamma}_j^{m'}$ при $P \in S_{m'}$; $\tilde{J}_m^\Omega = \left(\tilde{J}_{m, j}^\Omega \right)_{|M_P|}$ в $\tilde{J}_m(P) = \sum_{j \in M_P} \tilde{J}_{m, j}^\Omega \tilde{\gamma}_j^m$ при $P \in \Omega_m$; $\tilde{J}_m^{\Omega_M} = \left(\tilde{J}_{m, j}^{\Omega_M} \right)_{|M_P|}$ в $\tilde{J}_m^M(P) = \sum_{j \in M_P} \tilde{J}_{m, j}^{\Omega_M} \tilde{\chi}_j^m$ при $P \in \Omega_m$.

Блочный вектор \tilde{C} включает подвекторы $\tilde{C}_{m'}^S = \left(c_{m', j}^S \right)_{|M_P|}$; $\tilde{C}_m^\Omega = \left(c_{m, j}^\Omega \right)_{|M_P|}$; $\tilde{C}_m^0 = (0)_{|M_P|}$, где:

$$c_{m', j}^S = - \iint_{S_{m'}} \tilde{\gamma}_j^{m'} \sum_{k=1}^M \tilde{E}_{k, m'}^{\text{ind}} dS; \quad c_{m, j}^\Omega = - \iint_{\Omega_m} \tilde{\gamma}_j^m \sum_{k=1}^M \tilde{E}_{k, m}^{\text{ind}} d\Omega. \quad (19)$$

Применение процедур численного интегрирования [34] по $S_{m'}$ и Ω_m при вычислении элементов (13-19) позволяет определить искомый вектор коэффициентов $\tilde{J} = \mathbf{B}^{-1} \tilde{C}$ из (12). Следует отметить, что для корректного расчета элементов матрицы \mathbf{B} в виду сингулярности ядра $\Psi(\vec{r}, \vec{r}')$ для случаев $\vec{r} \rightarrow \vec{r}'$ используется

асимптотики функции Грина [35] $\Psi(\vec{r}, \vec{r}') = \frac{1}{4\pi} \left(R^{-1} - i\kappa + O(R) \right)$ и

[10] $\Psi(\vec{r}, \vec{r}') = \frac{1}{4\pi} \left(R^{-1} - i\kappa - 0,5\kappa^2 R + O(R) \right)$ при представлении интегралов (13)–(18) в виде суммы от регулярной и сингулярной частей с последующим вычислением последней части с учетом метода дискретных вихрей [14, 15].

Дальнейшая процедура расчета характеристик излучения ЗА по методу моментов является стандартной [4, 6].

5. Результаты моделирования и их обсуждение. Для определенного численного решения (12) в приближении БМ задачи

анализа ЗА (4) в Matlab проведено математическое моделирование. При этом с целью сравнительной оценки предпочтительности применения предложенной аппроксимации функциями (5) аналогичное решение задачи анализа ЗА также выполнено разбиением проводящей поверхности при формировании векторных базисных функций вида [31]. Параметры моделирования выбраны следующими: $\lambda = 0,05$ м; $M' = 1$; Ω'_1 — прямоугольник с высотой 14λ и шириной 30λ ; $Z_1(x, y) = 0,25(x^2 + y^2)/f$ при $f = 16\lambda$; системы координат $OXYZ$ и $O'_1X'_1Y'_1Z'_1$ совпадают; $M = 3$; Ω_1 , Ω_2 и Ω_3 — правильный треугольник, квадрат и круг соответственно, радиус вписанной окружности для которых равен λ ; $\vec{s}_1^{(o)} = \vec{s}_2^{(o)} = \vec{s}_3^{(o)} = (0 \ 0 \ -1)^T$; $x_1 = x_2 = x_3 = 0$, $y_1 = -1,5\lambda$, $y_2 = 0$, $y_3 = 1,5\lambda$, $z_1 = z_2 = z_3 = f$. На рисунке 2 представлены геометрия антенной системы и диаграмма направленности в H -плоскости ($\xi = 90^\circ$) анализируемой ЗА.

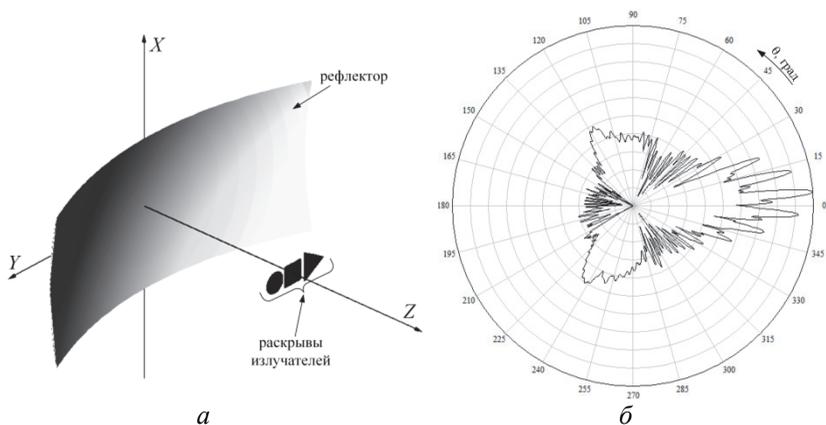


Рис. 2. Геометрия анализируемой ЗА (а) и реализуемая диаграмма направленности (б)

На рисунке 3 представлена зависимость среднеквадратического отклонения $\sigma = \sqrt{\iint_{\Omega'_1} (\vec{J}(x, y) - \vec{J}_p(x, y))^2 dx dy}$ между реальным распределением плотности электрического тока $\vec{J}_p(x, y)$ на

поверхности рефлектора и его рассчитываемым значением $\bar{J}(x, y)$ для различного порядка аппроксимации p при сравнении разработанного решения в приближении БМ с известным [31].

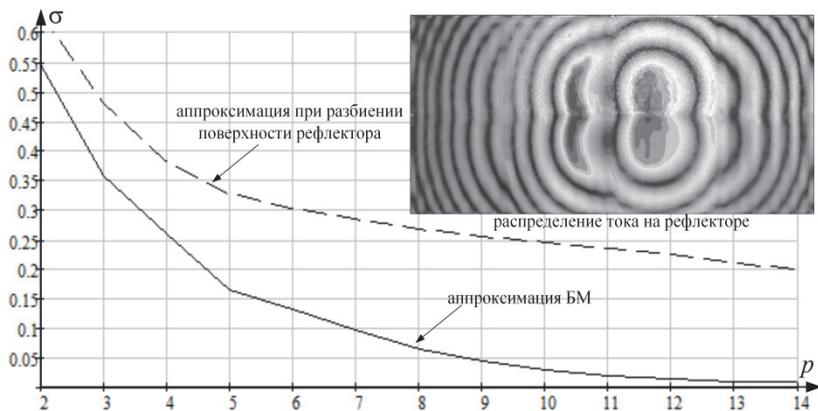


Рис. 3. Зависимость отклонения σ между $\bar{J}_p(x, y)$ и $\bar{J}(x, y)$ от порядка аппроксимации p

На рисунке 4 отражена зависимость σ от диаметра D раскрыва апертуры \mathcal{S}_k к числу узловых точек $|\mathbb{M}_p|$, выраженного в длинах волн. Следует отметить, что с целью обеспечения соизмеримых вычислительных затрат кусочно-полиномиальная аппроксимация по методу [31] выполнена при разбиении раскрыва рефлектора на два одинаковых треугольных элемента.

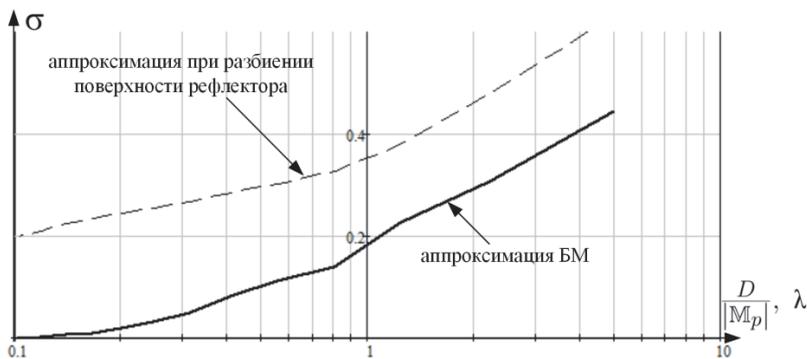


Рис. 4. Зависимость отклонения σ между $\bar{J}_p(x, y)$ и $\bar{J}(x, y)$ от $D/|\mathbb{M}_p|$, выраженного в λ

Полученные результаты моделирования (рисунки 3 и 4) определяют предпочтительность использования БМ при численном решении в приближении вариационных методов Галёркина и Ритца систем СИУ вида (4) при анализе антенных систем зеркального типа. При этом точность БМ в сравнении с существующими методами значительно увеличивается при росте порядка аппроксимации p (рисунок 3) при условии соизмеримости вычислительных затрат. Более низкая точность метода [31] для выполненного решения (рисунки 3 и 4) обусловлена тем, что разбиение раскрыва рефлектора малым числом элементарных подобластей приводит к значительным погрешностям на смежных границах между отдельными подобластями. Увеличение точности, сопоставимой с БМ, по методу [31] требует разбиения Ω'_1 на как можно большее число конечных элементов. Однако это при решении общей задачи (4) неизменно приводит к существенному росту вычислительных и емкостных затрат в сравнении с разработанным решением в приближении БМ.

6. Заключение. В целом заданные соотношения позволяют формировать эффективные алгоритмы численного решения задач, связанных с проектированием, оптимизацией и управлением зеркальных антенн с произвольным числом излучающих и отражающих элементов. Порядок аппроксимаций базисными векторными функциями $\vec{\gamma}_j^k(P)$ и $\vec{\chi}_j^k(P)$ для поверхностей \mathcal{S}_k предлагается задавать различным p_k в зависимости от отношения длины волны λ к диаметру раскрыва апертуры \mathcal{S}_k (рисунок 4) при определении числа узловых точек аппроксимации, характеризуемого мощностью множества $\left| \mathbb{M}_{p_k} \right|$ мультииндексов (7) с учетом правила $D/\left| \mathbb{M}_p \right| \leq \lambda$. Предпочтительность применения барицентрического подхода при решении систем сингулярных интегральных уравнений вида (4) заключается в удобном задании аппроксимации Ритца плотности поверхностных электрического и магнитного токов для всей области анализа в целом при учете граничных условий, уравнений Максвелла и уравнения непрерывности.

Литература

1. Проблемы антенной техники / под ред. Л.Д. Бахраха, Д.И. Воскресенского // М.: Радио и связь. 1989. 368 с.
2. *Архипов Н.С., Полянский И.С., Сомов А.М.* Анализ и структурно-параметрический синтез зеркальных антенн / под ред. А.М. Сомова // М.: Горячая линия телеком. 2017. 226 с.

3. Григорьев А.Д. Методы вычислительной электродинамики // М.: ФИЗМАТЛИТ. 2012. 432 с.
4. Gibson W.C. The method of moments in electromagnetics: second Edition // N.-Y.: Chapman and Hall/CRC. 2014. 450 p.
5. Ильинский А.С., Кравцов В.В., Свешников А.Г. Математические модели электродинамики // М.: Высшая школа. 1991. 224 с.
6. Незанов В.А., Клюев Д.С., Якунин В.С. Метод сингулярных интегральных уравнений в теории зеркальных антенн // Вестник СГАУ. 2010. № 2. С. 212–218.
7. Клюев Д.С., Соколова Ю.В. Электродинамический анализ зеркальных антенн самосогласованным методом // Журнал технической физики. 2014. Т. 84. № 9. С. 155–158.
8. Архипов Н.С., Полянский И.С., Степанов Д.Е. Бариецентрический метод в задачах анализа поля в регулярном волноводе с произвольным поперечным сечением // Антенны. 2015. № 1(212). С. 32–40.
9. Сомов А.М., Архипов Н.С., Полянский И.С., Степанов Д.Е. Расчет диаграммы направленности зеркальных антенн в приближении методов физической оптики и физической теории дифракции // Труды НИИР. 2015. № 2. С. 43–53.
10. Клюев Д.С. Электродинамическая теория зеркальных и полосковых антенн: диссертация на соискание ученой степени доктора физико-математических наук // Самара. 2012. 253 с.
11. Григорьев А.Д. Современные программные средства моделирования высокочастотных электромагнитных полей // Радиотехника и электроника. 2014. № 8(59). С. 804–808.
12. Peterson A.F., Mittra R. Computational methods for electromagnetics // Oxford University Press. 1998. 592 p.
13. Chen Z.N. et al. Handbook of Antenna Technologies // New York: Springer. 2016. 3470 p.
14. Лифанов И.К. Метод сингулярных интегральных уравнений и численный эксперимент // М.: ТОО «Янус». 1995. 520 с.
15. Вайникко Г.М., Лифанов И.К., Полтавский Л.Н. Численные методы в гиперсингулярных интегральных уравнениях и их приложения // М.: Янус-К. 2001. 508 с.
16. Борн М., Вольф Э. Основы оптики // М.: Наука. 1964. 368 с.
17. Боровиковский В.А., Кинбер Б.Е. Геометрическая теория дифракции // М.: Связь. 1978. 248 с.
18. Уфимцев П.Я. Теория дифракции краевых волн в электродинамике. Введение в физическую теорию дифракции // М.: БИНОМ. Лаборатория знаний. 2012. 372 с.
19. Russer P. Electromagnetics, microwave circuit and antenna design for communications engineering // N.-Y.: Morgan and Claypool. 2006. 123 p.
20. Jin J. The Finite Element Method in Electromagnetics: third edition // Wiley-IEEE Press. 2014. 876 p.
21. Christopoulos C. The Transmission-Line Modeling Method // Oxford: Morgan and Claypool. 2006. 124 p.
22. Jung B.H., Sakar T.K., Zhang Y. et al. Time and Frequency Domain Solutions of EM Problems Using Integral Equations and a Hybrid Methodology // New York: IEEE Press. 2010.
23. Sabariego R.V. The fast multipole method for electromagnetic field computation in numerical and physical hybrid systems: Ph.D. thesis // University of Liège. 2004.
24. Cai W., Yu Y., Yuan X.C. Singularity treatment and high-order RWG basis functions for integral equations of electromagnetic scattering // Int. J. Numerical Methods Eng. 2002. vol. 53. pp. 31–47.

25. *Полянский И.С.* Векторный барицентрический метод в вычислительной электродинамике // Труды СПИИРАН. 2017. № 2(51). С. 206–222.
26. *Корн Г., Корн К.* Справочник по математике для научных работников и инженеров // М.: Наука. 1970. 720 с.
27. *Полянский И.С.* Барицентрические координаты Пуассона–Римана // Труды СПИИРАН. 2016. № 6(49). С. 32–48.
28. *Полянский И.С.* Барицентрические координаты Пуассона для многомерной аппроксимации скалярного потенциала внутри произвольной области (Часть 1) // Вестник СГТУ. 2015. № 1(78). С. 30–36.
29. *Семёнов А.А.* Теория электромагнитных волн // М.: МГУ. 1968. 316 с.
30. *Colton D., Kress R.* Integral equation methods in scattering theory // Philadelphia: SIAM. 2013. 286 p.
31. *Chobanyan E., Ilie M., Notaros B.* Double-higher-order large-domain volume/surface integral equation method for analysis of composite wire-plate-dielectric antennas and scatterers // IEEE Transactions on antennas and propagation. 2013. vol. 61. no. 12. pp. 6051–6063.
32. *Graglia R.D., Peterson A.F., Boella M.* Higher-order techniques in computational electromagnetics // SciTech Publishing. 2016. 392 p.
33. *Родионов В.И.* О применении специальных многомерных сплайнов произвольной степени в числовом анализе // Вестник удмуртского университета. Математика. Механика. Компьютерные науки. 2010. № 4. С. 146–153.
34. *Архипов Н.С., Полянский И.С., Степанов Д.Е.* Представление отражающих поверхностей антенной системы в задачах анализа и синтеза зеркальных антенн методами физической оптики // Телекоммуникации. 2014. № 7. С. 15–21.
35. *Яковлев С.Л., Градусов В.А.* Об особенности функции Грина оператора Шрёдингера с потенциалами, сингулярными в начале координат // Вестник РУДН. 2014. № 1. С. 153–157.

Полянский Иван Сергеевич — к-т техн. наук, сотрудник, Академия Федеральной службы охраны Российской Федерации. Область научных интересов: математическое моделирование, динамические системы, дифференциальные уравнения, методы оптимизации, оптимальное управление, конформные отображения. Число научных публикаций — 117. van341@mail.ru, <http://www.mathnet.ru/rus/person117188>; Приборостроительная, 35, Орел, 302034; р.т.: +7(953) 618-71-00.

Пехов Юрий Сергеевич — сотрудник, Академия Федеральной службы охраны Российской Федерации. Область научных интересов: математическое моделирование, методы оптимизации, методы цифровой обработки сигналов, вычислительная электродинамика, теория антенн. Число научных публикаций — 4. leviathan.po4ta@gmail.com; Приборостроительная, 35, Орел, 302034; р.т.: +7(999)7550234.

I.S. POLANSKY, Y.S. PEHOV
**BARYCENTRIC METHOD IN THE SOLUTION OF SINGULAR
 INTEGRAL EQUATIONS OF ELECTRODYNAMIC THEORY OF
 REFLECTOR ANTENNAS**

Polansky I.S., Pehov Y.S. Barycentric Method in the Solution of Singular Integral Equations of Electrodynamics Theory of Reflector Antennas.

Abstract. The use of barycentric method in problems of analysis of radiation characteristics of reflector antennas is proposed. Formalization of the problem is done in the approximation of electrodynamic theory of the mirror and the strip antennas while keeping to a solution of the system of singular integral equations. In the production of barycentric method the authors propose to solve the generated system numerically using variation methods of Ritz and Galerkin. To improve the effectiveness of the solution in comparison with the known methods the approximation of Ritz for regions of interest (disclose customer reflectors and emitters) can be set without partitioning into elementary subdomains (finite elements). For a given approximation the initial problem is reduced to a system of linear equations. To determine the desirability of applying barycentric method, the authors consider a sample solution for the analysis of reflector antennas when compared with existing methods that form an approximation with basis functions by splitting the conductive surface of the antenna sub-areas of simple shapes.

Keywords: barycentric method, method of moments, singular integral equation, mirror antenna.

Polansky Ivan Sergeevich — Ph.D., researcher, The Academy of Federal Security Guard Service of the Russian Federation. Research interests: mathematical modeling, dynamical systems, differential equations, optimization methods, optimal control, conformal mapping. The number of publications — 117. van341@mail.ru, <http://www.mathnet.ru/rus/person117188>; 35, Priborostroitel'naya Street, Orel, 302034, Russia; office phone: +7(953) 618-71-00.

Pehov Yuri Sergeevich — researcher, The Academy of Federal Security Guard Service of the Russian Federation. Research interests: mathematical modeling, optimization methods, methods of digital signal processing, computational electrodynamics, the theory of antennas. The number of publications — 4. leviathan.po4ta@gmail.com; 35, Priborostroitel'naya Street, Orel, 302034, Russia; office phone: +7(999)7550234.

References

1. *Problemy antennoj tehniki. Pod red. L.D. Bahraha, D.I. Voskresenskogo* [Problems antenna technology. Edited by L.D. Bahrah, D.I. Voskresenskij]. Radio i svjaz. 1989. 368 p. (In Russ.).
2. Arkhipov N.S., Polansky I.S., Somov A.M. *Analiz i strukturno-parametricheskij sintez zerkal'nyh anten* [Analysis of structural-parametric synthesis of reflector antennas]. Moscow: Hotline Telekom. 2017. 226 p. (In Russ.).
3. Grigorev A.D. *Metody vychislitel'noj elektrodinamiki* [Methods for computational electrodynamic]. Moscow: FIZMATLIT. 2012. 432 p. (In Russ.).
4. Gibson W.C. *The method of moments in electromagnetics: second Edition*. N.-Y.: Chapman and Hall/CRC. 2014. 450 p.

5. Ilinskij A.S., Kravcov V.V., Sveshnikov A.G. *Matematicheskie modeli jelektrodinamiki* [Mathematical models of electrodynamics]. Moscow: Vysshaja shkola. 1991. 224 p. (In Russ.).
6. Neganov V.A., Kljuev D.S., Jakunin V.S. [The method of singular integral equations in the theory of reflector antennas]. *Vestnik SGAU – Herald of Samara state aerospace University*. 2010. vol. 2. pp. 212–218. (In Russ.).
7. Kljuev D.S., Sokolova Yu.V. [Electrodynamic analysis of reflector antennas the self-consistent method]. *Zhurnal tehnichej fiziki – Journal of technical physics*. 2014. vol. 9(84). pp. 155–158. (In Russ.).
8. Arkhipov N.S., Polansky I.S., Stepanov D.E. [Barycentric method in problems of analysis of the field in a regular waveguide with arbitrary cross section]. *Antenny – Antenna*. 2015. vol. 1(212). pp. 32–40. (In Russ.).
9. Somov A.M., Arhipov N.S., Poljanskij I.S., Stepanov D.E. [Calculation of the directivity pattern of reflector antennas in the approximation methods of physical optics and physical theory of diffraction]. *Trudy NIIR – Works SRIR*. 2015. vol. 2. pp. 43–53. (In Russ.).
10. Kljuev D.S. *Jelektrodinamicheskaja teorija zerkalnyh i poloskovyh antenn: dissertacija na soiskanie uchenoj stepeni doktora fiziko-matematicheskijh nauk* [Electrodynamic theory of the mirror and strip antennas: dissertation on competition of a scientific degree of the doctor of physical and mathematical Sciences]. Samara. 2012. 253 p. (In Russ.).
11. Grigorev A.D. [Modern software tools for modeling high-frequency electromagnetic fields]. *Radiotekhnika i elektronika – Technology and electronics*. 2014. vol. 8(59). pp. 804–808. (In Russ.).
12. Peterson A.F., Mittra R. *Computational methods for electromagnetic*. Oxford University Press. 1998. 592 p.
13. Chen Z.N. et al. *Handbook of Antenna Technologies*. New York: Springer. 2016. 3470 p.
14. Lifanov I.K. *Metod singularnyh integralnyh uravnenij i chislennyj jeksperiment* [Method of singular integral equations and numerical experiment]. Moscow: TOO «Janus». 1995. 520 p. (In Russ.).
15. Vajnikko G.M., Lifanov I.K., Poltavskij L.N. *Chislennye metody v gipersingularnyh integralnyh uravnenijah i ih prilozhenija* [Numerical methods in hypersingular integral equations and their applications]. Moscow: Janus-K. 2001. 508 p. (In Russ.).
16. Born M., Volf Je. *Osnovy optiki* [Principles of optics]. Moscow: Nauka. 1964. 368 p. (In Russ.).
17. Borovikovskij V.A., Kinber B.E. *Geometricheskaja teorija difrakcii* [Geometrical theory of diffraction]. Moscow: Svjaz. 1978. 248 p. (In Russ.).
18. Ufimcev P.Ya. *Teorija difrakcii kraevykh voln v jelektrodinamike. Vvedenie v fizicheskiju teoriju difrakcii* [Theory of diffraction edge waves in electrodynamics. Introduction to the physical theory of diffraction]. Moscow: BINOM. Laboratorija znaniy. 2012. 372 p. (In Russ.).
19. Russer P. *Electromagnetics, microwave circuit and antenna design for communications engineering*. Edited by C.F. Balanis. N.-Y.: Morgan and Claypool. 2006. 123 p.
20. Jin J. *The Finite Element Method in Electromagnetics: third edition*. Wiley-IEEE Press. 2014. 876 p.
21. Christopoulos C. *The Transmission-Line Modeling Method*. Oxford: Morgan and Claypool. 2006. 124 p.
22. Jung B.H., Sakar T.K., Zhang Y. et. al. *Time and Frequency Domain Solutions of EM Problems Using Integral Equations and a Hybrid Methodology*. New York: IEEE Press. 2010.

23. Sabariego R.V. The fast multipole method for electromagnetic field computation in numerical and physical hybrid systems: Ph.D. thesis. University of Liège. 2004.
24. Cai W., Yu Y., Yuan X.C. Singularity treatment and high-order RWG basis functions for integral equations of electromagnetic scattering. *Int. J. Numerical Methods Eng.* 2002. vol. 53. pp. 31–47.
25. Polansky I.S. [Vector barycentric method in computational electrodynamics]. *Trudy SPIIRAN – SPIIRAS Proceedings.* 2017. vol 2(51). pp. 206–222. (In Russ.).
26. Korn G., Korn K. *Spravochnik po matematike dlja nauchnyh rabotnikov i inzhenerov* [Handbook on mathematics for researchers and engineers]. Moscow: Nauka. 1970. 720 p. (In Russ.).
27. Polansky I.S. [Barycentric coordinates of Poisson–Riemann]. *Trudy SPIIRAN – SPIIRAS Proceedings.* 2016. vol 6(49). pp. 32–48. (In Russ.).
28. Polansky I.S. [Barycentric coordinates for the multidimensional Poisson approximation of the scalar potential inside an arbitrary region (Part 1)]. *Vestnik SGTU – Herald SGTU.* 2015. vol. 1(78). pp. 30–36. (In Russ.).
29. Semjonov A.A. *Teorija jelektromagnitnyh voln* [Theory of electromagnetic waves]. Moscow: MGU. 1968. 316 p. (In Russ.).
30. Colton D., Kress R. Integral equation methods in scattering theory. *Philadelphia: SIAM.* 2013. 286 p.
31. Chobanyan E., Ilie M., Notaros B. Double-higher-order large-domain volume/surface integral equation method for analysis of composite wire-plate-dielectric antennas and scatterers. *IEEE Transactions on antennas and propagation.* 2013. vol. 61. no. 12. pp. 6051–6063.
32. Graglia R.D., Peterson A.F., Boella M. Higher-order techniques in computational electromagnetics. SciTech Publishing. 2016. 392 p.
33. Rodionov V.I. [Finite element method for radio engineers and electrical engineers]. *Vestnik udmurtskogo universiteta. Matem. Meh. Kompjut. nauki. – University Udmurtia. Math. Mechanics. Computer science Proceedings.* 2010. vol 4. pp. 146–153. (In Russ.).
34. Arhipov N.S., Poljanskij I.S., Stepanov D.E. [Representation of reflecting surfaces of the antenna system in the tasks of analysis and synthesis of reflector antennas by the methods of physical optics]. *Telekommunikacii – Telecommunications.* 2014. vol. 7. pp. 15–21. (In Russ.).
35. Yakovlev S.L., Gradusov V.A. [On peculiarities of the green's function of the schrödinger operator with potentials singular at the origin]. *Vestnik RUDN – Herald RIFP.* 2014. vol. 1. pp. 153–157. (In Russ.).

Signed to print 29.09.2017

Printed in Publishing center GUAP, 67, B. Morskaya, St. Petersburg, 190000, Russia

The journal is registered in Russian Federal Agency for Communications
and Mass-Media Supervision, certificate ПИ № ФС77-41695 dated August 19, 2010 г.
Subscription Index П5513, Russian Post Catalog

Подписано к печати 29.09.2017. Формат 60×90 1/16. Усл. печ. л. 15,2. Заказ № 420.

Тираж 150 экз., цена свободная

Отпечатано в Редакционно-издательском центре ГУАП, 190000, Санкт-Петербург, Б. Морская, д. 67

Журнал зарегистрирован Федеральной службой по надзору в сфере связи
и массовых коммуникаций,
свидетельство ПИ № ФС77-41695 от 19 августа 2010 г.

Подписной индекс П5513 по каталогу «Почта России»

РУКОВОДСТВО ДЛЯ АВТОРОВ

Взаимодействие автора с редакцией осуществляется через личный кабинет на сайте журнала «Труды СПИИРАН» <http://www.proceedings.spiiras.nw.ru>. При регистрации авторам рекомендуется заполнить все предложенные поля данных.

Подготовка статьи ведется с помощью текстовых редакторов MS Word 2007 и выше. Объем основного текста – от 15 до 25 страниц включительно. Формат страницы документа – А5 (148 мм ширина, 210 мм высота); ориентация – портретная; все поля – 20 мм. Верхний и нижний колонтитулы страницы – пустые. Основной шрифт документа – Times New Roman, основной кегль (размер) шрифта – 10 pt. Переносы разрешены. Абзацный отступ устанавливается размером в 10 мм. Межстрочный интервал – одинарный. Номера страниц не проставляются.

В основную часть допускается помещать рисунки, таблицы, листинги и формулы. Правила их оформления подробно рассмотрены на нашем сайте в разделе «Руководство для авторов».

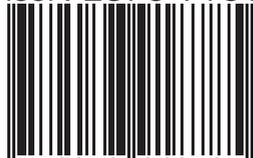
AUTHOR GUIDELINES

Interaction between each potential author and the Editorial board is realized through the personal account on the website of the journal "Proceedings of SPIIRAS" <http://www.proceedings.spiiras.nw.ru>. At the registration the authors are requested to fill out all data fields in the proposed form.

The submissions should be prepared using MS Word 2007 text editor or higher versions, at that, only manuscripts in *.docx format will be considered. The text of the paper in the main part of it should be from 15 – 25 pages of A5 size that is 210 X 148 mm; orientation – portrait; all margins – 20 mm. The font of the main paper text is Times New Roman of 10 pt font size. The pages' headers and footers should be empty; indentation – 10 mm; line spacing – single; pages are not numbered; hyphenations are allowed.

Certain figures, tables, listings and formulas are allowed in the main section, and their typography is considered by the paper template in more detail in journal web.

ISSN 2078-9181



9 772078 918785 >

