

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

ИНФОРМАЦИОННО- УПРАВЛЯЮЩИЕ СИСТЕМЫ

НАУЧНЫЙ ЖУРНАЛ

3(106)/2020

3(106)/2020

PEER REVIEWED JOURNAL

INFORMATSIONNO- UPRAVLIAIUSHCHIE SISTEMY (INFORMATION AND CONTROL SYSTEMS)

Founder

«Information and Control Systems», Ltd.

PublisherSaint-Petersburg State University
of Aerospace Instrumentation**Editor-in-Chief**

M. Sergeev

Dr. Sc., Professor, Saint-Petersburg, Russia

Deputy Editor-in-Chief

E. Krouk

Dr. Sc., Professor, Moscow, Russia

Executive secretary

O. Muravtsova

Editorial Board

S. Andreev

Dr. Sc., Tampere, Finland

V. Anisimov

Dr. Sc., Professor, Saint-Petersburg, Russia

B. Bezruchko

Dr. Sc., Professor, Saratov, Russia

N. Blaunstein

Dr. Sc., Professor, Beer-Sheva, Israel

C. Christodoulou

PhD, Professor, Albuquerque, New Mexico, USA

A. Dudin

Dr. Sc., Professor, Minsk, Belarus

I. Dumer

PhD., Professor, Riverside, USA

M. Favorskaya

Dr. Sc., Professor, Krasnoyarsk, Russia

L. Fortuna

PhD, Professor, Catania, Italy

A. Fradkov

Dr. Sc., Professor, Saint-Petersburg, Russia

A. Hramov

Dr. Sc., Professor, Innopolis, Russia

L. Jain

PhD, Professor, Canberra, Australia

V. Khimenko

Dr. Sc., Professor, Saint-Petersburg, Russia

G. Maltsev

Dr. Sc., Professor, Saint-Petersburg, Russia

G. Matvienko

Dr. Sc., Professor, Tomsk, Russia

A. Myllari

PhD, Professor, Grenada, West Indies

Y. Podoplyokin

Dr. Sc., Professor, Saint-Petersburg, Russia

K. Samouylov

Dr. Sc., Professor, Moscow, Russia

J. Seberry

PhD, Professor, Wollongong, Australia

A. Shalyto

Dr. Sc., Professor, Saint-Petersburg, Russia

A. Shepeta

Dr. Sc., Professor, Saint-Petersburg, Russia

Yu. Shokin

RAS Academician, Dr. Sc., Novosibirsk, Russia

A. Smirnov

Dr. Sc., Professor, Saint-Petersburg, Russia

T. Sutikno

PhD, Associate Professor, Yogyakarta, Indonesia

Z. Yuldashev

Dr. Sc., Professor, Saint-Petersburg, Russia

R. Yusupov

RAS Corr. Member, Dr. Sc., Professor, Saint-Petersburg, Russia

A. Zeifman

Dr. Sc., Professor, Vologda, Russia

Editor: A. Larionova**Proofreader:** T. Zvertanovskaia**Design:** M. Chernenko, Y. Umnikina**Layout and composition:** Y. Umnikina**Contact information**

The Editorial and Publishing Center, SUAI

67, B. Morskaia, 190000, St. Petersburg, Russia

Website: <http://i-us.ru/en>, e-mail: i-us.spb@gmail.com

Tel.: +7 - 812 494 70 02

INFORMATION PROCESSING AND CONTROL**Pyataev A. S., Favorskaya M. N.** Classification of tree stand condition using visual data based on fuzzy logic 2**Volkov V. Yu.** Adaptive multi-threshold object selection in remote sensing images 12**INFORMATION AND CONTROL SYSTEMS****Geidarov P. Sh.** Algorithm for calculating synapse weights of the first layer of a neural network on the base of metric recognition methods. Part 2 25**Arseniev V. N., Khomonenko A. D., Yadrenkin A. A.** Weighed ranking of aprioristic and experimental data in control system functioning efficiency estimation problem with Pascal-distributed number of tests 39**SYSTEM AND PROCESS MODELING****Doronina Yu. V., Skatkov A. V.** Cascade-hierarchical modeling in analyzing the dynamics of complex system resource characteristics 48**INFORMATION SECURITY****Le D. T., Dao M. H., Nguyen Q. L. T.** Comparison of machine learning algorithms for DDoS attack detection in SDN 59**INFORMATION CODING AND TRANSMISSION****Moldovyan D. N., Moldovyan A. A., Gurianov D. Yu.** Blind signature protocols based on hidden discrete logarithm problem 71**INFORMATION CHANNELS AND MEDIUM****Burkov A. A., Shneer S. V., Turlikov A. M.** Lower bound for average delay in unblocked random access algorithm with orthogonal preambles 79**INFORMATION ABOUT THE AUTHORS** 86

3(106)/2020

ИНФОРМАЦИОННО-
УПРАВЛЯЮЩИЕ
СИСТЕМЫ

РЕЦЕНЗИРУЕМОЕ ИЗДАНИЕ

Учредитель
ООО «Информационно-управляющие системы»
Издатель
Санкт-Петербургский государственный университет
аэрокосмического приборостроения
Главный редактор
М. Б. Сергеев,
д-р техн. наук, проф., Санкт-Петербург, РФ
Зам. главного редактора
Е. А. Крук,
д-р техн. наук, проф., Москва, РФ
Ответственный секретарь
О. В. Муравцова
Редакционная коллегия:
С. Д. Андреев,
д-р техн. наук, Тампере, Финляндия
В. Г. Анисимов,
д-р техн. наук, проф., Санкт-Петербург, РФ
Б. П. Безручко,
д-р физ.-мат. наук, проф., Саратов, РФ
Н. Блаунштейн,
д-р физ.-мат. наук, проф., Беэр-Шева, Израиль
Л. С. Джайн,
д-р наук, проф., Канберра, Австралия
А. Н. Дудин,
д-р физ.-мат. наук, проф., Минск, Беларусь
И. И. Думер,
д-р наук, проф., Риверсайд, США
А. И. Зейфман,
д-р физ.-мат. наук, проф., Вологда, РФ
К. Кристоделу,
д-р наук, проф., Альбукерке, Нью-Мексико, США
Г. Н. Мальцев,
д-р техн. наук, проф., Санкт-Петербург, РФ
Г. Г. Матвиенко,
д-р физ.-мат. наук, проф., Томск, РФ
А. А. Мюллери,
д-р наук, профессор, Гренада, Вест-Индия
Ю. Ф. Подоплёкин,
д-р техн. наук, проф., Санкт-Петербург, РФ
К. Е. Самуилов,
д-р техн. наук, проф., Москва, РФ
Д. Себерри,
д-р наук, проф., Волонгонг, Австралия
А. В. Смирнов,
д-р техн. наук, проф., Санкт-Петербург, РФ
Т. Сутикнуо,
д-р наук, доцент, Джокьякарта, Индонезия
М. Н. Фаворская,
д-р техн. наук, проф., Красноярск, РФ
Л. Фортуна,
д-р наук, проф., Катания, Италия
А. Л. Фрадков,
д-р техн. наук, проф., Санкт-Петербург, РФ
В. И. Хищенко,
д-р техн. наук, проф., Санкт-Петербург, РФ
А. Е. Храмов,
д-р физ.-мат. наук, Иннополис, РФ
А. А. Шалыто,
д-р техн. наук, проф., Санкт-Петербург, РФ
А. П. Шепета,
д-р техн. наук, проф., Санкт-Петербург, РФ
Ю. И. Шокин,
акад. РАН, д-р физ.-мат. наук, проф., Новосибирск, РФ
З. М. Юлдашев,
д-р техн. наук, проф., Санкт-Петербург, РФ
Р. М. Юсупов,
чл.-корр. РАН, д-р техн. наук, проф., Санкт-Петербург, РФ
Редактор: А. Г. Ларионова
Корректор: Т. В. Звертановская
Дизайн: М. Л. Черненко, Ю. В. Умницына
Компьютерная верстка: Ю. В. Умницына
Адрес редакции: 190000, Санкт-Петербург,
Б. Морская ул., д. 67, ГУАП, РИЦ
Тел.: (812) 494-70-02, эл. адрес: ius.spb@gmail.com,
сайт: http://i-us.ru

Журнал зарегистрирован в Министерстве РФ по делам печати,
телерадиовещания и средств массовых коммуникаций.
Свидетельство о регистрации ПИ № 77-12412 от 19 апреля 2002 г.
Перерегистрирован в Роскомнадзоре.
Свидетельство о регистрации ПИ № ФС77-49181 от 30 марта 2012 г.
© Коллектив авторов, 2020

ОБРАБОТКА ИНФОРМАЦИИ И УПРАВЛЕНИЕ

Пятаев А. С., Фаворская М. Н. Классификация состояния деревьев
по визуальным данным на основе нечеткой логики 2

Волков В. Ю. Адаптивная многопороговая селекция объектов
на изображениях в системах дистанционного зондирования 12

ИНФОРМАЦИОННО-УПРАВЛЯЮЩИЕ СИСТЕМЫ

Гейдаров П. Ш. Алгоритм вычисления значений весов синапсов
первого слоя нейронной сети на основе метрических методов
распознавания. Часть 2 25

Арсеньев В. Н., Хомоненко А. Д., Ядренкин А. А. Взвешенный учет
априорной и опытной информации в задаче оценивания эффективности
функционирования системы управления при распределении числа
испытаний по закону Паскаля 39

МОДЕЛИРОВАНИЕ СИСТЕМ И ПРОЦЕССОВ

Доронина Ю. В., Скاتков А. В. Каскадно-иерархическое моделиро-
вание в задачах анализа динамики ресурсных характеристик
сложных систем 48

ЗАЩИТА ИНФОРМАЦИИ

Le D. T., Dao M. H., Nguyen Q. L. T. Comparison of machine learning
algorithms for DDoS attack detection in SDN 59

КОДИРОВАНИЕ И ПЕРЕДАЧА ИНФОРМАЦИИ

Молдовян Д. Н., Молдовян А. А., Гурьянов Д. Ю. Протоколы слепой
цифровой подписи на основе скрытой задачи дискретного
логарифмирования 71

ИНФОРМАЦИОННЫЕ КАНАЛЫ И СРЕДЫ

Burkov A. A., Shneer S. V., Turlikov A. M. Lower bound for average delay
in unblocked random access algorithm with orthogonal preambles 79

СВЕДЕНИЯ ОБ АВТОРАХ

86

Журнал входит в БД SCOPUS, в RSCI на платформе Web of Science
и в Перечень рецензируемых научных изданий, в которых должны быть опубликованы
основные научные результаты диссертаций на соискание ученой степени
кандидата наук, на соискание ученой степени доктора наук.

Сдано в набор 06.05.20. Подписано в печать 19.06.20. Формат 60×84^{1/8}.
Гарнитура SchoolBookC. Печать цифровая.
Усл. печ. л. 10,4. Уч.-изд. л. 14,2. Тираж 1000 экз (1-й завод 50 экз.). Заказ № 180.

Оригинал-макет изготовлен в редакционно-издательском центре ГУАП.
190000, Санкт-Петербург, Б. Морская ул., 67.
Отпечатано с готовых диапозитивов в редакционно-издательском центре ГУАП.
190000, Санкт-Петербург, Б. Морская ул., 67.

УДК 004.932

doi:10.31799/1684-8853-2020-3-2-11

Классификация состояния деревьев по визуальным данным на основе нечеткой логики

А. С. Пятаев^{а, б}, аспирант, orcid.org/0000-0001-5489-8555

М. Н. Фаворская^а, доктор техн. наук, профессор, orcid.org/0000-0002-2181-0454, favorskaya@sibsau.ru

^аСибирский государственный университет науки и технологий им. академика М. Ф. Решетнёва, Красноярский рабочий пр., 31, Красноярск, 660037, РФ

^бфилиал ФБУ «Рослесозащита» – «ЦЗЛ Красноярского края», Академгородок ул., 50А/№ 2, Красноярск, 660036, РФ

Введение: достоверная оценка состояния деревьев непосредственно влияет на планирование экономических показателей использования лесных ресурсов и экологических мер по защите леса, поэтому корректное определение санитарного состояния леса является актуальной задачей. В настоящее время специалист-лесопатолог принимает решение о том, к какой категории состояния следует отнести исследуемое дерево и участок леса в целом, на основе визуального осмотра, руководствуясь субъективными представлениями об особенностях деревьев. **Цель:** разработка метода классификации состояния деревьев по степени густоты кроны, годичному приросту ветвей, усыханию ветвей и отпаду коры на основе нечеткой логики путем создания характеристических функций для лингвистических переменных «Густота кроны», «Годичный прирост ветвей», «Усыхание ветвей», «Отпад коры». **Результаты:** предложен метод классификации состояния деревьев на примере сосны. Метод заключается в предварительной обработке изображений, включающей удаление объектов фона, извлечении текстурных признаков в виде расширенных бинарных шаблонов и применении разработанного контроллера на основе нечеткой логики. Предложены четыре вида лингвистических переменных с соответствующими термами. Для них вначале задаются характеристические функции в табличном виде, которые затем аппроксимируются гладкими функциями. Нечеткий логический контроллер позволяет получить объективную оценку состояния кроны дерева. Экспериментальные исследования подтверждают эффективность разработанного метода. **Практическая значимость:** использование интеллектуальной системы классификации состояния деревьев по визуальным данным окажет существенную поддержку специалистам при обследовании насаждений. Предложенный в работе метод позволяет повысить качество проводимых обследований состояния леса, минимизировать влияние человеческого фактора, а также правильно оценить и спланировать комплекс мероприятий по защите лесных насаждений.

Ключевые слова – классификация, состояние деревьев, нечеткая логика, нечеткий контроллер, обработка изображений, текстурный анализ.

Для цитирования: Пятаев А. С., Фаворская М. Н. Классификация состояния деревьев по визуальным данным на основе нечеткой логики. *Информационно-управляющие системы*, 2020, № 3, с. 2–11. doi:10.31799/1684-8853-2020-3-2-11

For citation: Pyataev A. S., Favorskaya M. N. Classification of tree stand condition using visual data based on fuzzy logic. *Informatsionno- upravliaiushchie sistemy* [Information and Control Systems], 2020, no. 3, pp. 2–11 (In Russian). doi:10.31799/1684-8853-2020-3-2-11

Введение

В Российской Федерации установлена шкала категорий состояния деревьев, включающая следующие категории: I — без признаков ослабления, II — ослабленные, III — сильно ослабленные, IV — усыхающие, V — свежий сухостой (текущего года) и VI — старый сухостой (прошлых лет) [1]. Интегральная оценка состояния деревьев проводится по комплексу визуальных признаков (густоте и цвету кроны, наличию и доли усохших ветвей в кроне, состоянию коры и др.).

Использование интеллектуальной системы оценки состояния деревьев может оказать существенную поддержку специалистам при обследовании лесных насаждений. От корректности определения санитарного состояния леса зависит планирование экономических показателей и экологических мер по защите леса, включающих различные санитарно-оздоровительные меропр

ятия, поэтому корректное определение санитарного состояния леса является актуальной задачей. В настоящее время специалист-лесопатолог, руководствуясь субъективными представлениями об особенностях деревьев, на основе визуального осмотра принимает решение, к какой категории состояния следует отнести исследуемое дерево и участок леса в целом. При этом степень густоты кроны имеет определяющее значение в данном вопросе.

Задача является сложной, так как, несмотря на наличие шкалы категорий состояния деревьев, отнесение конкретного дерева к той или иной категории происходит с большой долей субъективизма и зависит прежде всего от опыта специалиста. Требуется разработка интеллектуальных решений, основанных, в частности, на нечеткой логике, которая позволяет формализовать процесс классификации состояния деревьев по объективным визуальным признакам.

Анализ существующих методов определения состояния деревьев

Плотность кроны является важнейшим показателем для оценки состояния дерева. Как правило, деревья фотографируют специалисты-лесопатологи для получения наиболее информативных снимков. Методы обработки таких изображений основаны на разных способах сегментации.

Допустима оценка состояния кроны дерева с использованием фрактальных признаков. В работе [2] рассчитывались две фрактальные размерности — силуэта дерева DS и силуэта контура DO (разность этих показателей обозначается DSO), по которым оценивался показатель прозрачности кроны для различных пород деревьев.

В работе [3] исследованы несколько способов обнаружения листьев с учетом особенностей изображений деревьев на фоне ярко освещенного неба и наличия облаков. Показано, что метод Отцу в применении к RGB-изображениям приводит к существенному пропуску мелких деталей. Лучшим способом является построение гистограмм, особенно по компоненте насыщенность — Hue — в цветовом пространстве HSI (Hue, Saturation, Intensity). Текстуриный анализ, представленный признаками амплитуды, направления и периодичности, совместно с контурным анализом позволил авторам провести кластеризацию регионов на основе метода *C-means*. Следует отметить, что те же авторы [4] опубликовали связанную статью, где предложили метод автоматического извлечения границ кроны дерева с применением сплайн-функций. Позже была предпринята попытка проводить мониторинг состояния крон деревьев [5] с использованием двух мер прозрачности кроны, а именно компактности и фрактальной размерности DSO [2]. Мониторинг выполняется на основе анализа построенных карт распределения прозрачности в разные моменты времени.

В соответствии с разработанной методикой сотрудники Южной научно-исследовательской станции Лесной службы США (the U.S. Forest Service Southern Research Station) разработали программу *ForestCrowns* [6, 7], которая рассчитывает прозрачность кроны (или коэффициент пропускания света) по цифровым изображениям, полученным от цифровых фотоаппаратов со стандартными объективами или объективами типа «рыбий глаз» (полусферическими). При этом некоторые фотографии можно использовать для оценки прозрачности крон отдельных деревьев. Программа позволяет анализировать набор фотографий. Выходными данными являются значения коэффициента прозрачности для указанных пользователем областей изображения. Сравнительные оценки состояния кроны приводятся в работе [8].

В статье [9] предлагается метод расчета коэффициента прозрачности кроны по бинарным изображениям деревьев, основанный на автоматическом определении трех показателей: плотности верхней части кроны, плотности макроотверстий и плотности микроотверстий. Вначале извлекается силуэт дерева, затем формируются срезы кроны и строятся профили глубины. Индекс прозрачности кроны вычисляется как псевдовзвешенное среднее по эмпирической формуле, учитывающей все три показателя плотности. На основе полученных значений изображения объединяются в кластеры.

С развитием методов глубокого обучения появляются работы, оценивающие состояние деревьев по новым технологиям. Так, в [10] классификация видов деревьев и оценка объема запаса древесины осуществляются по изображениям, полученным от цифровой камеры. Семантическая сегментация крон и стволов выполняется глубокой сетью UNET, предварительно обученной по модели VGG16. В результате определяются породы деревьев и подсчитывается их количество. Однако в этом исследовании не учитывается состояние деревьев. В статье [11] представлен метод обнаружения крон деревьев на цветных изображениях, полученных с использованием средств дистанционного зондирования (беспилотных летательных аппаратов), на основе глубокого обучения.

Оригинальный метод измерения общей площади листьев в кронах деревьев на основе моделирования многоплатформенных лидарных данных позволяет количественно оценить степень перекрытия крон деревьев при различных стратегиях сканирования [12]. Были построены пять моделей трехмерных виртуальных деревьев. Для построения сканированной поверхности листа из точечного облака использовался алгоритм триангуляции Делоне с автоматическим адаптивным выбором порога, а площадь листа оценивалась по количеству сканированных точек.

Отметим, что связанной задачей является сегментация кроны дерева [13], которая в настоящее время выполняется с использованием визуальных и лидарных данных [14–16]. Краткий обзор показывает, что исследования по оценке состояния деревьев, проводимые с помощью наземных средств получения изображений, не прекращаются и требуют дальнейшего совершенствования.

Метод классификации состояния деревьев

Метод определения состояния дерева по визуальным данным состоит из трех этапов. На первом этапе удаляют объекты фона, применяя пороговую обработку. Далее по результатам тек-

стурного анализа определяют показатель густоты кроны. На третьем этапе проводят оценку состояния дерева на основе нечеткой логики и относят его к одной из категорий.

Удаление объектов фона

В реальных условиях на фотографиях обследуемых деревьев присутствуют посторонние объекты фона, что существенно затрудняет классификацию степени густоты кроны дерева, поэтому для успешного дальнейшего анализа требуется отделить изображение дерева от фона. Удаляют объекты фона с помощью пороговой обработки с глобальным порогом, который определяется на основе градиента яркости пикселей изображения [17]. Градиент $\nabla I(x, y)$ вычисляется следующим образом:

$$\nabla I(x, y) = \max(|I_x(x, y)|, |I_y(x, y)|),$$

где $I(\cdot)$ — функция, характеризующая изображение; (x, y) — координаты;

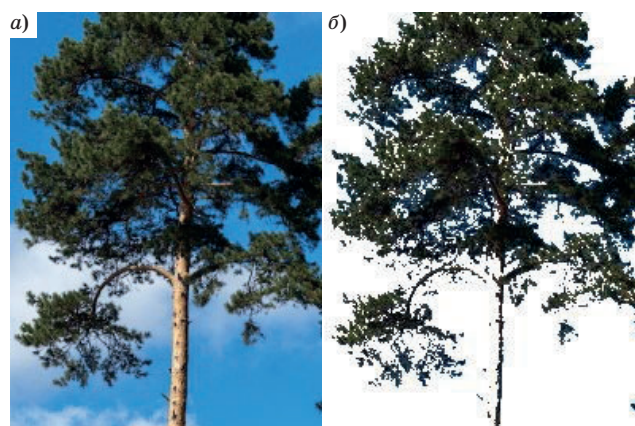
$$I_x(x, y) = I_x(x + 1, y) - I_x(x - 1, y);$$

$$I_y(x, y) = I_y(x, y + 1) - I_y(x, y - 1).$$

Пороговое значение T вычисляется по формуле

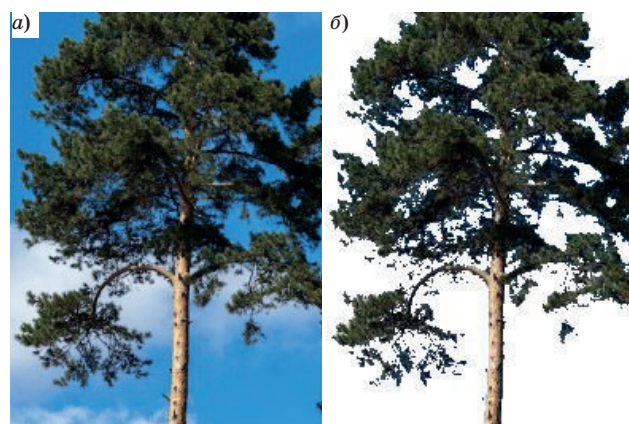
$$T = \frac{\sum_x \sum_y I(x, y) \nabla I(x, y)}{\sum_x \sum_y \nabla I(x, y)}.$$

При этом работа алгоритма удаления фона может существенно зависеть от выбора цветовой модели. Так, использование цветовой модели



■ **Рис. 1.** Удаление объектов фона на основе градиента яркости пикселей: *a* — исходное изображение; *б* — объект интереса

■ **Fig. 1.** Removal of background objects based on intensity gradient: *a* — original image; *б* — object of interest



■ **Рис. 2.** Удаление объектов фона на основе градиента яркости и цветового тона пикселей: *a* — исходное изображение; *б* — объект интереса

■ **Fig. 2.** Removal of background objects based on intensity gradient and color tone: *a* — original image; *б* — object of interest

YUV в некоторых случаях дает некорректный результат. На рис. 1, *a* и *б* приведен пример такой ситуации, где за фон принимается ярко освещенная часть ствола.

Использование цветовой модели HSB (Hue, Saturation, Brightness) позволило устранить этот недостаток (рис. 2, *a* и *б*). Сегментация изображения проводится по двум параметрам: цветовому тону (Hue) и яркости (Brightness).

При этом функция изображения с удаленными объектами фона $I^*(\cdot)$ имеет вид

$$I^*(x, y) = \begin{cases} I(x, y) & \text{при } H(x, y) < T_H \vee B(x, y) < T_B \\ 256 & \text{в противном случае} \end{cases}$$

Таким образом, для удаления объектов фона применен метод пороговой обработки с глобальным порогом с помощью цветовой модели HSB.

Текстурный анализ

Для определения степени густоты исследуемого фрагмента кроны использованы расширенные локальные бинарные шаблоны (Extended Local Binary Patterns — ELBP), которые вычисляются следующим образом:

$$ELBP(p) = \sum_{n=0}^p s(I_n - I_c) \cdot 2^n, \quad (*)$$

где p — число точек в окрестности ELBP; $s(I_n - I_c) = 1$, если $(I_n - I_c) \geq 0$, и $s(I_n - I_c) = 0$ в противном случае; I_n и I_c — значения яркости текущего и центрального пикселей, представляющие собой значение компоненты Y цветового пространства YUV [18]. Как видно из выражения (*), вычисление расширенных локальных бинарных шабло-

нов происходит аналогично вычислению классических локальных бинарных шаблонов. Отличие расширенных локальных бинарных шаблонов от классических заключается в способе построения гистограмм по полученным бинарным строкам. ELBP позволяют учесть такие особенности изображения, как концы линий, грани, углы и пятна с использованием для каждой из таких особенностей отдельного столбца гистограммы. Для всех неравномерных шаблонов при построении гистограммы отводится отдельный столбец.

В качестве меры различия гистограмм целесообразно рассчитать расстояние Кульбака — Лейблера

$$D_{K,L}(f, g) = \sum_{m=1}^{p(p-1)+3} f_m \ln \frac{f_m}{g_m},$$

где f и g — гистограммы первого и второго изображения; m — номер столбца.

Решение о степени густоты кроны исследуемого дерева принимается после подсчета количества блоков (фрагментов) вычисленной густоты с учетом их расположения: на нижней, верхней или средней части дерева. Для более точного отнесения изображения исследуемой текстуры к тому или иному классу степени густоты кроны следует учитывать масштаб и размер, занимаемый деревом на изображении.

Оценка по глобальным признакам на основе нечеткой логики

Следующим этапом определения степени густоты кроны является сопоставление определенного на предыдущем этапе процента густоты кроны со степенью густоты кроны, указанной в Постановлении Правительства РФ от 20.05.2017 № 607 «О Правилах санитарной безопасности в лесах» [1]. Сложность такого сопоставления состоит в размытости понятий, определяющих категорию состояния обследуемого дерева. Для установления такого соответствия в настоящей работе был использован подход, основанный на применении нечеткой логики. Для каждой категории состояния дерева определены следующие лингвистические переменные: «Густота кроны», «Годичный прирост ветвей», «Усыхание ветвей», «Отпад коры» (табл. 1).

Были выделены границы и вид термов лингвистических переменных. Изначально характеристические функции лингвистических переменных задаются в табличном виде (табл. 2–5 для четырех типов лингвистических переменных). В дальнейшем характеристические функции лингвистических переменных «Густота кроны» и «Годичный прирост ветвей», заданные в табличном виде, аппроксимируются гладкими функциями [19] как наиболее значимые параме-

■ **Таблица 1.** Лингвистические переменные и их термы [19]

■ **Table 1.** Linguistic variables and their terms [19]

Переменная	Термы
Густота кроны	Густая, разреженная, ажурная, сильно ажурная
Годичный прирост ветвей	Нормальный, уменьшен, слабый, очень слабый или отсутствует
Усыхание ветвей	Отсутствует, засохли отдельные ветви, усыхание до 2/3 кроны, усыхание более 2/3 кроны
Отпад коры	Отсутствует, частичный, полный

■ **Таблица 2.** Узловые точки значений термов лингвистической переменной «Густота кроны»

■ **Table 2.** Node points of term values of linguistic variable “Crown density”

Густота кроны, %	Густая	Разреженная	Ажурная	Сильно ажурная
100	1	0	0	0
95	0,8	0,2	0	0
90	0,3	0,6	0	0
85	0,1	1	0	0
80	0	1	0	0
75	0	0,6	0,3	0
70	0	0,2	0,6	0
65	0	0	1	0
60	0	0	1	0
55	0	0	0,6	0,2
50	0	0	0,3	0,3
45	0	0	0,2	0,6
40	0	0	0	0,8
35	0	0	0	1
30	0	0	0	1
25	0	0	0	1
20	0	0	0	1
15	0	0	0	1
10	0	0	0	1
5	0	0	0	1
0	0	0	0	1

тры для определения состояния дерева. Графики аппроксимированных термов для четырех типов лингвистических переменных представлены на рис. 3–6.

■ **Таблица 3.** Узловые точки значений термов лингвистической переменной «Годичный прирост ветвей»

■ **Table 3.** Node points of term values of linguistic variable “Annual branches growth”

Прирост, мм	Нормальный	Уменьшен	Слабый	Очень слабый
0	0	0	0	1
5	0	0	0	1
10	0	0	0	1
15	0	0	0,2	0,8
20	0	0,1	0,5	0,5
25	0	0,2	1	0,2
30	0	0,5	0,5	0,1
35	0	0,8	0,2	0
40	0,2	1	0	0
45	0,3	0,8	0	0
50	0,5	0,5	0	0
55	0,7	0,2	0	0
60	0,8	0,1	0	0
65	0,9	0	0	0
70	1	0	0	0
75	1	0	0	0
80	1	0	0	0

■ **Таблица 4.** Узловые точки значений термов лингвистической переменной «Усыхание ветвей» [19]

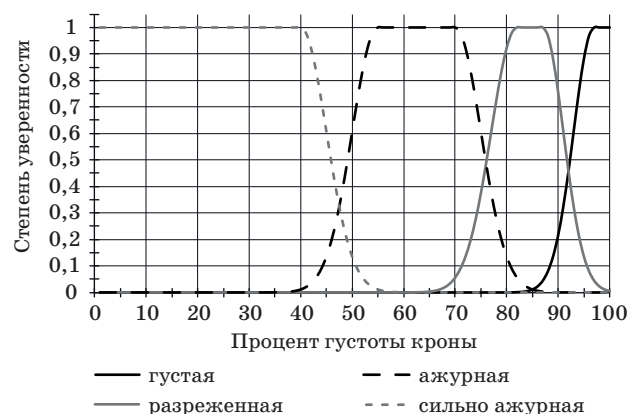
■ **Table 4.** Node points of term values of linguistic variable “Shrinking branches” [19]

Усыхание ветвей, %	Засохли отдельные ветви	Усыхание до 2/3 кроны	Усыхание более 2/3 кроны
0	1	0	0
10	0,5	0,5	0
20	0	1	0
30	0	1	0
40	0	1	0
50	0	1	0
60	0	1	0
70	0	0,5	0,5
80	0	0	1
90	0	0	1
100	0	0	1

■ **Таблица 5.** Узловые точки значений термов лингвистической переменной «Отпад коры»

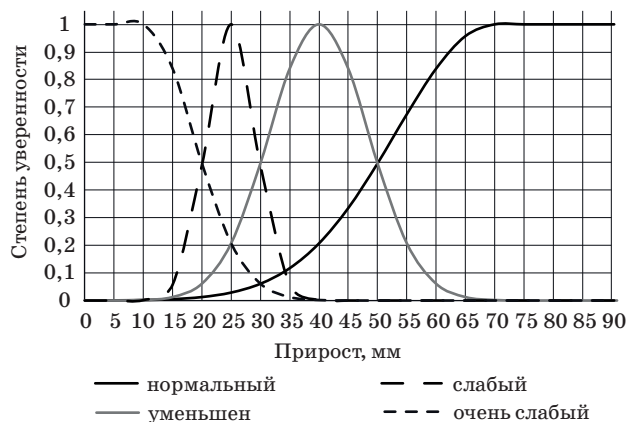
■ **Table 5.** Node points of term values of linguistic variable “Bark falling”

Отпад коры, %	Полный	Частичный	Отсутствует
0	0	0	1
10	0	0,5	0,5
20	0	1	0
30	0	1	0
40	0	1	0
50	0	1	0
60	0	1	0
70	0	1	0
80	0	1	0
90	0,5	0,5	0
100	1	0	0



■ **Рис. 3.** Графики аппроксимированных термов лингвистической переменной «Густота кроны»

■ **Fig. 3.** Graphs of approximated terms of linguistic variable “Crown density”



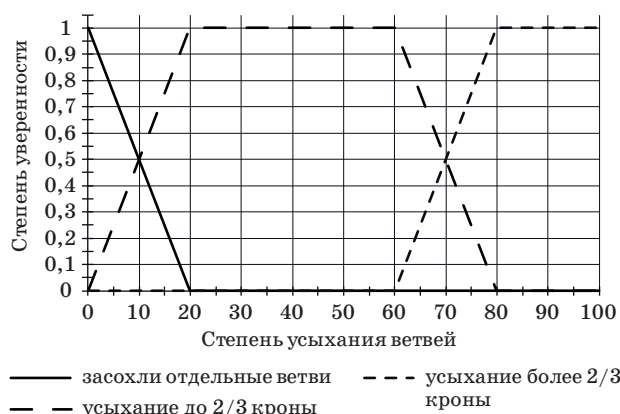
■ **Рис. 4.** Графики термов лингвистической переменной «Годичный прирост ветвей»

■ **Fig. 4.** Graphs of approximated terms of linguistic variable “Annual branches growth”



■ **Рис. 5.** Графики термов лингвистической переменной «Усыхание ветвей»

■ **Fig. 5.** Graphs of approximated terms of linguistic variable “Shrinking branches”



■ **Рис. 6.** Графики термов лингвистической переменной «Отпад коры»

■ **Fig. 6.** Graphs of approximated terms of linguistic variable “Bark falling”

На основе характеристики категорий состояний хвойных пород, определенных постановлением правительства РФ [1], построена база нечетких правил. Набор нечетких правил, позволяющий сделать заключение о категории состояния обследуемого дерева, имеет следующий вид.

1. IF («Густота кроны» = «Густая») AND («Прирост» = «Нормальный») AND («Усыхание ветвей» = «Отсутствует») AND («Отпад коры» = «Отсутствует») THEN («Состояние» = «Без признаков ослабления»).

2. IF («Густота кроны» = «Разреженная») AND («Прирост» = «Уменьшен») AND («Усыхание ветвей» = «Отдельные ветви засохли») AND («Отпад коры» = «Отсутствует») THEN («Состояние» = «Ослабленное»).

3. IF («Густота кроны» = «Ажурная») AND («Прирост» = «Слабый») AND («Усыхание ветвей» = «Усыхание до 2/3 кроны») AND («Отпад

коры» = «Отсутствует») THEN («Состояние» = «Сильно ослабленное»).

4. IF («Густота кроны» = «Сильно ажурная») AND («Прирост» = «Слабый» OR «Прирост» = «Отсутствует») AND («Усыхание ветвей» ≥ «Усыхание более 2/3 кроны») AND («Отпад коры» = «Отсутствует») THEN («Состояние» = «Усыхающее»).

5. IF («Густота кроны» = «Сильно ажурная») AND («Прирост» = «Отсутствует») AND («Усыхание ветвей» ≥ «Усыхание более 2/3 кроны») AND («Отпад коры» = «Частичный») THEN («Состояние» = «Свежий сухой»).

6. IF («Густота кроны» = «Отсутствует») AND («Прирост» = «Отсутствует») AND («Усыхание ветвей» ≥ «Усыхание более 2/3 кроны») AND («Отпад коры» = «Частичный» OR «Отпад коры» = «Полный») THEN («Состояние» = «Старый сухой»).

Нечеткие правила, включающие иные комбинации значений лингвистических переменных, выходят за рамки характеристики категорий состояний хвойных пород. Чтобы не получить отсутствующее заключение, применен следующий подход: выбирается не одно значение по каждой лингвистической переменной, а их множество, упорядоченное по степени уверенности. Рассмотрим такой случай, когда система определила следующие значения лингвистических переменных (табл. 6).

Если взять значения переменных с максимальной степенью уверенности, то полученное решающее правило выйдет за рамки характеристик категорий состояния. Поэтому для оценки категории состояния дерева с использованием построенной базы нечетких правил в качестве фазсифицированных входных переменных выбираются близкие по степени уверенности значения (табл. 7).

Таким образом, густота кроны будет выбрана «разреженная», несмотря на меньшую степень

■ **Таблица 6.** Полученные значения лингвистических переменных, распределенные по степени уверенности

■ **Table 6.** The obtained values of linguistic variables distributed by degree of confidence

Переменная	Значение (степень уверенности)
Густота кроны	Густая (0,55); разреженная (0,45)
Годичный прирост ветвей	Уменьшен (0,7); нормальный (0,3)
Усыхание ветвей	Засохли отдельные ветви (0,75); усыхание до 2/3 кроны (0,25)
Отпад коры	Отсутствует (1)

■ **Таблица 7.** Значения лингвистических переменных, соответствующие решающему правилу № 2

■ **Table 7.** Values of linguistic variables corresponding to the decision rule no. 2

Переменная	Значение
Густота кроны	Разреженная
Годичный прирост ветвей	Уменьшен
Усыхание ветвей	Засохли отдельные ветви кроны
Отпад коры	Отсутствует

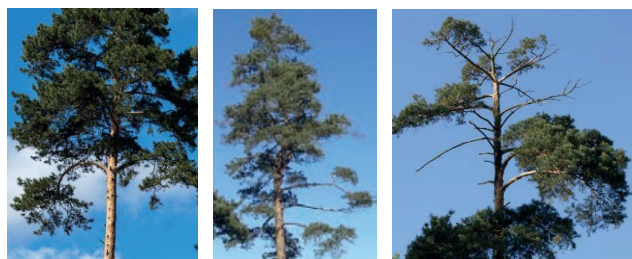
принадлежности четкого входного значения, поскольку такая комбинация значений лингвистических переменных уже соответствует решающему правилу № 2.

Данный способ позволяет настроить нечеткий логический контроллер для принятия решений об отнесении исследуемого дерева к определенной категории состояния.

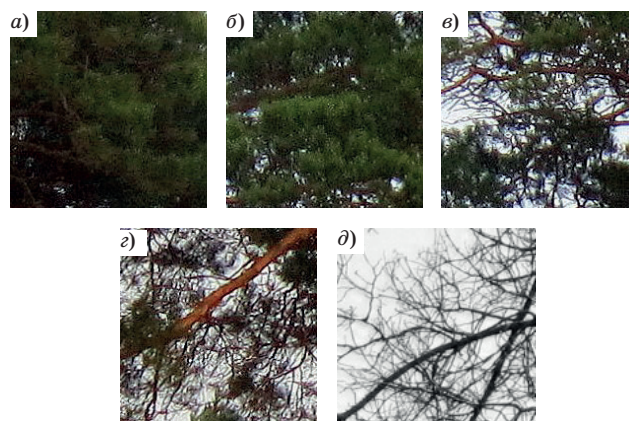
Экспериментальные исследования

Экспериментальные исследования качества работы алгоритма определения категории состояния дерева проведены на примере сосны, так как именно сосна является типичным представителем хвойных пород бореальной зоны. При проведении экспериментальных исследований использовано 228 изображений сосны различных размеров. Минимальный размер изображения составил 396×452 пикселей. Примеры использованных изображений приведены на рис. 7.

Для исследования эффективности оценки степени густоты кроны с помощью предложенного метода был подготовлен набор данных, содержащий фрагменты размером 50×50 пикселей. Фрагменты обрабатывались в соответствии с предложенным методом. Примеры изображений фрагментов кроны различной густоты показаны на рис. 8, а–д.



■ **Рис. 7.** Примеры использованных изображений сосны
 ■ **Fig. 7.** Examples of pine images

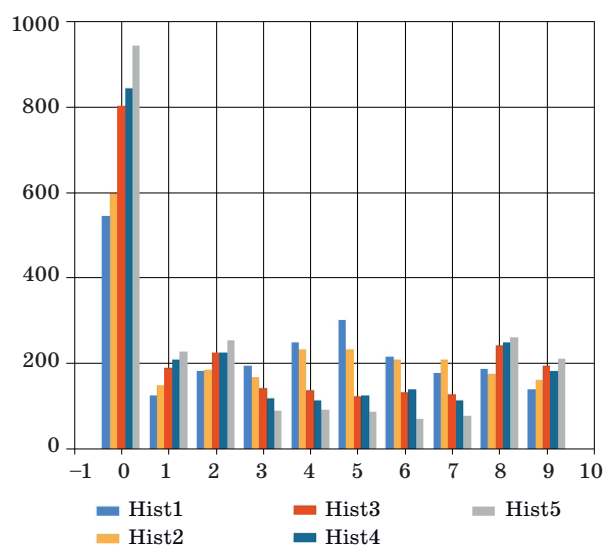


■ **Рис. 8.** Фрагменты кроны сосны разной густоты: а — густая; б — разреженная; в — ажурная; г — сильно ажурная; д — отсутствующая

■ **Fig. 8.** Fragments of a pine crown with different density: а — thick; б — sparse; в — openwork; г — high openwork; д — missing

Гистограммы для различных фрагментов кроны, полученные при использовании ELBP, показаны на рис. 9. Условное обозначение «Hist1» соответствует фрагменту кроны на рис. 8, а, «Hist2» соответствует фрагменту кроны на рис. 8, б и т. д. Как видно из рис. 9, при уменьшении густоты кроны происходит увеличение количества неравномерных шаблонов ELBP, т. е. таких шаблонов, которые отвечают за уменьшение контуров на изображении исследуемой текстуры.

Данные, полученные на этапе оценки густоты кроны на основе гистограммного подхода, а так-



■ **Рис. 9.** ELBP-гистограммы для разных фрагментов кроны

■ **Fig. 9.** ELBP-histograms for different crown fragments

же другие характеристики дерева, такие как степень усыхания ветвей, отпада коры и т. д., поступают на вход нечеткого логического контроллера. В качестве параметров работы нечеткого контроллера использованы импликация Мамани, метод дефаззификации — первый максимум, так как именно эти настройки показали наиболее близкий к экспертному мнению специалиста — лесопатолога — результат классификации исследуемого дерева. В табл. 8 приведены результаты классификации тестовых изображений деревьев, выполненной с помощью предложенного метода.

Результаты, приведенные в табл. 8, сопоставлялись с оценкой эксперта. Лучшие результаты получились для категорий I — без признаков ослабления, V — свежий сухостой (текущего года) и VI — старый сухостой. Пограничные состояния между категориями II — ослабленные и III — сильно ослабленные трудно различимы, а пограничные состояния категорий III и IV (усыхающие) отличить, используя только изображения, практически невозможно. Поэтому значения показателя точности классификации для этих категорий получились самыми низкими. Категории V — свежий сухостой (текущего года) и VI — старый сухостой (прошлых лет) разделимы, так как на старом сухостое живая хвоя отсутствует полностью, а на свежем сухостое присутствует серая, желтая или красно-бурая хвоя.

Следует отметить, что предложенный метод основан на обработке 2D-изображений, и, следовательно, восстановить 3D-структуру дерева не представляется возможным. Однако метод показывает хорошие результаты классификации и

■ **Таблица 8.** Результаты классификации тестовых изображений, %

■ **Table 8.** Classification results of test images, %

Категория	Верная классификация	Ошибка I рода	Ошибка II рода
I	99,24	0,76	2,04
II	86,11	13,89	12,74
III и IV	83,44	16,56	10,15
V	98,42	1,58	1,11
VI	98,70	1,30	2,15

способен формировать объективную оценку состояния дерева, не предъявляя большие требования к вычислительным ресурсам.

Заключение

В работе представлен метод оценки категории санитарного состояния дерева на примере сосны. Метод включает удаление фона методом пороговой обработки, текстурный анализ на базе расширенных бинарных шаблонов и применение контроллера на нечеткой логике. Кусочно-линейные графики лингвистических переменных аппроксимированы гладкими функциями: проведена сплайн-аппроксимация функциями Гаусса. Экспериментальные исследования подтверждают эффективность разработанного алгоритма.

Литература

1. Постановление Правительства Российской Федерации «О Правилах санитарной безопасности в лесах» от 20 мая 2017 года № 607. <http://docs.cntd.ru/document/436736467> (дата обращения: 9.03.2020).
2. Mizoue N. Fractal analysis of tree crown images in relation to crown transparency. *Journal of Forest Planning*, 2001, vol. 7, no. 2, pp. 79–87. doi.org/10.20659/jfp.7.2_79
3. Clark N. A., Lee S.-M., Araman P. A. Finding a good segmentation strategy for tree crown transparency estimation. *19th Biennial Workshop on Color Photography, Videography and Airborne Imaging for Resource Assessment*, 2003, Logan, Utah, pp. 1–13.
4. Lee S.-M., Clark N. A., Araman P. A. Automated methods of tree boundary extraction and foliage transparency estimation from digital imagery. *19th Biennial Workshop on Color Photography, Videography and Airborne Imaging for Resource Assessment*, 2003, Logan, Utah, pp. 1–10.

5. Clark N., Lee S.-M. Digital photo monitoring for tree crown foliage change evaluation. *The 15th Central Hardwood Forest Conference*, 2007, pp. 46–54.
6. Winn M. F., Lee S. M., Araman P. A. ForestCrowns: a software tool for analyzing ground based digital photographs of forest canopies. *The 18th Central Hardwoods Forest Conference — GTR-NRS-P-117*, 2012, pp. 35–41.
7. Winn M. F., Palmer A. J., Lee S.-M., Araman P. A. ForestCrowns: A transparency estimation tool for digital photographs of forest canopies. *e-Gen. Tech. Rep. SRS-215*. Asheville, NC: U.S. Department of Agriculture Forest Service, Southern Research Station, 2016, 10 p.
8. Nakajima H., Kume A., Ishida M., Ohmiya T., Mizoue N. Evaluation of estimates of crown condition in forest monitoring: comparison between visual estimation and automated crown image analysis. *Annals of Forest Science*, 2011, vol. 68, pp. 1333–1340. doi.org/10.1007/s13595-011-0132-9
9. Borianne P., Subsol G., Caraglio Y. Automated efficient computation of crown transparency from tree

- silhouette images. *Computers and Electronics in Agriculture*, 2017, vol. 133, pp. 108–118. doi.org/10.1016/j.compag.2016.12.011
10. Liu J., Wang X., Wang T. Classification of tree species and stock volume estimation in ground forest images using Deep Learning. *Computers and Electronics in Agriculture*, 2019, vol. 166, pp. 105012–105012-10. doi.org/10.1016/j.compag.2019.105012
 11. Weinstein B. G., Marconi S., Bohlman S. A., Zare A., White E. P. Cross-site learning in deep learning RGB tree crown detection. *Ecological Informatics*, 2020, vol. 56, pp. 101061–101061-9. doi.org/10.1016/j.ecoinf.2020.101061
 12. Yun T., Cao L., An F., Chen B., Xue L., Weizheng Lid, Pincebourde S., Smith M. J., Eichhorn M. P. Simulation of multi-platform LiDAR for assessing total leaf area in tree crowns. *Agricultural and Forest Meteorology*, 2019, vol. 276–277, pp. 107610–107610-10. doi.org/10.1016/j.agrformet.2019.06.009
 13. Holmgren J., Lindberg E. Tree crown segmentation based on a geometric tree crown model for prediction of forest variables. *Canadian Journal of Remote Sensing*, 2013, vol. 39, pp. S86–S98. doi:10.5589/m13-025
 14. Favorskaya M., Tkacheva A., Danilin I. M., Medvedev E. M. Fusion of airborne LiDAR and digital photography data for tree crowns segmentation and measurement. In: *Intelligent Interactive Multimedia Systems and Services*/ E. Damiani, R. J. Howlett, L. C. Jain, L. Gallo, G. De Pietro (eds). Springer International Publishing Switzerland, 2015, SIST, vol. 40, pp. 191–201. doi.org/10.1007/978-3-319-19830-9_18
 15. Favorskaya M. N., Jain L. C. Data fusion for evaluation of woodland parameters. In: *Handbook on Advances in Remote Sensing and Geographic Information Systems: Paradigms and Applications in Forest Landscape Modeling*/ M. N. Favorskaya, L. C. Jain (eds). Springer International Publishing, 2017, ISRL, vol. 122, pp. 111–137. doi.org/10.1007/978-3-319-19830-9_18
 16. Lindberg E., Holmgren J. Individual tree crown methods for 3D data from remote sensing. *Current Forestry Reports*, 2017, vol. 3, no. 1, pp. 19–31. doi:10.1007/s40725-017-0051-6
 17. Пятаев А. С. Сегментация дерева на изображении методом неполной пороговой обработки с глобальным порогом. *Решетневские чтения*, 2018, т. 2, с. 291–293.
 18. Favorskaya M., Pyataeva A., Popov A. Texture analysis in watermarking paradigms. *Procedia Computer Science*, 2017, vol. 112, pp. 1460–1469. doi.org/10.1016/j.procs.2017.08.019
 19. Pyataev A. S., Pyataeva A. V., Redkin A. Y. Tree state category identification for boreal area conifers using global features estimation by fuzzy logic approach. *Proceedings of the International Conference Information Technology and Nanotechnology. Session Data Science*, Polish-Japanese Institute of Information Technology, Poland, 2019, vol. 2391, pp. 211–215.

UDC 004.932

doi:10.31799/1684-8853-2020-3-2-11

Classification of tree stand condition using visual data based on fuzzy logicA. S. Pyataev^{a,b}, Post-Graduate Student, orcid.org/0000-0001-5489-8555M. N. Favorskaya^a, Dr. Sc., Tech., Professor, orcid.org/0000-0002-2181-0454, favorskaya@sibsau.ru^aReshetnev Siberian State University of Science and Technology, 31, Krasnoyarsky Rabochy Av., 660037 Krasnoyarsk, Russian Federation^bBranch of FBI «Russian Centre of Forest Protection» — «Centre of Forest Protection of Krasnoyarskiy Krai», 50A/no. 2, Akademgorodok St., 660036, Krasnoyarsk, Russian Federation

Introduction: A reliable assessment of tree condition directly affects the planning of economic indicators for the use of forest resources and ecological actions for forest protection. Therefore, the correct evaluation of the sanitary state of forest is very important. At present, the decisions that forest pathologists make about classifying trees or forest areas are based on visual inspection and their subjective knowledge about the tree features. **Purpose:** Development of a method for classifying the condition of a tree in terms of its crown density degree and other features, based on fuzzy logic with characteristic functions for linguistic variables such as “Crown density”, “Annual branch growth”, “Bark falling off” or “Shrinking branches”. **Results:** The proposed method classifies the tree condition using pine as an example. The method consists in preliminary image processing, including the removal of background objects, extraction of texture features as extended binary patterns, and application of a specially designed controller based on fuzzy logic. We propose four types of linguistic variables, with their respective terms. For these variables, characteristic functions are specified in tabular form and then approximated by smooth functions. A fuzzy logic controller allows you to obtain an objective assessment of the tree crown condition. Experimental studies confirm the effectiveness of the developed method. **Practical relevance:** The intelligent system of classifying the tree condition according to visual data can provide a significant support to plantation survey specialists. The proposed method allows you to improve the quality of forest monitoring, minimize the influence of human factor, and organize the forest protection in the best possible way.

Keywords — classification, tree condition, fuzzy logic, fuzzy controller, image processing, texture analysis.

For citation: Pyataev A. S., Favorskaya M. N. Classification of tree stand condition using visual data based on fuzzy logic. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2020, no. 3, pp. 2–11 (In Russian). doi:10.31799/1684-8853-2020-3-2-11

References

1. Postanovlenie Pravitelstva Rossiyskoi Federatzii "O Pravilakh sanitarnoi bezopasnosti v lesakh" ot 20 maya 2017 N 607 [Decree of the Government of the Russian Federation "On the Rules of Sanitary Safety in Forests"]. Available at: <http://docs.cntd.ru/document/436736467> (accessed 9 February 2020).
2. Mizoue N. Fractal analysis of tree crown images in relation to crown transparency. *Journal of Forest Planning*, 2001, vol. 7, no. 2, pp. 79–87. doi.org/10.20659/jfp.7.2_79
3. Clark N. A., Lee S.-M., Araman P. A. Finding a good segmentation strategy for tree crown transparency estimation. *19th Biennial Workshop on Color Photography, Videography and Airborne Imaging for Resource Assessment*, 2003, Logan, Utah, pp. 1–13.
4. Lee S.-M., Clark N. A., Araman P. A. Automated methods of tree boundary extraction and foliage transparency estimation from digital imagery. *19th Biennial Workshop on Color Photography, Videography and Airborne Imaging for Resource Assessment*, 2003, Logan, Utah, pp. 1–10.
5. Clark N., Lee S.-M. Digital photo monitoring for tree crown foliage change evaluation. *The 15th Central Hardwood Forest Conference*, 2007, pp. 46–54.
6. Winn M. F., Lee S. M., Araman P. A. ForestCrowns: a software tool for analyzing ground based digital photographs of forest canopies. *The 18th Central Hardwoods Forest Conference — GTR-NRS-P-117*, 2012, pp. 35–41.
7. Winn M. F., Palmer A. J., Lee S.-M., Araman P. A. ForestCrowns: A transparency estimation tool for digital photographs of forest canopies. *e-Gen. Tech. Rep. SRS-215*. Asheville, NC: U.S. Department of Agriculture Forest Service, Southern Research Station, 2016, 10 p.
8. Nakajima H., Kume A., Ishida M., Ohmiya T., Mizoue N. Evaluation of estimates of crown condition in forest monitoring: comparison between visual estimation and automated crown image analysis. *Annals of Forest Science*, 2011, vol. 68, pp. 1333–1340. doi.org/10.1007/s13595-011-0132-9
9. Borianne P., Subsol G., Caraglio Y. Automated efficient computation of crown transparency from tree silhouette images. *Computers and Electronics in Agriculture*, 2017, vol. 133, pp. 108–118. doi.org/10.1016/j.compag.2016.12.011
10. Liu J., Wang X., Wang T. Classification of tree species and stock volume estimation in ground forest images using Deep Learning. *Computers and Electronics in Agriculture*, 2019, vol. 166, pp. 105012–1–105012-10. doi.org/10.1016/j.compag.2019.105012
11. Weinstein B. G., Marconi S., Bohlman S. A., Zare A., White E. P. Cross-site learning in deep learning RGB tree crown detection. *Ecological Informatics*, 2020, vol. 56, pp. 101061–1–101061-9. doi.org/10.1016/j.ecoinf.2020.101061
12. Yun T., Cao L., An F., Chen B., Xue L., Weizheng Lid, Pincebourde S., Smith M. J., Eichhorn M. P. Simulation of multi-platform LiDAR for assessing total leaf area in tree crowns. *Agricultural and Forest Meteorology*, 2019, vol. 276–277, pp. 107610–1–107610-10. doi.org/10.1016/j.agrformet.2019.06.009
13. Holmgren J., Lindberg E. Tree crown segmentation based on a geometric tree crown model for prediction of forest variables. *Canadian Journal of Remote Sensing*, 2013, vol. 39, pp. S86–S98. doi:10.5589/m13-025
14. Favorskaya M., Tkacheva A., Danilin I. M., Medvedev E. M. Fusion of airborne LiDAR and digital photography data for tree crowns segmentation and measurement. In: *Intelligent Interactive Multimedia Systems and Services*. E. Damiani, R. J. Howlett, L. C. Jain, L. Gallo, G. De Pietro (eds). Springer International Publishing Switzerland, 2015, SIST, vol. 40, pp. 191–201. doi.org/10.1007/978-3-319-19830-9_18
15. Favorskaya M. N., Jain L. C. Data fusion for evaluation of woodland parameters. In: *Handbook on Advances in Remote Sensing and Geographic Information Systems: Paradigms and Applications in Forest Landscape Modeling*. M. N. Favorskaya, L. C. Jain (eds). Springer International Publishing, 2017, ISRL, vol. 122, pp. 111–137. doi.org/10.1007/978-3-319-19830-9_18
16. Lindberg E., Holmgren J. Individual tree crown methods for 3D data from remote sensing. *Current Forestry Reports*, 2017, vol. 3, no. 1, pp. 19–31. doi:10.1007/s40725-017-0051-6
17. Pyataev A. S. Tree segmentation in image using the partial threshold processing method with global threshold. *Reshetnevskie chteniya*, 2018, vol. 2, pp. 291–293 (In Russian).
18. Favorskaya M., Pyataeva A., Popov A. Texture analysis in watermarking paradigms. *Procedia Computer Science*, 2017, vol. 112, pp. 1460–1469. doi.org/10.1016/j.procs.2017.08.019
19. Pyataev A. S., Pyataeva A. V., Redkin A. Y. Tree state category identification for boreal area conifers using global features estimation by fuzzy logic approach. *Proceedings of the International Conference Information Technology and Nanotechnology. Session Data Science*, Polish-Japanese Institute of Information Technology, Poland, 2019, vol. 2391, pp. 211–215.

УДК 004.932.2:621.396.96

doi:10.31799/1684-8853-2020-3-12-24

Адаптивная многопороговая селекция объектов на изображениях в системах дистанционного зондирования

В. Ю. Волков^{а, б}, доктор техн. наук, профессор, orcid.org/0000-0001-8552-4775, vl_volk@mail.ru

^аСанкт-Петербургский государственный университет аэрокосмического приборостроения, Б. Морская ул., 67, Санкт-Петербург, 190000, РФ

^бСанкт-Петербургский государственный электротехнический университет «ЛЭТИ», Профессора Попова ул., 5, Санкт-Петербург, 197376, РФ

Введение: обнаружение, селекция и анализ объектов интереса на цифровых изображениях являются важнейшей задачей в системах дистанционного зондирования и технического зрения. Известные методы порогового обнаружения и селекции объектов не используют результаты обработки, поэтому не обеспечивают низкую вероятность ложных тревог и плохо сохраняют форму выделяемых объектов. Имеется довольно мало результатов количественной оценки качества таких алгоритмов как на модельных, так и на реальных изображениях. **Цель:** исследование эффективности алгоритмов обнаружения, выделения и локализации объектов интереса, использующих селекцию объектов по геометрическим характеристикам, в условиях априорной неопределенности относительно свойств объектов и фона при сохранении формы селективируемых объектов. **Результаты:** получены и исследованы характеристики алгоритмов обнаружения и выделения объектов интереса на тестовых моделях монохромных изображений. Реализованные программно, алгоритмы используют многопороговую обработку, которая дает набор двоичных срезов. Это позволяет проводить морфологическую обработку объектов на каждом срезе в целях анализа их геометрических характеристик и последующего отбора по геометрическим критериям с учетом эффекта перколяции, который приводит к изменению площади и фрагментации объектов. В результате анализа этих изменений устанавливается адаптивный порог обнаружения для каждого из выбранных объектов. Селекция позволяет значительно снизить количество ложных срабатываний при обнаружении и установить пороги более низкого уровня, что повышает вероятность правильного обнаружения интересующих объектов. Представлены характеристики обнаружения и результаты обработки тестовых моделей изображений, а также результаты селекции объектов на реальном телевизионном и радиолокационном изображении, подтверждающие эффективность рассмотренных алгоритмов. **Практическая значимость:** алгоритмы способны более эффективно селективировать объекты на изображениях различной природы, получаемых в системах дистанционного зондирования, системах исследования материалов и в системах медицинской диагностики. Их микропроцессорная реализация существенно проще, чем реализация универсальных обучаемых алгоритмов на базе нейронных сетей.

Ключевые слова – многопороговая обработка, селекция объектов, перколяция, адаптивный порог, селекция признаков, критерии сегментации.

Для цитирования: Волков В. Ю. Адаптивная многопороговая селекция объектов на изображениях в системах дистанционного зондирования. *Информационно-управляющие системы*, 2020, № 3, с. 12–24. doi:10.31799/1684-8853-2020-3-12-24

For citation: Volkov V. Yu. Adaptive multi-threshold object selection in remote sensing images. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2020, no. 3, pp. 12–24 (In Russian). doi:10.31799/1684-8853-2020-3-12-24

Введение

Проблемы обнаружения, селекции и локализации объектов интереса на изображениях различной природы весьма актуальны и интенсивно изучаются. Эти задачи решаются телевизионными, инфракрасными, лазерными и радиолокационными системами наблюдения с синтезированными апертурами [1–5].

Сегментация изображения на отдельные объекты обычно основана на таких характеристиках, как однородность интенсивности и соответствие цветов. Основными свойствами объектов по отношению к фону являются их компактность и изолированность. Региональные методы основаны на свойствах областей и на предположении, что соседние пиксели в пределах области объекта имеют более или менее сходные значения [4].

Подробный обзор методов сегментации объектов, которые достаточно универсальны для различных приложений анализа изображений, без ограничения системами дистанционного зондирования, микроскопической или биомедицинской визуализации, приведен в работе [5].

Многопороговая обработка связана с преобразованием исходного монохромного изображения в набор бинарных слоев. В случае достаточно большого числа пороговых значений можно предположить, что такое преобразование не приводит к потере информации. В то же время обработка двоичных изображений проще и быстрее, чем обработка изображений в оттенках серого. Различные применения многопороговой обработки для сегментации изображений были рассмотрены в многочисленных работах, например [6–13]. Как правило, многопороговая сегмента-

ция основана на свойствах гистограммы исходного изображения, причем последним шагом является выбор оптимального порогового значения. Свойства объектов интереса и результаты их селекции при этом никак не учитываются и не используются в алгоритме обработки.

В данной статье развивается и количественно анализируется оригинальный подход к анализу монохромных изображений, основанный на многопороговой обработке с учетом эффекта перколяции (просачивания) [14–16]. Специфической особенностью подхода является то, что сегментация и отбор осуществляются на основе апостериорной информации после получения бинарных слоев для ряда пороговых значений. Основная идея заключается в том, чтобы исключить предварительное обучение, а параметры алгоритма селекции объектов настраивать для каждого изображения, а затем и для каждого объекта в отдельности. Такой подход позволяет связать свойства объекта на соседних бинарных слоях и построить трехмерную иерархическую структуру для последующей сегментации.

Эффект перколяции связан с просачиванием пустых пикселей в область, занимаемую объектом, при увеличении пороговых значений, что в конечном итоге приводит к его разрушению и появлению новых объектов из образовавшихся фрагментов [14, 15]. Вычисляя изменения площади для каждого объекта, можно ввести условия его существования на некоторой совокупности порогов и определить трехмерную иерархическую структуру, где третьим измерением будет значение порога. После 3D-реконструкции все интересующие объекты могут быть отобраны с использованием различных критериев, таких как коэффициент перколяции, геометрические характеристики или параметры текстуры [16].

Селекция объектов на основе многопороговой обработки

Чтобы реализовать селекцию объекта, необходимо описать и учесть при обработке ожидаемые свойства объекта. Как правило, наблюдается острая нехватка информации об объектах, за исключением типичных размеров и некоторых предположений о площади, периметре, форме и ориентации. Новая идея состоит в том, чтобы выбрать и установить оптимальное пороговое значение на основе результатов селекции объектов в многопороговой структуре для достижения наилучшего результата, опираясь на апостериорную информацию. Этот подход был первоначально предложен в работе [17] для выделения мелко-масштабных объектов.

Площадь объекта является очень эффективным признаком для выделения объектов при

многопороговой обработке [16]. Сложность реализации алгоритмов заключается в необходимости указывать абсолютные значения площади в пикселях. Обычно можно исключить из рассмотрения небольшие объекты, которые возникают из фона или являются мелкими фрагментами крупных объектов, подвергшихся разрушению на этом пороге. Это позволяет упростить алгоритм и ускорить вычисления. Однако реализация многомасштабной обработки требует инвариантных признаков объектов.

В идеале для каждого объекта интереса требуется свое пороговое значение, и такие адаптивные локальные пороги могут быть сформированы с помощью локальных (скользящих) окон, в пределах которых фон считается однородным [18]. Такие методы требуют априорного знания размера и формы интересующего объекта, в противном случае возникают значительные искажения формы объекта. При малых размерах окон по отношению к размерам объекта возникает эффект дифференцирования формы, в результате чего подавляются центральные области объектов. Кроме того, использование локального окна, которое охватывает объект и прилегающий участок фона, приводит к потере разрешения близко расположенных объектов и подавлению одного объекта соседними объектами, которые попадают в область этого окна.

Рассматриваемый здесь подход основан на многопороговой обработке. Один из методов предполагает установление порога для каждой категории интересующих объектов, который выбирается по заданному критерию [16, 17, 19]. В этом случае могут использоваться различные параметры для описания категории объектов, такие как размер объекта или его форма. Инвариантные параметры, такие как отношение площади периметра к площади, коэффициент удлинения описывающего эллипса и другие геометрические или текстурные характеристики, более удобны для анализа разномасштабных изображений. На каждом бинарном слое селекционируются объекты, удовлетворяющие заданным геометрическим свойствам, а порог бинаризации для таких объектов устанавливается таким образом, чтобы получить максимальное количество выделенных объектов данной категории (или их пикселей) с учетом требуемого сохранения формы объектов [19]. Этот процесс может быть автоматизирован, что приводит к адаптивным методам установки порога.

Однако такой способ не подходит для объектов, различающихся по площади и интенсивности. Инвариантные геометрические признаки в значительной степени преодолевают вышеуказанное ограничение. Одним из таких признаков является отношение периметра объекта в квад-

рате к площади объекта: $P_S = P^2/4\pi S$. Этот признак характеризует компактность объекта [16]. Нормирующий коэффициент 4π вводится для того, чтобы обеспечить единичное значение коэффициента для наиболее компактного объекта, которым является круг. Шумовые и фоновые объекты обычно характеризуются «фрактальными границами», поэтому их значения P_S намного превышают единицу. Объекты интереса, отличающиеся компактной формой, имеют меньшие значения коэффициентов, чем шумовые объекты, что позволяет обнаруживать и выделять их на шумном фоне.

Другим геометрическим инвариантом является коэффициент протяженности главной оси описывающего эллипса: $P_L = \pi L^2/4S$. В этом случае используется более сглаженная оценка протяженности границы объекта по сравнению с периметром. Для круга он также равен единице [16].

При использовании коэффициентов P_S и P_L пороговый уровень может быть установлен для каждого из выбранных объектов по минимальному значению этого коэффициента на определенном бинарном срезе. В этом случае определяются адаптивные локальные пороги. Эти методы позволяют значительно снизить количество ложных срабатываний при обнаружении, а также использовать пороги более низкого уровня, что повышает вероятность правильного обнаружения интересных объектов. Основными трудностями на этом пути является определение набора пороговых уровней, относящихся к данному объекту, и задание правил формирования новых объектов из отколовшихся фрагментов. Эти трудности были успешно преодолены в работах [16, 17, 19], где, однако, не было количественных оценок получаемых результатов обнаружения и выделения объектов.

В данной статье исследуются характеристики качества алгоритмов, которые программно реализованы на основе идей многопороговой селекции с использованием указанных геометрических признаков.

Иерархическая структура на основе многопороговой обработки с учетом эффекта перколяции

Чтобы выбрать локальный порог, необходимо установить отношения между отдельными бинарными слоями и определить, что данный пиксель принадлежит одному и тому же или новому объекту, который появляется на новом бинарном слое вместо предыдущего из-за эффекта фрагментации при увеличении порогового уровня. Алгоритм должен устанавливать связи между пикселями с одинаковыми координатами на раз-

ных бинарных слоях. После введения некоторого параметра, определяющего заданное соотношение между площадями объекта на разных слоях, может быть сформирована трехмерная иерархическая структура. Она строится на базе одного бинарного (белого) объекта, полученного при нулевом пороговом значении и совпадающего по площади со всем изображением. Для решения этой задачи используется структура, основанная на эффекте перколяции, исследование которой начато в работе [16].

Рассмотрим бинаризацию монохромного изображения $I(x, y)$, где I — интенсивность, а x, y — координаты пикселей, с помощью глобального порога T . В результате получается бинарный слой (срез) B_T : $\{B_T = 1, \text{ если } I(x, y) \geq T; B_T = 0, \text{ если } I(x, y) < T\}$, в котором подмножество единиц представляет интересующие объекты (передний план), такие как здания, сооружения, транспортные средства, береговая линия, а подмножество нулей относится к фону, который определяется ландшафтом области наблюдения. Увеличение порога приводит к «просачиванию» нулей на бинарных изображениях, в результате которого площади объектов уменьшаются, а далее они фрагментируются (разваливаются) на части, каждая из которых может либо принадлежать исходному объекту, либо образовать новый объект на этом слое. Здесь использованы обозначения, введенные в работе [19]. Все моделируемые и исследуемые изображения представляются в цифровом формате в диапазоне интенсивностей от 0 до 255.

Начнем с нулевого порога $T = 0$, когда $I(x, y) > 0$ удовлетворяется для всего изображения, образуя таким образом бинарный слой, содержащий один глобальный изолированный объект площадью S_0 и занимающий полный размер изображения. Пусть порог T увеличивается на ΔT , так что некоторые пиксели появляются ниже нового порога ΔT , а сверху создается новый бинарный слой, удовлетворяющий $I(x, y) \geq \Delta T$. Если ΔT относительно невелико и из объекта исключена лишь малая часть пикселей, то глобальный изолированный объект остается единственным, несмотря на уменьшение его площади: $S_T < S_0$.

При дальнейшем увеличении порога доля пикселей с интенсивностью ниже порога становится достаточно большой для того, чтобы эти пиксели сливались вместе, образуя черные пятна в изображении. В конечном счете это приводит к образованию разрывов в исходном объекте и отделению от него фрагментов. Этот вид изменения фазового состояния известен как перколяция [14]. После разрушения исходного объекта на его месте формируется набор фрагментов, каждый из которых может быть его преемником. В этом случае необходимо решить, существует ли преемник исходного объекта вообще или исходный объект

уничтожен и все фрагменты, которые появляются, являются новыми объектами. Для этого вводится параметр, характеризующий устойчивость каждого изолированного объекта при изменении порога. Он связан со скоростью, с которой площадь изолированного объекта уменьшается при увеличении порога. Проводилось рассмотрение данного эффекта [16], но без количественной оценки его влияния на эффективность селекции объектов.

Для каждого из двух смежных слоев отношение $K_S = S_{T+\Delta T} / S_T$ характеризует долю связанных пикселей, хранящихся внутри изолированного объекта. Дальнейшее увеличение порога оставляет все большее число пикселей объекта ниже новых порогов $T + k\Delta T$, где $k = 0, 1, \dots$ есть целое число слоев. Значение K_S зависит от двух смежных значений порога, и его можно считать характеристикой ожидаемой стойкости (устойчивости площади) для объекта по мере увеличения порога. Его учет позволяет связать каждые два соседних бинарных слоя для выбранного объекта.

Пока отношение K_S равно или больше половины, объект в верхнем слое может рассматриваться как однозначный преемник объекта в нижнем слое, который оказывается его единственным предшественником или базовым объектом. Если базовый объект точно разделен пополам, то оба фрагмента, за вычетом пикселей, образующих промежутки, имеют меньшую площадь, чем $S_T/2$, и таким образом на этом месте появляются два новых объекта. Начальный бинарный слой для нового объекта формируется в момент просачивания предшественника, а затем новый объект накапливает дальнейшие бинарные представления для различных пороговых значений T_k до тех пор, пока будет выполняться условие $I(x, y) > T_k$. Пусть объект возникает при пороговом значении T , его базовая площадь S_T уменьшается по K слоям до тех пор, пока не произойдет одно из двух событий: 1) объект исчезает, т. е. $S_{T+(K+1)\Delta T} = 0$; 2) объект теряет устойчивость, т. е. на $(K + 1)$ -м слое отношение $S_{T+(K+1)\Delta T} / S_{T+K\Delta T}$ становится меньше заданного значения K_p (допустимый коэффициент устойчивости или персистентности). В последнем случае объект считается «погибшим», а его остатки (фрагменты) образуют новые объекты [16].

Изменение площади за время существования объекта $S_{T+K\Delta T} / S_T$ можно рассматривать как его коэффициент перколяции P_C . Значение порогового уровня T_K считается порогом перколяции T_C для данного объекта. Коэффициент перколяции частично характеризует текстуру поверхности объекта. Если объект имеет плоскую вершину с постоянным значением интенсивности $I(x, y) = \text{const}$, то он полностью исчезает за один пороговый шаг, а его коэффициент перколяции $P_C = 1$. В этом случае величина интенсивности

сама по себе не влияет на величину коэффициента перколяции, т. е. он оказывается инвариантным к таким преобразованиям, как сдвиг или масштабирование изображения. Пусть объект плавно меняет свою интенсивность, т. е. имеет небольшой градиент интенсивности. Тогда наследование между соседними слоями с возрастающим порогом будет сохраняться до тех пор, пока $S_{T+(K+1)\Delta T} \geq K_p S_{T+K\Delta T}$. Если объект «долгоживущий», то его коэффициент перколяции P_C может оказаться небольшим [16].

Для увеличения детализации можно задать $K_p = 1$, что является самым жестким требованием, когда потеря даже одного пикселя для исходного объекта приводит к образованию нового объекта. В этом случае каждый следующий возрастающий порог дает новую коллекцию объектов. Напротив, выбор $K_p = 0,5$ приводит к значительному уменьшению общего числа трехмерных объектов, охватывающих несколько бинарных слоев. Если установить значение K_p , меньшее 0,5, то каждый изолированный объект в нижнем слое может иметь более одного последовательного изолированного объекта в верхнем слое, что приводит к неоднозначности [19].

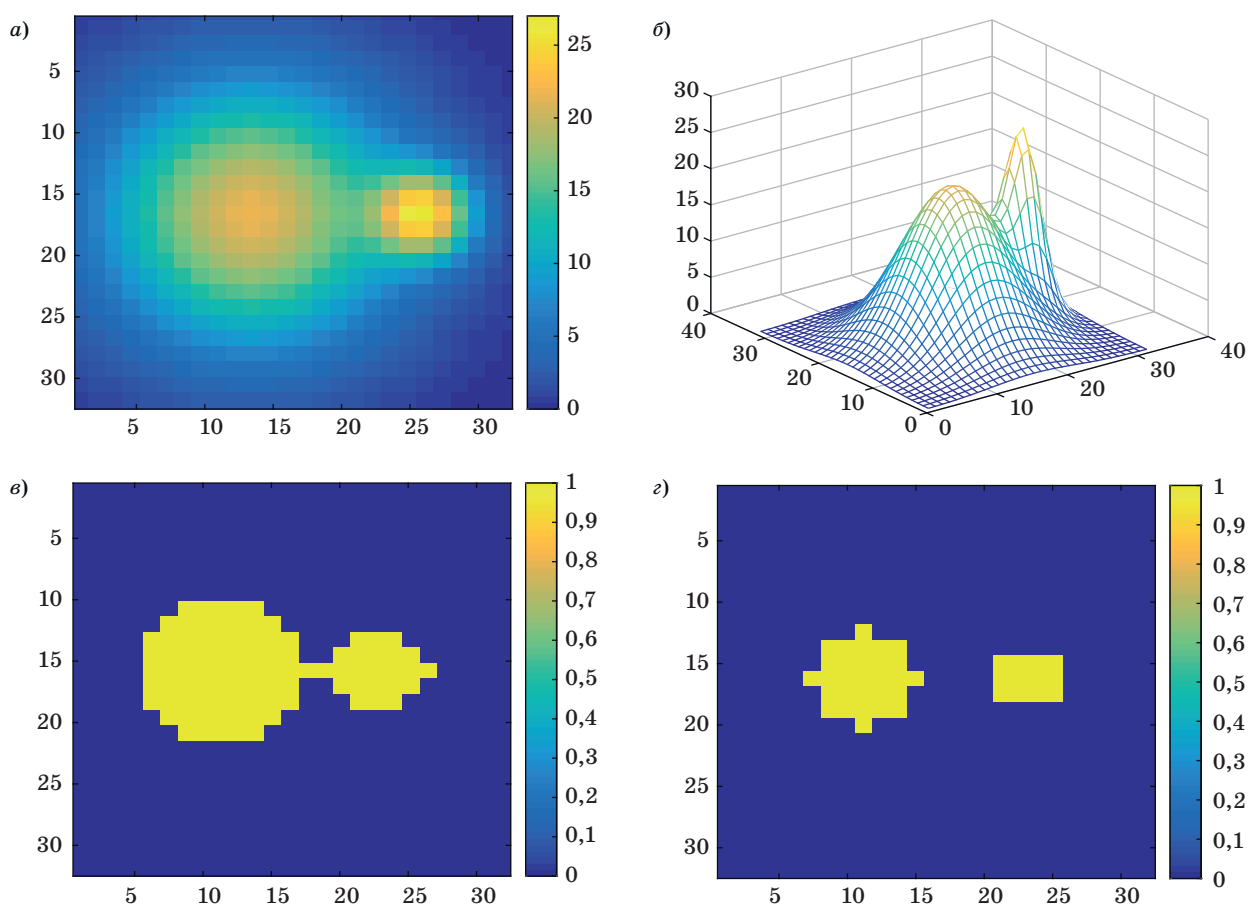
Результат бинаризации объекта при разных порогах представлен на рис. 1, а–г.

Исходный объект (при $T = 0$) с ростом порога медленно теряет свою площадь (его коэффициент персистентности K_S больше половины) и дробится (фрагментируется) на две части при пороге $T_1 = 17$ (см. рис. 1, г). Один из фрагментов (на рис. 1, г слева) исчезает после своего порога перколяции $T_{c1} = 21$. Его коэффициент перколяции P_C очень мал, он меньше 0,01. Второй фрагмент (на рис. 1, г справа) имеет несколько более высокий коэффициент перколяции, равный 0,12.

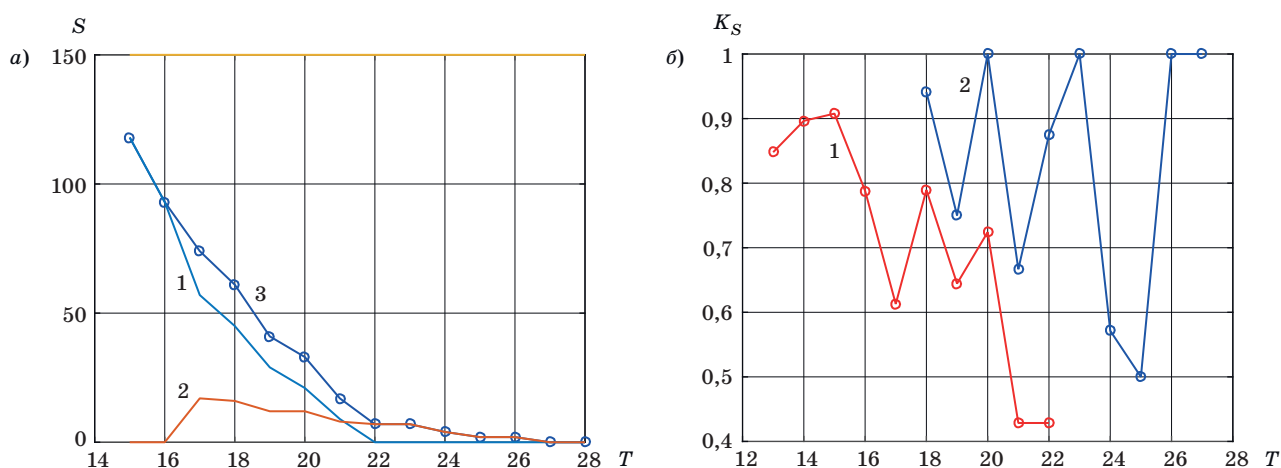
Зависимости площадей фрагментов и коэффициентов персистентности K_S показаны на рис. 2, а и б. Установка минимального значения K_p выше 0,5 уменьшает время жизни каждого из фрагментов.

Если изображение содержит только случайный шум, т. е. интенсивность беспорядочно распределена по всему изображению, то расположение пикселей, которые появляются ниже следующего порога, также является случайным. Теоретически хорошо известно [8], что для бесконечного размера изображения и чистого фонового шума коэффициент перколяции инвариантен к распределению шума и равен $P_C \approx 0,593$.

В результате такого подхода формируется трехмерная иерархическая структура, содержащая все выделенные объекты, в которой каждый пиксель больше не принадлежит одному бинарному слою, а может соответствовать нескольким бинарным слоям в структуре. На основе этой конструкции можно обобщить сегментацию изобра-



■ **Рис. 1.** Фрагментация объекта в результате эффекта перколяции при возрастании порога: *a* — исходный объект; *б* — интенсивность изображения; *в, г* — бинарные срезы
 ■ **Fig. 1.** Object fragmentation as a result of the percolation effect when the threshold increases: *a* — the source object; *б* — image intensity; *в, г* — binary slices



■ **Рис. 2.** Зависимости: *a* — площадей фрагментов от увеличения порога: *1* — первый фрагмент; *2* — второй фрагмент; *3* — общая площадь; *б* — коэффициентов персистентности фрагментов от порогового уровня: *1* — первый фрагмент; *2* — второй фрагмент
 ■ **Fig. 2.** Dependencies: *a* — change in the area of fragments with an increase in the threshold: *1* — the first fragment; *2* — the second fragment; *3* — total area; *б* — the object's persistence coefficients on the threshold level: *1* — the first fragment; *2* — the second fragment

жений и выделение объектов. Для этого вводятся ограничения по площади выделенных объектов и другим геометрическим признакам, характеризующим компактность и форму. Для каждого выбранного объекта, который проходит через несколько бинарных слоев в трехмерном представлении, оптимальный порог T_{opt} , соответствующий слою с его лучшим плоским представлением, может быть выбран по различным геометрическим или текстурным критериям. Частными примерами таких критериев являются области или диапазоны областей выбранных объектов [16, 19].

Аналогичную трехмерную структуру можно получить, понижая порог с его наивысшего уровня до нуля [20]. Это представление имеет то преимущество, что сначала выбираются наиболее интенсивные объекты, образующие изолированные кластеры. Здесь перколяция приводит к слиянию двух или более объектов, что также соответствует разрушению исходных объектов и появлению новых. Недостатком этого подхода является то, что все мелкомасштабные объекты, обнаруженные в верхних слоях, должны сохраняться до тех пор, пока не будет выяснено, являются ли они преемниками более крупных объектов с малыми коэффициентами перколяции и соответствуют ли их базовые размеры введенным ограничениям. Такой подход может увеличить вычислительную сложность алгоритма.

Селекция объектов с использованием адаптивного глобального порога

Глобальный порог хорошо работает при обнаружении и селекции группы однородных компактных (или протяженных) объектов на однородном шумовом фоне. Рассмотрим случай, когда изображение содержит ряд похожих объектов, которые необходимо выделить. На каждом бинарном слое выбираются объекты, удовлетворяющие заданным свойствам, а порог бинаризации для таких объектов выбирается таким образом, чтобы получить максимальное количество объектов данной категории (или их пикселей) с учетом требуемого сохранения формы объекта. Этот процесс может быть автоматизирован, что приводит к адаптивным пороговым методам.

Простейшая селекция основана на площади объектов. В некоторых случаях имеется априорная информация о типичном размере или площади объектов. Слишком мелкие объекты следует исключить из рассмотрения, что значительно снижает вычислительную сложность алгоритмов. Таким образом, минимальная площадь объекта S_{min} является одним из параметров настройки алгоритма.

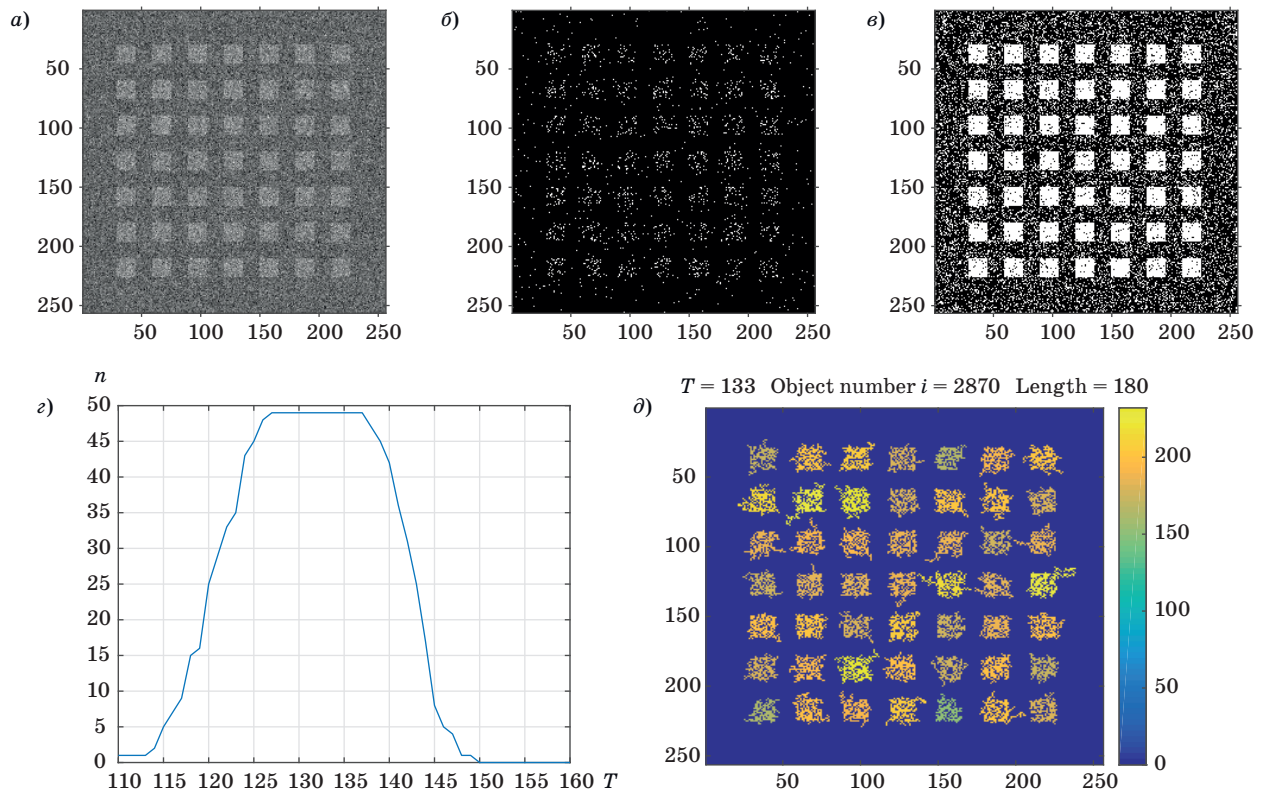
Результаты моделирования селекции по площади показаны на рис. 3. Исходное изображение (рис. 3, а) содержит 49 квадратных объектов размером 16×16 пикселей на стандартном гауссовом шумовом фоне. Отношение сигнал/шум (относительный сдвиг или отклонение математического ожидания) d в каждом пикселе равно 1,163. При высоких пороговых значениях некоторые объекты могут быть потеряны. Детектор с порогом Неймана — Пирсона дает вероятность ложной тревоги $F = 0,01$ в каждом пикселе (рис. 3, б), но вероятность обнаружения составляет только $D = 0,12$, поэтому объекты сильно фрагментированы и ни один не сохраняет свою квадратную форму.

Можно видеть два типа искажений и помех: пропадание точек и разрушение (фрагментация) объекта для порога высокого уровня (см. рис. 3, б); при более низких пороговых значениях форма объектов искажается фрактальным шумом, который существенно фрагментирует границы объектов (рис. 3, в). Фрактальные шумовые разрастания появляются и вне объектов.

Широко используемый детектор Отсу дает слишком много ложных тревог (рис. 3, в). Зависимость числа отсеleccionированных объектов от порогового значения показана на рис. 3, г. При селекции объектов по площади с ограничением $S_{min} = 150$ приемлемое искажение границ объектов достигается при пороговом значении $T = 133$ (рис. 3, д).

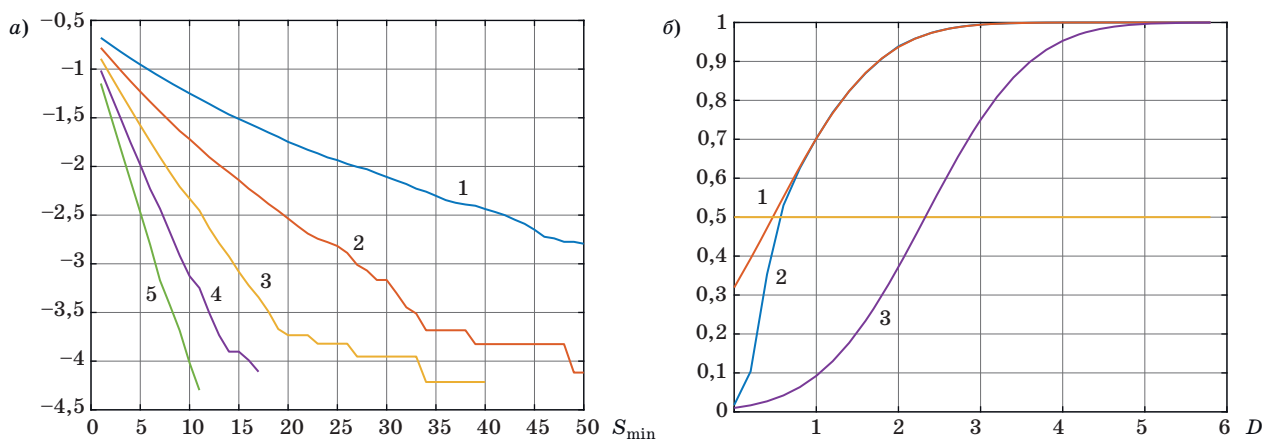
Вероятность ложной тревоги зависит от минимальной площади S_{min} обнаруженных объектов, как показано на рис. 4, а в логарифмическом масштабе. Кривые получены путем моделирования. Для заданного порога и для каждой области, превышающей по площади границу S_{min} (которая укладывается в пикселях вдоль горизонтальной оси), рассчитано количество превышений шума, которое нормировано к общей площади поля. Ось ординат представляет собой десятичный логарифм этого нормированного значения, который соответствует расчетной степени вероятности ложной тревоги.

Этот анализ используется для задачи обнаружения, в которой объект выделяется путем сравнения интенсивности в каждом пикселе с пороговым уровнем T_{NP} , удовлетворяющим критерию Неймана — Пирсона. Рассмотрим задачу обнаружения без учета формы объекта, которая обычно неизвестна, но ее сохранение желательно для последующей селекции. Пусть мы имеем сдвиг в математическом ожидании в области сигнала для гауссова шумового поля (см., например, рис. 3). При отсутствии селекции объектов вероятность обнаружения для каждого пикселя в сигнальном поле вычисляется по формуле $D = 1 - \Phi(t_{NP} - d)$, где Φ — интеграл вероят-



■ **Рис. 3.** Селекция квадратных объектов в гауссовом шуме: *a* — исходное изображение; *б* — порог по критерию Неймана — Пирсона; *в* — порог Отсу; *г* — зависимость числа отселектированных объектов от порога; *д* — результат селекции по площади с помощью адаптивного глобального порога при $S_{\min} = 150$ (цветовая шкала соответствует значению площади соответствующего объект)

■ **Fig. 3.** Selection of square objects in Gaussian noise: *a* — the original image; *б* — the threshold for the Neyman — Pearson criterion; *в* — the Otsu threshold; *г* — the dependence of the number of selected objects from the threshold; *д* — the result of selection by area using an adaptive global threshold (the color scale corresponds to the area of corresponding object)



■ **Рис. 4.** Характеристики обнаружения: *a* — зависимость степени превышения порога для шума (логарифма вероятности ложной тревоги) от площади S_{\min} ; кривые 1–5 соответствуют возрастающим пороговым уровням $T = 150, 155, 160, 165, 170$; *б* — характеристики обнаружения для каждого пикселя: 1, 3 — детектор Неймана — Пирсона для $F = 0,32$ и $F = 0,01$; 2 — селекция объектов с $S_{\min} = 150$

■ **Fig. 4.** Detection characteristics: *a* — dependence of the degree of exceeding the threshold for noise (logarithm of false alarm probability) on the S_{\min} area; curves 1–5 correspond to increasing threshold levels $T = 150, 155, 160, 165, 170$; *б* — detection characteristics for each pixel: curves 1, 3 — Neyman — Pearson detector for $F = 0,32$ and $F = 0,01$; curve 2 — selection of objects with $S_{\min} = 150$

ности Лапласа; $t_{NP} = T_{NP}/\sigma$ — нормированный к среднеквадратическому значению шума порог для получения заданной вероятности ложной тревоги F ; d — отношение сигнал/шум (дефлексия решающей статистики). Эта характеристика обнаружения показана кривой 3 на рис. 4, б для $F = 0,01$ и нормированного порога $t_{NP} = 2,326$. Ось x показывает отношение сигнал/шум d (дефлекцию). Пороговая дефлексия для обеспечения вероятности правильного обнаружения $D = 0,5$ равна $d(0,5) = 2,326$.

В случае селекции объектов предполагается, что область сигнала больше, чем параметр S_{\min} , который используется в алгоритме. Удаление небольших шумовых объектов из двойного изображения может уменьшить количество ложных тревог или порог обнаружения для заданного значения вероятности ложной тревоги. Следствием отбора и удаления мелких объектов является возможность снижения порогового уровня при сохранении низкой вероятности ложной тревоги при обнаружении полезных объектов. Чем больше площадь шумовых объектов, подлежащих удалению, тем ниже порог обнаружения может быть установлен при той же вероятности ложной тревоги. В то же время селекция позволяет сохранить низкое значение порога обнаружения без увеличения вероятности ложной тревоги, что в результате дает увеличение вероятности правильного обнаружения. Для $S_{\min} = 150$ в этой задаче селекция дает нормированный порог $t_{NP} = 0,47$ вместо $2,326$ в случае детектора Неймана — Пирсона. Понятно, что без селекции такой по-

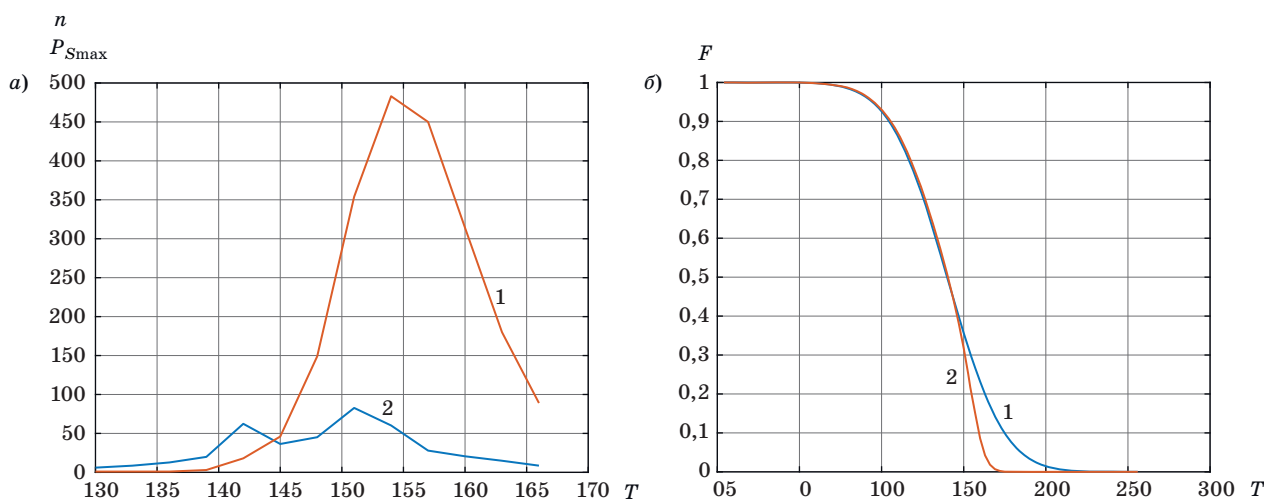
рог даст значительно большую вероятность ложной тревоги (0,32), что показано кривой 1 на рис. 4, б.

Характеристика обнаружения для каждого пикселя в случае селекции объектов проходит между этими двумя линиями (линия 2 на рис. 4, б). Результаты получены путем моделирования, поэтому ось ординат представляет оценку вероятности обнаружения. Пороговая дефлексия для вероятности правильного обнаружения $D = 0,5$ теперь равна 0,5, что обеспечивает увеличение отношения сигнал/шум примерно на 6 дБ. Если имеется информация о форме сигнальной области, то характеристика может быть улучшена применением накопления.

Стоит отметить, что адаптация порогового уровня по максимуму отсекалируемых объектов имеет смысл в тех ситуациях, когда сцена содержит достаточное их количество. Кроме того, этот метод дает несколько меньшее значение порогового уровня, если требовать сохранения формы. Эти недостатки можно попытаться устранить применением геометрических инвариантов в соответствии с подходом, предложенным в работе [16].

Селекция объектов с использованием адаптивного локального порога на основе геометрических критериев

Рассмотрим сначала бинаризацию гауссова шумового поля с ограничением $S_{\min} = 10$, которое может существенно уменьшить число



■ **Рис. 5.** Результаты обработки гауссова шума: а — число объектов n (кривая 1) и максимальный коэффициент удлинения периметра $P_{S_{\max}}$ после выделения области с $S_{\min} = 10$ (кривая 2); б — вероятности обнаружения шумовых выбросов при простой бинаризации (кривая 1) и после предварительного выделения области с $S_{\min} = 10$ (кривая 2) в зависимости от порога

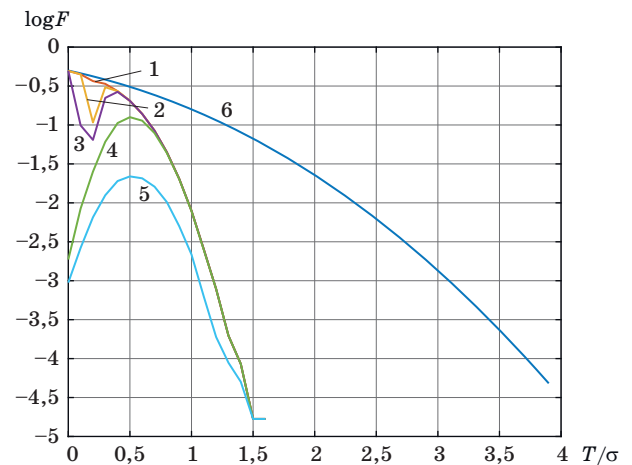
■ **Fig. 5.** Results of processing Gaussian noise: а — the number of objects n (curve 1) and the maximum coefficient of perimeter elongation $P_{S_{\max}}$ after selecting the area with $S_{\min} = 10$ (curve 2); б — the probability of detecting noise emissions during simple binarization (curve 1) and after preliminary selection of the area with $S_{\min} = 10$ (curve 2), depending on the threshold

изолированных объектов после бинаризации. Зависимость числа объектов n от порогового значения T показана на рис. 5, а (линия 1), где максимальное число объектов равно 500.

Кривая 2 (см. рис. 5, а) представляет зависимость максимального (по всем объектам в данном срезе) коэффициента удлинения периметра $P_{S_{max}}$ от порогового значения. Очевидно, что коэффициент $P_{S_{max}}$ достигает больших значений (здесь около 90) для шумовых объектов с фрактальными структурами. Предварительная селекция по площади уменьшает вероятность возникновения шумовых выбросов в области высоких порогов (рис. 5, б, кривая 2) по сравнению с простой бинаризацией (кривая 1). При этом преселекция практически не влияет на значения коэффициентов удлинения периметра P_S для остальных объектов.

Интересно исследовать эффективность ограничений на максимальное значение коэффициента удлинения периметра P_S . Очевидно, селекция по P_S приводит к дополнительному удалению шумовых объектов, имеющих слишком большие коэффициенты удлинения периметра. Ожидается, что эти объекты имеют фрактальные структуры и не представляют собой объекты интереса. Значение $P_{S_{max}}$ является вторым важным параметром для адаптивной селекции объектов. Алгоритм удаляет объекты с $P_S > P_{S_{max}}$, которые появляются при определенных значениях порога.

В результате селекции объектов по P_S кривые на рис. 6 приобретают провал при низких пороговых значениях. Как видно из анализа ограничений на коэффициент удлинения, алгоритм может уменьшить вероятность ложной тревоги в задаче обнаружения для некоторых значений порога. Это приводит к более эффективному обнаружению при этих значениях порога других объектов с меньшими значениями P_S , но выигрыш зависит от этого значения P_S , которое, в свою очередь, за-



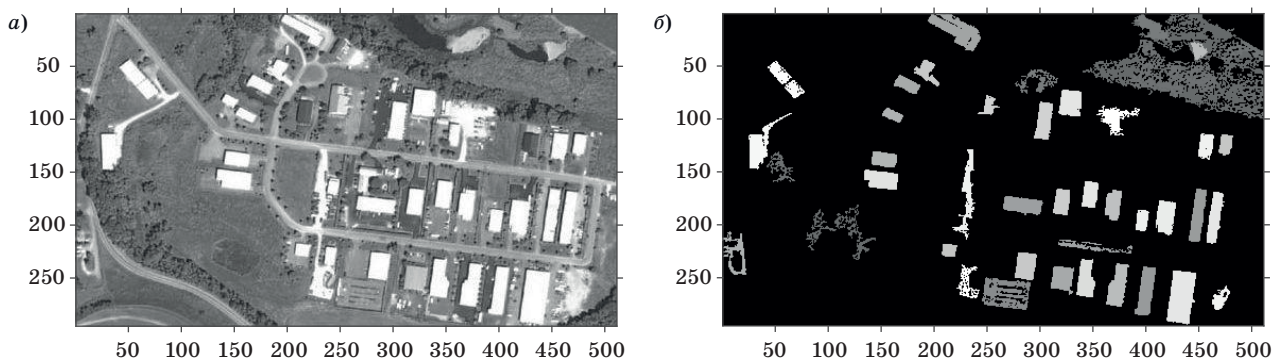
■ Рис. 6. Зависимости логарифма вероятности превышения (степени вероятности ложной тревоги) от нормированного порога для $S_{min} = 10$ и $P_{S_{max}} = 1000, 300, 100, 10, 3$ (кривые 1–5); кривая 6 — без селекции

■ Fig. 6. Dependences of the logarithm of the probability of exceeding (the degree of probability of a false alarm) on the normalized threshold for $S_{min} = 10$ and $P_{S_{max}} = 1000, 300, 100, 10, 3$ (curves 1–5); curve 6 — without selection

висит от порога. Таким образом, выигрыш существенно зависит от формы объекта и формы удаленных объектов. С другой стороны, данный вид селекции работает независимо от абсолютных размеров селектируемых объектов, т. е. данный метод обнаружения и выделения компактных объектов инвариантен к масштабу изображения.

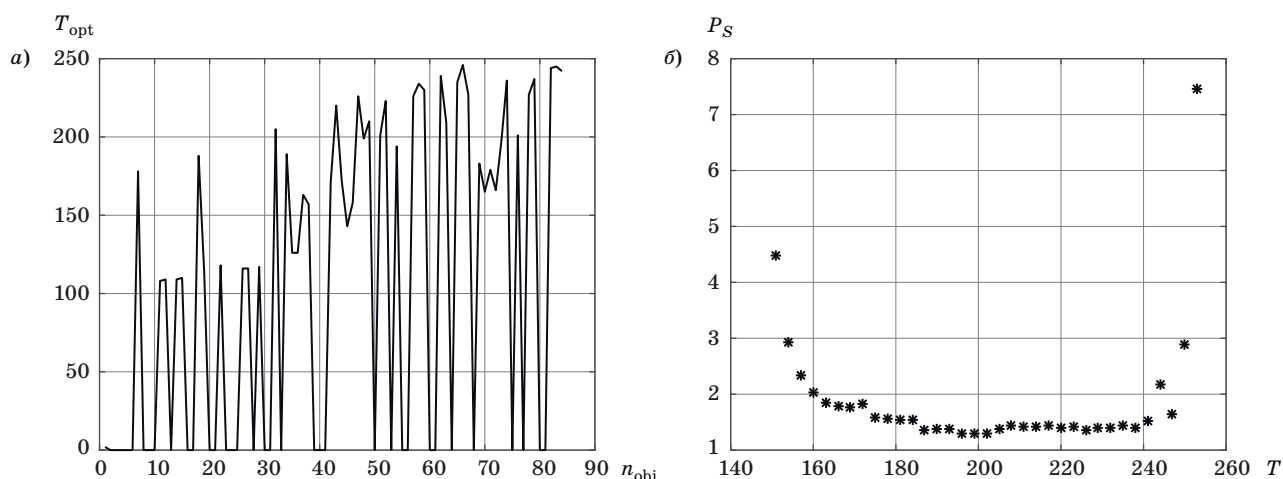
Практические примеры обработки дистанционных наблюдений

Рассмотрим телевизионное изображение (рис. 7, а). Задача состоит в том, чтобы выбрать и локализовать здания в этой сцене. Изображение

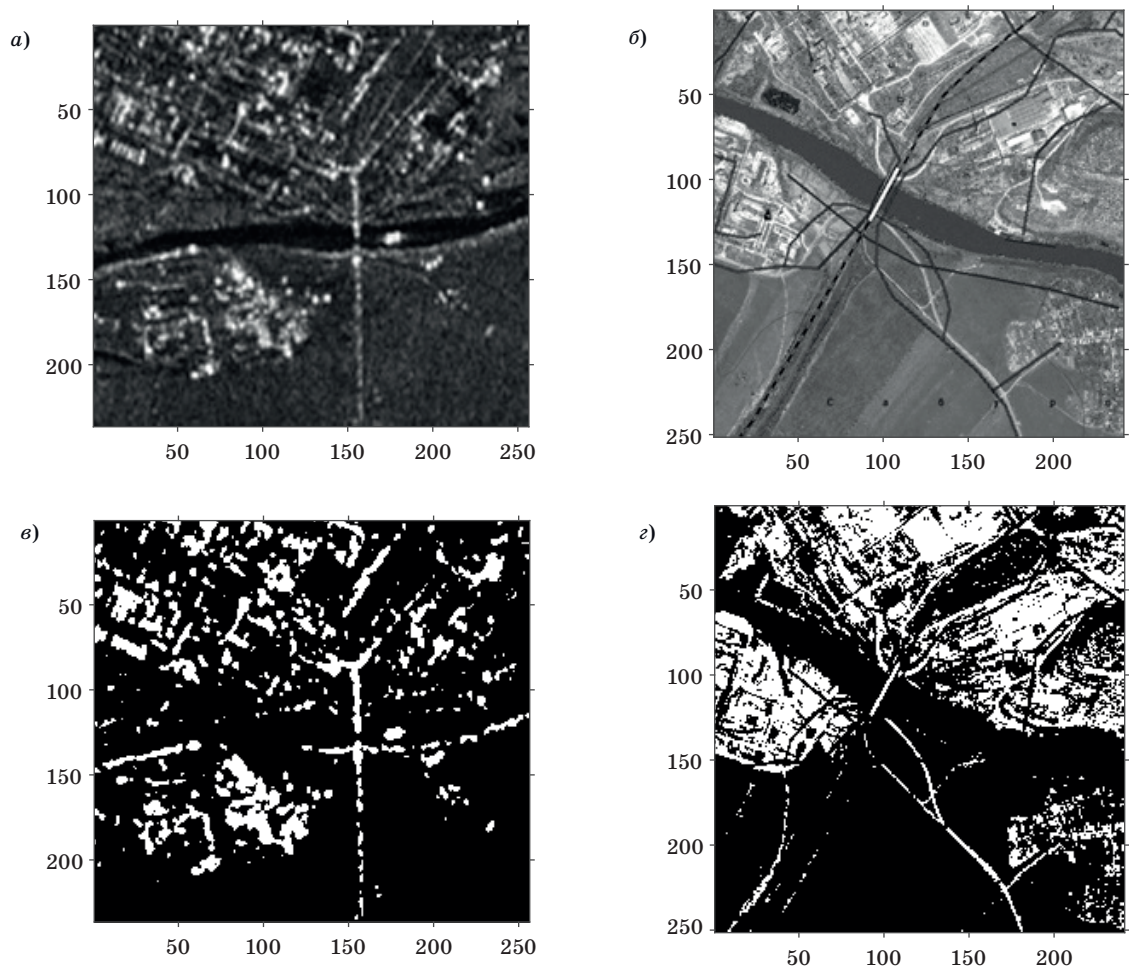


■ Рис. 7. Селекция объектов на телевизионном изображении при $K_p = 0,5$; $S_{min} = 150$ и $P_{S_{max}} = 50$: а — исходное изображение; б — результаты селекции, яркость объектов соответствует значениям адаптивного порога

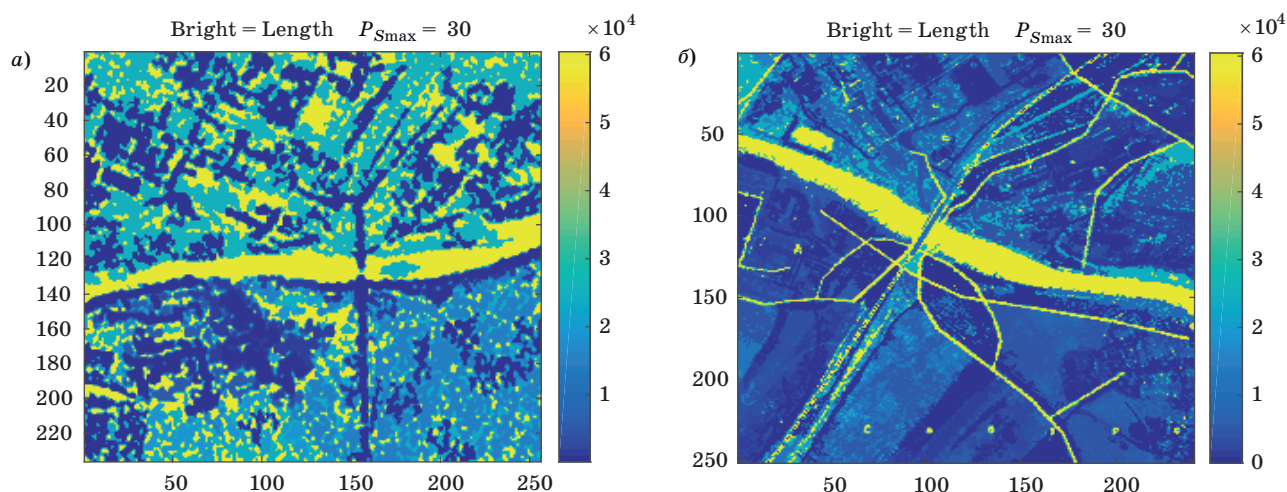
■ Fig. 7. Selection of objects in the TV image at $K_p = 0,5$; $S_{min} = 150$ and $P_{S_{max}} = 50$: а — original image; б — results of selection, brightness of objects corresponds to adaptive thresholds



■ **Рис. 8.** Адаптация пороговых уровней: *a* — зависимость адаптивных пороговых значений от номера объекта; *б* — типичная U-образная кривая оптимизации порога
 ■ **Fig. 8.** Threshold level adaptation: *a* — dependence of adaptive threshold values on the object's number; *б* — typical U-shaped curve for optimizing the threshold



■ **Рис. 9.** Анализируемые изображения и соответствующие бинарные срезы с порогом по Отсу: *a* — изображение, полученное с помощью радиолокатора с синтезированной апертурой; *б* — эскиз местности; *в* и *г* — соответствующие бинарные срезы
 ■ **Fig. 9.** Images analyzed and corresponding binary slices with Otsu threshold: *a* — image obtained using a synthesized aperture radar; *б* — sketch of the terrain; *в* and *г* — corresponding binary slices



■ **Рис. 10.** Результаты обработки с применением селекции объектов, представленные значениями площадей выделенных объектов: *a* — для изображения, полученного с помощью радиолокатора с синтезированной апертурой; *б* — для скетча местности

■ **Fig. 10.** The results of processing with object selection which are represented by area values: *a* — for image obtained using a synthesized aperture radar; *b* — for the sketch of terrain

содержит 82 объекта, которые селектируются с помощью локального адаптивного порога для каждого объекта. Предлагаемый способ применяется при $K_p = 0,5$; $S_{\min} = 150$ и $P_{S\max} = 50$. Результаты селекции представлены на рис. 7, б.

Адаптивный локальный порог вычисляется для каждого объекта с помощью процесса оптимизации, в ходе которого порог устанавливается по минимуму P_S . Оптимальные значения порогов для всех 82 объектов представлены на рис. 8, а, где ось x содержит номера изолированных объектов. Типичная U-образная кривая оптимизации представлена на рис. 8, б.

Радиолокационное изображение местности, полученное с помощью радиолокационной станции с синтезированной апертурной решеткой, и эскиз для этой области показаны на рис. 9, а и б. Часто проблема заключается в объединении фрагментов (задача сопоставления или регистрации изображений), и одним из возможных решений является селекция одних и тех же объектов на разных изображениях, которые затем используются для формирования опорных точек.

В этом случае глобальный порог работает очень хорошо. Бинарные изображения получены с помощью алгоритма Отсу (рис. 9, в, г). Большая часть объектов уничтожена.

Использование адаптивного локального порога с выделением объектов по площади с учетом коэффициента P_S позволяет получить ряд представлений для выделенных объектов, отличающихся отображаемым параметром. Результаты селекции представлены на рис. 10, а и б, где яркостная индикация соответствует площадям объектов.

Используемый метод позволяет получить набор различных бинарных срезов, содержащих геометрические признаки селектируемых объектов для решения задач сопоставления и совмещения изображений, формируемых разнородными датчиками.

Заключение

Исследованы характеристики алгоритмов адаптивной селекции компактных объектов на изображениях. Алгоритмы реализованы программно на основе начальной многопороговой обработки исходного цифрового изображения, в результате чего создается несколько бинарных слоев. На основе эффекта перколяции («просачивания» нулей в бинарном слое) устанавливаются связи между слоями, формируется и исследуется трехмерная иерархическая структура, где третьей координатой выступает пороговый уровень.

После селекции изолированных объектов в каждом из бинарных слоев находятся слои, содержащие его наилучшее представление с точки зрения используемого геометрического критерия, например минимума коэффициента удлинения периметра объекта. Ключевая идея алгоритма заключается в том, что решение основано на апостериорной информации о свойствах объектов, которые могут быть выбраны из каждого бинарного слоя, и в поиске наилучшего слоя с точки зрения геометрических свойств этих объектов. Используя эту информацию, можно успешно решать задачи адаптивной селекции, устанавливая локальный пороговый уровень для каждого

из объектов интереса, при этом сохранять форму объекта, несмотря на нестационарный фон.

Новые результаты заключаются в получении количественных характеристик эффективности при обнаружении компактных объектов и в исследовании влияния на них параметров алгоритмов.

Для тестовой задачи обнаружения квадратных объектов в гауссовом шуме получены зависимости вероятности ложной тревоги от минимальной площади селектируемых объектов и характеристики обнаружения для критерия Неймана — Пирсона. Получены оценки выигрыша в отношении сигнал/шум при селекции объ-

ектов по площади. Использование адаптивной селекции по коэффициенту удлинения периметра объекта также обеспечивает выигрыш в отношении сигнал/шум не менее 6 дБ.

Исследования качества селекции компактных (пятенных) объектов на типовых зашумленных модельных и реальных телевизионных, инфракрасных и радиолокационных изображениях показали работоспособность и эффективность рассмотренного алгоритма. При этом наблюдались минимальные искажения их границ объектов по сравнению с известными алгоритмами сегментации при довольно низких отношениях сигнал/шум.

Литература

1. Cheng G., Han J. A survey on object detection in optical remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2016, vol. 117, pp. 11–28.
2. Arias-Castro E., Grimmett G. R. Cluster detection in networks using percolation. *Bernoulli*, 2013, vol. 19(2), pp. 676–719.
3. Patil G. P., Taillie C. Upper level set scan statistic for detecting arbitrarily shaped hotspots. *Environmental and Ecological Statistics*, 2004, vol. 11, pp. 183–197.
4. Zhou W., Troy A. An object-oriented approach for analyzing and characterizing urban landscape at the parcel level. *International Journal of Remote Sensing*, 2008, vol. 29(11), pp. 3119–3135.
5. Gu H., Han Y., Yang Y., Li H., Liu Z., Soergel U., Blaschke T., Cui S. An efficient parallel multi-scale segmentation method for remote sensing imagery. *Remote Sensing*, 2018, vol. 10(4), pp. 590–608.
6. Shivahare B. D., Gupta S. K. Multilevel thresholding based image segmentation using whale optimization algorithm. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 2019, vol. 8, iss. 12, pp. 4602–4613.
7. Cuevas E., González A., Fausto F., Zaldívar D., Pérez-Cisneros M. Multithreshold segmentation by using an algorithm based on the behavior of locust swarms. *Mathematical Problems in Engineering*, 2015, article ID 805357, pp. 1–12.
8. Arora S., Acharya J., Verma A., Panigrahi P. K. Multilevel thresholding for image segmentation through a fast statistical recursive algorithm. *Pattern Recognition Letters*, 2008, vol. 29, pp. 119–125.
9. Yang Jie, Yang Yang, Yu Weiyu, Feng Jiuchao. Multi-threshold image segmentation based on K-means and firefly algorithm. *Proc. of 3rd Intern. Conf. on Multimedia Technology, ICMT-13*, 2013, Atlantis Press, pp. 134–142.
10. Priyanka P., Vasudevarao K., Sunitha Y., Sridhar B. A. Multilevel fuzzy threshold image segmentation method for industrial applications. *IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)*, 2017, vol. 12, iss. 2, ver. III, pp. 06–17.
11. Banimelhem O., Yahya Y. Multi-thresholding image segmentation using genetic algorithm. *Proc. IPCV, Las-Vegas, CSREA*, 2012, pp. 1–7.
12. Naidu M. S. R., Kumar P. R. Multilevel image thresholding for image segmentation by optimizing fuzzy entropy using Firefly algorithm. *Int. J. Eng. Technol*, 2017, vol. 9.2, pp. 472–488.
13. Kumari R., Gupta N., Kumar N. Image segmentation using improved genetic algorithm. *International Journal of Engineering and Advanced Technology (IJEAT)*, 2019, vol. 9, iss. 1, pp. 1784–1792.
14. Bunde A., Havlin S. *Fractals and disordered systems*. Springer-Verlag Berlin, Heidelberg, 1996. 408 p.
15. Langovoy M., Wittich O. Randomized algorithms for statistical image analysis and site percolation on square lattices. *Statistica Neerlandica*, 2013, vol. 67, pp. 337–353. doi:10.1111/stan.12010
16. Volkov V. Yu., Bogachev M. I., Kayumov A. R. Object selection in computer vision: from multi-thresholding to percolation based scene representation. In: *Intelligent Systems Reference Library: Computer Vision in Advanced Control Systems-5: Advanced Decisions in Technical and Medical Applications*. Ed. by M. N. Favorskaya, L. C. Jain. Springer, 2019. Pp. 161–194.
17. Волков В. Ю. Адаптивное выделение мелких объектов на цифровых изображениях. *Известия вузов России. Радиоэлектроника*, 2017, № 1, с. 17–28.
18. Rohling H. Ordered statistic CFAR technique — an overview. *Proc. of the 12th International Radar Symposium (IRS)*, 2011, pp. 631–638.
19. Волков В. Ю., Богачев М. И., Маркелов О. А. Многопороговая селекция объектов на изображениях в системах дистанционного наблюдения. *Радиопромышленность*, 2019, т. 29, № 2, с. 76–88. doi.org/10.21778/2413-9599-2019-29-2-76-88
20. Fan J., Meng J., Saberi A. A. Percolation framework of the Earth's topography. *Phys. Rev. E*, 2019, 99:022304. doi:10.1103/PhysRevE.99.022304

UDC 004.932.2:621.396.96

doi:10.31799/1684-8853-2020-3-12-24

Adaptive multi-threshold object selection in remote sensing imagesV. Yu. Volkov^{a,b}, Dr. Sc., Tech., Professor, orcid.org/0000-0001-8552-4775, vl_volk@mail.ru^aSaint-Petersburg State University of Aerospace Instrumentation, 67, B. Morskaya St., 190000, Saint-Petersburg, Russian Federation^bSaint-Petersburg Electrotechnical University «LETI», 5, Prof. Popov St., 197376, Saint-Petersburg, Russian Federation

Introduction: Detection, selection and analysis of objects of interest in digital images is a major problem for remote sensing and technical vision systems. The known methods of threshold detection and selection of objects avoid using the processing results, therefore not providing a low probability of false alarms, and not keeping the shape of the selected objects well enough. There are only few results from the studies about quantifying the quality of such algorithms on either model or real images. **Purpose:** Studying the effectiveness of algorithms for detecting, selecting, and localizing objects of interest using their geometric characteristics, when the object properties and background are a priori uncertain, and the shape of the selected objects is kept unchanged. **Results:** We have obtained and studied the characteristics of algorithms for detecting and selecting objects of interest on test models of monochrome images. These software-implemented algorithms use multi-threshold processing, providing a set of binary slices. This makes it possible to perform morphological processing of objects on each slice in order to analyze their geometric characteristics and then select them according to geometric criteria, taking into account the percolation effect which causes changes in the area, and fragmentation of the objects. As a result of analyzing these changes, an adaptive detection threshold is set for each of the selected objects. The selection allows you to significantly reduce the number of false positives during the detection and to use lower thresholds, increasing the correct detection probability. We present the detection characteristics and the results of test model processing, as well as the results of object selection on a real television and radar image, confirming the effectiveness of the considered algorithms. **Practical relevance:** The proposed algorithms can more effectively select objects on images of various nature obtained in remote sensing, material research or medical diagnostics systems. Their microprocessor implementation is much simpler than the implementation of universal trainable neural network algorithms.

Keywords — multi-threshold processing, object selection, percolation, adaptive threshold, feature selection, segmentation criteria.

For citation: Volkov V. Yu. Adaptive multi-threshold object selection in remote sensing images. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2020, no. 3, pp. 12–24 (In Russian). doi:10.31799/1684-8853-2020-3-12-24

References

- Cheng G., Han J. A survey on object detection in optical remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2016, vol. 117, pp. 11–28.
- Arias-Castro E., Grimmett G. R. Cluster detection in networks using percolation. *Bernoulli*, 2013, vol. 19(2), pp. 676–719.
- Patil G. P., Taillie C. Upper level set scan statistic for detecting arbitrarily shaped hotspots. *Environmental and Ecological Statistics*, 2004, vol. 11, pp. 183–197.
- Zhou W., Troy A. An object-oriented approach for analyzing and characterizing urban landscape at the parcel level. *International Journal of Remote Sensing*, 2008, vol. 29(11), pp. 3119–3135.
- Gu H., Han Y., Yang Y., Li H., Liu Z., Soergel U., Blaschke T., Cui S. An efficient parallel multi-scale segmentation method for remote sensing imagery. *Remote Sensing*, 2018, vol. 10(4), pp. 590–608.
- Shivhare B. D., Gupta S. K. Multilevel thresholding based image segmentation using whale optimization algorithm. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 2019, vol. 8, iss. 12, pp. 4602–4613.
- Cuevas E., González A., Fausto F., Zaldívar D., Pérez-Cisneros M. Multithreshold segmentation by using an algorithm based on the behavior of locust swarms. *Mathematical Problems in Engineering*, 2015, article ID 805357, pp. 1–12.
- Arora S., Acharya J., Verma A., Panigrahi P. K. Multilevel thresholding for image segmentation through a fast statistical recursive algorithm. *Pattern Recognition Letters*, 2008, vol. 29, pp. 119–125.
- Yang Jie, Yang Yang, Yu Weiyu, Feng Jiuchao. Multi-threshold image segmentation based on K-means and firefly algorithm. *Proc. of 3rd Intern. Conf. on Multimedia Technology, ICMT-13*, 2013, Atlantis Press, pp. 134–142.
- Priyanka P., Vasudevarao K., Sunitha Y., Sridhar B. A. Multilevel fuzzy threshold image segmentation method for industrial applications. *IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)*, 2017, vol. 12, iss. 2, ver. III, pp. 06–17.
- Banimelhem O., Yahya Y. Multi-thresholding image segmentation using genetic algorithm. *Proc. IPCV*, Las-Vegas, CSREA, 2012, pp. 1–7.
- Naidu M. S. R., Kumar P. R. Multilevel image thresholding for image segmentation by optimizing fuzzy entropy using Firefly algorithm. *Int. J. Eng. Technol.*, 2017, vol. 9.2, pp. 472–488.
- Kumari R., Gupta N., Kumar N. Image segmentation using improved genetic algorithm. *International Journal of Engineering and Advanced Technology (IJEAT)*, 2019, vol. 9, iss. 1, pp. 1784–1792.
- Bunde A., Havlin S. *Fractals and disordered systems*. Springer-Verlag Berlin, Heidelberg, 1996. 408 p.
- Langovoy M., Wittich O. Randomized algorithms for statistical image analysis and site percolation on square lattices. *Statistica Neerlandica*, 2013, vol. 67, pp. 337–353. doi:10.1111/stan.12010
- Volkov V. Yu., Bogachev M. I., Kayumov A. R. *Object selection in computer vision: from multi-thresholding to percolation based scene representation*. In: *Intelligent Systems Reference Library: Computer Vision in Advanced Control Systems-5: Advanced Decisions in Technical and Medical Applications*. Ed. by M. N. Favorskaya, L. C. Jain. Springer, 2019. Pp. 161–194.
- Volkov V. Yu. Adaptive extraction of small objects in digital images. *Izvestiya vysshikh uchebnykh zavedenii Rossii. Radioelektronika*, 2017, no. 1, pp. 17–28 (In Russian).
- Rohling H. Ordered statistic CFAR technique — an overview. *Proc. of the 12th International Radar Symposium (IRS)*, 2011, pp. 631–638.
- Volkov V. Yu., Bogachev M. I., Markelov O. A. Multi-threshold object selection in images of remote sensing systems. *Radio industry (Russia)*, 2019, vol. 29(2), pp. 76–88 (In Russian). doi.org/10.21778/2413-9599-2019-29-2-76-88
- Fan J., Meng J., Saberi A. A. Percolation framework of the Earth's topography. *Phys. Rev. E*, 2019, 99:022304. doi:10.1103/PhysRevE.99.022304

Алгоритм вычисления значений весов синапсов первого слоя нейронной сети на основе метрических методов распознавания. Часть 2¹

П. Ш. Гейдаров^а, канд. техн. наук, доцент, orcid.org/0000-0002-3881-0629, plbaku2010@gmail.com

^аИнститут систем управления Национальной академии наук Азербайджана, Бахтияр Вагабзаде ул., 9, Баку, Az 1141, Азербайджан

Введение: метрические методы распознавания позволяют предварительно и строго определять структуры нейронных сетей прямого распространения, а именно количество нейронов, слоев и связей на основе начальных параметров задачи распознавания. Они также дают возможность на основе метрических выражений близости аналитически вычислять веса синапсов нейронов сети. Процедура настройки для данных сетей включает в себя последовательное аналитическое вычисление значения каждого веса синапса в таблице весов для нейронов нулевого или первого слоя, что позволяет уже на начальном этапе без применения алгоритмов обучения получить работоспособную нейронную сеть прямого распространения. Затем нейронные сети прямого распространения могут дообучаться известными алгоритмами обучения, что в целом ускоряет процедуры их создания и обучения. **Цель:** определить, сколько времени требует процесс вычисления значений весов и, соответственно, насколько является оправданным предварительное вычисление значений весов нейронной сети прямого распространения. **Результаты:** предложен и реализован алгоритм автоматизированного вычисления всех значений таблиц весов синапсов для нулевого и первого слоя применительно к задаче распознавания черно-белых однотонных изображений символов. Описание предлагаемого алгоритма приведено в программной среде Builder C++. Рассмотрена возможность оптимизировать процесс вычисления весов синапсов в целях ускорения всего алгоритма. Выполнена оценка затрачиваемого времени на вычисление этих весов для разных конфигураций нейронных сетей на основе метрических методов распознавания. Приведены примеры создания таблиц весов синапсов согласно рассмотренному алгоритму. Результаты вычисления таблиц показывают, что на процедуру аналитического вычисления весов нейронной сети потрачены считанные секунды, минуты, что никак не сравнимо со временем, необходимым для обучения нейронной сети. **Практическая значимость:** аналитическое вычисление значений весов нейронной сети позволяет существенно ускорить процедуру создания и обучения нейронной сети прямого распространения. На основе предложенного алгоритма может быть также реализован и алгоритм вычисления трехмерных таблиц весов для более сложных, черно-белых и цветных полутоновых, изображений.

Ключевые слова – нейронные сети, весовые и пороговые значения, нейрокомпьютер, алгоритмы обучения, программирование нейронных сетей.

Для цитирования: Гейдаров П. Ш. Алгоритм вычисления значений весов синапсов первого слоя нейронной сети на основе метрических методов распознавания. Ч. 2. *Информационно-управляющие системы*, 2020, № 3, с. 25–38. doi:10.31799/1684-8853-2020-3-25-38

For citation: Geidarov P. Sh. Algorithm for calculating synapse weights of the first layer of a neural network on the base of metric recognition methods. Part 2. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2020, no. 3, pp. 25–38 (In Russian). doi:10.31799/1684-8853-2020-3-25-38

Оценка времени вычисления таблиц весов

С увеличением размерности таблицы весов синапсов количество связей нейрона первого слоя увеличивается. Оно равно количеству ячеек в таблице весов и определяется по выражению $n_{\text{связей}} = CR$, где C — количество столбцов; R — количество строк таблицы весов. При этом время вычислений также увеличивается. Для оценки времени вычисления значения весов синапсов использовалась встроенная в C++ функция `clock()`, которая выводит время в миллисекундах в начале функции `Izmeneniye_X0_Y0()` и после завер-

шения вычисления таблицы весов. Результаты вычитаются, переводятся в секунды, и итоговое время вычисления таблицы весов выводится в статусную панель формы (см. рис. 4):

```
start = clock();
.....
.....
end = clock();
timework = end - start;
timework = timework/1000;
Form1->StatusBar1->SimpleText = "Время вычисления" + FloatToStr(timework) + "с";
```

Для таблиц весов, приведенных на рис. 8, было потрачено время, соответственно для каждого символа: $t_A = 0,078$ с и $t_C = 0,093$ с. Эти данные были получены на компьютере с использованием про-

¹ Окончание. Начало см. *Информационно-управляющие системы*, 2020, № 2, с. 20–30. doi:10.31799/1684-8853-2020-2-20-30

цессора со скоростью работы 2,8 ГГц и оперативной памятью 1 ГБ. При использовании более быстрых конфигураций время вычисления будет меньше.

Таблицы весов на рис. 8 являются таблицами весов синапсов нулевого слоя НС. Если в НС на основе метрического метода распознавания используется N эталонов, то понадобится в общем количестве N таблиц весов. Если реализуется НС без нулевого слоя, то поскольку для N эталонов требуется $n = N(N - 1)$ нейронов первого слоя (см. рис. 2), то столько же потребуется и таблиц весов, либо понадобится $n = N(N - 1)/2$ нейронов для сжатой схемы НС [26]. При этом одна таблица весов синапсов нейрона первого слоя $W_{i,j}^{(1)}$ определяется как разность двух таблиц весов нулевого слоя $W_i^{(0)}, W_j^{(0)}$:

$$W_{i,j}^{(1)} = W_i^{(0)} - W_j^{(0)}. \quad (16)$$

Например, если в выражении (16) i и j соответствуют символам «А» и «С», приведенным на рис. 8, то в НС без нулевого слоя таблица весов нейрона первого слоя $W_{A,C}^{(1)}$ будет выглядеть так, как показано на рис. 9.

Тогда время вычисления таблиц весов синапсов T для всех нейронов нулевого слоя определится как $T = n \times t_{cp}$, где n — количество нейронов первого или второго слоя; t_{cp} — среднее время вычисления одной таблицы весов синапсов. К примеру, если мы, исходя из рис. 8, определяем среднее время вычисления таблицы весов синапсов как $t_{cp} = (0,078 + 0,093)/2 = 0,0855$ с, то для задачи распознавания алфавитных символов с количеством распознаваемых образов, равным 30, в котором

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	21	15	11	7	3	1	0	0	0	0	1	4	8	12	18	25
1	21	15	9	5	2	1	0	0	0	0	1	2	5	10	16	24
2	23	15	9	4	1	0	0	-1	-1	-1	-1	1	4	8	13	20
3	19	13	8	4	1	0	-1	-2	-4	-4	-2	0	2	5	10	17
4	17	10	5	2	0	-1	-2	-5	-7	-8	-5	-2	0	4	9	16
5	13	8	4	1	-1	-4	-5	-7	-11	-12	-8	-5	-1	3	8	12
6	10	5	2	0	-2	-5	-9	-11	-14	-17	-13	-8	-4	0	4	8
7	9	4	1	-1	-4	-8	-12	-16	-21	-24	-18	-13	-8	-4	0	3
8	8	4	1	-1	-4	-9	-15	-21	-30	-31	-25	-18	-13	-9	-7	-4
9	5	2	1	-1	-4	-8	-12	-19	-24	-24	-17	-13	-8	-5	-3	0
10	4	1	0	-1	-2	-5	-10	-13	-16	-16	-13	-8	-5	-2	0	1
11	2	1	0	0	-1	-4	-5	-8	-9	-9	-8	-5	-2	-1	0	1
12	1	0	0	0	0	0	-1	-3	-3	-3	-3	-1	-1	0	0	1
13	0	0	0	0	1	4	3	3	3	3	3	0	0	0	0	1
14	-1	-1	0	1	2	5	9	9	9	9	5	2	1	0	0	-1
15	-5	-2	-1	1	4	8	13	16	16	13	8	5	2	1	-1	-2

■ **Рис. 9.** Таблица весов синапсов для нейрона первого слоя сети без нулевого слоя, определенных для пары эталонных печатных символов «А» и «С» с размерностью 16×16

■ **Fig. 9.** Table of weights for a neuron of the first layer of a neural network without a zero layer, defined for a pair samples printed characters “A” and “C” with a dimension of 16×16

для каждого образа используется в среднем по три эталона (что в целом составляет 90 эталонов), понадобится время вычисления $T = 90 \times t_{cp} \approx 7$ с для НС с нулевым слоем, $T = 90 \times 89 \times t_{cp} \approx 6,23$ мин для архитектуры расширенной НС без нулевого слоя (см. рис. 2) и $T = 90 \times 89 \times t_{cp} / 2 \approx 3$ мин для сжатой схемы НС.

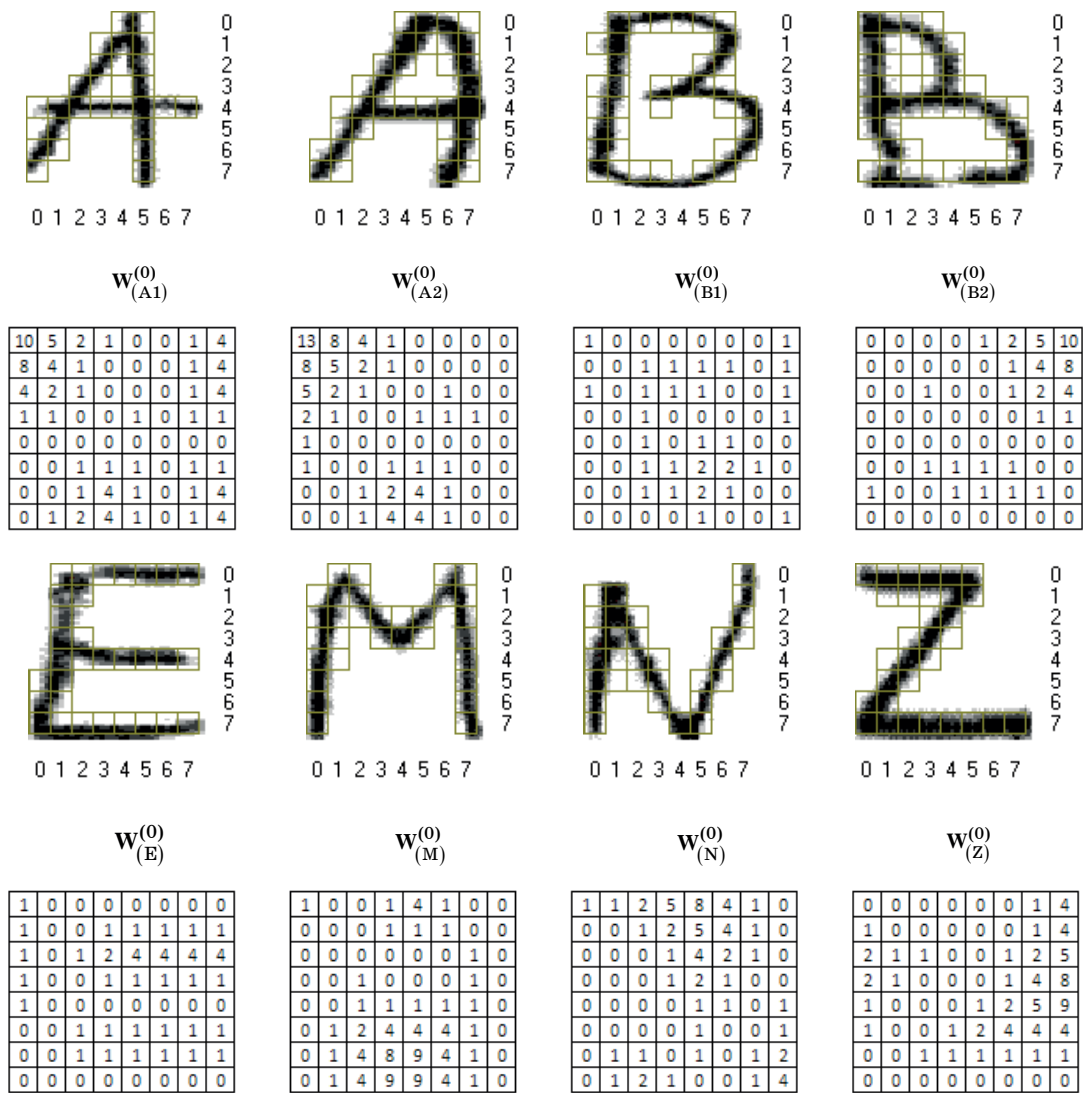
Нужно отметить, что при обучении НС классическими алгоритмами время требуется значительно большее, которое измеряется часами, а иногда и сутками. Таким образом, вычисление значения весов синапсов предлагаемым алгоритмом существенно выигрывает по времени, при этом создается уже работающая НС, которая может еще и доучаться классическими алгоритмами.

Примеры

Пример 1. Создание НС и применение вычисленных весовых значений синапсов для набора изображения букв

Приведем наглядный пример НС для задачи распознавания шести образов рукописных букв А, В, Е, М, N, Z (рис. 10). Для нашей задачи $N = 8$ (количество эталонов); образу символа «А» и «В» соответствует по два эталона, наименованные как А1, А2, В1, В2. Для наглядности и простоты примера будем также использовать схему НС с нулевым слоем (см. рис. 3).

Исходя из условия задачи количество нейронов нулевого слоя $n_0 = 8$; количество нейронов первого слоя $n_1 = 8 \times 7 = 56$; количество нейронов второго слоя равно количеству эталонов: $n_2 = 8$; количество нейронов третьего слоя равно количеству образов: $n_3 = 6$. Размерность таблицы весов синапсов примем равной 8×8 , и, таким образом, количество связей одного нейрона нулевого слоя будет равно 64. На рис. 10 приведены таблицы весов синапсов для каждого нейрона нулевого слоя, вычисленные описанным выше способом. Каждому нейрону нулевого слоя соответствует одна вычисленная таблица весов синапсов согласно их расположению на рис. 10, т. е. первому нейрону нулевого слоя соответствует таблица весов синапсов, определенная эталоном «А1», второму нейрону — эталоном «А2», третьему нейрону — эталоном «В1», четвертому нейрону — эталоном «В2», пятому нейрону — эталоном «Е», шестому нейрону — эталоном «М», седьмому нейрону — эталоном «N», восьмому нейрону — эталоном «Z». В этом же порядке определяются и выходы второго слоя. Выходы НС (выходы третьего слоя) определяются в следующем порядке: образу «А» соответствует 1-й выход, «В» — 2-й, «Е» — 3-й, «М» — 4-й, «N» — 5-й, «Z» — 6-й выход. В качестве входного тестируемого символа будем использовать значение рукописного символа



■ **Рис. 10.** Эталонные символы A1, A2, B1, B2, E, M, N, Z с размерностью матрицы разбиения изображений 8×8 и соответствующие им таблицы весов синапсов нейронов нулевого слоя
 ■ **Fig. 10.** Samples symbols A1, A2, B1, B2, E, M, N, Z with the dimension of the image partition matrix 8×8 and the corresponding tables of weights of neurons of the zero layer

ла «N», приведенного на рис. 11, а. Это означает, что в случае правильности узнавания входного символа НС 5-й выход НС должен активизироваться (стать равным 1), тогда как остальные выходы должны быть равны 0. На рис. 11, б приведена бинарная матрица входного тестируемого символа с рис. 11, а с размерностью также 8×8 . При применении тестируемого символа (см. рис. 11, а) определим формируемые значения состояний и активаций всех нейронов НС. На рис. 12 приведены таблицы умножений бинарной матрицы входного символа (см. рис. 11, а) на таблицы весов (см. рис. 10), а также значения состояния нейронов, определяемые как

$$Sw_i^{(0)} = \sum XW_i^{(0)}. \quad (17)$$

Здесь для нулевого слоя используются нейроны, у которых значения функции активации нейрона равны значению состояния нейрона:

$$f(Sw_i^{(0)}) = Sw_i^{(0)}. \quad (18)$$

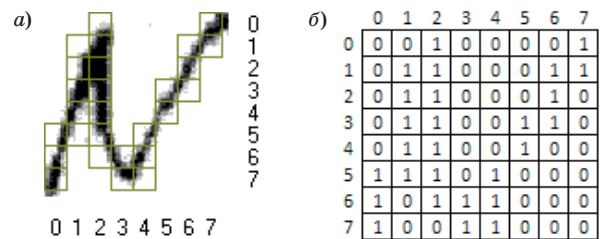
Значения функций состояний нейронов первого слоя определяются по выражению

$$Sw_{i,j}^{(1)} = Sw_i^{(0)} - Sw_j^{(0)}, \quad (19)$$

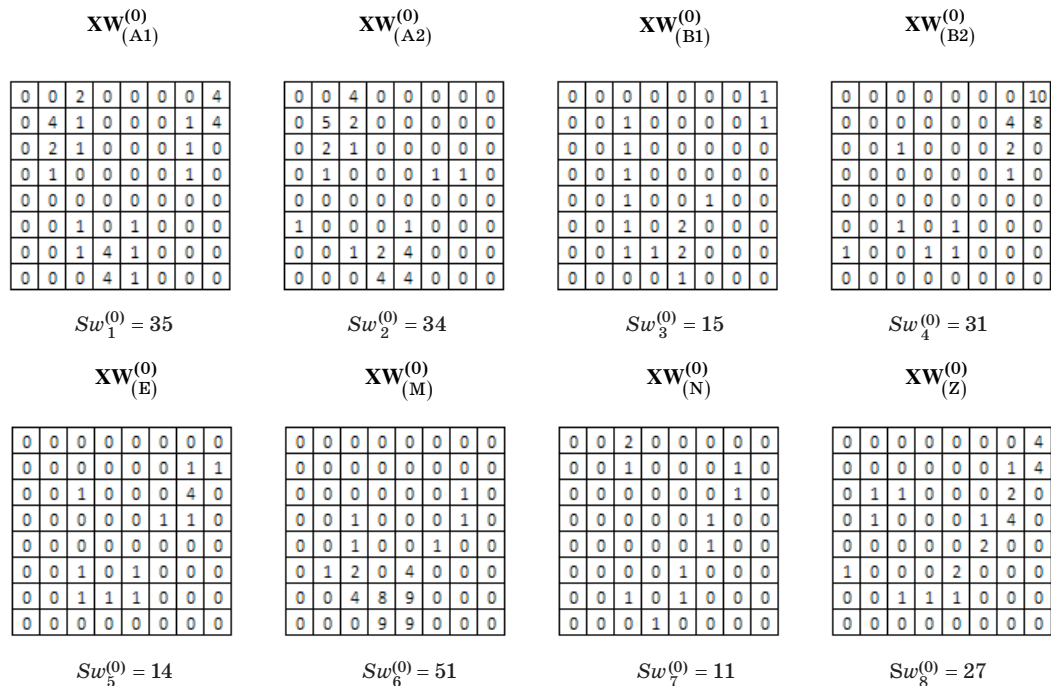
где i и j — порядковые номера попарно сравниваемых эталонов. В табл. 1 приведены значения функций состояний всех $8 \times 7 = 56$ нейронов первого слоя. Функции активации нейронов первого слоя $f(Sw_{i,j}^{(1)})$ определяются по выражению

$$\begin{cases} f(Sw_{i,j}^{(1)}) = 1, \text{ если } Sw_{i,j}^{(1)} < 0 \\ f(Sw_{i,j}^{(1)}) = 0, \text{ если } Sw_{i,j}^{(1)} \geq 0 \end{cases} \quad (20)$$

Значения всех выходов $f(Sw_{i,j}^{(1)})$ нейронов первого слоя приведены в табл. 2. Каждый нейрон второго слоя объединяет семь выходов нейронов первого слоя, в которых в выражении (19) индекс i соответствует одному и тому же эталону. Например, первый нейрон второго слоя соединен с семью нейронами первого слоя, где i в (19) и (20)



■ **Рис. 11.** Тестируемый рукописный символ «N» (а) и бинарная матрица X тестируемого символа (б)
 ■ **Fig. 11.** Test handwritten character “N” (a) and binary matrix of the test character X (b)



■ **Рис. 12.** Таблицы значений произведений бинарной матрицы X тестируемого (входного) символа с таблицами весов синапсов нейронов нулевого слоя W^0 и значения функций состояния $Sw_i^{(0)}$ нейронов нулевого слоя
 ■ **Fig. 12.** Tables of product values of the binary matrix X of the test (input) symbol with tables of weights of the zero layer neurons W^0 and values of the state functions $Sw_i^{(0)}$ of zero layer neurons

■ **Таблица 1.** Значения функций состояний $Sw_{i,j}^{(1)}$ (19) нейронов первого слоя
 ■ **Table 1.** Values of state functions $Sw_{i,j}^{(1)}$ (19) neurons of the first layer

Символ, i , $Sw_i^{(0)}$	Символ, j , $Sw_j^{(0)}$							
	A1, 1, 35	A2, 2, 34	B1, 3, 15	B2, 4, 31	E, 5, 14	M, 6, 51	N, 7, 11	Z, 8, 27
A1, 1, 35		1	20	4	21	-16	24	8
A2, 2, 34	-1		19	3	20	-17	23	7
B1, 3, 15	-20	-19		-16	1	-36	4	-12
B2, 4, 31	-4	-3	16		17	-20	20	4
E, 5, 14	-21	-20	-1	-17		-37	3	-13
M, 6, 51	16	17	36	20	37		40	24
N, 7, 11	-24	-23	-4	-20	-3	-40		-16
Z, 8, 27	-8	-7	12	-4	13	-24	16	

■ **Таблица 2.** Значения выходов нейронов первого, второго и третьего слоя
 ■ **Table 2.** The values of the outputs of the neurons of the first, second and third layer

Символ, <i>i</i>	Символ, <i>j</i>								Второй слой		Третий слой	
	A1, 1	A2, 2	B1, 3	B2, 4	E, 5	M, 6	N, 7	Z, 8	$Sw_k^{(2)}$	$f(Sw_k^{(2)})$	$Sw_{\text{ооп}}^{(3)}$	Y_{out}
	$f(Sw_{i,j}^{(1)})$											
A1, 1		1	20	4	21	-16	24	8	1	0	0 + 0 = 0	0
A2, 2	-1		19	3	20	-17	23	7	2	0		0
B1, 3	-20	-19		-16	1	-36	4	-12	5	0	0 + 0 = 0	0
B2, 4	-4	-3	16		17	-20	20	4	3	0		0
E, 5	-21	-20	-1	-17		-37	3	-13	6	0	0	0
M, 6	16	17	36	20	37		40	24	0	0	0	0
N, 7	-24	-23	-4	-20	-3	-40		-16	7	1	1	1
Z, 8	-8	-7	12	-4	13	-24	16		4	0	0	0

определяет эталон «A1»: $f(Sw_{A1,A2}^{(1)})$, $f(Sw_{A1,B1}^{(1)})$, $f(Sw_{A1,B2}^{(1)})$, $f(Sw_{A1,E}^{(1)})$, $f(Sw_{A1,M}^{(1)})$, $f(Sw_{A1,N}^{(1)})$, $f(Sw_{A1,Z}^{(1)})$. Соответственно, активный выход *i*-го нейрона второго слоя определяет схожесть входного распознаваемого элемента к эталону «A1». Аналогично и по другим нейронам второго слоя. Функции состояния нейрона второго слоя для расширенных схем рис. 2, 3 определяются следующим образом [26, 27]:

$$Sn_k^{(2)} = \alpha \sum_{j=1, j \neq k}^{N-1} f(Sn_{k,j}^{(1)}); \quad (21)$$

$$\begin{cases} f(Sn_k^{(2)}) = 1, \text{ если } Sn_k^{(2)} \geq \alpha(N-1) = H_k^{(2)} \\ f(Sn_k^{(2)}) = 0, \text{ если } Sn_k^{(2)} < \alpha(N-1) = H_k^{(2)} \end{cases}, \quad (22)$$

где *N* — количество эталонов; α — постоянная величина, определяющая значение веса $w_k^{(2)}$ входных связей нейронов второго слоя. В данном примере $\alpha = 1$; $H_k^{(2)} = \alpha(N - 1) = 7$ — порог нейрона второго слоя. Поскольку для данного примера количество эталонов равно восьми, то порог $H_k^{(2)} = 7$ для всех нейронов второго слоя. В табл. 2 показаны также значения функций состояний $Sw_k^{(2)}$ активации $f(Sw_k^{(2)})$ нейронов второго слоя согласно приведенным выражениям (21), (22). Можно видеть, что уже на этом этапе 7-й выход второго слоя, соответствующий эталону «N», является активным. В третьем слое НС выполняется объединение нейронов одного образа в одном нейроне. Для данного примера третий слой необходим для первого и второго образов «A» и «B», поскольку эти образы имеют по два эталона и их выходы объединяются в нейронах третьего слоя. Функции состояния $Sn_k^{(3)}$ и $f(Sn_k^{(3)})$ активации нейрона третьего слоя определяются следующим образом:

$$Sn_k^{(3)} = \sum_{j=1}^K (\alpha \times f(Sn_j^{(2)})); \quad (23)$$

$$\begin{cases} f(Sn_k^{(3)}) = 1, \text{ если } Sn_k^{(3)} > 0 \\ f(Sn_k^{(3)}) = 0, \text{ если } Sn_k^{(3)} \leq 0 \end{cases}, \quad (24)$$

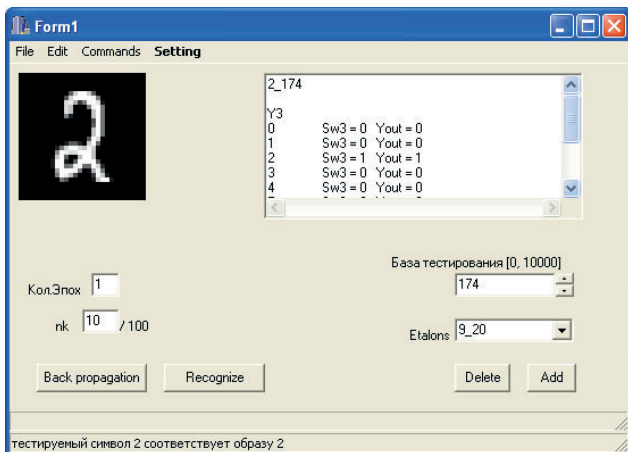
где *K* — количество эталонов в *k*-м образе; α — постоянное значение, определяющее значение веса связей нейронов третьего слоя $w_k^{(3)}$, для данного примера также $\alpha = 1$.

Из табл. 2 можно видеть, что в итоге для входного символа на рис. 11, *a* активизируется 5-й выход НС, соответствующий образу «N».

Пример 2. Создание и обучение НС на основе базы цифр MNIST

В данном примере НС создается и далее обучается на базе MNIST (рис. 13). Набор эталонов набирается из базы MNIST, размерность таблицы весов синапсов определяется также на основе размерности матрицы изображений базы MNIST, превышая ее в два раза (28 × 56). Дальнейшее дообучение сети выполняется также на базе MNIST. Для оценки эффективности полученной сети обучение той же НС выполняется и классическим способом со случайной инициализацией всех весов НС. Результаты сравниваются.

Напомним, что база MNIST состоит из обучающей базы, в которую входят 60 000 изображений рукописных цифр, и контрольной (тестовой) базы, состоящей из 10 000 образцов изображений цифр. Каждой базе также прилагается своя база наименований цифр, порядок расположения которых такой же, как и порядок расположения изображений цифр в обучающей и контрольной базах. Изображения цифр в базах описаны в виде матрицы цифр размерностью 28 × 28. Каждая



■ **Рис. 13.** Программный модуль для распознавания и обучения изображений базы MNIST, реализованный в среде Builder C++

■ **Fig. 13.** Software module for recognition and training of images of the MNIST base, implemented in the Builder C++ environment

цифра матрицы определяет значение тональности одного пикселя изображения цифры в диапазоне [0, 255].

В качестве эталонов было выбрано по три образца изображений цифр от каждого образа. Итого использовалось 30 эталонов (рис. 14). Эталоны были выбраны интуитивно из первых 250 изображений цифр контрольной базы MNIST. Над каждым выбранным эталоном приведено наименование изображения, например 2_35, в котором первая цифра 2 указывает образ, которому принадлежит изображение цифры, число 35 указывает порядковый номер данного изображения цифры в базе MNIST. Поскольку количество эталонов равно 30, то в соответствии с архитектурой НС, реализующих метрические методы распознавания, количество нейронов второго слоя будет равно также 30. Эталоны в рассматриваемой задаче расположены в такой же последовательности, как на рис. 14. Сначала столбец с изображениями цифры «0», потом цифры «1» и т. д. В соответствии с этим схема сети на рис. 2



■ **Рис. 14.** Выбранные эталоны из контрольной базы MNIST

■ **Fig. 14.** Selected samples from the MNIST control base

определяет и выходы второго слоя. Количество нейронов третьего слоя равно количеству распознаваемых образов цифр: $n^{(3)} = 10$. Каждый i -й выход третьего слоя определяет принадлежность распознаваемого элемента к i -му образу цифры. Порядок образов цифр определен последовательно от 0 до 9.

Количество нейронов первого слоя $n^{(1)} = 30 \times 29 = 870$. В процессе распознавания или обучения для каждого входного распознаваемого изображения составляется бинарная матрица, которая, в отличие от первого примера на рис. 11, б, будет состоять из двух частей. В первой части бинарной матрицы единицы определяют светлые пиксели изображения, значения которых >150 , и наоборот, нули определяют затемненные пиксели изображения, значения которых <150 . Другая часть матрицы зеркально противоположна, определяет активными ($= 1$) темные пиксели изображений (<150) и неактивными ($= 0$) светлые пиксели изображений (>150). Соответственно, размерность бинарной матрицы равна 28×56 . Каждая таблица весов синапсов нулевого слоя соразмерна с бинарной матрицей и состоит также из двух частей (рис. 15, а и б).

Соответственно, таблица весов синапсов первого слоя определяется на основе выражения (19).

Прежде чем начать обучение НС (см. рис. 2) алгоритмом *back propagation*, необходимо провести некоторые преобразования в сети. Поскольку значения весов синапсов первого слоя получаются большими, в основном располагаются в диапазоне [0, 100], что не рекомендуется для алгоритма *back propagation*, то каждое значение веса синапсов первого слоя делится на 100, что не меняет результативность полученной сети. На рис. 16, а и б приведен пример таблицы весов синапсов первого слоя, вычисленных для пары эталонов 2_172 и 5_102. Таким образом, для каждого нейрона первого слоя вычисляется 870 таблиц весов синапсов. При этом нужно отметить, что в данном примере при обучении НС нейроны нулевого слоя не используются, рассматриваться будет только полученная трехслойная сеть (см. рис. 2).

Нейронная сеть на рис. 2 является сетью прямого распространения, но при этом второй и третий слой не являются полносвязными. В работе [5] были приведены обобщенный алгоритм создания полносвязной НС и вычисления значений весов синапсов второго и третьего слоя, при которых сохраняется первоначальная логика работы сети. Но в простейшем случае из схемы на рис. 2 полносвязную НС можно получить добавлением всех недостающих связей второго и третьего слоя, значения весов синапсов которых будут приравнены к нулю. В этом случае логика работы сети на рис. 2 также не изменится. На рис. 17, а и б приведены значения весов связей нейронов второго и треть-

■ **Таблица 3.** Выходы третьего слоя НС с пороговой и сигмоидной функцией активации при распознавании символа 2_174 (см. рис. 13) из контрольной базы MNIST

■ **Table 3.** Outputs of the third layer of a neural network with a threshold and sigmoidal activation function upon recognition of the 2_174 symbol (see Fig. 13) from the MNIST control base

Функция активации					
пороговая			сигмоидная		
Выход	$Sw3$	Y_{out}	Выход	$Sw3$	Y_{out}
0	0	0	0	0,000270242677595017	0,500067560668988
1	0	0	1	3,43951083407716E-6	0,500000859877708
2	1	1	2	0,0526576299057448	0,513161366435398
3	0	0	3	0,0109778057331013	0,502744423871947
4	0	0	4	9,28464982946175E-6	0,500002321162457
5	0	0	5	0,000934588149375769	0,500233647020337
6	0	0	6	0,000250751415905857	0,500062687853648
7	0	0	7	1,20144980966343E-7	0,500000030036245
8	0	0	8	0,014626235919069	0,503656493794839
9	0	0	9	0,00149247436075622	0,50037311852093

■ **Таблица 4.** Результаты распознавания контрольной базы MNIST (10 000 символов) без обучения

■ **Table 4.** MNIST control base recognition results (10 000 characters) without learning

Функция активации					
пороговая			сигмоидная		
s	i	$p, \%$	s	i	$p, \%$
$s_0 = 834$	$i_0 = 980$	$p_0 = 85$	$s_0 = 962$	$i_0 = 980$	$p_0 = 98$
$s_1 = 968$	$i_1 = 1135$	$p_1 = 85$	$s_1 = 872$	$i_1 = 1135$	$p_1 = 76$
$s_2 = 530$	$i_2 = 1032$	$p_2 = 51$	$s_2 = 530$	$i_2 = 1032$	$p_2 = 51$
$s_3 = 454$	$i_3 = 1010$	$p_3 = 44$	$s_3 = 420$	$i_3 = 1010$	$p_3 = 41$
$s_4 = 410$	$i_4 = 982$	$p_4 = 41$	$s_4 = 487$	$i_4 = 982$	$p_4 = 49$
$s_5 = 411$	$i_5 = 892$	$p_5 = 46$	$s_5 = 421$	$i_5 = 892$	$p_5 = 47$
$s_6 = 586$	$i_6 = 958$	$p_6 = 61$	$s_6 = 151$	$i_6 = 958$	$p_6 = 15$
$s_7 = 556$	$i_7 = 1028$	$p_7 = 54$	$s_7 = 499$	$i_7 = 1028$	$p_7 = 48$
$s_8 = 773$	$i_8 = 974$	$p_8 = 79$	$s_8 = 298$	$i_8 = 974$	$p_8 = 30$
$s_9 = 750$	$i_9 = 1009$	$p_9 = 74$	$s_9 = 227$	$i_9 = 1009$	$p_9 = 22$
Итого: $s = 6272, i = 10\ 000, p = 62 \%$			Итого: $s = 4867, i = 10\ 000, p = 48 \%$		

но идентифицированных объектов контрольной базы MNIST отдельно для каждого образа цифры (s_j, p_j , где j — наименование образа), а также приводится общее количество изображений i_j для каждого j -образа в контрольной базе MNIST.

По данным табл. 4 можно видеть, что общее количество правильно идентифицированных изображений MNIST составило 62 % для НС с пороговой функцией активации и 48 % с сигмоидной

функцией активации. Падение до 48 % объясняется тем, что установленное правило проверки на выходе ($>0,51$) в некоторых случаях приводит к ошибкам активации одновременно нескольких выходов третьего слоя, а в некоторых случаях, наоборот, к тому, что на выходе НС нет активных выходов, в результате чего итоговое количество правильно узанных изображений уменьшается на 14 %. В действительности потенциал вычис-

ленных значений весов синапсов сохраняется и с сигмоидной функцией активации, что подтверждается также распознаванием контрольной базы MNIST НС с сигмоидной функцией активации и с проверкой на выходе НС по наибольшему значению выхода Y_i . В этом случае результат НС с сигмоидной функцией активации идентичен результату НС с пороговой функцией активации и равен также 62 %.

Обучение НС выполнялось стохастическим алгоритмом обратного распространения (*back propagation*) с использованием обучающей выборки MNIST, состоящей из 60 000 изображений. Поправки вносились после каждого представления нового объекта на входы сети и только в случае наличия ошибки распознавания на выходе НС. При обучении активность выхода оценивалась по правилу $>0,51$. Для каждого эксперимента обучения НС использовались три эпохи, из которых первые две обучались со скоростью $nk = 0,1$, а последняя со скоростью $nk = 0,02$. Ошибка обучения вычислялась для каждой эпохи по формуле

$$S_{err} = \frac{1}{2} \sum_{i=0}^P \sum_{k=0}^{N_{обп}} \left(y_k^{(corr)} - f(Sn_k^{(3)}) \right), \quad (26)$$

где $y_k^{(corr)}$ — принятое правильное значение k -го выхода третьего слоя, для активного выхода $y_k^{(corr)} = 0,7$, для неактивного $y_k^{(corr)} = 0,2$; P — количество неправильно идентифицированных изображений обучающей базы MNIST, для которых в процессе эпохи делались правки весов синапсов алгоритмом обратного распространения ошибки.

В табл. 5 приведены результаты обучений полученной НС как с предварительно вычисленными значениями весов синапсов НС, так и классическим образом — со случайной инициализацией весов синапсов. В табл. 5 можно видеть, что для случая обучения с предварительно вычисленными весами результаты по всем параметрам и на всех трех эпохах лучше по сравнению с результатами обучения НС со случайной инициализацией весов синапсов. При этом общее время, потраченное на обучение НС с предварительно вычисленными весами синапсов, на 2 ч 50 мин (499 – 329 = 170 мин) меньше.

В табл. 6 приведены результаты проверки полученной НС после каждой эпохи обучения на контрольной базе MNIST (10 000 изображений). Проверка выполнялась как по правилу $>0,51$

■ **Таблица 5.** Сравнение результатов обучения НС по обучающей выборке MNIST (60 000 изображений) для каждой эпохи обучения

■ **Table 5.** Comparison of the results of training the neural network in the training set MNIST (60 000 images) for each epohe of training

№ эпохи	Скорость обучения в эпохе	Обучение НС с предварительно вычисленными весами				Обучение НС со случайной инициализацией весов в диапазоне $[-0,5; 0,5]$			
		Количество узанных изображений	% узанных изображений	S_{err}	Время, мин	Количество узанных изображений	% узанных изображений	S_{err}	Время, мин
1	0,1	43 932	73	1199	159	35 370	59	1935	256
2	0,1	49 748	83	737	98	46 033	76	1051	139
3	0,02	52 285	87	545	72	49 195	82	784	104
Общее время обучения					329	Общее время обучения			499

■ **Таблица 6.** Сравнение результатов обучения НС с проверкой на контрольной выборке MNIST (10 000 изображений) для каждой эпохи обучения

■ **Table 6.** Comparison of the results of training a neural network with testing on a control sample MNIST (10 000 images) for each era of training

№ эпохи	Скорость обучения в эпохе	Обучение НС с предварительно вычисленными весами		Обучение НС со случайной инициализацией весов	
		Количество узанных изображений по правилу $>0,51$	Количество узанных изображений по наибольшему значению	Количество узанных изображений по правилу $>0,51$	Количество узанных изображений по наибольшему значению
1	0,1	8140	9145	7506	8894
2	0,1	8557	9282	8022	9116
3	0,02	8764	9449	8194	9256

на выходе третьего слоя, так и по наибольшему значению выхода третьего слоя. В табл. 6 также можно видеть, что на всех эпохах количество узанных изображений контрольной базы MNIST больше для случая обучения НС с предварительно вычисленными весами. Наибольшее количество узанных изображений было достигнуто после третьей эпохи обучения НС с предварительно вычисленными весами синапсов и с проверкой по наибольшему значению выходов третьего слоя, что составило 9449 изображений, 94 %. Результаты табл. 5 и 6 можно объяснить тем, что алгоритм обучения отталкивается от тех начальных результатов, которые НС уже имеет. То есть можно сказать, что чем больше начальная результативность НС, тем больше у алгоритма обучения возможности и времени для того, чтобы добраться до более лучших результатов обучений за одну эпоху. По этой же причине с лучшими начальными результатами (с вычисленными весами синапсов) алгоритму обучения требуется меньше времени на обучение сети. При этом можно также сказать, что с большим количеством выбранных эталонов (больше чем три на один образ) начальная результативность сети была бы также больше (>62 %), соответственно, и результаты, приведенные в табл. 5 и 6, могли быть еще более выраженными.

Выводы

Таким образом, процесс вычисления таблиц весов синапсов НС на основе метрических методов распознавания при помощи существующих вычислительных средств позволяет существенно ускорить процедуры создания и обучения НС. В зависимости от параметров эталонных изображений: размеров изображений, количества

пикселей, — а также в зависимости от размерности таблицы весов и общего количества эталонов время вычисления весов исчисляется считанными долями секунд и минут. Это означает, что процедура вычисления значений весов синапсов НС позволяет выполнить существенный скачок по времени в процессе создания и обучения НС. Наименьшее время вычисления значений весов будет свойственно для архитектуры НС с нулевым слоем (см. рис. 3), поскольку для этой архитектуры количество таблиц весов будет минимально и равно количеству используемых эталонов.

Нужно также отметить, что в качестве объектов распознавания могут использоваться не только однотонные черно-белые изображения, но и более сложные объекты, например, полутонные черно-белые и цветные изображения. В этих случаях используемые метрические меры близости могут быть более сложными. Для данного случая, как один из вариантов, таблицы весов для полутонных черно-белых изображений будут трехмерными, а для цветных изображений, помимо трехмерности таблиц, один образец-эталон цветного изображения может быть представлен несколькими черно-белыми эталонами в соответствии с количеством цветовой гаммы цветного изображения (RGB, CMYK и т. д.). В качестве выражения меры близости может быть, например, использовано выражение среднеквадратичной разности в трехмерном пространстве. При этом в рассмотренном выше алгоритме ячейки трехмерной таблицы весов будут перебираться не только по оси X , Y , но еще и по оси Z . Соответственно, для более сложных задач потребуются и больше эталонов и большая размерность таблиц весов синапсов. С увеличением этих параметров результативность НС будет также увеличиваться, но при этом будут увеличиваться и количество нейронов, и количество связей НС.

Литература

1. Азаров И. С., Петровский А. А. Формирование персональной модели голоса диктора с универсальным фонетическим пространством признаков на основе искусственной нейронной сети. *Труды СПИИРАН*, 2014, № 5, с. 12–15. <https://doi.org/10.15622/sp.36.8>
2. Будко Р. Ю., Старченко И. Б. Создание классификатора мимических движений на основе анализа электромиограммы. *Труды СПИИРАН*, 2016, № 3 (46), с. 76–89. <https://doi.org/10.15622/sp.46.7>
3. Осипов В. Ю., Никифоров В. В. Возможности рекуррентных нейронных сетей с управляемыми элементами по восстановлению потоков кадров. *Информационно-управляющие системы*, 2019, № 5, с. 10–17. doi.org/10.31799/1684-8853-2019-5-10-17
4. Ле Т. Ч. Сравнение нейронной сети СМАС и многослойной нейронной сети в задаче обнаружения DoS-атак. *Нейрокомпьютеры: разработка, применение*, 2016, № 7, с. 65–69.
5. Катасев А. С., Катасева Д. В., Кирпичников А. П. Распознавание рукописных символов на базе искусственной нейронной сети. *Вестник технологического университета*, 2018, т. 18, № 11, с. 173–176.
6. Дрокин И. С. Об одном алгоритме последовательной инициализации весов глубоких нейронных сетей и обучении ансамбля нейронных сетей. *Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления*, 2016, № 4, с. 66–74.
7. Лещенко Ю. Е. Ассоциативно-вербальные сети и искусственные нейронные сети как метод моделирования ментального лексикона индивида. *Гло-*

- бальный научный потенциал, 2014, № 10 (43), с. 154–156.
8. Гафуров Д. О., Гафуров О. М., Конторович В. А. Интерпретация данных геофизических исследований Талаканского нефтегазоконденсатного месторождения обучаемыми нейронными сетями, прогноз строения Осинского горизонта. *Технологии сейсморазведки*, 2014, № 4, с. 85–92.
 9. Владимирова Д. Б., Кокшарова А. А. Прогнозирование финансовых рынков искусственными нейронными сетями. *Наука и бизнес: пути развития*, 2014, № 3 (33), с. 42–46.
 10. Синчук О. Н., Бойко С. Н. Нейронные сети и управление процессом управления электроснабжением объектов от комбинированных электрических сетей. *Технічна електродинаміка*, 2014, № 5, с. 53–55.
 11. Шведов В. А., Ильин Е. С. Распознавания дорожных знаков сверточными нейронными сетями различной структуры, с применением графических процессоров CUDA. *Транспортная инфраструктура Сибирского региона: материалы Шестой международной научно-практической конференции*, Иркутск, 30 сентября–03 октября 2015 г., 2015, т. 2, с. 282–286.
 12. Туровский Я. А., Кургалин С. Д., Адаменко А. А. Сравнительный анализ программных пакетов для работы с искусственными нейронными сетями. *Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии*, 2016, № 1, с. 161–168.
 13. Лукина А. С., Некрасов М. В., Пакман Д. Н. Обработка телеметрической информации космического аппарата нейронными сетями на основе теории фильтров Калмана. *Тенденции развития науки и образования*, 2016, № 13-1, с. 43–45.
 14. Стасюк В. В. Прогнозирование чувствительности к удару взрывчатых веществ нейронными сетями с предварительной группировкой данных. *Фундаментальные исследования*, 2015, № 12–6, с. 1139–1143.
 15. Хусаинов А. Т. Оценка прогнозируемости системы поддержания пластового давления нейронными сетями на нефтяных месторождениях. *Академический журнал Западной Сибири*, 2016, т. 12, № 3 (64), с. 48.
 16. Бондарко В. М., Бондарко Д. В., Солнушкин С. Д., Чихман В. Н. Моделирование результатов психофизических экспериментов нейронными сетями. *Нейрокомпьютеры: разработка, применение*, 2018, № 5, с. 31–33.
 17. Sedov V. A., Sedova N. A. Modelling collision avoidance actions in closest approach zones by means of neural networks. *Asia-Pacific Journal of Marine Science & Education*, 2014, vol. 4, no. 2, pp. 104–111.
 18. Peng Shi, Fanbiao Li, Ligang Wu. Neural network-based passive filtering for delayed neutral-type semi-Markovian jump systems. *IEEE Transactions on Neural Networks and Learning Systems*, 2017, vol. 28, iss. 9, pp. 2101–2114.
 19. Chenggang Yan, Hongtao Xie, Dongbao Yang, Jian Yin, Yongdong Zhang, Qionghai Dai. Supervised hash coding with deep neural network for environment perception of intelligent vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 2018, vol. 19, iss. 1, pp. 284–295.
 20. Anusha N., Gregory K., Ronald S. F., Sergey Levine. Neural network dynamics for model-based deep reinforcement learning with model-free fine-tuning. *2018 IEEE International Conference on Robotics and Automation (ICRA)*, Brisbane, Australia, 2018, pp. 156–162.
 21. Wei He, Yuhao Chen, Zhao Yin. Adaptive neural network control of an uncertain robot with full-state constraints. *IEEE Transactions on Cybernetics*, 2016, vol. 46, iss. 3, pp. 620–629.
 22. Song Han, Xingyu Liu, Huizi Mao, Jing Pu, Arda van Pedram, Mark A. H., William J. Efficient inference engine on compressed deep neural network. *2016 ACM/IEEE 43rd Annual International Symposium on Computer Architecture (ISCA)*, Seoul, South Korea, 2016, pp. 304–309.
 23. William Chan, Navdeep Jaitly, Quoc Le, Oriol Vinyals. Listen, attend and spell: A neural network for large vocabulary conversational speech recognition. *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, 2016, pp. 107–115.
 24. Andre Esteva, Brett Kuperl, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau, Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 2017, vol. 542, pp. 115–118.
 25. Круглов В. В., Борисов В. В. *Искусственные нейронные сети. Теория и практика*. М., Горячая линия–Телеком, 2001. 328 с.
 26. Geidarov P. Sh. Neural networks on the basis of the sample method. *Automatic Control and Computer Sciences*, 2009, vol. 43, pp. 203–210. <https://doi.org/10.3103/S0146411609040063>
 27. Гейдаров П. Ш. Алгоритм реализации метода ближайшего соседа в многослойном перцептроне. *Труды СПИИРАН*, 2017, т. 51, с. 123–151. <https://doi.org/10.15622/sp.51.6>
 28. Биргрет И. А. *Техническая диагностика*. М., Машиностроение, 1978. 240 с.

UDC 007.519.7

doi:10.31799/1684-8853-2020-3-25-38

Algorithm for calculating synapse weights of the first layer of a neural network on the base of metric recognition methods. Part 2P. Sh. Geidarov^a, PhD, Tech., Associate Professor, orcid.org/0000-0002-3881-0629, plbaku2010@gmail.com^aAzerbaijan National Academy of Sciences Institute of Control Systems, 9, Bahtijar Vagabzade St., Az 1141, Baku, Azerbaijan

Introduction: Metric recognition methods make it possible to preliminarily and strictly determine the structures of feed-forward neural networks, namely, the number of neurons, layers, and connections based on the initial parameters of the recognition problem. They also make it possible to analytically calculate the synapse weights of network neurons based on metric expressions. The setup procedure for these networks includes a sequential analytical calculation of the values of each synapse weight in the weight table for neurons of the zero or first layer, which allows us to obtain a working feed-forward neural network at the initial stage without the use of training algorithms. Then feed-forward neural networks can be trained by well-known learning algorithms, which generally speeds up the process of their creation and training. **Purpose:** To determine how much time the process of calculating the values of weights requires and, accordingly, how reasonable it is to preliminarily calculate the weights of a feed-forward neural network. **Results:** An algorithm is proposed and implemented for the automated calculation of all values of synapse weight tables for the zero and first layers as applied to the task of recognizing black-and-white monochrome symbol images. The proposed algorithm is described in the Builder C++ software environment. The possibility of optimizing the process of calculating the weights of synapses in order to accelerate the entire algorithm is considered. The time spent on calculating these weights for different configurations of neural networks based on metric recognition methods is estimated. Examples of creating and calculating synapse weight tables according to the considered algorithm are given. According to them, the analytical calculation of the weights of a neural network takes just seconds or minutes, being in no way comparable to the time necessary for training a neural network. **Practical relevance:** Analytical calculation of the weights of a neural network can significantly accelerate the process of creating and training a feed-forward neural network. Based on the proposed algorithm, we can implement one for calculating three-dimensional weight tables for more complex images, either black-and-white or color grayscale ones.

Keywords — neural networks, weight and threshold values, neurocomputer, learning algorithms, programming of neural networks.

For citation: Geidarov P. Sh. Algorithm for calculating synapse weights of the first layer of a neural network on the base of metric recognition methods. Part 2. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2020, no. 3, pp. 25–38 (In Russian). doi:10.31799/1684-8853-2020-3-25-38

References

1. Azarov I. S., Petrovskij A. A. Training personal voice model of a speaker with unified phonetic space of features using artificial neural network. *SPIIRAS Proceedings*, 2014, no. 5, pp. 128–15 (In Russian). <https://doi.org/10.15622/sp.36.8>
2. Budko R. Yu., Starchenko I. B. Creation of the facial gestures dlassifier based on the electromyogram analysis. *SPIIRAS Proceedings*, 2016, no. 3 (46), pp. 76–89 (In Russian). <https://doi.org/10.15622/sp.46.7>
3. Osipov V. Yu., Nikiiforov V. V. Recurrent neural networks with controlled elements in restoring frame flows. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2019, no. 5, pp. 10–17 (In Russian). <https://doi.org/10.31799/1684-8853-2019-5-10-17>
4. Le T. Ch. The comparison of neural network CMAC and multilayer neural network in the task of detection of DoS attacks. *Neurokomputers*, 2016, no. 7, pp. 65–69 (In Russian).
5. Katasev A. S., Kataseva D. V., Kirpichnikov A. P. Handwritten character recognition based on artificial neural network. *Vestnik tehnologicheskogo universiteta*, 2018, vol. 18, no. 11, pp. 173–176 (In Russian).
6. Drokin I. S. About an algorithm for consistent weights initialization of deep neural networks and neural networks ensemble learning. *Vestnik of Saint Petersburg University Applied Mathematics. Computer Science. Control Processes*, 2016, no. 4, pp. 66–74 (In Russian).
7. Leshhenko Ju. E. Associative-verbal networks and artificial neural networks as a method of mental lexicon modeling. *Global Scientific Potential*, 2014, no. 10 (43), pp. 154–156 (In Russian).
8. Gafurov D. O., Gafurov O. M., Kontorovich V. A. Interpretation of data from geophysical studies of the Talakan oil and gas condensate field by trained neural networks, forecast of the structure of the Osinsky horizon. *Seismic Technologies*, 2014, no. 4, pp. 85–92 (In Russian).
9. Vladimirova D. B., Koksharova A. A. Financial Markets Forecasting Using Artificial Neural Networks. *Science and Business: Ways of Development*, 2014, no. 3 (33), pp. 42–46 (In Russian).
10. Sinchuk O. N., Bojko S. N. Neural networks and process control of power supply of objects from combined electric networks. *Tehnichna elektrodinamika*, 2014, no. 5, pp. 53–55 (In Russian).
11. Shvedov V. A., Il'in E. S. Recognition of road signs by convolutional neural networks of various structures, using cuda GPUs. *Materialy Shestoy mezhdunarodnoy nauchno-prakticheskoy konferentsii "Transportnaya infrastruktura Sibirskogo regiona"* [Proc. 6th Int. Sump. "Transport infrastructure of the Siberian region"], 2015, vol. 2, pp. 282–286 (In Russian).
12. Turovskij Ya. A., Kurgalin S. D., Adamenko A. A. Comparative analysis of software packages for working with artificial neural networks. *Vestnik Voronezhskogo gosudarstvennogo universiteta. Seriya: Sistemnyj analiz i informacionnye tekhnologii*, 2016, no. 1, pp. 161–168 (In Russian).
13. Lukina A. S., Nekrasov M. V., Pakman D. N. Processing telemetric information of a spacecraft by neural networks based on the theory of Kalman filters. *Tendencii razvitiya nauki i obrazovaniya*, 2016, no. 13-1, pp. 43–45 (In Russian).
14. Stasjuk V. V. Prediction of the explosives impact sensitivity by neural networks with preliminary grouping data. *Fundamental Research*, 2015, no. 12-6, pp. 1139–1143 (In Russian).
15. Husainov A. T. Predictability assessment of a reservoir pressure maintenance system by neural networks in oil fields. *Academic Journal of West Siberia*, 2016, vol. 12, no. 3 (64), p. 48 (In Russian).
16. Bondarko V. M., Bondarko D. V., Solnushkin S. D., Chihman V. N. Modeling the results of psychophysical experiments by neural networks. *Neurokomputers*, 2018, no. 5, pp. 31–33 (In Russian).
17. Sedov V. A., Sedova N. A. Modelling collision avoidance actions in closest approach zones by means of neural networks. *Asia-Pacific Journal of Marine Science & Education*, 2014, vol. 4, no. 2, pp. 104–111.
18. Peng Shi, Fanbiao Li, Ligang Wu. Neural network-based passive filtering for delayed neutral-type semi-Markovian jump systems. *IEEE Transactions on Neural Networks and Learning Systems*, 2017, vol. 28, iss. 9, pp. 2101–2114.
19. Chenggang Yan, Hongtao Xie, Dongbao Yang, Jian Yin, Yongdong Zhang, Qionghai Dai. Supervised hash coding with deep neural network for environment perception of intelligent vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 2018, vol. 19, iss. 1, pp. 284–295.

20. Anusha Nagabandi, Gregory Kahn, Ronald S. Fearing, Sergey Levine. Neural network dynamics for model-based deep reinforcement learning with model-free fine-tuning. *2018 IEEE International Conference on Robotics and Automation (ICRA)*, Brisbane, Australia, 2018, pp. 156–162.
21. Wei He, Yuhao Chen, Zhao Yin. Adaptive neural network control of an uncertain robot with full-state constraints. *IEEE Transactions on Cybernetics*, 2016, vol. 46, iss. 3, pp. 620–629.
22. Song Han, Xingyu Liu, Huizi Mao, Jing Pu, Ardavan Pedram, Mark A. Horowitz, William J. Efficient inference engine on compressed deep neural network. *2016 ACM/IEEE 43rd Annual International Symposium on Computer Architecture (ISCA)*, Seoul, South Korea, 2016, pp. 304–309.
23. William Chan, Navdeep Jaitl, Quoc Le, Oriol Vinyals. Listen, attend and spell: A neural network for large vocabulary conversational speech recognition. *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, 2016, pp. 107–115.
24. Andre Esteva, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau, Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 2017, vol. 542, pp. 115–118.
25. Kruglov V. V., Borisov V. V. *Iskusstvennyye neyronnyye seti. Teoriya i praktika* [Artificial neural networks. Theory and practice]. Moscow, Goriachaia liniia–Telekom Publ., 2001. 328 p. (In Russian).
26. Geidarov P. Sh. Neural networks on the basis of the sample method. *Automatic Control and Computer Sciences*, 2009, vol. 43, pp. 203–210. <https://doi.org/10.3103/S0146411609040063>
27. Geidarov P. Sh. An algorithm implementing the method of the nearest neighbor in a multi-layer perceptron. *SPIIRAS Proceedings*, 2017, vol. 51, pp. 123–151 (In Russian).
28. Birger I. A. *Tekhnicheskaya diagnostika* [Technical diagnostics]. Moscow, Mashinostroenie Publ., 1978. 240 p. (In Russian).

Уважаемые авторы!

При подготовке рукописей статей необходимо руководствоваться следующими рекомендациями.

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

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

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

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

Формулы набирайте в Word, не используя формульный редактор (Mathtype или Equation), при необходимости можно использовать формульный редактор; для набора одной формулы не используйте два редактора; при наборе формул в формульном редакторе знаки препинания, ограничивающие формулу, набирайте вместе с формулой; для установки размера шрифта никогда не пользуйтесь вкладкой Other..., используйте заводские установки редактора, не подгоняйте размер символов в формулах под размер шрифта в тексте статьи, не растягивайте и не сжимайте мышью формулы, вставленные в текст; в формулах не отделяйте пробелами знаки: + = -.

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

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

Иллюстрации предоставляются отдельными исходными файлами, поддающимися редактированию:

— рисунки, графики, диаграммы, блок-схемы предоставляйте в виде отдельных исходных файлов, поддающихся редактированию, используя векторные программы: Visio (*.vsd, *.vsdx); Coreldraw (*.cdr); Excel (*.xls); Word (*.docx); Adobe Illustrator (*.ai); AutoCad (*.dxf); Matlab (*.ps, *.pdf или экспорт в формат *.ai);

— если редактор, в котором Вы изготавливаете рисунок, не позволяет сохранить в векторном формате, используйте функцию экспорта (только по отношению к исходному рисунку), например, в формат *.ai, *.esp, *.wmf, *.emf, *.svg;

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

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

В редакцию предоставляются:

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

— экспертное заключение.

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

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

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

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

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

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

Более подробно правила подготовки текста с образцами изложены на нашем сайте в разделе «Правила для авторов».

Контакты

Куда: 190000, Санкт-Петербург,

Б. Морская ул., д. 67, ГУАП, РИЦ

Кому: Редакция журнала «Информационно-управляющие системы»

Тел.: (812) 494-70-02

Эл. почта: ius.spb@gmail.com

Сайт: www.i-us.ru

Взвешенный учет априорной и опытной информации в задаче оценивания эффективности функционирования системы управления при распределении числа испытаний по закону Паскаля

В. Н. Арсеньев^а, доктор техн. наук, профессор, orcid.org/0000-0001-6424-5320, vladar56@mail.ru

А. Д. Хомоненко^а, доктор техн. наук, профессор, orcid.org/0000-0003-0411-258X

А. А. Ядренкин^а, канд. техн. наук, доцент, orcid.org/0000-0003-2089-7460

^аВоенно-космическая академия им. А. Ф. Можайского, Ждановская наб., 13, Санкт-Петербург, 197198, РФ

Постановка проблемы: для получения устойчивых оценок показателей эффективности функционирования систем управления перспективных объектов необходимо провести испытания большого числа опытных образцов, что не всегда удается на практике. Повысить качество оценивания показателей эффективности системы можно путем совместной обработки информации, имеющейся до проведения натурных испытаний и представленной априорными оценками соответствующих показателей, и опытной информации, полученной по результатам испытаний опытных образцов. Для определения апостериорных оценок показателей эффективности, учитывающих результаты априорных и опытных исследований, необходимо найти их функциональную зависимость от априорной и опытной оценок показателей эффективности и определить ее параметры. **Цель:** комплексная обработка результатов априорных и опытных исследований системы управления и апостериорное оценивание показателей эффективности. **Результаты:** предложен метод оценивания эффективности функционирования системы управления путем комплексирования априорных и опытных оценок показателей эффективности, найденных до проведения испытаний системы и в процессе испытаний ограниченного числа опытных образцов. Показана возможность его применения, когда результаты априорных и опытных исследований представлены точечными оценками показателей эффективности и использование известных методов затруднено. Получены аналитические выражения для апостериорной оценки вероятности выполнения системой управления задачи функционирования и показателей, используемых для исследования влияния априорной информации на точность этой оценки и число испытаний. Работоспособность метода продемонстрирована на актуальном примере. В отличие от известных подходов, он учитывает близость априорных оценок к опытным. **Практическая значимость:** предложенный подход к оцениванию эффективности функционирования системы управления является достаточно общим, поскольку позволяет комплексировать информацию, полученную на различных этапах исследования системы, существенно повысить точность оценивания показателей эффективности и определить выигрыш в числе испытаний во всех случаях, когда результаты априорных исследований не противоречат опытным данным.

Ключевые слова — система управления, эффективность функционирования, априорная информация, опытные данные, выигрыш в точности, выигрыш в числе испытаний.

Для цитирования: Арсеньев В. Н., Хомоненко А. Д., Ядренкин А. А. Взвешенный учет априорной и опытной информации в задаче оценивания эффективности функционирования системы управления при распределении числа испытаний по закону Паскаля. *Информационно-управляющие системы*, 2020, № 3, с. 39–47. doi:10.31799/1684-8853-2020-3-39-47

For citation: Arseniev V. N., Khomonenko A. D., Yadrenkin A. A. Weighed ranking of aprioristic and experimental data in control system functioning efficiency estimation problem with Pascal-distributed number of tests. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2020, no. 3, pp. 39–47 (In Russian). doi:10.31799/1684-8853-2020-3-39-47

Введение

Для получения устойчивых оценок показателей эффективности функционирования систем управления (СУ) перспективных объектов необходимо иметь результаты испытаний достаточно большого числа опытных образцов. На практике, в силу известных причин, обеспечить требуемое число испытаний далеко не всегда представляется возможным. Одним из путей повышения качества оценивания показателей эффективности системы и уменьшения числа натурных испытаний является использование априорной информации, накопленной до проведения испытаний [1].

Всю имеющуюся информацию о качестве функционирования СУ условно можно разделить

на априорную и опытную. На объем априорной информации практически нет ограничений, но достоверность ее существенно зависит от адекватности используемых моделей. Опытные данные, как правило, отражают реальные характеристики СУ, но ограничены по объему. Повысить достоверность оценок показателей качества функционирования системы в такой достаточно часто встречающейся на практике ситуации можно путем комплексной обработки всей имеющейся априорной и опытной информации.

Данная задача не является новой, и имеется ряд методов, позволяющих при определенных условиях найти ее решение. Часть из них базируется на байесовском подходе к оцениванию, требующем наиболее полной априорной информа-

ции [2–9]. Основным ограничением этого подхода является неоднозначность, которую он допускает при выборе априорных распределений, и, как следствие, противоречивость (в ряде случаев) получаемых результатов. Кроме того, при его использовании не учитывается близость априорной информации к результатам испытаний. В методе линейного объединения априорных и опытных оценок для определения параметров модели необходимо, чтобы обе оценки являлись несмещенными и были известны их дисперсии (при многомерных показателях качества — корреляционные матрицы) [3, 5]. Эти условия не всегда выполняются. Сложность применения метода коррелированных процессов связана с необходимостью определять взаимную корреляционную матрицу выходных сигналов модели и системы [5, 10]. В методе параметрических функций используются аддитивные и мультипликативные модели объединения априорных и опытных оценок [5, 11, 12]. Поскольку параметры моделей в общем случае нелинейно зависят от моментов априорной оценки, то применение их на практике часто затруднено. Решения задачи апостериорного оценивания при отсутствии априорного распределения в некоторых частных случаях приведены в работах [13, 14].

Следует отметить, что качество априорной информации может значительно влиять не только на достоверность получаемых апостериорных оценок характеристик СУ, но и на число испытаний. Это становится особенно актуальным на заключительных этапах создания перспективных объектов, когда проводятся дорогостоящие испытания их опытных образцов.

Хорошим актуальным аналогом рассматриваемой задачи может служить тестирование на основе конечных автоматов (детерминированных и недетерминированных), используемое в интерактивных дискретных системах при построении тестовых последовательностей.

В частности, для примера здесь можно назвать синтез проверяющих тестов для телекоммуникационных протоколов [15]. В статье [16] рассматривается проблема синтеза проверяющих последовательностей для недетерминированных спецификаций, возникающих в компьютерных приложениях, а в работе [17] — вопросы построения адаптивной проверяющей и адаптивной различающей последовательностей. В статье [18] показано, что класс автоматов без слияний содержит множество используемых в различных приложениях детерминированных автоматов.

Еще одними близкими и актуальными аналогами решаемой нами задачи, на наш взгляд, служат задачи: тестирования надежности для критических информационных систем на осно-

ве оценивания случайных состояний [19]; тестирования программного обеспечения на этапе его разработки с помощью моделирования [20]; математического моделирования тестирования надежности программного обеспечения с несовершенной отладкой [21]. При этом в работах [15–21] не рассматриваются детально задачи выбора параметров, определяющих веса априорной и опытной информации при оценивании искомых показателей.

Постановка задачи апостериорного оценивания показателей эффективности системы

Полагается, что испытание опытного образца прошло успешно, если была достигнута цель функционирования СУ. Вероятность того, что СУ решит поставленные задачи при одном испытании, остается постоянной и равна p .

Если случайная величина $\hat{X} \geq 0$ представляет число независимых испытаний, которые необходимо провести для того, чтобы ровно m из них были успешными, то закон распределения ее определяется вероятностями того, что число испытаний будет равно x :

$$P(\hat{X} = x; p) = C_{x-1}^{m-1} p^m (1-p)^{x-m}, \\ x = m, m+1, m+2, \dots, \quad (1)$$

где C_{x-1}^{m-1} — число сочетаний из $x-1$ по $m-1$.

Данная формула описывает распределение Паскаля (отрицательное биномиальное распределение) [22]. Известны соотношения, связывающие математическое ожидание $M_{\hat{X}}$ и дисперсию $D_{\hat{X}}$ необходимого числа испытаний \hat{X} с вероятностью p :

$$M_{\hat{X}} = m/p; D_{\hat{X}} = m(1-p)/p^2. \quad (2)$$

Для получения опытной оценки вероятности проведены испытания опытных образцов, включающие N_0 серий. Испытания в каждой серии проводились до получения m успешных результатов. Число испытаний в 1-й серии составило x_1 , во 2-й — x_2 и т. д., в N_0 -й серии — x_{N_0} . Полагается, что результаты испытаний известными методами приведены к единым условиям и представляют выборку из одной генеральной совокупности.

На этапах, предшествующих испытаниям опытных образцов, получена априорная (расчетная) оценка p_p вероятности того, что цель будет достигнута при одном испытании. Она может быть пересчитана при изменении условий испытаний.

Необходимо найти апостериорную оценку p_a этой вероятности, учитывая априорную информацию и результаты испытаний.

Сформулированная задача, по сути дела, сводится к обоснованию структуры и выбору параметров функциональной зависимости апостериорной оценки p_a от априорной p_p и опытной p_o оценок вероятности p . Для ее решения предлагается использовать метод приоритета опытной информации [23, 24].

Получение апостериорной оценки вероятности достижения цели

В методе приоритета опытной информации сначала по результатам испытаний определяется оценка максимального правдоподобия (опытная оценка) p_o вероятности p [2].

Для этого по выборке $x_i, i = 1, N_o$ составляется функция правдоподобия

$$\prod_{i=1}^{N_o} P(\hat{X} = x_i; p) = \prod_{i=1}^{N_o} C_{x_i-1}^{m-1} p^m (1-p)^{x_i-m}. \quad (3)$$

Опытная оценка p_o должна обеспечивать максимум этой функции. Для определения оценки p_o в данном случае можно использовать необходимое условие максимума функции правдоподобия или логарифма этой функции, поскольку закон распределения Паскаля удовлетворяет условию регулярности [22]. Необходимое условие максимума имеет вид

$$\left. \frac{\partial \sum_{i=1}^{N_o} \ln [C_{x_i-1}^{m-1} p^m (1-p)^{x_i-m}]}{\partial p} \right|_{p=p_o} = 0$$

или

$$mN_o/p_o + mN_o/(1-p_o) - \sum_{i=1}^{N_o} x_i / (1-p_o) = 0.$$

Решением последнего уравнения является опытная оценка вероятности того, что СУ достигнет цели при одном испытании:

$$p_o = mN_o / \sum_{i=1}^{N_o} x_i. \quad (4)$$

Функцию правдоподобия (3) можно преобразовать к виду

$$\begin{aligned} & \prod_{i=1}^{N_o} C_{x_i-1}^{m-1} p^m (1-p)^{x_i-m} = \\ & = \left(\prod_{i=1}^{N_o} C_{x_i-1}^{m-1} \right) p^{mN_o} (1-p)^{\sum_{i=1}^{N_o} x_i - mN_o}. \end{aligned}$$

Подстановка в последнюю формулу суммы $\sum_{i=1}^{N_o} x_i = mN_o/p_o$, полученной из уравнения (4), дает новое выражение для функции (3)

$$L(p_o, p) = C_o p^{mN_o} (1-p)^{mN_o(1/p_o-1)}, \quad (5)$$

где постоянная величина $C_o = \prod_{i=1}^{N_o} C_{x_i-1}^{m-1}$.

Рассматривается статистическая гипотеза $H: p = p_p$ о том, что вероятность достижения цели при одном испытании равна ее априорной оценке. Для проверки этой гипотезы используется величина отношения правдоподобия, которая, фактически, характеризует близость априорной информации к опытным данным. Она вычисляется по формуле

$$\begin{aligned} v^* &= \frac{L(p_o, p)|_{p=p_p}}{L(p_o, p)|_{p=p_o}} = \frac{L(p_o, p_p)}{L(p_o, p_o)} = \\ &= \frac{C_o p_p^{mN_o} (1-p_p)^{mN_o(1/p_o-1)}}{C_o p_o^{mN_o} (1-p_o)^{mN_o(1/p_o-1)}} = \\ &= \left(\frac{p_p}{p_o} \right)^{mN_o} \left(\frac{1-p_p}{1-p_o} \right)^{mN_o(1/p_o-1)}. \quad (6) \end{aligned}$$

Отсюда видно, что $0 \leq v^* \leq 1$, причем максимальное значение $v^* = 1$ достигается при совпадении априорной и опытной оценок, т. е. при $p_p = p_o$. В этом случае их доли в апостериорной оценке полагаются равными. При значительном отличии оценок p_p и p_o величина отношения правдоподобия (6) будет мала ($v^* \approx 0$) и вес априорной информации в апостериорной оценке будет близок к нулю.

Качество опытной оценки p_o зависит от числа испытаний N_o . Для определения качества априорной оценки вводится величина

$$N_p = v^* N_o. \quad (7)$$

Тогда при справедливости гипотезы $H: p = p_p$ об однородности априорной и опытной информации функция

$$L(p_p, p) = C_p p^{mN_p} (1-p)^{mN_p(1/p_p-1)}, \quad (8)$$

где $C_p = \text{const}$, может рассматриваться как функция, характеризующая распределение априорной оценки p_p , а произведение функций (5) и (8)

$$\begin{aligned} L(p) &= L(p_p, p) L(p_o, p) = \\ &= C_p C_o p^{m(N_p+N_o)} (1-p)^{m[N_p(1/p_p-1)+N_o(1/p_o-1)]} \quad (9) \end{aligned}$$

— как функция, характеризующая распределение обеих оценок: p_o и p_p .

Для определения апостериорной оценки p_a вероятности p используется необходимое условие максимума функции (9) $\partial \ln L(p) / \partial p|_{p=p_a} = 0$, которое дает уравнение

$$m(N_o + N_p) / p_a - m[N_p(1/p_p - 1) + N_o(1/p_o - 1)] / (1 - p_a) = 0.$$

Решением полученного уравнения является апостериорная оценка вероятности p того, что цель будет достигнута при одном испытании СУ:

$$p_a = \frac{p_o p_p (N_o + N_p)}{N_o p_p + N_p p_o} = \frac{p_o p_p (1 + v^*)}{p_p + v^* p_o} = \frac{p_o p_p}{\frac{1}{1 + v^*} p_p + \frac{v^*}{1 + v^*} p_o}. \quad (10)$$

Отсюда видно, что апостериорная оценка является нелинейной функцией от взвешенных значений априорной и опытной оценок вероятности p . Из формулы (10) также следует, что процедура проверки гипотезы H об однородности всей имеющейся информации не является обязательной. Действительно, при $v^* \approx 1$, т. е. при близости априорной и опытной оценок ($p_p \approx p_o$), апостериорная оценка $p_a \approx p_o \approx p_p$. При $v^* \approx 0$, т. е. при существенном отклонении результатов априорных исследований от данных, полученных в процессе испытаний, апостериорная оценка близка к опытной ($p_a \approx p_o$) и априорная информация фактически игнорируется. При $0 \leq v^* \leq 1$ опытная и априорная оценки входят в апостериорную оценку с разными весами, причем вес опытной информации всегда превышает вес априорной. Например, если в формуле (6) положить $mN_o = 20$, то при $p_p = 0,5$ и $p_o = 0,9$ отношение правдоподобия $v^* = 0,0003$ и априорная информация, в силу ее существенного отличия от опытных данных, очень слабо влияет на апостериорную оценку: $p_a = 0,8998$. При $p_p = 0,8$ и $p_o = 0,9$ отношение правдоподобия $v^* = 0,4425$ и влияние априорной информации становится существенным. Это видно и по апостериорной оценке $p_a \approx 0,86$. При дальнейшем возрастании отношения правдоподобия (при приближении априорной информации к опытным данным, содержащим достоверную информацию о характеристиках СУ) вес априорной информации в апостериорной оценке также возрастает. Так, при $p_p = 0,85$ и $p_o = 0,9$ отношение правдоподобия $v^* = 0,7850$ и $p_a \approx 0,88$, а при $p_p = 0,9$ и $p_o = 0,9$ отношение правдоподобия $v^* = 1$ и $p_a = p_o = p_p = 0,9$.

Если априорная информация отсутствует, то апостериорная оценка вероятности, как видно из формулы (10), будет совпадать с оценкой максимального правдоподобия p_o , полученной по результатам натуральных испытаний опытных образцов системы.

Определение выигрыша в качестве оценивания

Из формулы (7) следует, что априорная оценка p_p может рассматриваться как оценка, полученная по некоторой гипотетической выборке, включающей результаты $\tilde{N}_p = E[N_p] = E[v^* N_o]$ серий испытаний, где $E[\cdot]$ — функция округления до ближайшего целого числа. Тогда можно полагать, что апостериорная оценка p_a получена по результатам $N_a = N_o + \tilde{N}_p$ серий испытаний, и рассматривать величину

$$\delta_{ч.и} = E[v^* N_o], \quad (11)$$

как число дополнительных серий испытаний, которое дает учет априорной информации в процедуре оценивания вероятности p .

Из последней формулы видно, что $\delta_{ч.и}$ зависит от близости априорной оценки к опытной, т. е. от величины отношения правдоподобия, определяемой по формуле (6). Чем ближе априорная информация к опытной, тем больше выигрыш в числе испытаний.

В качестве показателя, характеризующего выигрыш в точности оценивания, используется относительная величина

$$\delta_T = \frac{D_{p_o} - D_{p_a}}{D_{p_o}} 100\%, \quad (12)$$

где D_{p_o} и D_{p_a} — дисперсии оценок p_o и p_a соответственно.

Приближенное значение дисперсии D_{p_o} опытной оценки p_o вероятности того, что цель будет достигнута при одном испытании СУ, можно получить, воспользовавшись известным распределением оценки p_o , которое согласно [25] имеет вид

$$P(\hat{p}_o = p) = C_{mN_o/p_o+1}^{mN_o+1} p^{N_o m+2} (1-p)^{mN_o/p_o - mN_o}, \\ p_o = 1, \frac{mN_o}{mN_o+1}, \frac{mN_o}{mN_o+2}, \dots \quad (13)$$

Зная закон распределения, можно найти моменты опытной оценки p_o . Поскольку $p_o = mN_o / \sum_{i=1}^{N_o} x_i$,

то, обозначив $s_o = \sum_{i=1}^{N_o} x_i = mN_o / p_o$, найдем выражение для начального момента k -го порядка оценки p_o

$$M[\hat{p}_o^k] = \sum_{p_o} p_o^k P(\hat{p}_o = p) =$$

$$= \sum_{p_o} p_o^k C_{mN_o/p_o+1}^{mN_o+1} p_o^{N_o m+2} (1-p_o)^{mN_o/p_o - mN_o} =$$

$$= \sum_{s_o} (N_o m/s_o)^k C_{s_o+1}^{N_o m+1} p_o^{N_o m+2} (1-p_o)^{s_o - N_o m},$$

$$s_o = N_o m, N_o m + 1, N_o m + 2, \dots$$

Для приближенного вычисления начального момента k -го порядка оценки p_o может использоваться формула

$$M[\hat{p}_o^k] \approx \sum_{s_o} (N_o m/s_o)^k C_{s_o+1}^{N_o m+1} p_o^{N_o m+2} (1-p_o)^{s_o - N_o m}.$$

Тогда приближенное значение дисперсии опытной оценки

$$D_{p_o} \approx M[\hat{p}_o^2] - (M[\hat{p}_o])^2. \quad (14)$$

В силу сделанных предположений, как уже отмечалось выше, апостериорную оценку p_a вероятности p можно рассмотреть как оценку, полученную по результатам $N_a = N_o + \tilde{N}_p$ серий испытаний. В этом случае формула для определения приближенного значения дисперсии D_{p_a} апостериорной оценки p_a аналогична формуле (14), т. е.

$$D_{p_a} \approx M[\hat{p}_a^2] - (M[\hat{p}_a])^2, \quad (15)$$

где

$$M[\hat{p}_a^k] \approx \sum_{s_a} \left(\frac{N_a m}{s_a}\right)^k C_{s_a+1}^{N_a m+1} p_a^{N_a m+2} (1-p_a)^{s_a - N_a m},$$

$$s_a = N_a m, N_a m + 1, N_a m + 2, \dots; k = 1, 2.$$

Подстановка значений D_{p_o} и D_{p_a} , полученных по формулам (14), (15), в правую часть выражения (12) позволяет сравнить приближенные значения дисперсий апостериорной оценки и опытной оценки, полученной только по результатам натуральных испытаний опытных образцов, и исследовать влияние априорной информации на точность оценивания.

Оценки математического ожидания $M_{\hat{X}}$ и дисперсии $D_{\hat{X}}$ случайной величины \hat{X} , учитывающие результаты априорных исследований и натуральных испытаний, определяются в соответствии с формулами (2) путем подстановки в них вместо вероятности p ее апостериорной оценки p_a .

Пример применения метода апостериорного оценивания

Одной из основных подсистем автоматизированной СУ полетами самолетов крупного аэро-

дрома является система дальнего обнаружения и управления. В ее состав входит радиолокатор дальнего обнаружения, предназначенный для обнаружения воздушных объектов. Вероятность получения отметки об объекте на экране радиолокатора при одном обороте антенны равна p .

Объект считается обнаруженным, если от него получено две отметки ($m = 2$). По результатам моделирования процесса обнаружения воздушных объектов получена расчетная оценка $p_p = 0,6$ вероятности p .

Были проведены четыре испытания радиолокатора ($N_o = 4$) в одинаковых условиях. При первом испытании для обнаружения воздушного объекта потребовалось четыре оборота антенны ($x_1 = 4$), при втором — шесть оборотов ($x_2 = 6$), при третьем — пять ($x_3 = 5$), при четвертом — четыре ($x_4 = 4$).

Требуется определить:

1) апостериорную вероятность того, что для обнаружения объекта придется сделать не более пяти оборотов антенны;

2) минимальное число оборотов антенны, необходимое для обнаружения воздушного объекта с вероятностью не менее 0,8.

Решение первой задачи

Вероятность того, что для обнаружения объекта придется сделать не более пяти оборотов антенны, можно найти по формуле

$$P(2 \leq \hat{X} \leq 5; p) = \sum_{i=2}^5 C_5^i p^i (1-p)^{5-i} =$$

$$= 1 - \sum_{i=0}^1 C_5^i p^i (1-p)^{5-i}.$$

На ее основе определяется априорное значение вероятности

$$P_p = P(2 \leq \hat{X} \leq 5; p_p) = \sum_{i=2}^5 C_5^i p_p^i (1-p_p)^{5-i} =$$

$$= \sum_{i=2}^5 C_5^i \cdot 0,6^i \cdot 0,4^{5-i} \approx 0,91.$$

По данным натуральных испытаний радиолокатора рассчитывается оценка максимального правдоподобия вероятности получения отметки об объекте на экране радиолокатора при одном обороте антенны [см. формулу (4)]

$$p_o = \frac{2 \cdot 4}{4 + 6 + 5 + 4} = \frac{8}{19} \approx 0,42.$$

Соответствующая ей опытная оценка вероятности того, что для обнаружения объекта придется сделать не более пяти оборотов антенны, имеет вид

$$P_o = P(2 \leq \hat{X} \leq 5; p_o) = \sum_{i=2}^5 C_5^i p_o^i (1-p_o)^{5-i} = \\ = \sum_{i=2}^5 C_5^i \cdot 0,42^i \cdot 0,58^{5-i} \approx 0,70.$$

Отношение правдоподобия, характеризующее близость результатов моделирования к данным натурных испытаний, рассчитывается по формуле (6):

$$v^* \approx 0,29.$$

Тогда, в соответствии с формулой (10), апостериорная оценка вероятности получения отметки об объекте на экране радиолокатора при одном обороте антенны $p_a \approx 0,45$, а выигрыш в числе испытаний согласно формуле (11) $\delta_{ч.и} = 1$, т. е. апостериорную оценку p_a вероятности p можно рассматривать как оценку, полученную по пяти испытаниям радиолокатора.

По формулам (14) и (15) получены приближенные значения дисперсий опытной и апостериорной оценок вероятности p : $D_{p_o} \approx 0,0113$ и $D_{p_a} \approx 0,0101$. Их подстановка в формулу (12) показала, что учет априорной информации позволил повысить точность оценивания примерно на 10 %.

Апостериорная оценка вероятности того, что для обнаружения объекта придется сделать не более пяти оборотов антенны:

$$P_a = P(2 \leq \hat{X} \leq 5; p_a) = \sum_{i=2}^5 C_5^i p_a^i (1-p_a)^{5-i} = \\ = \sum_{i=2}^5 C_5^i \cdot 0,45^i \cdot 0,55^{5-i} \approx 0,75.$$

Отсюда видно, что $P_o < P_a < P_p$, причем оценка P_a ближе к опытной оценке P_o , чем к априорной P_p . Это связано с относительно небольшим значением отношения правдоподобия v^* , определяющего согласно формуле (10) вес априорной информации (p_p) в апостериорной оценке p_a вероятности p .

Решение второй задачи

Минимальное число оборотов антенны $X_{0,8}$, необходимое для обнаружения объекта с вероятностью не менее 0,8, зависит от вероятности p , т. е. $X_{0,8} = X_{0,8}(p)$, и может быть найдено из выражения $P(2 \leq \hat{X} \leq X_{0,8}; p) \geq 0,8$, которое эквива-

$$\text{лентно условию } \sum_{i=2}^{X_{0,8}} C_{X_{0,8}}^i p^i (1-p)^{X_{0,8}-i} \geq 0,8.$$

Минимальные значения числа оборотов антенны, обеспечивающие удовлетворение этого

условия при оценках p_p , p_o и p_a вероятности p , равны, соответственно: $X_{0,8}(p_p) = 4$, $X_{0,8}(p_o) = 7$ и $X_{0,8}(p_a) = 6$.

Таким образом, при ограниченных испытаниях радиолокатора учет априорной информации позволил скорректировать выводы о качестве его функционирования и уточнить требования к скорости вращения антенны.

Заключение

Предложенный подход к оцениванию вероятности того, что СУ решит поставленные задачи при однократном ее применении, может быть использован для решения ряда других практических задач. Это обусловлено широким распространением закона распределения Паскаля. Так, объем выборки, необходимой для получения наперед заданного числа объектов с определенными свойствами при их случайном извлечении из генеральной совокупности, подчинен этому закону. Он используется в задачах оптимального резервирования элементов, в теории стрельбы, при решении задач, связанных с исследованием долговечности системы, имеющей определенное число резервных элементов, при определении количества особей данного вида в выборках из биологических популяций, в статистике несчастных случаев и заболеваний, а также в других приложениях. Полученные выше результаты, на наш взгляд, могут быть использованы при оценивании состояний сложных объектов с помощью инвариантов [26], при планировании испытаний и оценивании характеристик сложных систем [27], при оценивании состояний информационных систем на основе использования функции энтропии [28].

Финансовая поддержка

Работа частично поддержана грантом МОН РК: проект № AP05133699 «Исследование и разработка инновационных информационно-телекоммуникационных технологий с использованием современных кибертехнических средств для интеллектуальной транспортной системы города».

Financial support

The work was partially supported by the grant of the MES RK: project No. AP05133699 «Research and development of innovative information and telecommunication technologies using modern cybertechnical means for the city's intelligent transport system».

Литература

1. Буренок В. М., Найденов В. М. Испытательная база: выход из кризиса. *Воздушно-космическая оборона*, 2009, № 1 (44), с. 1–25.
2. Пугачев В. С. *Теория вероятностей и математическая статистика*. М., Физматлит, 2002. 496 с.
3. Городецкий В. И., Дмитриев А. К., Марков В. М., Петухов Г. Б., Юсупов Р. М. *Элементы теории испытаний и контроля технических систем* / под ред. Р. М. Юсупова. Л., Энергия, 1978. 192 с.
4. Фроленков К. В. Уточнение оценок вероятностей при локальном апостериорном выводе алгебраической байесовской сети в случае неточного свидетельства. *Труды СПИИРАН*, 2013, № 1(24), с. 152–164.
5. Рожков В. А. *Теория и методы статистического оценивания вероятностных характеристик случайных величин и функций с гидрометеорологическими примерами*. Кн. 1. СПб., Гидрометеиздат, 2001. 340 с.
6. Сайпулаева Г. А., Дандамаев А. У. Разработка методики априорной и апостериорной оценки трудозатрат в системе технического обслуживания и ремонта. *Системные технологии*, 2016, № 3(20), с. 43–52.
7. Храменков А. С., Ярмолик С. Н. Алгоритм последовательного распознавания радиолокационных объектов, обеспечивающий минимизацию среднего риска принимаемых решений. *Доклады Белорусского государственного университета*, 2016, № 2(96), с. 37–43.
8. Тулупьев А. Л. Апостериорные оценки вероятностей в алгебраических байесовских сетях. *Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления*, 2012, № 2, с. 51–59.
9. Бондаренко В. А., Ярица А. И. Сравнительный анализ априорной и апостериорной оценок точности плановой геодезической сети с помощью программного комплекса «Россия–Беларусь», разработанного в Полоцком государственном университете. *Вестник Полоцкого государственного университета*. Серия F: Строительство. Прикладные науки, 2014, № 16, с. 92–95.
10. Пугачев В. Н. *Комбинированные методы определения вероятностных характеристик*. М., Сов. радио, 1973. 256 с.
11. Буряк Ю. И., Скрынников А. А. Повышение степени обоснованности принимаемых решений в системе распознавания за счет использования априорной информации. *Научный вестник Московского государственного технического университета гражданской авиации*, 2015, № 220 (10), с. 47–54.
12. Мадера А. Г. Метод определения вероятностей прогнозируемых событий при принятии решений. *Искусственный интеллект и принятие решений*, 2016, № 2, с. 38–45.
13. Ардашов А. А., Арсеньев В. Н., Силантьев С. Б. Метод обработки неоднородной статистической информации о характеристиках точности системы управления. *Информационно-управляющие системы*, 2015, № 5, с. 55–59. doi:10.15217/issn1684-8853.2015.5.55
14. Арсеньев В. Н., Силантьев С. Б., Ядренкин А. А. Использование априорной информации для коррекции модели потока событий в сложной системе. *Изв. вузов. Приборостроение*, 2017, т. 60, № 5, с. 391–397.
15. Chow T. S. Testing software design modelled by finite state machines. *IEEE Trans. Software Eng.*, 1978, vol. 4, no. 3, pp. 178–187.
16. Petrenko A., Simão A., Yevtushenko N. Generating checking sequences for nondeterministic finite state machines. *Proceedings of the ICST*, 2012, pp. 310–319.
17. Kushik N., El-Fakih K., Yevtushenko N. Adaptive homing and distinguishing experiments for nondeterministic finite state machines. *Lecture Notes in Computer Science*, 2013, vol. 8254, pp. 33–48.
18. Günçen C., Inan K., Türker U.C., Yenigün H. The relation between preset distinguishing sequences and synchronizing sequences. *Formal Aspects of Computing*, 2014, vol. 26, no. 6, pp. 1153–1167.
19. Jie M., Honlin Z., Wenbo X., Jin L. Reliability testing methods for critical information system based on state random. *2011 International Conference on Information Communication and Management, IPC-SIT*, 2011, vol. 1, IACSIT Press, Singapore, pp. 28–32.
20. Xin W., Feng-Yan H., Zheng Q. Software reliability testing data generation approach based on a mixture model. *Information Technology Journal*, 2010, vol. 9, no. 5, pp. 1038–1043.
21. Cai K.-Y., Caob. P., Dongc. Z., Liu. K. Mathematical modeling of software reliability testing with imperfect debugging. *Computers and Mathematics with Applications*, 2010, vol. 59, no. 10, pp. 3245–3285.
22. Королюк В. С., Портенко Н. И., Скороход А. В., Турбин А. Ф. *Справочник по теории вероятностей и математической статистике*. М., Наука. 1985. 640 с.
23. Арсеньев В. Н., Лабетский П. В. Метод апостериорного оценивания характеристик системы управления летательного аппарата. *Изв. вузов. Приборостроение*, 2014, т. 57, № 10, с. 23–28.
24. Arseniev V. N., Adadurov S. E., Gerasimenko P. V., Degtyarev V. G. Correction of models of disturbing perturbances at research of complex system properties. *2017 Proceedings of 2017 20th IEEE International Conference on Soft Computing and Measurements, SCM*, 2017, pp. 184–188.
25. Арсеньев В. Н. *Новые методы принятия решений при ограниченных экспериментальных данных*. СПб., ВКА им. А. Ф. Можайского, 1999. 90 с.
26. Хомоненко А. Д., Яковлев Е. Л. К оценке состояния сложных объектов с помощью инвариантов.

Материалы 3-й Российской конференции «Математическое моделирование и информационные технологии», Екатеринбург, 16 ноября 2016 г., CEUR Workshop Proceedings, т. 1825, с. 44–49.

27. Арсеньев В. Н., Силантьев С. Б., Хомоненко А. Д., Адагуров С. Е. Определение вероятности выполнения задачи сложной системой при ограниченном объеме опытной информации. *Международная конференция по мягким вычислениям и измерениям*, 2018, т. 1, с. 43–46.

28. Смагин В. А., Хомоненко А. Д. Использование функции энтропии для определения состояний информационной системы. *Семинар «Компьютерная наука и инженерия» в рамках 5-й Международной научно-методической конференции «Проблемы математической и естественно-научной подготовки в инженерном образовании»*, Санкт-Петербург, 8–9 ноября 2018 г., CEUR Workshop Proceedings, т. 234, с. 53–60.

UDC 519.2

doi:10.31799/1684-8853-2020-3-39-47

Weighed ranking of aprioristic and experimental data in control system functioning efficiency estimation problem with Pascal-distributed number of tests

V. N. Arseniev^a, Dr. Sc., Tech., Professor, orcid.org/0000-0001-6424-5320, vladar56@mail.ru

A. D. Khomonenko^a, Dr. Sc., Tech., Professor, orcid.org/0000-0003-0411-258X

A. A. Yadrenkin^a, PhD, Tech., Associate Professor, orcid.org/0000-0003-2089-7460

^aA. F. Mozhaiskii Military Space Academy, 13, Zhdanovskaia Emb., 197198, Saint-Petersburg, Russian Federation

Introduction: In order to steadily estimate the efficiency of control systems for new objects, a great number of prototypes should be tested, which is not always possible in practice. The estimation quality can be improved by joint processing of the a priori information you have before the tests by analyzing certain indicators, and the data obtained from the tests. To estimate the efficiency a posteriori, taking into account both the a priori knowledge and the test results, you have to find their functional dependence on each of them, and specify the parameters of this dependence. **Purpose:** Integrated processing of the results from both aprioristic and experimental research of a control system, and obtaining posterior estimations of the efficiency indices. **Results:** A control system efficiency estimation method is proposed, which integrates the aprioristic and experimental estimations of the efficiency indices obtained a priori and during a limited number of tests of system prototypes. It can be used when the results of aprioristic research and the tests are presented by point estimations of the efficiency indices, and the most common methods are difficult to apply. We present analytical expressions for posterior estimation of the probability that the system will perform its task, along with the indicators which are used to study the influence of the aprioristic information on the estimation accuracy and number of tests. The working capacity of the method is illustrated by a real-life example. This approach, unlike others, takes into account how close the aprioristic estimations are to the experimental ones. **Practical relevance:** The proposed approach is universal enough, as it allows you to integrate the information obtained at various stages of studying the system, and essentially improve the efficiency estimation accuracy, specifying the gain in the number of tests in all the cases when the aprioristic research results are in consonance with the experimental data.

Keywords — control system, efficiency of functioning, aprioristic information, experimental data, gain in accuracy, gain in number of tests.

For citation: Arseniev V. N., Khomonenko A. D., Yadrenkin A. A. Weighed ranking of aprioristic and experimental data in control system functioning efficiency estimation problem with Pascal-distributed number of tests. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2020, no. 3, pp. 39–47 (In Russian). doi:10.31799/1684-8853-2020-3-39-47

References

- Burenok V. M., Naidyonov V. M. Test base: recovery from the crisis. *Vozdushno-kosmicheskaya oborona*, 2009, no. 1(44), pp. 18–25 (In Russian).
- Pugachev V. S. *Teoriya veroyatnostey i matematicheskaya statistika* [Probability Theory and Mathematical Statistics]. Moscow, Fizmatlit Publ., 2002. 496 p. (In Russian).
- Gorodeckiy V. I., Dmitriev A. K., Markov V. M., Petukhov G. B., Yusupov R. M. *Elementy teorii ispytaniy i kontrolja tehnikeskikh sistem* [Elements of the theory of testing and control of technical systems]. Ed. by R. M. Yusupov. Leningrad, Energiia Publ., 1978. 192 p. (In Russian).
- Frolenkov K. V. Specification of estimations of probabilities at local a posteriori conclusion of algebraic Bayesian networks in case of the inexact certificate. *SPIIRAS Proceedings*, 2013, no. 1(24), pp. 152–164 (In Russian).
- Rojkov V. A. *Teoriya i metody statisticheskogo ocenivaniya veroyatnostnykh harakteristik sluchajnykh velichin i funkciy s gidrometeorologicheskimi primerami*. Kn. 1 [The theory and methods of statistical estimation of probabilistic characteristics of random variables and functions with hydrometeorological examples. Vol. 1]. Saint-Petersburg, Gidrometeoizdat Publ., 2001. 340 p. (In Russian).
- Saypulaeva G. A., Dandamaev A. U. Working of a technique aprioristic and a posteriori estimations of expenditures of labour in maintenance service and repair system. *Sistemnye tehnologii*, 2016, no. 3(20), pp. 43–52 (In Russian).
- Hramenkov A. S., Iarmolik S. N. Algorithm of consecutive recognition of the radar-tracking objects, providing minimisation of average risk of accepted decisions. *Doklady Belorusskogo gosudarstvennogo universiteta*, 2016. no. 2 (96), pp. 37–43 (In Russian).
- Tulupye A. L. A posteriori estimates of probabilities in algebraic Bayesian networks. *Vestnik Sankt-Peterburgskogo universiteta. Prikladnaya matematika. Informatika. Processy upravleniya*, 2012, no. 2, pp. 51–59 (In Russian).
- Bondarenko V. A., Iaritsa A. I. Comparative the analysis aprioristic and a posteriori estimations of accuracy of a planned geodetic network by means of the program complex “Rossija–Belarus” developed in the Polotsk state university. *Vestnik Polotskogo gosudarstvennogo universiteta. Seriya F: Stroitel'stvo. Prikladnye nauki*, 2014, no. 16, pp. 92–95 (In Russian).
- Pugachev V. N. *Kombinirovannye metody opredeleniya veroyatnostnykh harakteristik* [Combined methods of determina-

- tion of probability characteristics]. Moscow, Sovetskoe radio Publ., 1973. 256 p (In Russian).
11. Buriak J. I., Skrinnikov A. A. Increase of degree of validity of accepted decisions in system of recognition at the expense of use of the aprioristic information. *Nauchnyj vestnik Moskovskogo gosudarstvennogo tehničeskogo universiteta grajdanskoj aviacii*, 2015, no. 220 (10), pp. 47–54 (In Russian).
 12. Madera A. G. Method of definition of probabilities of predicted events at decision-making. *Iskusstvennyj intellekt i prinjatje reshenij*, 2016, no. 2, pp. 38–45 (In Russian).
 13. Ardashov A. A., Arseniev V. N., Silantiev S. B. Processing miscellaneous statistical information about control system accuracy characteristics. *Informatcionno-upravljajushchie sistemy*. [Information and Control Systems]. 2015, no. 5, pp. 55–60 (In Russian). doi:10.15217/issn1684-8853.2015.5.55
 14. Arseniev V. N., Silantev S. B., Yadrenkin A. A. Use of the aprioristic information for specification of the events stream model in the complex system. *Journal of Instrument Engineering*, 2017, vol. 60, no. 5, pp. 391–397 (In Russian).
 15. Chow T. S. Testing software design modelled by finite state machines. *IEEE Trans. Software Eng.*, 1978, vol. 4, no. 3, pp. 178–187.
 16. Petrenko A., Simão A., Yevtushenko N. Generating checking sequences for nondeterministic finite state machines. *Proceedings of the ICST*, 2012, pp. 310–319.
 17. Kushik N., El-Fakih K., Yevtushenko N. Adaptive homing and distinguishing experiments for nondeterministic finite state machines. *Lecture Notes in Computer Science*, 2013, vol. 8254, pp. 33–48.
 18. Günçen C., Inan K., Türker U. C., Yenigün H. The relation between preset distinguishing sequences and synchronizing sequences. *Formal Aspects of Computing*, 2014, vol. 26, no. 6, pp. 1153–1167.
 19. Jie M., Honlin Z., Wenbo X., Jin L. Reliability testing methods for critical information system based on state random. *2011 International Conference on Information Communication and Management, IPCSIT*, 2011, vol. 1, IACSIT Press, Singapore, pp. 28–32.
 20. Xin W., Feng-Yan H., Zheng Q. Software reliability testing data generation approach based on a mixture model. *Information Technology Journal*, 2010, vol. 9, no. 5, pp. 1038–1043.
 21. Cai K.-Y., Caob. P., Dongc. Z., Liu. K. Mathematical modeling of software reliability testing with imperfect debugging. *Computers and Mathematics with Applications*, 2010, vol. 59, no. 10, pp. 3245–3285.
 22. Koroljuk V. S., Portenko N. I., Skorokhod A. V., Turbin A. F. *Spravochnik po teorii veroyatnostej i matematicheskoj statistike* [Handbook on probability theory and mathematical statistics]. Moscow, Nauka Publ., 1985. 640 p. (In Russian).
 23. Arseniev V. N., Labeckij P. V. Method a posteriori estimations of characteristics of a control system of the flying machine. *Journal of Instrument Engineering*, 2014, vol. 57, no. 10, pp. 23–28 (In Russian).
 24. Arseniev V. N., Adadurov S. E., Gerasimenko P. V., Degtiarev V. G. Correction of models of disturbing perturbances at research of complex system properties. *2017 Proceedings of 20th IEEE International Conference on Soft Computing and Measurements, SCM*, 2017, pp. 184–188.
 25. Arseniev V. N. *Novie metody prinyatiya reshenij pri ogranichennyh eksperimentalnyh dannyh* [New methods of decision-making on the limited experimental data]. Saint-Petersburg, A. F. Mozhaiskii Military Space Academy Publ., 1999. 90 p. (In Russian).
 26. Khomonenko A. D., Yakovlev E. L. State estimation of complex object via invariants). *Proceedings of 3rd Russian Conference “Mathematical Modeling and Information Technologies”*, Yekaterinburg, November 16, 2016, CEUR Workshop Proceedings, Vol. 1825, pp. 44–49.
 27. Arseniev V. N., Silantev S. B., Khomonenko A. D., Adadurov S. E. *Opreделение veroyatnosti vypolneniya zadachi slozhnoy sistemoy pri ogranichenom ob’eme opytной informatsii*. [Definition of Probability of Task Execution by Complex System at a Limited Volume of Experimental Information]. In: *Mezhdunarodnaya konferentsiya po myagkim vychisleniyam i izmereniyam* [International Conference on Soft Computing and Measurement], *SCM*, 2018, vol. 1. Pp. 43–46. (In Russian).
 28. Smagin V. A., Khomonenko A. D. Using entropy function for definition states of information system. *Workshop “Computer Science and Engineering” in the framework of the 5th International Scientific-Methodical Conference “Problems of Mathematical and Natural-Scientific Training in Engineering Education”*, Saint-Petersburg, November 8–9, 2018, CEUR Workshop Proceedings, vol. 2341, pp. 53–60.

УВАЖАЕМЫЕ АВТОРЫ!

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

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

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

Каскадно-иерархическое моделирование в задачах анализа динамики ресурсных характеристик сложных систем

Ю. В. Доронина^а, доктор техн. наук, доцент, orcid.org/0000-0003-4699-025X, apkSev@yandex.ru

А. В. Скатков^а, доктор техн. наук, профессор, orcid.org/0000-0002-5678-9587

^аСевастопольский государственный университет, Севастополь, Университетская ул., 33, 299053, РФ

Постановка проблемы: описание динамики ресурсных характеристик сложных систем обуславливает необходимость декомпозиции оптимизационных задач и детализацию уровней представления систем, что порождает полимодельность. В задачах, где требуется построение полимодельных комплексов, имеющих сложные связи между моделями, применение только иерархической топологии комплексов моделей не обеспечивает взаимосвязь факторов динамики моделирования и учет объема данных на каждом иерархическом уровне и не отражает параметрическую изменчивость экспериментов на каждом уровне. **Цель:** разработка методологии построения полимодельных комплексов в рамках каскадно-иерархических структур и формирование оценочного функционала с учетом принципа его перманентности для моделирования ресурсных характеристик сложных систем при исследовании динамики деградационных отказов и пополнения ресурсов. **Методы:** системный анализ и структурный синтез моделей, встроенные модели Маркова с квазипоглощающим состоянием на локальном уровне полимодельного комплекса. **Результаты:** сформулирован принцип перманентности оценочного функционала, на основе которого реализован подход к решению задач поддержки принятия решений по ресурсной обеспеченности сложных технических систем путем определения стратегии обслуживания. В общем случае каскадно-иерархическая схема моделирования сформирована в трехмерном функциональном пространстве: структура системы – вид (уровень) модели – каскады планов моделирования. Предложенная методология моделирования в рамках каскадно-иерархической топологии полимодельного комплекса на основе принципа перманентности оценочного функционала позволила сохранить константность представимости системных характеристик при фазовом укрупнении пространства моделирования и тем самым рационализировать планирование экспериментов. **Практическая значимость:** результаты исследований использованы при разработке и анализе динамики ресурсной обеспеченности функционирования сложных технических систем. Предложенный подход реализует возможность в параметрическом модельном пространстве строить планы исследований сложных систем и на основе принципа перманентности оценочного функционала сохранить константность представимости системных характеристик при фазовом укрупнении пространства моделирования.

Ключевые слова – полимодельный комплекс, каскадно-иерархические модели, оценочный функционал, динамика ресурсов, сложные технические системы, модели Маркова.

Для цитирования: Доронина Ю. В., Скатков А. В. Каскадно-иерархическое моделирование в задачах анализа динамики ресурсных характеристик сложных систем. *Информационно-управляющие системы*, 2020, № 3, с. 48–58. doi:10.31799/1684-8853-2020-3-48-58

For citation: Doronina Yu. V., Skatkov A. V. Cascade-hierarchical modeling in analyzing the dynamics of complex system resource characteristics. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2020, no. 3, pp. 48–58 (In Russian). doi:10.31799/1684-8853-2020-3-48-58

Введение

В задачах анализа характеристик сложных систем исследователю в большинстве случаев приходится осмысливать причинно-следственные связи, разворачивающиеся в многомерном координатном пространстве. При построении полимодельных комплексов (ПМК), имеющих сложные связи между отдельными моделями, если применять только иерархическую топологию (монотопологию) моделей, могут появиться следующие проблемы:

1) сложность обеспечения взаимосвязи факторов динамики моделирования и учета объема данных на каждом иерархическом уровне;

2) сложность отражения параметрической изменчивости экспериментов на каждом уровне;

3) упрощение модельной организации на каждом иерархическом уровне,

— что может привести к невозможности оценить качество модели и требуемый объем данных для моделирования [1, 2].

Исследование ресурсных характеристик сложных технических систем (СТС) реализуется в процессах анализа и синтеза этих систем на основе аналитического или имитационного подходов и часто сопряжено с проблемой размерности, заключающейся как в объеме требуемой априорной информации для моделирования, так и в структурной сложности отображений (моделей), реализующих процесс исследования [2–5]. При построении комплексов моделей для анализа ресурсных характеристик СТС должны решаться задачи:

— передачи данных результатов работы одной модели в другую, что приводит к росту промежуточных данных;

— нивелирования сложности оптимизационных задач и неопределенности характеристик функционирования СТС путем декомпозиции и детализации уровней представления моделей;

— снижения трудности в оценке качества комплексов моделей, порожденных их структурной сложностью и топологическим многообразием.

С одной стороны, полимодельность порождается необходимостью моделирования на большом числе групп переменных, с другой стороны, требуется построение ПМК, в том числе для оценки качества самих моделей [4, 6, 7]. То есть полимодельность является функционально-операционной, что связано с разнотипностью задач моделирования на различных уровнях.

Полимодельные комплексы в большинстве своем представлены исследователями в рамках иерархических структур, хотя можно утверждать, что ПМК может формироваться различными способами, например, посредством объединения результатов каждой модели, входящей в комплекс, или каскадным наращиванием и уточнением результатов моделирования [4, 5, 8].

Исследованиями качественного построения комплексов моделей занимались многие авторы, например, в работе [4] констатируется, что остается нерешенной проблема оценивания качества моделей, упорядочения и организации взаимодействия различных классов моделей ПМК, обоснованного синтеза новых моделей либо выбора из числа уже существующих моделей наиболее предпочтительных, предназначенных для решения конкретных прикладных задач [9–13].

Таким образом, моделирование ресурсных характеристик и уточнение топологии модельных комплексов являются актуальными для отраслевых предприятий.

Принцип перманентности оценочного функционала как основа реализации каскадно-иерархического подхода

Для решения задачи оценки остаточного ресурса будем использовать следующую структурную иерархию СТС: детали (уровень *A*) — узлы (уровень *B*) — агрегаты (уровень *C*) — система в целом (уровень *СТС*). Обозначим остаточные ресурсы, требуемые на каждом уровне иерархии, как *RA*, *RB*, *RC*, *RСТС*. Для определенного класса систем уровни *A*, *B*, *C*, *СТС* являются естественными структурными единицами, например, для систем в области машиностроения или информационно-управляющих систем [14].

С использованием графовых отображений можно определить некоторую структуру, основанную на определенной детализации СТС, которая учитывает требуемый или остаточный ресурс на каждом элементном уровне: $G^N(RN, M)$, где *RN* — вершины, соответствующие остаточному ресурсу уровня *N*; *M* — дуги, отражающие влияние остаточного ресурса каждого уровня на остаточный ресурс СТС в целом. Для графа G_k^A уровня *A* дуга *M* определит влияние величины остаточного ресурса на ресурс вышестоящего уровня: $RA_k \rightarrow RB_j$, $k = 1, K$, $j = 1, J$. Аналогично для уровней *B*, *C* и *СТС*: $G_i^B : RB_i \rightarrow RC_j$; $G_j^C : RC_j \rightarrow RСТС$.

Требования к ресурсам на каждом уровне иерархии СТС (рис. 1) определяются следующими компонентами: $RСТС(t_0, X, D, E)$, где t_0 — начальный момент времени оценки остаточного ресурса СТС; *X*, *D*, *E* — некоторые конструктивные, технологические и энергетические характеристики остаточного ресурса СТС соответственно. Множество характеристик остаточного ресурса СТС может быть расширено, но в данном случае не оказывает принципиального влияния на результаты исследования. Принимая во внимание изменчивость распределения ресурсов СТС на некотором отрезке времени жизни системы $[0, T]$, с учетом t_{k-1} и $t_k \in [0, T]$, $t_{k-1} < t_k < t_{k+1}$ можно формализовать динамику изменения остаточного ресурса для каждого уровня иерархии: $N(t_{k-1}) \rightarrow RN(t_k)$, $N \in \{A, B, C, СТС\}$. Для одного уровня иерархии СТС рассматриваются два возможных варианта формального представления требуемого остаточного ресурса: относительно целостного системного объекта либо относительно множества элементов, составляющих этот объект, учитывая принцип аддитивности. Для одноуровневой одноэлементной системы в момент времени $t_{k-1} < t_k < t_{k+1}$ требуемый остаточный ресурс может быть определен как

$$R1(t_k) = \min\{(\alpha_0 R0(t_{k-1}) + \alpha_1 \psi_1(v_k, t_k) + \alpha_2 \varphi_1(u_k, t_k) + \alpha_3 \xi_1(t_{k-1}) + \alpha_4 \mu_1(t_k) + \alpha_5 \varepsilon_1)\}, \quad (1)$$

где $R0(t_k)$, $R1(t_{k-1})$ — остаточные ресурсы 0-го и 1-го уровней в соответствующие моменты времени (t_k), (t_{k-1}); $\psi_1(v_k, t_k)$ — функция, описывающая убывание ресурса вследствие воздействия условий внешней среды, в которой функционирует СТС, отражающая влияние этих условий на ресурсные характеристики СТС; $\varphi_1(u_k, t_k)$ — функция, описывающая убывание ресурса вследствие различных режимов работы (рабочих циклов СТС); $\xi_1(t_k)$ — функция, описывающая убывание ресурса вследствие естественных причин деградации СТС (старение, коррозия, износ, усталость), отражающая их влияние на ресурсные характеристики [8]; $\mu_1(t_k)$ — функция, описывающая рост ресурса вследствие про-

ведения профилактических работ (обслуживание СТС) с учетом работы персонала по регламенту и ошибок персонала; ε_1 — случайная функция, отражающая интенсивность влияния неопределенности условий функционирования на ресурсные характеристики СТС; $\alpha_0 - \alpha_5$ — коэффициенты, определяющие степень воздействия указанных функций на ресурсные характеристики СТС.

Оценка гарантируемого остаточного ресурса для одноуровневой многоэлементной системы (1) может быть записана в виде

$$R1(t_k) = \min_{j \in [N=1]} \{(\alpha_0 R0(t_{k-1}) + \alpha_1 \psi_1(v_k, t_k) + \alpha_2 \varphi_1(u_k, t_k) + \alpha_3 \xi_1(t_{k-1}) + \alpha_4 \mu_1(t_k) + \alpha_5 \varepsilon_1)\}. \quad (2)$$

Для двух уровней иерархии СТС второй уровень может быть только агрегатом, т. е. иметь нулевую аддитивность. С учетом выражений (1) и (2) для двухуровневой СТС в тех же предположениях получим

$$\left\{ \begin{aligned} R1(t_k) &= \min_{j \in [N=1]} \{(\alpha_0 R0_j(t_{k-1}) + \alpha_1 \psi_j(v_k, t_k) + \alpha_2 \varphi_j(u_k, t_k) + \alpha_3 \xi_j(t_{k-1}) + \alpha_4 \mu_j(t_k) + \alpha_5 \varepsilon_1)\}; \\ R2(t_2) &= \min_{j \in [N=2]} \{ \min_{j \in [N=2]} \{ \alpha_j R1_j(t_{k-1}) \}; \\ &(\alpha_2 R2(t_{k-1}) + \alpha_3 \psi_2(v_k, t_k) + \alpha_4 \varphi_2(u_k, t_k) + \alpha_5 \xi_2(t_{k-1}) + \alpha_6 \mu_2(t_k) + \alpha_7 \varepsilon_2)\}. \end{aligned} \right. \quad (3)$$

В общем случае запишем

$$RN(t_k) = \min_{j \in N} \{ \alpha_j R_j [N-1](t_{k-1}) \};$$

$$(\alpha_2 RN(t_{k-1}) + \alpha_3 \psi_N(v_k, t_k) + \alpha_4 \varphi_N(u_k, t_k) + \alpha_5 \xi_N(t_{k-1}) + \alpha_6 \mu_N(t_k) + \alpha_7 \varepsilon_N), \quad (4)$$

где $R[N-1](t_{k-1})$ — остаточный ресурс предшествующего иерархического уровня (в связи с тем, что уровень может быть представлен множеством элементов, справедливо: $\min_{j \in N} \{ \alpha_j R_j [N-1](t_{k-1}) \}$ при $j \in N$); $RN(t_{k-1})$ — остаточный ресурс предшествующего временного отрезка для $RN(t_k)$; $\psi_N(v_k, t_k)$ — функция, описывающая убывание ресурса вследствие воздействия условий внешней среды, в которой функционирует СТС, отражающая влияние этих условий на ресурсные характеристики для N -го уровня СТС; $\varphi_N(u_k, t_k)$ — функция, описывающая убывание ресурса вследствие различных режимов работы (рабочих циклов СТС) для N -го уровня СТС; $\xi_N(t_{k-1})$ — функция, описывающая убывание ресурса вследствие естественных причин деградации СТС для N -го уровня СТС; $\mu_N(t_k)$ — функция, описывающая рост ресурса вследствие проведения профилактических работ (обслуживание СТС) с учетом работы персонала по регламенту

и ошибок персонала для N -го уровня СТС; ε_N — случайная функция, описывающая убывание ресурса вследствие влияния неопределенности условий функционирования на ресурсные характеристики СТС; $\alpha_2 - \alpha_7$ — коэффициенты, определяющие степень воздействия указанных функций на ресурсные характеристики СТС.

Из двух сравниваемых величин, подлежащих минимизации в выражениях (3) и (4), как правило, минимальным оказывается второе слагаемое, т. е. $RN(t_k)$ в простейшем случае может быть определен как $RN(t_k) = \alpha_2 RN(t_{k-1}) + Z(\cdot)$, где $Z(\cdot) = \alpha_3 \psi_N(v_k, t_k) + \alpha_4 \varphi_N(u_k, t_k) + \alpha_5 \xi_N(t_{k-1}) + \alpha_6 \mu_N(t_k) + \alpha_7 \varepsilon_N$.

Принцип перманентности оценочного функционала (ППОФ) как основы реализации каскадно-иерархического подхода. Определяющее системное свойство осуществления каскадно-иерархического представления СТС — сохранение в рамках суперпозиции отображений ряда последовательных уровней детализации сквозной единой системной характеристики для категории оценочного функционала при исследовании качества функционирования (работоспособности) системы. В этом и состоит идея соблюдения ППОФ в рамках каскадно-иерархического подхода к топологическому описанию ПМК.

Определение ресурсных характеристик СТС на основе $RN(t_k)$ в выражении (1) соответствует ППОФ через каскадно-иерархическую топологию ПМК.

Для некоторого малого $\Delta t > 0$ и заданных величин остаточного ресурса в начальный момент времени: $RA(t_0) = ra$, $RB(t_0) = rb$, $RC(t_0) = rc$, $RCTC(t_0) = rctc$ для $t_{k-1}, t_k \in [0, T]$; $t_{k-1} < t_k < t_{k+1}$, остаточные ресурсы СТС в рамках иерархии структурных единиц могут быть представлены системой обыкновенных линейных дифференциальных уравнений в предположении, что $ra, rb, rc, rctc$ дифференцируемы, а функции, стоящие справа, — непрерывны на рассматриваемом временном интервале:

$$\left\{ \begin{aligned} \frac{dRA(t_k)}{dt} &= \alpha_1^A RA(t_{k-1}) + \alpha_3^A \psi_A(v_k, t_k) + \alpha_4^A \varphi_A(u_k, t_k) + \alpha_5^A \xi_A(t_{k-1}) + \alpha_6^A \mu_A(t_k) + \alpha_7^A \varepsilon_A \\ \frac{dRB(t_k)}{dt} &= \beta_2^B RB(t_{k-1}) + \beta_3^B \psi_B(v_k, t_k) + \beta_4^B \varphi_B(u_k, t_k) + \beta_5^B \xi_B(t_{k-1}) + \beta_6^B \mu_B(t_k) + \beta_7^B \varepsilon_B \\ \frac{dRC(t_k)}{dt} &= c_2^C RC(t_{k-1}) + c_3^C \psi_C(v_k, t_k) + c_4^C \varphi_C(u_k, t_k) + c_5^C \xi_C(t_{k-1}) + c_6^C \mu_C(t_k) + c_7^C \varepsilon_C \\ \frac{dRCTC(t_k)}{dt} &= \tau_2^{CTC} RCTC(t_{k-1}) + \tau_3^{CTC} \psi_{CTC}(v_k, t_k) + \tau_4^{CTC} \varphi_{CTC}(u_k, t_k) + \tau_5^{CTC} \xi_{CTC}(t_{k-1}) + \tau_6^{CTC} \mu_{CTC}(t_k) + \tau_7^{CTC} \varepsilon_{CTC} \end{aligned} \right. , \quad (5)$$

где $\alpha_{(\cdot)}^A, \beta_{(\cdot)}^B, c_{(\cdot)}^C, \tau_{(\cdot)}^{CTC}$ — параметры, определяющие степень воздействия соответствующих функций на ресурсные характеристики СТС.

Система (5) является описанием ресурсных зависимостей и может быть структурно усложнена за счет введения нелинейности в локальных областях, а также может быть представлена в безразмерной форме с сохранением той же нотации. Существование решения системы (5) гарантируется теоремой Коши и возможно по конечно-разностной схеме с учетом начальных условий (распределения ресурса в момент времени t_0) при задании неопределенных компонент детерминированными прогностическими значениями исходя из конкретных условий работы СТС. Поскольку функции ψ_N, ϕ_N, ξ_N являются неотрицательными, выпуклыми, следовательно, правые части не только непрерывны, но и удовлетворяют условию Липшица. Для системы (5) гарантируется существование решения и его единственность для заданных начальных условий.

Применение ППОФ для одноуровневой системы позволяет сформулировать суперпозицию отображения детализации единой системной характеристики категории оценочного функционала качества функционирования (работоспособности) системы в двух формах (для агрегированной и структурно аддитивной системы).

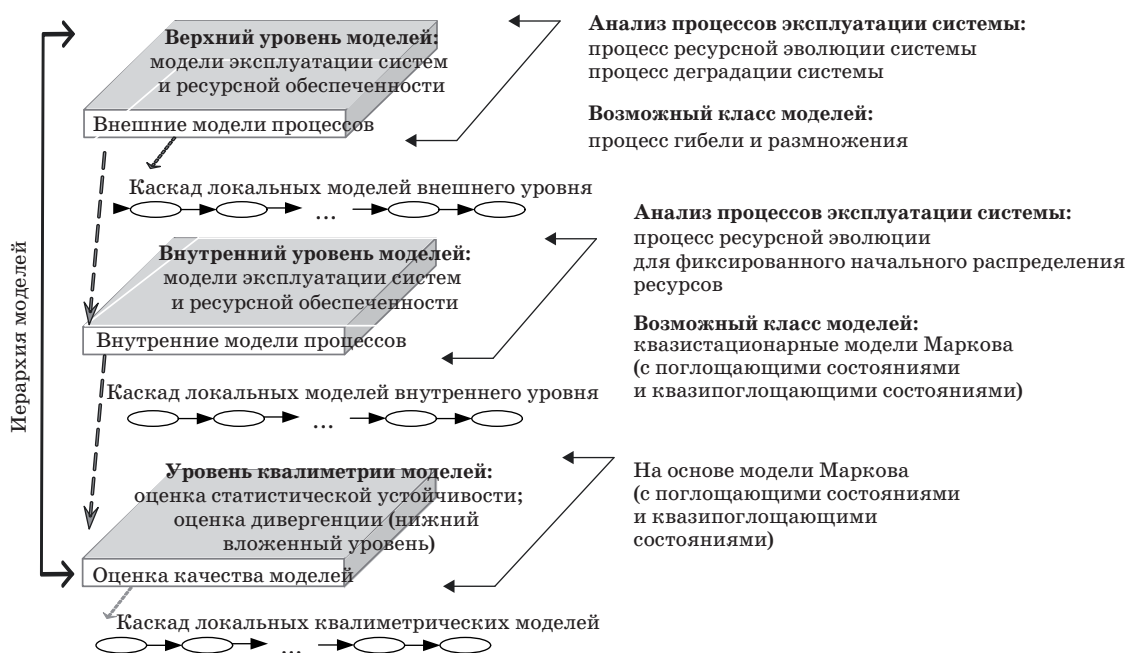
Для двухуровневой системы ППОФ позволяет сформулировать сохранение при суперпозиции отображений двух последовательных уровней детализации сквозной единой системной харак-

теристики категории оценочного функционала качества функционирования (работоспособности) системы при переходе от структурной аддитивности (на первом уровне) к агрегированию (на втором).

При решении задачи объект и предмет исследования представлены комплексом исследуемых процессов, что порождает формирование ПМК. Структурная схема иерархии моделей в ПМК (рис. 1) содержит три основных уровня: верхний и внутренний уровни, связанные с моделями эксплуатации и ресурсной обеспеченности, и уровень квалитметрии моделей. Данное структурное представление иерархии и каскадов моделей в ПМК иллюстрирует накопление степеней свободы моделирования, например, в рамках квалитметрического анализа формируется накапливаемая сложность согласно плану моделирования $\Pi(\cdot)$, что отражено в каскадных моделях текущего уровня.

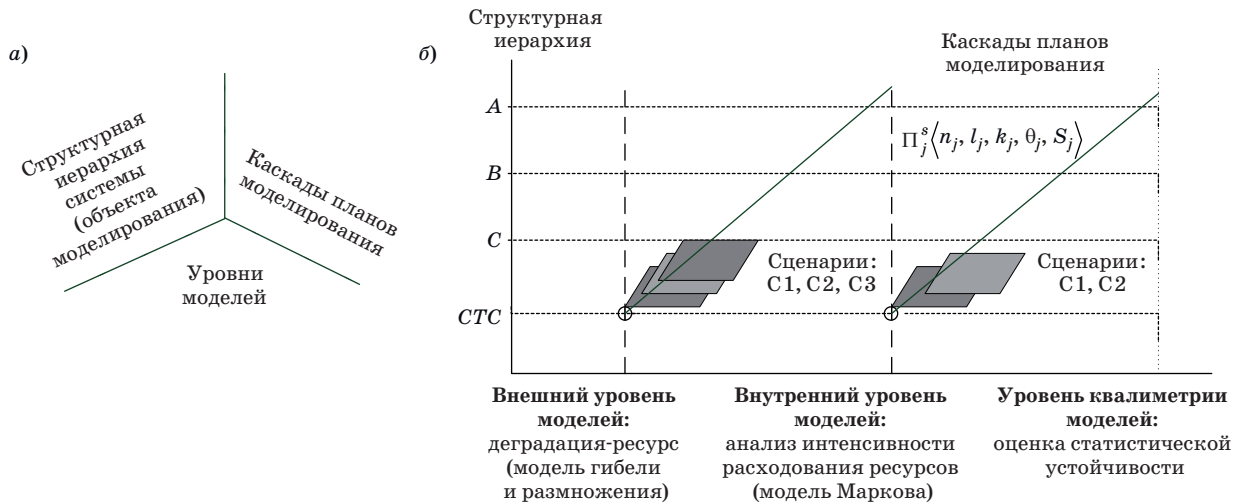
В задачах анализа динамики ресурсных характеристик СТС необходимо иметь обширный экспериментальный материал, что в ряде случаев затруднительно или невозможно. Следовательно, как на стадии создания, так и на стадии эксплуатации СТС ресурсный анализ может быть осуществлен на основе имитационного моделирования (ИМ) [15–18].

Применение скрытых моделей Маркова (СММ) к анализу надежности и прогнозированию характеристик СТС в ИМ широко распространено. Результаты моделирования не во всех случаях



■ **Рис. 1.** Структурное представление иерархии и каскадов моделей в ПМК

■ **Fig. 1.** Structural representation of hierarchy and cascades of models in a multi-model complex



■ **Рис. 2.** Каскадно-иерархическое обобщенное структурное представление: *a* — пространств модели — структур — планов моделирования; *b* — полимодельного комплекса для анализа ресурсных характеристик СТС
 ■ **Fig. 2.** Cascade-hierarchical generalized structural representation: *a* — model spaces — structures — modeling plans; *b* — a multi-model complex for analyzing the resource characteristics of the СТС

оказываются репрезентативными и достоверными, в связи с этим целесообразно развивать механизмы решения задач анализа динамики СТС с учетом оценки качества решений в рамках СММ.

При исследовании ресурсных характеристик СТС на основе ИМ с учетом ППОФ в ряде случаев каскадное представление моделей может быть осуществлено последовательностью планов экспериментов $\Pi_n \langle \cdot \rangle \rightarrow \Pi_{n+1} \langle \cdot \rangle$. На рис. 2, *a* и *b* схематически показаны пространство модели — структуры — планы моделирования и обобщенная схема каскадно-иерархического моделирования при анализе ресурсных характеристик СТС.

Планы имитационных экспериментов могут быть представлены в виде кортежа $\Pi \langle n, l, k, \theta \rangle$, где n — скаляр, задающий число сгенерированных переходов СММ, $n = |N|$; l_m — длина m -й цепи Маркова, $l = |L|$; k — кратность запусков модели, $k = |K|$; $\theta_{n,l,k}$ — многомерная матрица, которая может содержать результаты первичной статистики. Таким образом, если обозначить D — множество допустимых планов ($J = |D|$), то $\Pi \langle n_j, l_j, k_j, \theta_j \rangle$ — реализуемый план. Таким образом, каскадная составляющая ПМК представлена разным функциональным наполнением, имеющим, соответственно, накапливаемую сложность согласно планам $\Pi \langle \cdot \rangle$.

На основе принципа перманентности оценочного функционала как основы реализации каскадно-иерархического подхода осуществлено объединение в рамках каскадно-иерархического структурного представления моделей в виде пространства модели — структуры — планы ПМК для анализа ресурсных характеристик СТС.

Пример реализации каскадно-иерархического подхода при моделировании ресурсных характеристик СТС

Целью моделирования является формирование ПМК, на основе которого необходимо провести оценку ресурсных характеристик СТС, что соответствует вводу верхнего и вложенного подуровней и нескольких сценариев, определенных лицом, принимающим решения (ЛПР), как условия функционирования СТС на каждом из уровней.

Рассмотрим процесс деградации некоторой технической системы (уровень иерархии: СТС), уравниваемый циклами пополнения ресурса: деградационный процесс — пополнение ресурса (ДП-ПР) [13]. Схема моделирования может быть реализована на основе обобщенного опыта стохастического моделирования и с учетом классификации модельных схем, предложенных

в работе [5]: $schema P_m : \langle S_0, S_1, \dots, S_n \rangle \xrightarrow{M_x^{I,K}} \|P_{ij}\|$, где P_m — дискретно-стохастическая модель (P -схема), модель с ожидаемым результатом в виде матрицы; \rightarrow будем обозначать отображение, осуществляемое имитационной моделью $M_x^{I,K}$, где I, K — индексы иерархии и каскада ПМК соответственно.

Процесс ДП-ПР СТС может быть представлен моделью Маркова с дискретными состояниями: S_0 — начальное состояние системы; S_n — n -е состояние, в котором оказывается система после

n -го перехода; S_d — некоторое состояние деградации СТС (квазипоглощающее (КП)), после которого при достижении предельно низкого уровня обеспеченности ресурсом $P_d(t)$ система переходит в предельное поглощающее состояние S_p ; λ_{ij} — интенсивность перехода из состояния i в состояние j .

Согласно известной аналитической схеме решения данной задачи на основе модели гибели и размножения, вероятности пребывания СТС в состояниях можно записать системой дифференциальных уравнений Колмогорова с учетом нормировки и при условии, что начальное распределение вероятностей имеет вид $P_0(t) = 1, P_1(t) = 0, \dots, P_d(t) = 0$. Финальное значение вероятности пребывания СТС в состоянии P_0 с КП-состоянием S_d определяется по известным формулам (модель гибели и размножения).

Для рассматриваемого процесса ДП-ПР СТС в течение некоторого времени t согласно схеме моделирования *schema* P_m получено следующее семейство кривых, отражающих изменение распределения вероятностей пребывания в состояниях относительно сценариев изменения интенсивности λ_{ij} (т. е. от равновесия веток модели ДП-ПР) (табл. 1, рис. 3). Для удобства записи переменных в табл. 1 введены обозначения цепочек переходов и их интенсивностей: $\alpha_i = (\lambda_{01} \rightarrow \lambda_{12} \rightarrow \lambda_{23} \rightarrow \lambda_{34}) | (\lambda_{10} \leftarrow \lambda_{21} \leftarrow \lambda_{32} \leftarrow \lambda_{43})$, где α_i формирует три сценария S_j , описывающих различные ситуации по ДП-ПР. Шаг изменения $\lambda_{ij} \rightarrow \lambda_{jk}$ принят $\Delta = 0,2$.

Анализ результатов моделирования, представленных на рис. 3, позволил сделать ряд выводов.

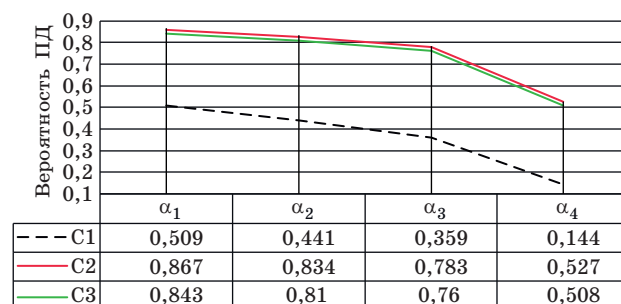
1. При описании условий моделирования потребовалось вводить дополнительные переменные и константы (α_i, Δ, C_j), что усложнило модель верхнего уровня. Это соответствует вводу вложенного подуровня, в котором произведена заме-

на переменных; выбраны три сценария и четыре различных коэффициента α_i .

2. С точки зрения содержательного аспекта моделирования следует отметить, что сценарии изменения параметров модели значительно влияют на вероятность пребывания СТС в КП-состоянии S_d . При известных интенсивностях λ_{ij} нахождение вероятности не составит труда, но в ситуации, когда λ_{ij} не заданы, моделирование возможно только на основе сценарного подхода.

3. Модель верхнего уровня позволила получить предварительные оценки распределения вероятностей при принятом исходном распределении ресурса СТС. Для сценариев S_1, S_2 средние изменения ΔP_d составили 0,35.

Следовательно, для модели внутреннего уровня, отражающей функционирование системы с учетом остаточного ресурса, целесообразно учитывать полученную на верхнем уровне оценку ΔP_d . Согласно (5), требуемый ресурс для СТС в момент времени t_k для верхнего уровня моделирования может быть определен следующим образом:



■ **Рис. 3.** Диаграмма оценок вероятностей пребывания СТС в КП-состоянии S_d от соотношения интенсивностей λ_{ij} процессов в модели верхнего уровня

■ **Fig. 3.** Chart of estimates of the probability of CTS stay in the quasi-absorbing state, S_d depends on the ratio λ_{ij} of the process intensities in the upper-level model

■ **Таблица 1.** Значения изменения интенсивностей λ_{ij} переходов СТС для модели верхнего уровня

■ **Table 1.** Values for changing transition intensities λ_{ij} CTS for the top-level model

Описание сценария	Интенсивность в сценариях			
	$\alpha_i = (\lambda_{01} \rightarrow \dots \lambda_{34}) (\lambda_{10} \leftarrow \dots \lambda_{43}), \Delta = 0,2$			
	α_1	α_2	α_3	α_4
Сценарий С1: постепенное снижение интенсивности ПР на фоне снижения ДП	(0,5–0,9) / (0,7–0,1)	(0,4–0,9) / (0,7–0,1)	(0,3–0,9) / (0,7–0,1)	(0,1–0,7) / (0,7–0,1)
Сценарий С2: постепенное повышение интенсивности ДП и ПР	(0,5–0,9) / (0,1–0,7)	(0,4–0,9) / (0,1–0,7)	(0,3–0,9) / (0,1–0,7)	(0,1–0,7) / (0,1–0,7)
Сценарий С3: рост интенсивности ДП при неизменной интенсивности ПР	(0,5–0,99) / 0,1	(0,4–0,99) / 0,1	(0,3–0,9) / 0,1	(0,1–0,7) / 0,1

$$RCTC(t_k) = \alpha_2 RCTC(t_{k-1}) + \alpha_3 \psi_{CTC}(v_k, t_k) + \alpha_4 \varphi_{CTC}(u_k, t_k) + \alpha_5 \xi_{CTC}(t_{k-1}) + \alpha_6 \mu_{CTC}(t_k) + \alpha_7 \varepsilon_{CTC} = \alpha_2 RCTC(t_{k-1}) + K(\cdot) + \alpha_6 \mu_{CTC}(t_k), \quad (6)$$

где $K(\cdot) = K(\alpha_3 \psi_{CTC}(v_k, t_k) + \alpha_4 \varphi_{CTC}(u_k, t_k) + \alpha_5 \xi_{CTC}(t_{k-1}) + \alpha_7 \varepsilon_{CTC})$; $RCTC(t_{k-1})$ — требуемый ресурс для агрегатов СТС; $\mu_{CTC}(t_k)$ — функция обслуживания СТС, которая наращивает ее ресурс $RCTC(t_0) + \Delta RCTC(t_k)$, вследствие чего вероятность пребывания СТС в КП P_d снижается. С учетом трех сценариев C_j , описывающих различные ситуации по ДП-ПР, имеет место система уравнений как частный случай систем (4) и (5):

$$\begin{cases} RCTC^{C1}(t_k) = \alpha_2^{C1} RCTC^{C1}(t_{k-1}) + \vartheta^{C1}(\cdot) + \alpha_7^{C1} \mu_{CTC}^{C1}(t_k) \\ RCTC^{C2}(t_k) = \alpha_2^{C2} RCTC^{C2}(t_{k-1}) + \vartheta^{C2}(\cdot) + \alpha_7^{C2} \mu_{CTC}^{C2}(t_k) \\ RCTC^{C3}(t_k) = \alpha_2^{C3} RCTC^{C3}(t_{k-1}) + \vartheta^{C3}(\cdot) + \alpha_7^{C3} \mu_{CTC}^{C3}(t_k) \end{cases}, \quad (7)$$

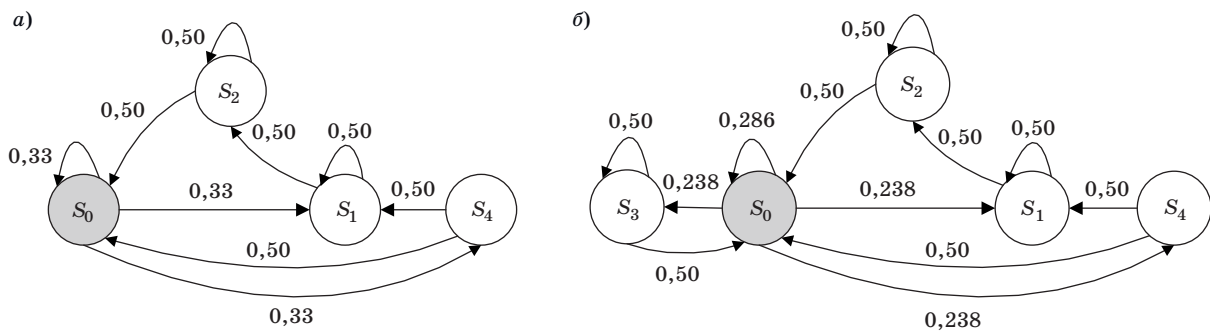
где коэффициенты $\alpha_{(\cdot)}^{C1}$, $\alpha_{(\cdot)}^{C2}$, $\alpha_{(\cdot)}^{C3}$ определяются условиями функционирования СТС по каждому из сценариев.

Решение системы (7) может быть реализовано любым известным методом, учитывая полученные оценки ΔP_d , однако в связи со сложностью технического представления система (7) не всегда может быть эффективным инструментом.

На основе фазового укрупнения модельного представления СТС для вложенной модели ПМК, связанной с оценкой расходования ресурсов в зависимости от сценария интенсивности расходования ресурсов в жизненном цикле СТС, всюду далее рассматривается вложенная модель Маркова (ВММ) без поглощающих состояний в рамках стационарного (квазистационарного) режимов.

ВММ строится по схеме *schema A*: $\|P_{ij}\| \rightarrow \|P_{ij}^{M_m}\|$ в следующем фазовом пространстве состояний системы: $S = \{S_0, S_1, S_2, S_3, S_4\}$, где $\|P_{ij}\|$ — матрица переходов ВММ; $\langle S_0, S_1, S_2, S_3, S_4 \rangle$ — последовательность состояний ВММ; k — число запусков модели, за которое система из первоначального состояния перейдет в терминальное состояние, описываемое матрицей $\|P_{ij}^k\|$; $\|P_{ij}^{M_m}\|$ — «восстановленная» матрица переходных вероятностей, содержащая статистические оценки P_{ij} ; S_0 — рабочее состояние СТС, S_1 — состояние отказа СТС, S_2 — состояние пополнения ресурса системы после отказа СТС, S_3 — состояние профилактики СТС (в модели не фиксируется, работает или простаивает система в момент пополнения ресурса), S_4 — состояние сбоя СТС (это состояние, из которого может быть осуществлен переход либо в состояние отказа, вслед за которым требуется пополнение ресурса, либо в начальное состояние без пополнения ресурса; таким образом, S_4 — это состояние, при котором пополнение ресурса СТС не требуется [7, 19, 20]). На рис. 4 приведены графы переходов такой СТС.

Введем коэффициент расходования ресурса как нормированный показатель, отражающий интенсивность динамики ресурсов различных типов в j -м состоянии СТС, RCI_j (Resource Consumption Index). Тогда общий потребляемый ресурс СТС может быть определен как $\sum_{j=1}^J n \cdot RCI_j$, где n — частота захода в каждое состояние ВММ. С учетом длительности локального жизненного цикла получим $\sum_{j=1}^J n \cdot RCI_j \cdot t$. Рассмотрим план $\Pi_j^s \langle n_j, l_j, k_j, \theta_j, Z_j \rangle$, где Z_j — интенсивность расходования ресурсов в j -м состоянии СТС. В табл. 2 приведены результаты моделирования



■ **Рис. 4.** Граф переходов для состояний СТС с учетом остаточного ресурса для модели внутреннего уровня: *a* — без пополнения ресурса; *b* — с пополнением ресурса, S_3
 ■ **Fig. 4.** Transition graph for CTS states with a residual resource for an internal-level model: *a* — without adding resources; *b* — with the completion of the resource, S_3

■ **Таблица 2.** Результаты моделирования RCI_j затрат при эксплуатации СТС в зависимости от разных сценариев интенсивности расходования ресурсов в модели внутреннего уровня

■ **Table 2.** Results of RCI_j simulation of costs during the operation of the СТС depending on different scenarios of expenditure intensity resources in the internal level model

Состояние системы	Значение RCI_j		Интенсивность расходования ресурсов		Общий израсходованный ресурс	
	C_1	C_2	C_1	C_2	C_1	C_2
S_0	0,5	0,2	36,4	14,56	56,94	40,66
S_1	0,2		9,38	9,384		
S_2	0,2		8,39	8,392		
S_3	0,1		2,62	5,24		
S_4	0,01		0,15	3,084		

$\sum_{j=1}^J n \cdot RCI_j$. Под сценарием $C_i, i = \|C\|$ в предлож-

женном методе текущего уровня моделирования понимается событие или группа событий, приводящих СТС в КП-состояние, что связано с изменением интенсивности расходования ресурсов:

$C_i :< P_{ij}^{M_m} | c(S_i), RCI_i >$, где $P_{ij}^{M_m}$ — элемент матрицы $\|P_{ij}^{M_m}\|$; $c(S_i)$ — стоимость пребывания СТС

в каждом S_i ; RCI_j — коэффициент расходования ресурсов. В табл. 2 рассмотрены два сценария: $C_1 : RCI_j$ при разных $c(S_i)$, определенных ЛПП, и $C_2 : RCI_j$ при равных $c(S_i)$ для всех S_i . Сценарий C_1 соответствует ситуации, когда ЛПП производит ранжирование $c(S_i)$, что определяется различными уровнями обеспеченности профилактики или отказа СТС. Для C_2 характерно нейтральное отношение ЛПП к стоимости пребывания в каждом состоянии системы S_i .

Полученные результаты моделирования (см. табл. 2) отражают зависимость RCI от различных сценариев обслуживания. Для $S_0(C_1)$ при $c(S_0(C_1)) = 0,5$ относительно $S_0(C_2)$ при $c(S_0(C_2)) RCI_j$ отличается в 2,5 раза.

В отличие от модели верхнего уровня число сценариев, отражающих варианты интенсивности расходования ресурсов, другое: три в модели верхнего уровня и два в модели внутреннего, — что приводит к сложной интерпретации постановки экспериментов при моделировании. С учетом двух сценариев C_1, C_2 , описывающих различные ситуации расходования ресурса на основе коэффициента RCI_j , можно записать систему уравнений как частный случай системы (4) по аналогии с (7).

Исследование динамики ресурсных характеристик систем обуславливает необходимость декомпозиции оптимизационных задач и детализацию уровней представления исследуемых систем, порождающую полимодельность, что, в свою очередь, для решения прикладных задач значительно усложняет процесс моделирования [20, 22–29]. Авторский подход позволил на основе предложенной методологии формализовать каскадно-иерархический подход и обеспечить упорядоченность построения ПМК для прикладной области, формализацию оценок и с учетом ППОФ получить комплексы оценок в каждом каскаде всех уровней иерархии.

Заключение

В задачах анализа ресурсных показателей СТС часто возникает проблема формулировки обобщенных выводов функционирования системы, а для более объективной оценки необходим больший объем данных. Этот фактор требует построения ПМК, имеющих сложные связи между моделями (при передаче данных из одной модели в другую). В этом случае применение одного, например иерархического, типа моделей в ПМК может привести к ряду проблем, связанных со сложностью идентификации и учета взаимосвязи факторов динамики моделирования, учета объема данных на каждом иерархическом уровне, наличием параметрической изменчивости экспериментов на каждом уровне.

В предложенной авторами методологии каскадно-иерархическая схема моделирования сформирована в трехмерном параметрическом пространстве: структура системы — вид (уровень) модели — каскады планов моделирования. Каскадно-иерархическая топология ПМК позволила минимизировать указанные проблемы, а на основе сформулированного принципа перманентности оценочного функционала сохранить константность зависимости системных характеристик при фазовом укрупнении пространства моделирования. Дальнейшее развитие исследований планируется в направлении исследования их топологических параметров и формирования схем принятия решений на разных уровнях и типах задач.

Финансовая поддержка

Исследование выполнено при финансовой поддержке Российского фонда фундаментальных исследований (гранты № 19-29-06015/19, 19-29-06023/19).

Литература

1. Ашимов А. А., Гейда А. С., Лысенко И. В., Юсупов Р. М. Эффективность функционирования и другие операционные свойства систем: задачи и метод оценивания. *Труды СПИИРАН*, 2018, № 5(60), с. 241–270. doi.org/10.15622/sp.60.9
2. Соколов Б. В., Бураков В. В., Микони С. В., Юсупов Р. М. Методологические и методические основы теории оценивания качества моделей и полимодельных комплексов. *Информация и космос*, 2018, № 3, с. 36–43.
3. Panella I., Hardwick G. Model oriented system design applied to commercial aircraft secondary flight control systems. In: Obaidat M., Zren T., Rango F. (eds). *Simulation and Modeling Methodologies, Technologies and Applications. SIMULTECH 2017. Advances in Intelligent Systems and Computing*. Springer, Cham, 2019, vol. 873. Pp. 55–76.
4. Микони С. В., Соколов Б. В., Юсупов Р. М. *Квалиметрия моделей и полимодельных комплексов*. М., РАН, 2018. 314 с.
5. Доронина Ю. В., Скатков А. В. Анализ статистической устойчивости стационарных марковских моделей. *Труды СПИИРАН*, 2019, № 5(18), с. 1119–1148. doi 10.15622/sp.2019.18.5.1119-1148
6. Degiannakis S., Floros C. Methods of volatility estimation and forecasting. In: *Modelling and Forecasting High Frequency Financial Data*. Palgrave Macmillan, London, 2015. Pp. 58–109.
7. Волкова В. Н., Козлов В. Н., Магер В. Е., Черненко Л. В. Классификация методов и моделей в системном анализе. *Сборник докладов XX Международной конференции по мягким вычислениям и измерениям (SCM-2017)*, Санкт-Петербург, 24–26 мая 2017 г., СПбГЭТУ (ЛЭТИ), СПб., 2017, с. 223–226.
8. Okhtilev M. Yu., Gnidenko A. S., Alferov V. V., Salukhov V. I., Nazarov D. I. Methods and algorithms of integrated modeling of complex technical objects in dynamically changing conditions. *Proceedings of the International Scientific Conference MMET NW*, 2018, pp. 282–284.
9. Дедков В. К. Методика построения модели прогнозирования надежности «стареющего» технического объекта. *Труды Международного симпозиума «Надежность и качество»*, 2009, № 1, с. 366–368.
10. Kunit Park. *Fundamentals of Probability and Stochastic Processes with Applications to Communications*. New York, Springer, 2018. 273 p.
11. Rabiner L. R., Wilpon J. G., and Juang B. H. A model-based connected digit recognition system using either hidden Markov models or templates. *Computer Speech and Language*, 1986, vol. 1, no. 2, pp. 167–197.
12. Chen J., Ma C., Song D. Time to failure estimation based on degradation model with random failure threshold. *International Journal of Reliability and Safety*, 2016, vol. 10, no. 2, pp. 145–157.
13. Гончаров А. М., Чащин С. В., Прохоров М. А. Подход к решению задачи оценивания устойчивого функционирования информационной системы на примере центра обработки данных. *T-Comm — Телекоммуникации и Транспорт*, 2017, т. 11, № 4, с. 20–25.
14. Цвиркун А. Д. *Основы синтеза структуры сложных систем*. М., Наука, 1982. 200 с.
15. Zelentsov V. A., Potryasaev S. A., Pimanov I. J., Nemyukin S. A. Creation of intelligent information flood forecasting systems based on service oriented architecture. *Advances in Intelligent Systems and Computing*, 2016, vol. 466, pp. 371–381. doi:10.1007/978-3-319-33389-2_35
16. Скатков А. В., Воронин Д. Ю., Скатков И. А. Особенности моделирования деградационных отказов первичных измерителей систем мониторинга. *Системы контроля окружающей среды*, 2017, № 7(27), с. 48–56.
17. Кузнецов А. В. Распределение ограниченных ресурсов в системе с устойчивой иерархией (на примере перспективной системы военной связи). *Управление большими системами*, 2017, вып. 66, с. 68–93.
18. Новиков Д. А. Аналитическая сложность и погрешность решения задач управления организационно-техническими системами. *Автоматика и телемеханика*, 2018, № 5, с. 107–118.
19. *New Frontiers in Information and Production Systems Modeling and Analysis — Incentive Mechanisms, Competence Management, Knowledge-based Production*. Series “Intelligent Systems Reference Library”. Ed. by P. Rozewsky, D. Novikov, O. Zaikin, N. Bakhtadze. Berlin, Springer, 2016. 268 p. doi:10.1007/978-3-319-23338-3
20. Zhukova N., Vodyaho A., Levonevskiy D., Simonenko A. The method of data transformation for modeling technical objects. *Proceedings of the XXI IEEE International Conference on Soft Computing and Measurements (SCM'2018)*, 2018, pp. 450–453.
21. Gnidenko A. S., Zelentsov V. A., Kulakov A. U. Hierarchical polymodel complex of combined planning of transport and logistics systems. *Proceedings of the International Scientific Conference MMET NW*, 2018, pp. 276–279.
22. Кононова Я. Г., Шубин М. В., Ермаков Е. А. Ремонт авиационной техники военного назначения в рамках государственно-частного партнерства. *Проблемы экономики и менеджмента*, 2016, № 5 (57). <https://cyberleninka.ru/article/n/remont-aviatsionnoy-tehniki-voennogo-naznacheniya-v-ramkah-gosudarstvenno-chastnogo-partnyorstva> (дата обращения: 23.11.2019).
23. Киселев Ю. В., Мотиенко А. И., Басов О. О., Саитов И. А. Структурно-функциональная модель интеллектуальной инфокоммуникационной системы. *Научно-технический вестник информационных технологий, механики и оптики*, 2018, т. 18, № 6, с. 1034–1046.
24. Баранов В. В., Цвиркун А. Д. Управление развитием: структурный анализ, задачи, устойчивость. *Автоматика и телемеханика*, 2018, № 10, с. 55–75.

25. Батенков К. А. Общие подходы к анализу и синтезу структур сетей связи. *Современные проблемы телекоммуникаций: материалы Российской научно-технической конференции*, 2017, с. 19–23.
26. Нестерук Ф. Г. К организации интеллектуальной защиты информации. *Труды СПИИРАН*, 2009, № 10, с. 148–159. <https://doi.org/10.15622/sp.10.10>
27. Новиков Д. А. Иерархические модели военных действий. *Управление большими системами*, 2012, вып. 37, с. 25–62.
28. Месарович М., Мако Д., Такахара И. *Теория иерархических многоуровневых систем*. М., Мир, 1973. 344 с.
29. Majedi M., Osman K. A novel architectural design model for enterprise systems: evaluating enterprise resource planning system and enterprise application integration against service oriented architecture. *Proceedings of 3rd International Conference on Pervasive Computing and Applications*, Alexandria, Egypt, IEEE Computer Society, 2008, pp. 116–121.

UDC 004.052

doi:10.31799/1684-8853-2020-3-48-58

Cascade-hierarchical modeling in analyzing the dynamics of complex system resource characteristicsYu. V. Doronina^a, Dr. Sc., Tech., Associate Professor, orcid.org/0000-0003-4699-025X, apkSev@yandex.ruA. V. Skatkov^a, Dr. Sc., Tech., Professor, orcid.org/0000-0002-5678-9587^aSevastopol State University, 33, Universitetskaya St., 299053, Sevastopol, Russian Federation

Introduction: Describing the dynamically changing resource characteristics of a complex system makes it necessary to decompose the optimization problems and itemize the system representation levels. This leads to multi-model approaches. In problems which require the construction of multi-model complexes with sophisticated links between the models, the sole use of hierarchical topology of model complexes does not guarantee that the modeling dynamics factors are connected and the data volume is taken into account at each hierarchical level. It also does not reflect the parametric variability of experiments at each level. **Purpose:** Developing a technique for constructing multi-model complexes within cascade-hierarchical structures, and forming the evaluation functionality taking into account the principle of its permanence for modeling the resource characteristics of complex systems when studying the dynamics of degradation failures and replacement of resources. **Methods:** System analysis and structural synthesis of models; embedded Markov models with quasi-absorbing state at the local level of a multi-model complex. **Results:** The principle of permanent evaluation functionality is formulated. It allows you to implement an approach to the problems of decision support for resource provision in complex technical systems by determining the service strategy. In general, a cascade-hierarchical modeling scheme is formed in a three-dimensional functional space: system structure — model type (level) — modeling plan cascades. The proposed modeling methodology within the cascade-hierarchical topology of a multi-model complex on the base of the permanent evaluation functionality principle allows you to keep the constancy of the representability of system characteristics under the phase enlargement of the modeling space, and thereby rationalize the planning of your experiments. **Practical relevance:** The results of the research were used in the development and analysis of the dynamics of resource security of complex technical systems. The proposed approach makes it possible to formulate research plans for complex systems in a parametric model space.

Keywords — multi-model complex, cascade-hierarchical models, evaluation functionality, resource dynamics, complex technical systems, Markov models.

For citation: Doronina Yu. V., Skatkov A. V. Cascade-hierarchical modeling in analyzing the dynamics of complex system resource characteristics. *Informatsionno-upravliayushchie sistemy* [Information and Control Systems], 2020, no. 3, pp. 48–58 (In Russian). doi:10.31799/1684-8853-2020-3-48-58

References

- Ashimov A. A., Geida A. S., Lysenko I. V., Yusupov R. M. System functioning efficiency and other system operational properties: research problems, evaluation method. *SPIIRAS Proceedings*, 2018, no. 5(60), pp. 241–270 (In Russian). doi: [10.15622/sp.60.9](https://doi.org/10.15622/sp.60.9)
- Sokolov B. V., Burakov V. V., Mikoni S. V., Yusupov R. M. Methodological and methodical bases of the theory of estimation of quality of models and polymodel complexes. *Informatsiya i kosmos*, 2018, no. 3, pp. 36–43 (In Russian).
- Panella L., Hardwick G. *Model oriented system design applied to commercial aircraft secondary flight control systems*. In: Obaidat M., Zren T., Rango F. (eds). *Simulation and Modeling Methodologies, Technologies and Applications. SIMULTECH 2017. Advances in Intelligent Systems and Computing*. Springer, Cham, 2019, vol. 873. Pp. 55–76.
- Mikoni S. V., Sokolov B. V., Jusupov R. M. *Kvalimetriya modelej i polimodel'nyh kompleksov* [Qualimetry of models and polymodel complexes]. Moscow, Rossijskaya akademiya nauk Publ., 2018. 314 p. (In Russian).
- Doronina Yu. V., Skatkov A. V. Statistical stability analysis of stationary Markov models. *SPIIRAS Proceedings*, 2019, no. 5(18), pp. 1119–1148 (In Russian). doi: [10.15622/sp.2019.18.5.1119-1148](https://doi.org/10.15622/sp.2019.18.5.1119-1148)
- Degiannakis S., Floros C. *Methods of Volatility Estimation and Forecasting*. In: *Modelling and Forecasting High Frequency Financial Data*. Palgrave Macmillan, London, 2015. Pp. 58–109.
- Volkova V. N., Kozlov V. N., Mager V. E., Chernenkaya L. V. Classification of methods and models in system analysis. *Sbornik dokladov XX Mezhdunarodnoj konferencii po myagkim vychisleniyam i izmereniyam (SCM-2017)* [Reports of the XX International Conference on Soft Issues and Measurements (SCM-2017)], Saint-Petersburg, 2017, pp. 223–226 (In Russian).
- Okhtilev M. Yu., Gnidenko A. S., Alferov V. V., Salukhov V. I., Nazarov D. I. Methods and algorithms of integrated modeling of complex technical objects in dynamically changing conditions. *Proceedings of the International Scientific Conference MMET NW*, 2018, pp. 282–284.
- Dedkov V. K. Method of constructing a model for predicting the reliability of an “aging” technical object. *Trudy Mezhdunarodnogo simpoziuma “Nadezhnost’ i kachestvo”* [Pro-

- ceedings of the International Symposium “Reliability and quality”], 2009, vol. 1, pp. 366–368 (In Russian).
10. Kunil Park. *Fundamentals of Probability and Stochastic Processes with Applications to Communications*. New York, Springer, 2018. 273 p.
 11. Rabiner L. R., Wilpon J. G., and Juang B. H. A model-based connected digit recognition system using either hidden Markov models or templates. *Computer Speech and Language*, 1986, vol. 1, no. 2, pp. 167–197.
 12. Chen J., Ma C., Song D. Time to failure estimation based on degradation model with random failure threshold. *International Journal of Reliability and Safety*, 2016, vol. 10, no. 2, pp. 145–157.
 13. Goncharov A. M., Chashchin S. V., Prokhorov M. A. Approach to solving the problem of assessing the sustainable functioning of the information system at the data processing center. *T-Comm*, 2017, vol. 11, no. 4, pp. 20–25 (In Russian).
 14. Tsvirkun A. D. *Osnovy sinteza struktury slozhnykh sistem* [Fundamentals of synthesis of the structure of complex systems]. Moscow, Nauka Publ., 1982. 200 p. (In Russian).
 15. Zelentsov V. A., Potryasaev S. A., Pimanov I. J., Nemykin S. A. Creation of intelligent information flood forecasting systems based on service oriented architecture. *Advances in Intelligent Systems and Computing*, 2016, vol. 466, pp. 371–381. doi:10.1007/978-3-319-33389-2_35
 16. Skatkov A. V., Voronin D. Yu., Skatkov I. A. Features of simulation of degradation failures of primary meters of monitoring systems. *Environmental Control Systems*, 2017, vol. 7 (27), pp. 48–56 (In Russian).
 17. Kuznetsov A. V. Allocation of limited resources in a system with a stable hierarchy (on the example of prospective military communications system). *UBS*, 2017, iss. 66, pp. 68–93 (In Russian).
 18. Novikov D. A. Analytical complexity and error of solving problems of management of organizational and technical. *Automation and Remote Control*, 2018, no. 5, pp. 107–118 (In Russian).
 19. *New Frontiers in Information and Production Systems Modeling and Analysis — Incentive Mechanisms, Competence Management, Knowledge-based Production*. Series “Intelligent Systems Reference Library”. Ed. by P. Rozewsky, D. Novikov, O. Zaikin, N. Bakhtadze. Berlin, Springer, 2016. 268 p. doi:10.1007/978-3-319-23338-3
 20. Zhukova N., Vodyaho A., Levonevskiy D., Simonenko A. The method of data transformation for modeling technical objects. *Proceedings of the XXI IEEE International Conference on Soft Computing and Measurements (SCM’2018)*, 2018, pp. 450–453.
 21. Gnidenko A. S., Zelentsov V. A., Kulakov A. U. Hierarchical polymodel complex of combined planning of transport and logistics systems. *Proceedings of the International Scientific Conference MMET NW*, 2018, pp. 276–279.
 22. Kononova Y. G., Shubin M. V., Ermakov E. A. Repair of military aviation equipment in the public-private partnership. *Problemy ekonomiki i menedzhmenta*, 2016, no. 5 (57). Available at: <https://cyberleninka.ru/article/n/remont-avitsionnoy-tehniki-voennogo-naznacheniya-v-ramkah-gosudarstvenno-chastnogo-partnyorstva> (accessed 21 November 2019) (In Russian).
 23. Kiselev Yu. V., Motienko A. I., Basov O. O., Saitov I. A. Structural-functional model of intelligent infocommunication system. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2018, iss. 18, vol. 6, pp. 1034–1046 (In Russian).
 24. Baranov V. V., Tsvirkun A. D. Development control: structural analysis, problems, stability. *Automation and Remote Control*, 2018, vol. 79(10), pp. 1780–1796 (In Russian).
 25. Batenkov K. A. To the question of assessing the reliability of bipolar and multipolar networks. *Materialy Rossijskoj nauchno-tehnicheskoy konferencii “Sovremennye problemy telekommunikacij”* [Proceedings of Russian Scientific Conference “Modern Problems of Radio Electronics”], 2017, pp. 604–608 (In Russian).
 26. Nesteruk P. G. To organization of intellectual protection of the information. *SPIIRAS Proceedings*, 2009, no. 10, pp. 148–159 (In Russian). <https://doi.org/10.15622/sp.10.10>
 27. Novikov D. A. Hierarchical models of combat. *UBS*, 2012, iss. 37, pp. 25–62 (In Russian).
 28. Mesarovich M., Mako D., Takahara I. *Teoriya ierarhicheskikh mnogourovnevnykh sistem* [Theory of Hierarchical Multilevel Systems]. Moscow, Mir Publ., 1973. 344 p. (In Russian).
 29. Majedi M., Osman K. A novel architectural design model for enterprise systems: evaluating enterprise resource planning system and enterprise application integration against service oriented architecture. *Proceedings of 3rd International Conference on Pervasive Computing and Applications*, Alexandria, Egypt, IEEE Computer Society, 2008, pp. 116–121.

УВАЖАЕМЫЕ АВТОРЫ!

Научная электронная библиотека (НЭБ) продолжает работу по реализации проекта SCIENCE INDEX. После того как Вы регистрируетесь на сайте НЭБ (<http://elibrary.ru/defaultx.asp>), будет создана Ваша личная страничка, содержание которой составят не только Ваши персональные данные, но и перечень всех Ваших печатных трудов, имеющих в базе данных НЭБ, включая диссертации, патенты и тезисы к конференциям, а также сравнительные индексы цитирования: РИНЦ (Российский индекс научного цитирования), h (индекс Хирша) от Web of Science и h от Scopus. После создания базового варианта Вашей персональной страницы Вы получите код доступа, который позволит Вам редактировать информацию, помогая создавать максимально объективную картину Вашей научной активности и цитирования Ваших трудов.

UDC 004.056.5

doi:10.31799/1684-8853-2020-3-59-70

Comparison of machine learning algorithms for DDoS attack detection in SDN

D. T. Le^a, PhD, Lecturer, orcid.org/0000-0003-3735-0314, letranduc@dut.udn.vn

M. H. Dao^a, Master Student, orcid.org/0000-0002-0998-6078

Q. L. T. Nguyen^a, M. Sc., Lecturer, orcid.org/0000-0003-4578-9925

^aThe University of Danang – University of Science and Technology, Information Technology Faculty, 54 Nguyen Luong Bang, 550000, Da Nang, Vietnam

Introduction: Distributed denial-of-service (DDoS) has become a common attack type in cyber security. Apart from the conventional DDoS attacks, software-defined networks also face some other typical DDoS attacks, such as flow-table attack or controller attack. One of the most recent solutions to detect a DDoS attack is using machine learning algorithms to classify the traffic. **Purpose:** Analysis of applying machine learning algorithms in order to prevent DDoS attacks in software-defined network. **Results:** A comparison of six algorithms (random forest, decision tree, naive Bayes, support vector machine, multilayer perceptron, k -nearest neighbors) with accuracy and process time as the criteria has shown that a decision tree and naive Bayes are the most suitable algorithms for DDoS attack detection. As compared to other algorithms, they have higher accuracy, faster processing time and lower resource consumption. The main features that identify malicious traffic compared to normal one are the number of bytes in a flow, time flow, Ethernet source address, and Ethernet destination address. A flow-table attack can be detected easier than a bandwidth attack, as all the six algorithms can predict this type with a high accuracy. **Practical relevance:** Important features which play a supporting role in correct data classification facilitate the development of a DDoS protection system with a smaller dataset, focusing only on the necessary data. The algorithms more suitable for machine learning can help us to detect DDoS attacks in software-defined networks more accurately.

Keywords – DDoS, machine learning algorithms, flow-table attack, bandwidth attack.

For citation: Le D. T., Dao M. H., Nguyen Q. L. T. Comparison of machine learning algorithms for DDoS attack detection in SDN. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2020, no. 3, pp. 59–70. doi:10.31799/1684-8853-2020-3-59-70

Introduction

Nowadays, software defined network (SDN) is becoming increasingly popular due to the benefits it provides, such as scalability, flexibility, monitoring [1]. SDN architecture separates the network control from forwarding devices and enables the controller to become directly programmable. The controller processes the packets and decides whether the packets will be forwarded in the switch or dropped. Due to its centralized nature, the controller can get a global view of the network, and it helps the network administrators to adjust the network traffic flow dynamically [2]. Besides, for network components to interact with each other, several application programming interfaces were developed with this network model, typically the OpenFlow (OF) protocol [3].

However, the SDN network also faces many security threats [4]. When management becomes centralized, it will be easier for the administration, but it will also be easier to be collapsed under attacks. One of the attacks that have the most devastating effect on an SDN network is the distributed denial-of-service (DDoS) attack [5]. It is explained based on the distinct characteristics of the SDN network. In the SDN network, besides conventional DDoS attacks by taking up network resources, causing the

system to be paralyzed, we also face other types of attacks. For example, instead of attacking with a large number of large packets to occupy bandwidth, the attacker will constantly flood the network with strange packets so that the controller is forced to create new rules for these packets and write them in flow-table. Then, the table on the switch will increase until there is no more space to new rules, and as a consequence, the time to respond to each new requisition increased [6].

There have been several proposed solutions to solve this problem. For example, drop packets, block port, redirection, control bandwidth, deep packet inspection, network reconfiguration, and topology change; each solution has its advantages and disadvantages [5]. However, for the above attack mitigation techniques to be effective, SDN needs to implement effective DDoS attack detection techniques. The paper [5] introduces several methods for detecting DDoS attacks, such as using entropy [7], traffic pattern analysis [8], connection rate [9], machine learning [6, 10]. Among them, DDoS detection techniques using machine learning have received much attention in the computational intelligence community [11].

This technique is not new. There have been many studies considering the ability of machine learning to classify traffic on the SDN environment as in

[12–15]. However, due to the variety of algorithms, each machine learning algorithm (simplified here by ML-algorithms) has its own approach to the problem, maybe appropriate, maybe not, but it gives us more options to solve the problem, as well as to pick out the algorithm that best suits the goal of detecting DDoS attack.

In this paper, besides focusing on how to apply machine learning to detect DDoS in the SDN environment, we will implement six different ML-algorithms, making comparisons based on some criteria to expand the choice and finding the optimal solution. These algorithms are random forest (RF), decision tree (DT), naive Bayes (NB), support vector machine (SVM), multilayer perceptron (MLP), k-nearest neighbors (KNN), all supported by Python libraries.

Related works

In [16], Braga et al. proposed a lightweight method for DDoS attack detection based on traffic flow features. This method is implemented over a NOX-based network, where OF switches keep Flow Tables with statistics about all active flows. This system monitors NOX switches at regular intervals and uses self-organizing maps to classify the traffic as normal or malicious.

The authors in [17] introduced a deep learning based multi-vector DDoS detection system in a SDN environment. A DDoS detection system that incorporates stacked autoencoder based deep learning approach in an SDN environment was implemented. The authors evaluated its performance on a dataset that consists of normal Internet traffic and various DDoS attacks. However, as every packet has to be collected for extracting features, this approach may limit the performance of the controller in large networks.

In [18], Giotis et al. combined an OpenFlow and sFlow for anomaly detection to reduce processing overhead in native OF statistics collection. It designs a modular mechanism that permits anomaly detection and mitigation on SDN environments, including collector, anomaly detection and anomaly mitigation. It leverages the packet sampling capability of sFlow to acquire scalability improvements and to reduce the required communication between switches and OF controllers. However, as the implementation was based on flow sampling using sFlow, false-positive was quite high in attack detection.

In [19], Ashraf et al. aimed to handle intrusion and DDoS attacks in the SDN environment applying machine learning techniques. However, they only analyzed various machine learning techniques, such as support vector machine, fuzzy logic, decision tree, neural networks, and Bayesian networks

(BayesNet), which can be used to detect DDoS attacks in the networking system and no further explanation of how to detect and mitigate DDoS attack was given.

Kokila et al. in [13] explored the possibility of launching DDoS attacks and detection of DDoS using the SVM classifier. The experiments are carried out using the DARPA dataset. They suggested that the use of a support vector machine for detection of DDoS with a previously trained dataset will give the least false-positive results compared with other machine learning techniques.

Dao et al. in [20] presented a solution based on the IP-filtering technique to defeat DDoS attacks. The proposed scheme analyzes user behaviour and uses it to assign the timeouts for the flow entries. Long timeouts are used for trusted users' flows, while a short timeout is assigned for malicious ones. It works well when the attack traffic is not very massive. However, this solution drops all malicious traffic, which may be problematic for false-positive flows.

In [21], Nanda et al. propose using machine learning algorithms, trained on historical network attack data, to identify the potential malicious connections and potential attack destinations. They used four ML-algorithms: DT, BayesNet, decision table and NB to predict the host that will be attacked based on the historical data. The SDN controller uses the prediction results to define security rules to protect the potentially vulnerable hosts and restrict the access of potential attackers by blocking the entire subnet.

Our paper is motivated by Santos's paper [6], in which the authors managed to exploit different kinds of machine learning algorithms to avoid three types of DDoS attacks (controller attack, flow table attack, and bandwidth attack). However, they only focused on the typical attack type of SDN networks. In our paper, we are going to consider both conventional and typical DDoS attacks. We also add more ML-algorithms as well as modify some parameters to make a comprehensive comparison and try to find out appropriate algorithms for detecting DDoS attacks in the SDN environment.

Machine learning algorithms for DDoS detection

In this paper, we will implement six different ML-algorithms, making comparisons based on some criteria to expand the choice and finding the optimal solution. These algorithms are RF, DT, NB, SVM, MLP, KNN [22].

Decision tree. The DT is one of the classification techniques, which performs classification through a learning tree. In the tree, each node represents a feature (attribute) of a data, all branches repre-

sent the conjunctions of features that lead to classifications, and each leaf node is a class label. The unlabeled sample can be classified by comparing its feature values with the nodes of the DT. The DT has many advantages, such as intuitive knowledge expression, simple implementation, and high classification accuracy. However, due to its instability, even a small change in the training dataset can result in significant changes in the DT-model.

Random forest. The RF-algorithm, also known as random decision forest, can be used for classification and regression tasks. A RF consists of many DTs. This algorithm works well on the large training dataset and reduces instability (relative to DT). However, it has low training speed. The steps to classify a new data sample by using a RF-algorithm

are: a) put the data sample to each tree in the forest; b) each tree gives a classification result, which is the tree’s “vote”; c) the data sample will be classified into the class, which has the most votes.

k-nearest neighbors. The KNN is a supervised learning technique, where the classification of a data sample is determined based on the k nearest neighbors of that unclassified sample. The process of the KNN-algorithm is very simple: if most of the KNN belong to a specific class, the unclassified sample will be classified into that class. This algorithm is simple to implement but computationally expensive due to the distance calculation of each training data sample to classify a new sample.

Naive Bayes uses Bayesian theory that predicts the type of the unknown samples based on

■ **Table 1. Hyperparameters and ML-algorithms**

Models	Hyperparameters	Description
Decision tree	criterion	The function to measure the quality of a split
	splitter	The strategy used to choose the split at each node
	min_samples_split	The minimum number of samples required to split an internal node
	min_samples_leaf	The minimum number of samples required to be at a leaf node
Random forest	n_estimators	The number of trees in the forest
	criterion	The function to measure the quality of a split
	min_samples_split	The minimum number of samples required to split an internal node
	min_samples_leaf	The minimum number of samples required to be at a leaf node
Naive Bayes	var_smoothing	A portion of the largest variance of all features that is added to variances for calculation stability
k-nearest neighbor	n_neighbors	The number of neighbors to use
	weights	Weight function used in prediction
	leaf_size	Leaf size passed to BallTree or KDTree
	p	Power parameter for the Minkowski metric
	metric	The distance metric to use for the tree
	algorithm	Auto between: ball_tree, kd_tree, brute
Support vector machine	Kernel	Specifies the kernel type to be used in the algorithm
	Gamma	Kernel coefficient for ‘rbf’, ‘poly’ and ‘sigmoid’
	C	Regularization parameter
	Tol	Tolerance for stopping criterion
	max_iter	Hard limit on iterations within solver
Multilayer perceptron	hidden_layer_sizes	The ith element represents the number of neurons in the ith hidden layer
	activation	Activation function for the hidden layer
	solver	The solver for weight optimization
	alpha	L2 penalty (regularization term) parameter
	max_iter	The maximum number of iterations
	tol	Tolerance for optimization
	max_fun	The maximum number of loss function calls

prior probability using the training samples. The Bayesian classification model relies on statistical analysis and Bayesian theory that consists of the Bayesian learning. The NB-algorithm operates by segregating the training set into an attribute vector and a decision variable. The algorithm also assumes that every member of the attribute vector independently acts on the decision variables.

Support vector machine. SVM is another popular supervised learning method, which has been widely used in classification and pattern recognition. The basic idea of SVM is to map the input vectors into a high-dimensional feature space. This mapping is achieved by applying different kernel functions, such as linear, polynomial and radial based function (RBF). The objective of SVM is to find a separating hyperplane in the feature space to maximize the margin between different classes. The disadvantage of this algorithm is hard to train large datasets because the training is computationally expensive.

Multilayer perceptron. The MLP is a class of feedforward artificial neural network and has been widely adopted neural network for intrusion detection in conventional systems. An MLP consists of at least three layers of nodes: an input layer, a hidden layer, and an output layer. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. MLP utilizes a supervised learning technique called backpropagation for training. Its multiple layers and non-linear activation distinguish MLP from a linear perceptron. It can distinguish data that is not linearly separable.

Each algorithm has its own strengths. We will test each algorithm as well as compare them together to select the optimal algorithm for the detection of DDoS attacks in the SDN network. Table 1 shows the hyperparameters used in this experiment associated with the respective machine learning algorithm.

The studied features used to build the model for the algorithms are shown in Table 2.

The number of studied features is up to 14, which is almost all of the data that we can get from the flow-table through OpenFlow Switch. Among the above features, not all features help to detect abnormal and normal traffic classification. However, at this time, it is not known which features will play a decisive role in detecting DDoS attacks, so the models will be built based on all these features. At the end of the experiment, we can review and take out the important features table to find out which features will play a decisive role in this case. Because the characteristics of each algorithm are different, important features are particular parameters that can only be derived when studying DT and RF. However, these features still help identify the fea-

■ **Table 2.** Description of studied features

№	Features	Description
0	Byte_count	Number of bytes in a flow
1	Cookie	Opaque controller-issued identifier
2	Eth_src	Ethernet source address
3	Eth_dst	Ethernet destination address
4	Duration_nsec	Time flow has been alive in nanoseconds
5	Duration_sec	Time flow has been alive in seconds
6	Hard_timeout	Max time before discarding (seconds)
7	Idle_timeout	Idle time before discarding (seconds)
8	In_port	Port ID
9	Max_len	Max length to send to the controller
10	Packet_count	Number of packets in the flows
11	Priority	The priority level of a flow entry
12	Port	Output port
13	Table_id	The ID of the table to put the flow in
14	Type	Type of action

tures needed to reduce data in building models of other algorithms.

Goals and implementation plan

Goals

The primary purpose of our experiment is to find ways to apply machine learning to detect DDoS attacks in SDN networks. Besides, another goal that we are aiming at in this paper is to compare different ML-algorithms as a solution to the problem, because each algorithm has its own characteristics.

Based on [23], the basic criteria to evaluate a model in detecting abnormal traffic such as DDoS include accuracy, data quality, correctness, and efficiency. In this paper, since all the tests take place in a simulation environment, and there is always a difference between simulation data and actual collected data, we will not perform an evaluation based on data quality.

Usually, with classification problems for any model of machine learning, accuracy criterion is a suitable criterion for evaluation. It indicates how much percentage of a model's accuracy is rated, which makes it easy to visualize. For processing

time, this is the time for a model to classify a flow from input into normal traffic or abnormal traffic. In other words, the accuracy and processing time criteria will represent the efficacy and efficiency of the machine learning model, respectively. Also, to evaluate the algorithms more objectively, we rely on information from the receiver operating characteristic (ROC) curve (correctness) [6] to be able to choose the suitable model.

Planning

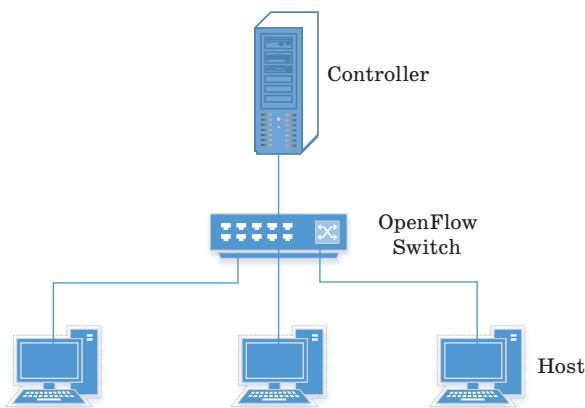
Network architecture

The entire experiment is carried out on Ubuntu 18.04 virtual machine VMware with hardware configuration Core i5-5200U (2.2 GHz, 4 cores, 4 processors), 2 GB of RAM, and 40 GB hard drive space. Mininet VN (version 2.3.0d6) is used for creating the SDN network with an RYU controller (version 4.32), one OpenFlow Switch and three hosts, as shown in Fig. 1.

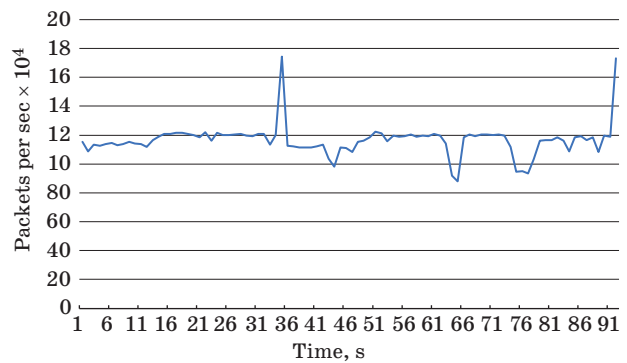
Methods of attacks

In addition to conventional attacks to a host or a group of hosts such as UDP flood, ping flood or smurf attack (collectively called bandwidth attacks), SDN network can also be attacked by new DDoS attack types due to its own structural characteristics such as controller attack, flow-table attack [6]. Therefore, to be able to study objectively and more fully, we will try the bandwidth attack and the flow-table attack in SDN at the same time.

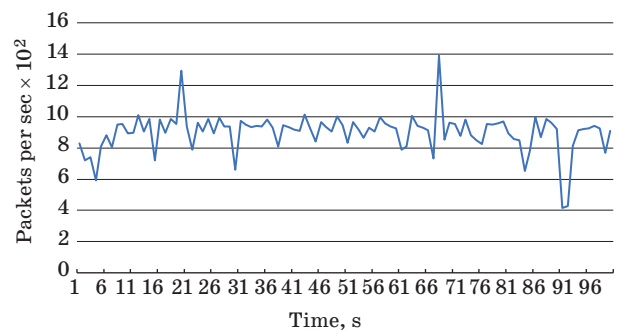
For flow-table attack, we will use Scapy tool to continuously send packets from different addresses (there are 20,000 randomly generated Ips saved in a file) to the attacked device. When the switch receives this type of packet, it creates a new rule and adds to its flow-table. As the number of incoming packets grows, the number of entries in flow-table increases and leads to overloading, causing a delay in responding to other requests from the controller.



■ Fig. 1. SDN network architecture for the experiment



■ Fig. 2. Bandwidth flooding traffic



■ Fig. 3. Normal traffic

For bandwidth attack, we will focus on taking up the network bandwidth by continuously flooding large packets (256–512 KB) to hosts in the network. We will combine many different types, including ICMP flood, TCP SYN flood and UDP flood. We use the hping3 tool to get the best results. Unlike flow-table attack when it only focuses on increasing the number of entries in the flow-table as quickly as possible, for this attack method instead of focusing on the number of attacker machines, we focus on packet-flood rate (pps — packets per second) and packet size to take up the network bandwidth.

Figure 2 shows the average network traffic under a DDoS attack with bandwidth attack (pps ~ 120,000).

We can see a huge difference comparing to normal traffic shown in Fig. 3. Normal traffic is simulated using Scapy. UDP, HTTP, ICMP packets are continuously created for sending inside the network, generating reasonable traffic (pps ~ 1.000).

Experiment Execution

Data Collection

To collect the necessary data from the switch’s flow-table, we built a separate module. This module is responsible for reading entries in the flow-table every second, recording the required information into a data file, and labelling it.

We then use `hping3` from one of the three hosts inside the network (see Fig. 1) to create a bandwidth attack on the remaining hosts. Similarly, we use `Scapy` to create flow-table attack as well as normal traffic.

In the end, the data collected is 7500 data for each type (2500 data of the dataflow table for bandwidth attack, 2500 for flow-table attack, and 2500 for normal traffic). Based on this data set, we will create two separate datasets: train dataset and test dataset for the next process.

Building model

We use the training dataset prepared above to build the models. After that, we will check with the test dataset to get the best results. To avoid model overfitting (especially for NB, KNN, SVM, and MLP), standardization of the values of the features was applied to the data using *StandardScaler* in *scikit-learn* [24].

Next is the process of tuning hyperparameters. We use the *GridSearch* technique (from *Sklearn* Library) to find the best hyperparameters set. It helps to build a suitable model that is highly effective.

After having obtained a reasonable hyperparameter set, during the next training period, we use cross-validation to avoid algorithms' overfitting with the training dataset. Specifically, we use *StratifiedKfold* [24] with ten folds and then evaluate the returned results, from which the conclusion is made.

Results

After the process of tuning hyperparameter, we obtain the following parameters, as in Table 3.

The main objective of this study is to apply machine learning to detect DDoS attacks, compare algorithms, and build a model that can classify as many types of traffic as possible. Therefore, from the initial data (7500 data of dataflow table, 2500 of each type), we will create six different datasets, including train/test datasets for normal traffic and bandwidth attack traffic (ratio 1:1); train/test datasets for normal traffic and flow-table attack traffic (ratio 1:1); finally, train/test datasets for all three traffic types at once (ratio 1:1:1).

To evaluate the accuracy of the ML-algorithms for each attack simulated, we use the following formula [6]:

$$\text{Accuracy} = \frac{\text{Number of correct classifications}}{\text{Total of samples}}$$

The accuracy is a statistical value that determines how close our ML-algorithm is to the ideal. If

■ **Table 3.** The best hyperparameters set for each ML-algorithm

Model	Hyperparameter	Value
Random forest	n_estimators	10
	criterion	Gini
	min_samples_split	2
	min_samples_leaf	1
Decision tree	Criterion	Gini
	Splitter	Best
	min_samples_split	2
	min_samples_leaf	1
Naive Bayes	var_smoothing	1e-9
k-nearest neighbors	n_neighbors	3
	weights	Uniform
	leaf_size	30
	p	2
	metric	Minkowski
	algorithm	Auto
Support vector machine	Kernel	Rbf
	Gamma	Auto
	C	1e+5
	Tol	1e-3
	Max_iter	-1
Multilayer perceptron	hidden_layer_sizes	(5,)
	activation	Relu
	solver	Lbfgs
	alpha	1e-3
	max_iter	2000
	Tol	1e-4
	max_fun	2000

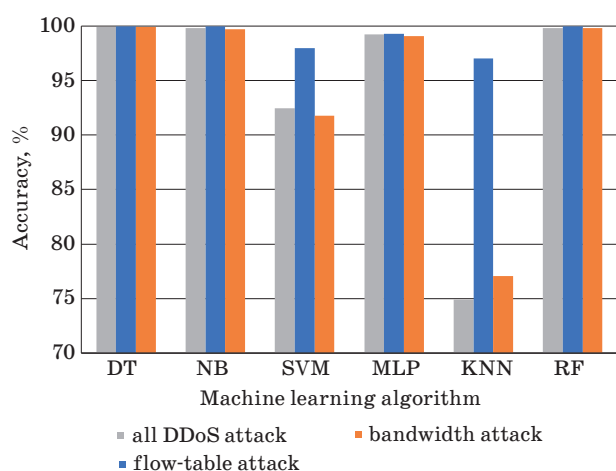
this value is 1 (100 %), it means that the algorithm has no error and classifies the data perfectly.

Figure 4 shows a comparison of the accuracy of ML-algorithms.

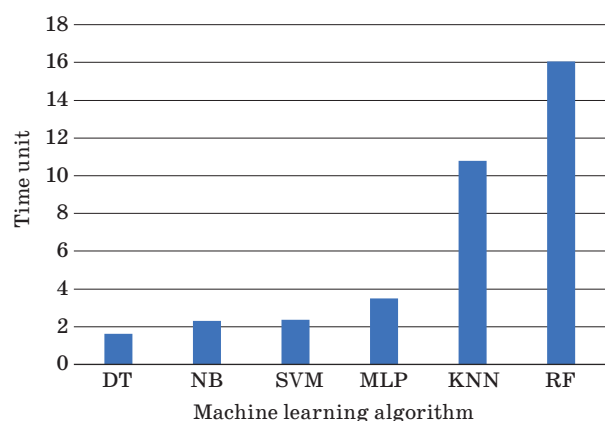
It is easy to realize that for the current problem, the lazy learner algorithm — KNN is entirely inappropriate. It gives results with low accuracy for bandwidth attack. For flow-table attack, although the results are quite good, it is inferior to the remaining algorithms.

The two algorithms SVM and MLP are useful algorithms for this problem with high accuracy. However, SVM still has many errors in the classification of bandwidth attack.

The other three algorithms are RF, DT, and NB are excellent algorithms with an almost absolute precise classification capability.



■ Fig. 4. Accuracies of ML-algorithms for each DDoS attack



■ Fig. 5. Relative time to process

Besides accuracy, we also use process time for comparison. Fig. 5 presents this comparison.

As we can see, the two algorithms RF and KNN, take too much time to process compared to other algorithms. Among the remaining four algorithms, although the DT has a faster processing speed than NB, SVM, and MLP, all four algorithms show that they are consistent with the traffic classification.

Besides, in this paper, we also build ROC curve graphs in the model evaluation phase by cross-validation for each algorithm in each attack.

The ROC curve represents a relation between true positive rate — represented by the percentage of data classified as malicious that is really malicious and false positive rate — the percentage of data classified as normal, but that is malicious. This curve is very used in the machine learning to choose a good point for the classifiers, given by the point above the central curve in which the distance between them is maximum.

To determine a good model, we need to consider the shape of the curve as well as the rate of false

prediction and the rate of omission depending on the characteristics of each specific case.

In the graphs, we also show the area under curve (AUC) metric, that is, the area under the curve. When this metric is higher, then the classification is better. Fig. 6, *a-f* presents the ROC curve for all algorithms.

Analysis and interpretation

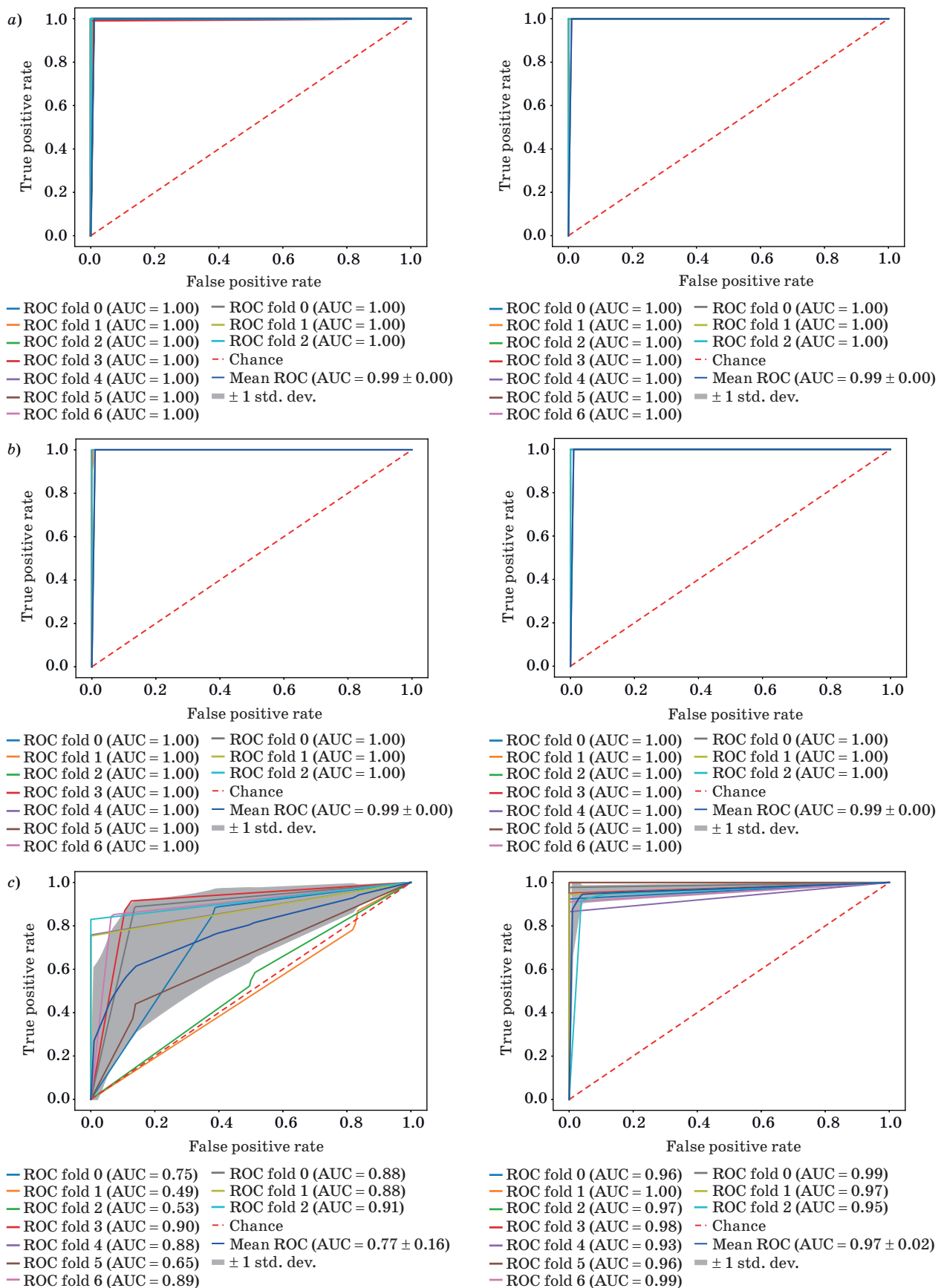
Based on the results we have obtained above, we can say that the type of algorithm lazy learner — KNN is not appropriate for DDoS detection. Because the accuracy prediction rate is low, and it takes a lot of processing time. The reason may be due to the similarity between the traffics at the start of the attack. It leads to misjudging the results because the evaluation is based on nearby neighbors. At the same time, for KNN, the process to make predictions always takes place when new data is received, meaning it requires a longer time to calculate and produce results.

The two algorithms, SVM and MLP, are useful algorithms, capable of applying in detecting DDoS attack with high accuracy and the processing speed is not slow, which is at an average level. However, looking at ROC curve with cross-validation, we found that there are still quite large errors for some data groups; this is relatively understandable because, for SVM, when the noise appears, the hyper-plane cannot divide the data exactly, but most are still acceptable.

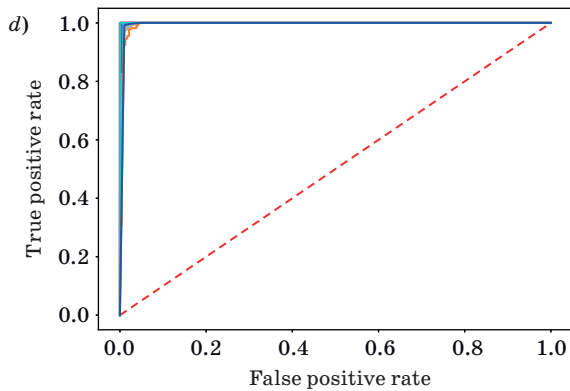
The other three algorithms, RF, DT, and NB, are all excellent algorithms with almost exact classification ability, showing the suitability of these algorithms for the classification of anomalies traffic and normal traffic. Nevertheless, although the RF has almost absolute accuracy, it takes a lot of time to process. The number of trees (10 trees) explains this. Each tree consumes a particular time, resulting in significantly increased time. Therefore, when we need high processing speed and low resource consumption, RF will not be appreciated as DT and NB algorithms.

In contrast to RF, both DT and NB algorithms have very fast processing speed. For the DT-algorithm, the fast processing speed is explained by its advantage. After training to build a decision, the next classification of this algorithm will not need much calculation. For NB, this is always considered an easy algorithm to implement and train even with small data sets. NB is a lightweight algorithm, but the results are still very good.

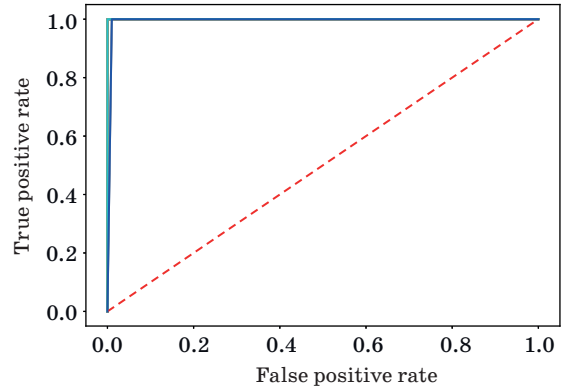
With such results, DT and NB are suitable algorithms for the problem of detecting DDoS attacks in the SDN network.



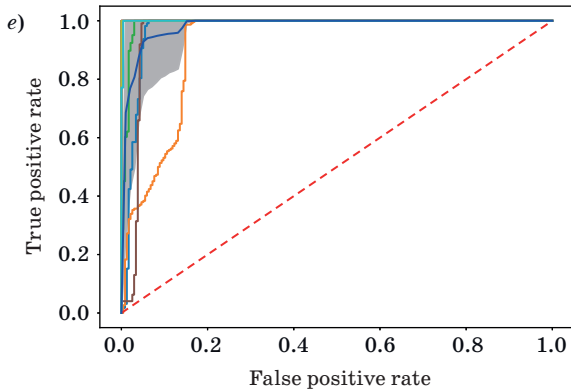
■ Fig. 6. ROC curve for DT-algorithm (a); RF-algorithm (b); KNN-algorithm (c); MLP-algorithm (d); SVM-algorithm (e);



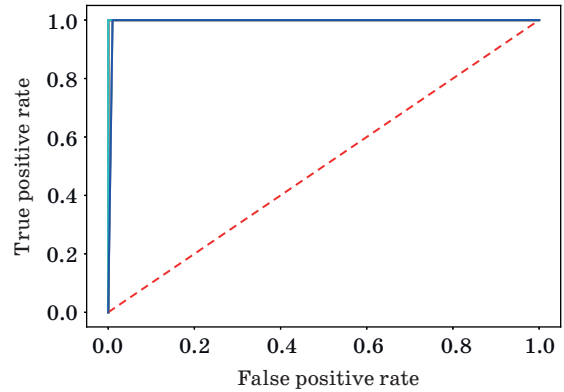
ROC fold 0 (AUC = 0.75) ROC fold 0 (AUC = 0.88)
 ROC fold 1 (AUC = 0.49) ROC fold 1 (AUC = 0.88)
 ROC fold 2 (AUC = 0.53) ROC fold 2 (AUC = 0.91)
 ROC fold 3 (AUC = 0.90) Chance
 ROC fold 4 (AUC = 0.88) Mean ROC (AUC = 0.77 ± 0.16)
 ROC fold 5 (AUC = 0.65) ± 1 std. dev.
 ROC fold 6 (AUC = 0.89)



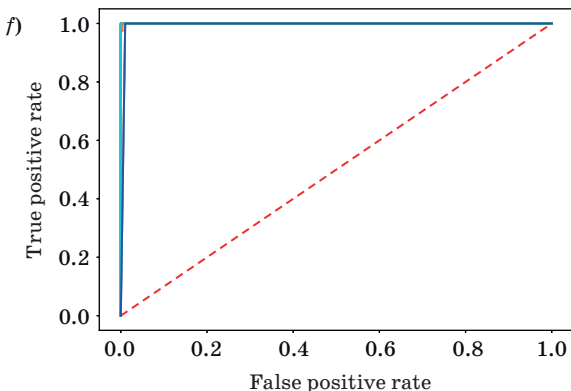
ROC fold 0 (AUC = 0.96) ROC fold 0 (AUC = 0.99)
 ROC fold 1 (AUC = 1.00) ROC fold 1 (AUC = 0.97)
 ROC fold 2 (AUC = 0.97) ROC fold 2 (AUC = 0.95)
 ROC fold 3 (AUC = 0.98) Chance
 ROC fold 4 (AUC = 0.93) Mean ROC (AUC = 0.97 ± 0.02)
 ROC fold 5 (AUC = 1.00) ± 1 std. dev.
 ROC fold 6 (AUC = 0.96)



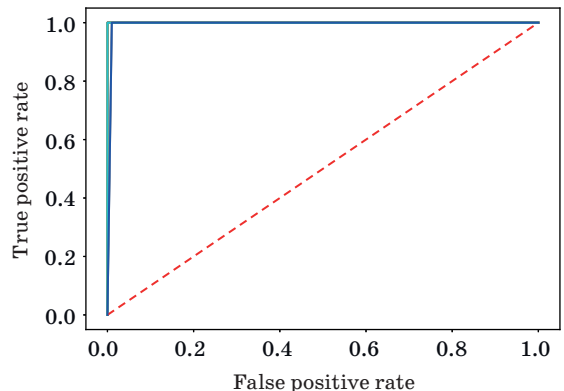
ROC fold 0 (AUC = 0.97) ROC fold 0 (AUC = 1.00)
 ROC fold 1 (AUC = 0.92) ROC fold 1 (AUC = 1.00)
 ROC fold 2 (AUC = 0.99) ROC fold 2 (AUC = 1.00)
 ROC fold 3 (AUC = 1.00) Chance
 ROC fold 4 (AUC = 1.00) Mean ROC (AUC = 0.98 ± 0.03)
 ROC fold 5 (AUC = 0.96) ± 1 std. dev.
 ROC fold 6 (AUC = 1.00)



ROC fold 0 (AUC = 1.00) ROC fold 0 (AUC = 1.00)
 ROC fold 1 (AUC = 1.00) ROC fold 1 (AUC = 1.00)
 ROC fold 2 (AUC = 1.00) ROC fold 2 (AUC = 1.00)
 ROC fold 3 (AUC = 1.00) Chance
 ROC fold 4 (AUC = 1.00) Mean ROC (AUC = 0.99 ± 0.00)
 ROC fold 5 (AUC = 1.00) ± 1 std. dev.
 ROC fold 6 (AUC = 1.00)



ROC fold 0 (AUC = 1.00) ROC fold 0 (AUC = 1.00)
 ROC fold 1 (AUC = 1.00) ROC fold 1 (AUC = 1.00)
 ROC fold 2 (AUC = 1.00) ROC fold 2 (AUC = 1.00)
 ROC fold 3 (AUC = 1.00) Chance
 ROC fold 4 (AUC = 1.00) Mean ROC (AUC = 0.99 ± 0.00)
 ROC fold 5 (AUC = 1.00) ± 1 std. dev.
 ROC fold 6 (AUC = 1.00)

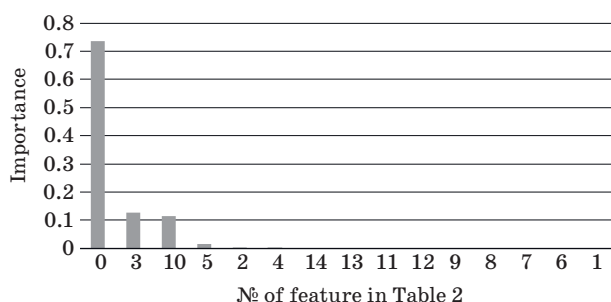


ROC fold 0 (AUC = 1.00) ROC fold 0 (AUC = 1.00)
 ROC fold 1 (AUC = 1.00) ROC fold 1 (AUC = 1.00)
 ROC fold 2 (AUC = 1.00) ROC fold 2 (AUC = 1.00)
 ROC fold 3 (AUC = 1.00) Chance
 ROC fold 4 (AUC = 1.00) Mean ROC (AUC = 0.99 ± 0.00)
 ROC fold 5 (AUC = 1.00) ± 1 std. dev.
 ROC fold 6 (AUC = 1.00)

NB-algorithm (f) under bandwidth attack (left) and flow-table attack (on right)

In terms of attack type, the data shows that flow-table attack is a more recognizable type of attack than a bandwidth attack using the above algorithms.

We also try to check the features, which play a supporting role (important features) to the process of correctly classifying data for the DT-algorithm. Features that play the essential role are *byte_count*, *duration_sec*, *packet_count*, and some other features. It is also to be expected because, with DDoS attacks, the large amount of incoming packets makes *packet_count* change a lot compared to nor-



■ Fig. 7. Importance of each feature

■ Table 4. Important features for classification

Position	Feature	Description
1 st	Byte_count	Number of bytes in a flow
2 nd	Duration_sec	Time flow has been alive in seconds
3 rd	Packet_count	Number of packets in the flows
4 th	Eth_src	Ethernet source address
5 th	Duration_nsec	Time flow has been alive in nanoseconds
6 th	Eth_dst	Ethernet destination address

mal traffic's *packet_count*. Same with *byte_count*. However, most of the features that we collect from OpenFlow Switch are almost useless. The result is shown in Fig. 7.

Table 4 presents the features according to its importance for the detection of the anomalies.

Conclusion

The benefits of SDN network to overcome the drawbacks of a traditional network model are undisputed, but there are also certain limitations. For example, the entire system will collapse if the controller receives a DDoS attack and cannot respond to other valid requests. But based on what we have tested above and the results we have obtained, the application of ML-algorithms to detect these DDoS attacks is entirely possible, and gives very good results.

We tested six different ML-algorithms including RF, NB, KNN, SVM, MLP, and DT, to classify different types of traffic, including normal traffic, bandwidth attack traffic, and flow-table attack traffic. We have proved that DT, as well as NB, are suitable algorithms for DDoS attack detection (high accuracy and fast processing time, consume less resource compared to other algorithms).

Besides, we also pointed out that the main features that identify malicious traffic compared to normal traffic. It will make it easier to build a DDoS protection system with a more compact dataset, focusing only on the data needed.

Furthermore, we realized that flow-table attack is a more easily discovered attack than Bandwidth attack, as all six algorithms can predict this type with high accuracy. The efficiency of bandwidth attack detection is lower, so we need to focus more on this type to improve the predictive results.

In future work, we will focus on data quality criteria by comparing the results of detection between simulation dataset and real dataset.

References

1. Kreutz D., Ramos F. M. V., Verissimo P. E., Rothenberg C. E., Azodolmolky S., & Uhlig S. Software-defined networking: A comprehensive survey. *Proceedings of the IEEE*, 2015, no. 103(1), pp. 14–76. doi:10.1109/jproc.2014.2371999
2. Farhady H., Lee H. & Nakao A. Software-defined networking. *Computer Networks*, 2015, no. 81(81), pp. 79–95. doi:10.1016/j.comnet.2015.02.014
3. McKeown N., Anderson T., Balakrishnan H., Parulkar G., Peterson L., Rexford J., Shenker S. & Turner J. OpenFlow: Enabling innovation in campus networks. *Acm Special Interest Group on Data Communication*, 2008, no. 38(2), pp. 69–74. doi:10.1145/1355734.1355746
4. Ahmad Ijaz, Namal S., Ylianttila M. & Gurtov A. Security in software defined networks: A Survey. *IEEE Communications Surveys and Tutorials*, 2015, no. 17(4), pp. 2317–2346. doi:10.1109/COMST.2015.2474118
5. Bawany N. Z., Shamsi J. A. & Salah K. DDoS attack detection and mitigation using SDN: Methods, practices, and solutions. *Arabian Journal for Science and Engineering*, 2017, no. 42(2), pp. 425–441. doi:10.1007/s13369-017-2414-5
6. Santos R., Souza D., Santo W., Ribeiro A. & Moreno E. Machine learning algorithms to detect DDoS at-

- tacks in SDN. *Concurrency and Computation: Practice and Experience*, 2019, e5402. doi:10.1002/cpe.5402
7. Swami R., Dave M. & Ranga V. Defending DDoS against software defined networks using entropy. *2019 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU)*, 2019, pp. 1–5. doi:10.1109/IoT-SIU.2019.8777688
 8. Yen T. F. & Reiter M. K. Traffic aggregation for malware detection. *International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment*, 2008, pp. 207–227. doi:10.1007/978-3-540-70542-0_11
 9. Mehdi S. A., Khalid J. & Khayam S. A. Revisiting traffic anomaly detection using software defined networking. *International Workshop on Recent Advances in Intrusion Detection*, 2011, pp. 161–180. doi:10.1007/978-3-642-23644-0_9
 10. Elsayed M. S., Le-Khac N., Dev S., & Jurcut A. D. Machine-learning techniques for detecting attacks in SDN. *2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT)*, 2019, pp. 277–281. doi:10.1109/ICCSNT47585.2019.8962519
 11. Xiaoqiong X., Hongfang Y. & Kun Y. DDoS attack in software defined networks: a survey. *ZTE Communications*, 2017, no. 15(3), pp. 13–19.
 12. Nam T. M., Phong P. H., Khoa T. D., Huong T. T., Nam P. N., Thanh N. H., Thang L. X., Tuan P. A. & Loi V. D. Self-organizing map-based approaches in DDoS flooding detection using SDN. *2018 International Conference on Information Networking (ICOIN)*, 2018, pp. 249–254. doi:10.1109/icoin.2018.8343119
 13. Kokila R. T., Selvi S. T. & Govindarajan K. DDoS detection and analysis in SDN-based environment using support vector machine classifier. *2014 Sixth International Conference on Advanced Computing (ICoAC)*, 2014, pp. 205–210. doi:10.1109/icoac.2014.7229711
 14. Yang L. & Zhao H. DDoS attack identification and defense using SDN based on machine learning method. *2018 15th International Symposium on Pervasive Systems, Algorithms and Networks (I-SPAN)*, 2018, pp. 174–178. doi:10.1109/i-span.2018.00036
 15. Deepa V., Sudar K.M. & Deepalakshmi P. Detection of DDoS attack on SDN control plane using hybrid machine learning techniques. *2018 International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 2018, pp. 299–303. doi:10.1109/ics-sit.2018.8748836
 16. Braga R., Mota E. & Passito A. Lightweight DDoS flooding attack detection using NOX/OpenFlow. *IEEE Local Computer Network Conference*, 2010, pp. 408–415. doi:10.1109/LCN.2010.5735752
 17. Niyaz Q., Sun W., & Javaid A. Y. A deep learning based DDoS detection system in Software-Defined Networking (SDN). *ICST Transactions on Security and Safety*, 2017, no. 4(12), p. 153515. doi:10.4108/eai.28-12-2017.153515
 18. Giotis K., Argyropoulos C., Androulidakis G., Kalogeras D. & Maglaris V. Combining OpenFlow and sFlow for an effective and scalable anomaly detection and mitigation mechanism on SDN environments. *Computer Networks*, 2014, vol. 62, pp. 122–136. doi:10.1016/j.bjp.2013.10.014
 19. Javed A. & Latif S. Handling intrusion and DDoS attacks in Software Defined Networks using machine learning techniques. *2014 National Software Engineering Conference*, 2014, pp. 55–60. doi:10.1109/nsec.2014.6998241
 20. Dao N. N., Park J., Park M. & Cho S. A feasible method to combat against DDoS attack in SDN network. *2015 International Conference on Information Networking (ICOIN)*, 2015, pp. 309–311. doi:10.1109/icoin.2015.7057902
 21. Nanda S., Zafari F., DeCusatis C., Wedaa E. & Yang B. Predicting network attack patterns in SDN using machine learning approach. *2016 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN)*, 2016, pp. 167–172. doi:10.1109/NFV-SDN.2016.7919493
 22. Xie J., Yu F. R., Huang T., Xie R., Liu J., Wang C. & Liu Y. A survey of machine learning techniques applied to software defined networking (SDN): Research issues and challenges. *IEEE Communications Surveys & Tutorials*, no. 21(1), 2018, pp. 393–430. doi:10.1109/comst.2018.2866942
 23. Bhuyan M. H., Bhattacharyya D. K. & Kalita J. K. *Network Traffic Anomaly Detection and Prevention: Concepts, Techniques, and Tools*. Springer, 2017. 263 p. doi:10.1007/978-3-319-65188-0
 24. Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V. & Vanderplas J. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 2011, no. 12, pp. 2825–2830.

УДК 004.056.5

doi:10.31799/1684-8853-2020-3-59-70

Сравнение алгоритмов машинного обучения при обнаружении DDoS-атак в программно-определяемых сетяхД. Ч. Ле^а, канд. техн. наук, лектор, orcid.org/0000-0003-3735-0314, letranduc@dut.udn.vnМ. Х. Дао^а, магистрант, orcid.org/0000-0002-0998-6078К. Л. Т. Нгуэн^а, магистр, лектор, orcid.org/0000-0003-4578-9925^аУниверситет Дананга — Университет науки и техники, факультет информационных технологий, 54, Нгуэн Лунг Банг, Дананг, 550000, Вьетнам

Введение: распределенная атака типа «отказ в обслуживании» (DDoS) стала популярным типом атак в кибербезопасности. Помимо обычных DDoS-атак, программно-определяемые сети сталкиваются с некоторыми другими типичными DDoS-атаками, такими как атака с использованием потоковой таблицы и атака контроллера. Одним из самых последних решений для обнаружения DDoS-атак является использование алгоритмов машинного обучения для классификации трафика. **Цель:** анализ применения алгоритмов машинного обучения для предотвращения DDoS-атак программно-определяемых сетей. **Результаты:** сравнение шести алгоритмов (случайный лес, дерево решений, наивный байесовский метод, машина опорных векторов, многослойный перцептрон, k-ближайшие соседи) по критериям точность и время обработки показало, что дерево решений, как и наивный байесовский, являются лучшими алгоритмами для обнаружения DDoS-атак (высокая точность и быстрое время обработки, меньшее потребление ресурсов по сравнению с другими алгоритмами). Указаны и проанализированы основные функции, которые идентифицируют вредоносный трафик по сравнению с обычным трафиком: количество байтов в потоке, поток времени, Ethernet-адрес источника, Ethernet-адрес назначения. По результатам исследований сделан вывод, что атака с использованием таблицы потоков является более легкой для обнаружения, чем атака по пропускной способности. **Практическая значимость:** основные функции, которые играют вспомогательную роль в процессе правильной классификации данных, облегчают создание системы защиты от DDoS-атак с более компактным набором данных, включающим только необходимые данные. Алгоритмы, которые более подходят для машинного обучения, помогут точнее обнаруживать DDoS-атаки в программно-определяемых сетях.

Ключевые слова — DDoS, алгоритмы машинного обучения, атака по таблице потоков, атака по пропускной способности.

Для цитирования: Le D. T., Dao M. H., Nguyen Q. L. T. Comparison of machine learning algorithms for DDoS attack detection in SDN. *Информационно-управляющие системы*, 2020, № 3, с. 59–70. doi:10.31799/1684-8853-2020-3-59-70

For citation: Le D. T., Dao M. H., Nguyen Q. L. T. Comparison of machine learning algorithms for DDoS attack detection in SDN. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2020, no. 3, pp. 59–70. doi:10.31799/1684-8853-2020-3-59-70

ПАМЯТКА ДЛЯ АВТОРОВ

Поступающие в редакцию статьи проходят обязательное рецензирование.

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

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

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

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

Протоколы слепой цифровой подписи на основе скрытой задачи дискретного логарифмирования

Д. Н. Молдовян^а, канд. техн. наук, научный сотрудник, orcid.org/0000-0001-5039-7198, mdn.spectr@mail.ru

А. А. Молдовян^а, доктор техн. наук, профессор, orcid.org/0000-0001-5480-6016

Д. Ю. Гурьянов^б, канд. техн. наук, доцент, orcid.org/0000-0002-2923-4965

^аСанкт-Петербургский институт информатики и автоматизации РАН, 14-я линия В.О., 39, Санкт-Петербург, 199178, РФ

^бГосударственный университет морского и речного флота им. адмирала С. О. Макарова, Двинская ул., 5/7, Санкт-Петербург, 198035, РФ

Введение: существенный прогресс в развитии квантовых вычислений выдвинул проблему построения постквантовых двухключевых криптографических алгоритмов и протоколов, т. е. криптосхем, которые были бы стойкими к атакам с использованием квантовых компьютеров. На основе скрытой задачи дискретного логарифмирования разработаны практические постквантовые схемы цифровой подписи. Представляет интерес разработка на ее основе протоколов слепой подписи. **Цель:** разработка протоколов слепой подписи на основе вычислительной трудности скрытой задачи дискретного логарифмирования. **Метод:** применение ослепляющих множителей, которые вносятся клиентом в ходе протокола слепой подписи при передаче подписанту параметров, необходимых для формирования слепой подписи. **Результаты:** предложен способ внесения ослепляющих множителей двух типов, левых и правых, что обеспечивает возможность реализовать протоколы слепой подписи на основе алгоритмов цифровой подписи с проверочным уравнением, задаваемым в некоммутативных алгебраических структурах. С применением этого способа разработаны новые протоколы слепой подписи, основанные на вычислительной трудности скрытой задачи дискретного логарифмирования. В качестве алгебраического носителя разработанных протоколов использованы конечные некоммутативные ассоциативные алгебры двух типов: с глобальной двухсторонней единицей и с большим множеством глобальных левых единиц. **Практическая значимость:** предложенные протоколы обладают достаточно высокой производительностью и удобны для программной и аппаратной реализации.

Ключевые слова — постквантовые криптосхемы, компьютерная безопасность, защита информации, электронная цифровая подпись, слепая подпись, задача дискретного логарифмирования, некоммутативные ассоциативные алгебры, вычислительно сложная задача.

Для цитирования: Молдовян Д. Н., Молдовян А. А., Гурьянов Д. Ю. Протоколы слепой цифровой подписи на основе скрытой задачи дискретного логарифмирования. *Информационно-управляющие системы*, 2020, № 3, с. 71–78. doi:10.31799/1684-8853-2020-3-71-78

For citation: Moldovyan D. N., Moldovyan A. A., Gurianov D. Yu. Blind signature protocols based on hidden discrete logarithm problem. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2020, no. 3, pp. 71–78 (In Russian). doi:10.31799/1684-8853-2020-3-71-78

Введение

В связи с ожиданием появления в ближайшее десятилетие практически доступного квантового компьютера, для которого известны полиномиальные алгоритмы решения задачи факторизации [1, 2] и задачи дискретного логарифмирования (ЗДЛ) [3, 4], в последнее время активно ведутся исследования по разработке постквантовых двухключевых криптосхем [5–7], т. е. криптографических алгоритмов и протоколов с открытым ключом, которые являются вычислительно эффективными при их реализации на обычных компьютерах и имеют сверхполиномиальную стойкость к атакам с использованием квантовых компьютеров.

Одним из направлений в области постквантовой криптографии является разработка постквантовых двухключевых криптосхем на основе

вычислительной трудности скрытой задачи дискретного логарифмирования (СЗДЛ) [8, 9]. Это направление представляет практический интерес не только для применения в постквантовую эру, но также и в настоящее время, поскольку криптосхемы данного типа имеют сравнительно высокую производительность и могут составить конкуренцию криптосхемам, основанным на вычислительной трудности ЗДЛ на эллиптической кривой.

В зависимости от типа разрабатываемой криптосистемы с открытым ключом применяются разные формы СЗДЛ. Различные версии протоколов согласования секретного ключа по открытому каналу, основанных на вычислительной трудности СЗДЛ, рассматриваются в работах [8–10]. Алгоритм открытого шифрования описан в [9], а схемы электронной цифровой подписи (ЭЦП) — в работах [11–13]. В ряде важных практических

приложений (например, системах тайного электронного голосования и электронных денег) уникальную роль играют протоколы слепой ЭЦП, позволяющие решать задачу обеспечения неотслеживаемости (анонимности) пользователей. В связи с этим значительный интерес представляет разработка протоколов слепой ЭЦП, основанных на вычислительной сложности СЗДЛ.

В данной статье представлены два протокола слепой цифровой подписи, разработанные с использованием известных схем ЭЦП, основанных на вычислительной сложности СЗДЛ. Для построения протоколов слепой подписи использован известный механизм [14, 15] внесения и удаления ослепляющих множителей в схемах ЭЦП, основанных на ЗДЛ в циклической конечной группе. Однако указанный механизм был модифицирован с учетом особенностей СЗДЛ, формулируемой в конечных некоммутативных ассоциативных алгебрах (КНАА). Новые внесенные элементы построения схемы слепой подписи связаны с тем, что один из двух используемых ослепляющих множителей играет роль левого операнда операции умножения, а второй — роль правого операнда.

Скрытая задача дискретного логарифмирования

Типовая ЗДЛ возникает при построении криптосхем с открытым ключом в конечных циклических группах простого порядка ω , в которых открытый ключ Y' вычисляется по формуле

$$Y' = G^x, \quad (1)$$

где G — элемент, являющийся генератором группы и имеющий порядок ω ; требуется вычислить неизвестное целое число $0 < x < \omega$, которое называется личным секретным ключом пользователя, являющегося владельцем открытого ключа Y' , т. е. пользователя, который выбрал случайное число x и по нему вычислил значение Y' .

Для задания СЗДЛ требуется использовать конечные алгебраические структуры, в которых содержится очень большое число различных конечных циклических групп в качестве подмножеств ее элементов, причем эти подмножества не пересекаются. Например, СЗДЛ формулируется в конечных некоммутативных группах [7] и в КНАА [9, 13]. Отличие СЗДЛ, используемой для построения схем ЭЦП, от обычной ЗДЛ состоит в том, что открытый ключ формируется в виде элементов Y и Z из двух разных циклических групп, каждая из которых отлична от группы, генерируемой всевозможными степенями элемента G . При этом в качестве маскирующих

операций, отображающих элементы Y' и G в элементы Y и Z , выбираются операции автоморфного [8] или гомоморфного [9, 12] отображения. Благодаря взаимной коммутативности таких маскирующих операций с операцией возведения в степень обеспечивается корректность процедуры проверки подлинности ЭЦП по открытому ключу (Y, Z) в случае, если схема ЭЦП в базовой циклической группе, генерируемой элементом G , работает корректно при использовании открытого ключа (Y', G) . Изучение известных протоколов ЭЦП, основанных на вычислительной сложности СЗДЛ, показывает, что они представляют собой модифицированные версии протоколов ЭЦП, основанных на вычислительной сложности ЗДЛ и построенных в явно заданной конечной циклической группе простого порядка. При этом модифицирование состоит в маскировании элементов базовой циклической группы, в которой выполняется базовая операция возведения в степень достаточно большого размера, которая вносит основной вклад в обеспечение стойкости криптосхемы. Важной задачей, имеющей теоретическое и практическое значение, является исследование вычислительной сложности решения различных форм СЗДЛ. Существенным вкладом в этом направлении являются работы [16–18], выполненные для СЗДЛ, заданной в конечной алгебре кватернионов.

Типы алгебраических носителей схем ЭЦП на основе СЗДЛ

Задание КНАА

Рассмотрим m -мерное векторное пространство над конечным простым полем $GF(p)$, где p — простое число. Его элементами являются векторы, которые могут быть записаны в виде упорядоченных наборов элементов поля $GF(p)$, т. е. в виде $\mathbf{A} = (a_0, a_1, \dots, a_{m-1})$, где a_i — координаты, или в виде всевозможных сумм однокомпонентных векторов, т. е. в виде $\mathbf{A} = a_0\mathbf{e}_0 + a_1\mathbf{e}_1 + \dots + a_{m-1}\mathbf{e}_{m-1}$, где \mathbf{e}_i — базисные векторы. В векторном пространстве определены операция сложения векторов и операция умножения вектора на скаляр. Векторное пространство с дополнительно определенной операцией векторного умножения (\circ), т. е. умножения произвольных двух векторов $\mathbf{A} = \sum_{i=0}^{m-1} a_i\mathbf{e}_i$ и $\mathbf{B} = \sum_{i=0}^{m-1} b_i\mathbf{e}_i$, которая является дистрибутивной слева и справа относительно операции сложения, называется m -мерной алгеброй.

Операция векторного умножения с указанным свойством может быть задана по следующей формуле:

$$\mathbf{A} \circ \mathbf{B} = \sum_{j=0}^{m-1} \sum_{i=0}^{m-1} a_i b_j \mathbf{e}_i \circ \mathbf{e}_j, \quad (2)$$

в которой предполагается подстановка вместо каждого произведения упорядоченных пар базисных векторов некоторого однокомпонентного вектора в соответствии с так называемой таблицей умножения базисных векторов. После такой подстановки правая часть формулы (2) в общем случае представляет собой m -мерный вектор, т. е. элемент рассматриваемого векторного пространства. При использовании таблиц умножения базисных векторов, задающих некоммутативную ассоциативную операцию умножения векторов, получаем КНАА.

Схема ЭЦП в КНАА с большим множеством глобальных левых единиц

В статье [19] описана схема ЭЦП, заданная в четырехмерной КНАА, содержащей p^2 глобальных левых единиц. Операция умножения в данной алгебре задана по таблице умножения базисных векторов (табл. 1). Множество глобальных левых единиц описывается формулой

$$\mathbf{L} = (l_0, l_1, l_2, l_3) = (d, h, 1-d, -h), \quad (3)$$

где $d, h = 0, 1, \dots, p-1$.

Каждая глобальная левая единица \mathbf{L} задает гомоморфные отображения рассматриваемой алгебры следующих двух типов [19]. Операция гомоморфного отображения первого типа определяется по формуле

$$\varphi_{\mathbf{L}}(\mathbf{X}) = \mathbf{X} \circ \mathbf{L}, \quad (4)$$

где переменная \mathbf{X} пробегает все значения в алгебре. Операция гомоморфного отображения второго типа определяется по формуле

$$\psi_{\mathbf{L}}(\mathbf{X}) = \mathbf{B} \circ \mathbf{X} \circ \mathbf{A}, \quad (5)$$

где фиксированные векторы \mathbf{A} и \mathbf{B} удовлетворяют условию $\mathbf{A} \circ \mathbf{B} = \mathbf{L}$ и переменная \mathbf{X} пробегает все значения в алгебре. Прототипом схемы ЭЦП, ос-

■ **Таблица 1.** Задание операции умножения в КНАА p^2 глобальных левых единиц [19], где μ — квадратичный невычет в $GF(p)$

■ **Table 1.** Setting the operation of multiplication in KNAA p^2 global left units [19], where μ — is a quadratic non-residue in $GF(p)$

\circ	e_0	e_1	e_2	e_3
e_0	e_0	e_1	e_2	e_3
e_1	e_1	μe_0	e_3	μe_2
e_2	e_0	e_1	e_2	e_3
e_3	e_1	μe_0	e_3	μe_2

нованной на вычислительной трудности СЗДЛ и предложенной в работе [19], является схема ЭЦП Шнорра [20]. В качестве характеристики конечного поля $GF(p)$, над которым задана КНАА, выбирается 256-битовое число вида $p = eq$, где e — небольшое четное число (например, имеющее разрядность от 2 до 16 бит) и q — простое число большого размера. Можно легко сгенерировать значение p такое, что структурный коэффициент $\mu = 2$ будет квадратичным невычетом по модулю p .

Процедура формирования открытого ключа описывается следующим образом.

1. Выбрать случайные четырехмерные векторы \mathbf{A} , \mathbf{A}' и \mathbf{N} , всевозможные степени каждого из которых генерируют циклическую группу порядка q , причем указанные три вектора являются попарно не перестановочными.

2. Используя формулу (3), вычислить две случайные глобальные левые единицы \mathbf{L} и \mathbf{L}' .

3. Вычислить вектор \mathbf{B}' как решение уравнения $\mathbf{A}' \circ \mathbf{B}' = \mathbf{L}'$.

4. Вычислить вектор \mathbf{B} как решение уравнения $\mathbf{A} \circ \mathbf{B} = \mathbf{L}$.

5. Вычислить порядок ω циклической группы, генерируемой вектором \mathbf{B}' .

6. Вычислить вектор \mathbf{T} как решение уравнения $\mathbf{A} \circ \mathbf{T} = \mathbf{B}'^{\omega-1}$.

7. Сгенерировать равновероятное неотрицательное число $x < q$ и вычислить вектор $\mathbf{Y} = \mathbf{B} \circ \mathbf{N}^x \circ \mathbf{A} \circ \mathbf{L}$.

8. Вычислить вектор $\mathbf{Z} = \mathbf{B}' \circ \mathbf{N} \circ \mathbf{A}'$.

9. Взять векторы \mathbf{Y} , \mathbf{Z} и \mathbf{T} в качестве открытого ключа.

Отметим, что векторы \mathbf{A} , \mathbf{A}' , \mathbf{B} , \mathbf{B}' , \mathbf{L} , \mathbf{L}' , \mathbf{N} и число x являются секретными элементами. Однако только значения \mathbf{B} , \mathbf{A}' , \mathbf{N} и x являются элементами личного секретного ключа, поскольку только они нужны для вычисления ЭЦП.

Вычисление значения подписи выполняется по следующему алгоритму.

1. Выбрать случайное натуральное число $k < q$ и вычислить вектор $\mathbf{V} = \mathbf{B} \circ \mathbf{N}^k \circ \mathbf{A}'$.

2. Вычислить первый элемент ЭЦП $v = F_h(M, \mathbf{V})$, где F_h — некоторая специфицированная хэш-функция.

3. Вычислить второй элемент ЭЦП $s: s = k + xv \pmod q$.

Цифровой подписью к электронному документу M является пара чисел (v, s) .

Процедура проверки подлинности подписи (v, s) включает следующие шаги.

1. Вычислить вектор $\mathbf{V}' : \mathbf{V}' = \mathbf{Y}^{-v} \circ \mathbf{T} \circ \mathbf{Z}^s$.

2. Вычислить значение хэш-функции v' от документа M , к которому присоединен вектор $\mathbf{V}' : v' = F_h(M, \mathbf{V}')$.

3. Если $v' = v$, то подпись (v, s) принимается как подлинная. В противном случае подпись отвергается.

Схема ЭЦП в КНАА с глобальной двухсторонней единицей

В работе [13] предложена схема ЭЦП, заданная в четырехмерной КНАА с глобальной двухсторонней единицей E , в которой операция векторного умножения определена по таблице умножения базисных векторов (табл. 2). Рассмотрим некоторые результаты [13], используемые при построении указанной схемы ЭЦП.

Глобальная двухсторонняя единица E может быть вычислена по формуле

$$E = \left(\frac{1}{\lambda - 1}, \frac{1}{1 - \lambda}, \frac{1}{1 - \lambda}, \frac{\lambda}{\lambda - 1} \right). \quad (6)$$

Если для некоторых векторов $A = (a_0, a_1, a_2, a_3)$ векторное уравнение вида $X \circ A = E$ имеет единственное решение $X = A^{-1}$, то векторное уравнение $A \circ X = E$ также имеет единственное решение $X = A^{-1}$. Вектор A^{-1} называется обратным по отношению к вектору A . Признаком обратимости вектора $A = (a_0, a_1, a_2, a_3)$ является неравенство вида

$$a_1 a_2 \neq a_0 a_3. \quad (7)$$

Векторы $G = (g_0, g_1, g_2, g_3)$, координаты которых удовлетворяют равенству $g_1 g_2 = g_0 g_3$, являются необратимыми. Рассматриваемая алгебра включает некоторые подмножества элементов, которые действуют на некоторый фиксированный необратимый вектор G и всевозможные его степени G^k как локальные левые единицы. Также можно выделить множества локальных правых единиц.

Множество локальных левых единиц L , соответствующих вектору G , описывается формулой [13]

$$L = (l_0, l_1, l_2, l_3) = \left(d, \frac{g_0 - (\lambda g_0 + g_2)d}{g_0 + g_2}, h, \frac{g_2 - (\lambda g_0 + g_2)h}{g_0 + g_2} \right), \quad (8)$$

■ **Таблица 2.** Умножение базисных векторов в четырехмерной КНАА с глобальной двухсторонней единицей ($\lambda \neq 1$) [13]

■ **Table 2.** Multiplication of basis vectors in four-dimensional KNAА with global two-sided unit ($\lambda \neq 1$) [13]

\circ	e_0	e_1	e_2	e_3
e_0	λe_0	λe_1	e_0	e_1
e_1	e_0	e_1	e_0	e_1
e_2	λe_2	λe_3	e_2	e_3
e_3	e_2	e_3	e_2	e_3

где $d, h = 0, 1, \dots, p - 1$. Множество (8) локальных левых единиц содержит в себе $p^2 - p$ обратимых векторов и p необратимых векторов рассматриваемой алгебры.

Множество локальных правых единиц R , соответствующих всевозможным натуральным степеням необратимого вектора G , описывается формулой [13]

$$R = (r_0, r_1, r_2, r_3) = \left(d, h, \frac{g_0 - (\lambda g_0 + g_1)d}{g_0 + g_1}, \frac{g_1 - (\lambda g_0 + g_1)h}{g_0 + g_1} \right), \quad (9)$$

где $d, h = 0, 1, \dots, p - 1$. В множестве (9) локальных правых единиц содержатся обратимые и необратимые векторы рассматриваемой КНАА. Количество первых равно $p^2 - p$, а вторых — p .

Пересечение множеств (8) и (9) задает p различных двухсторонних локальных единиц для векторов вида G^k при произвольном натуральном числе $k \geq 1$, из которых необратимым элементом рассматриваемой КНАА является единственный вектор, обозначаемый как E_G и представляющий собой единичный элемент циклической группы, генерируемой всевозможными степенями вектора G . Значение локальной двухсторонней единицы E_G можно найти по формуле $E_G = G^\omega$, где ω — порядок вектора G , являющийся делителем числа $p^2 - 1$. Из выражений (8) и (9) в статье [13] выводится следующее выражение, которое имеет значительно меньшую вычислительную сложность по сравнению с операцией возведения в степень ω :

$$E_G = \left(k, \frac{g_1 k}{g_0}, \frac{g_0 - (\lambda g_0 + g_1)k}{g_0 + g_1}, h, \frac{g_0 g_1 - (\lambda g_0 + g_1)g_1 k}{g_0^2 + g_0 g_1} \right), \quad (10)$$

где $k = g_0^2 (\lambda g_0^2 + g_0 g_1 + g_0 g_2 + g_1 g_2)^{-1}$.

В качестве скрытой циклической группы, в которой выполняется базовая операция возведения в степень, схема ЭЦП [13] использует группу, генерируемую необратимым вектором G . Легко видеть, что для произвольного обратимого вектора R из множества (9) локальных правых единиц и произвольного натурального значения k справедливо равенство $R \circ G^k = (R \circ G)^k$. Взаимная коммутативность операции умножения слева на правую единицу и операции возведения в степень в базовой циклической группе позволяют использовать первую в качестве маскирующей операции при задании СЗДЛ.

Другой тип используемой маскирующей операции задается по формуле $Z = R' \circ G \circ A$, где

R' — обратимый вектор из множества локальных правых единиц (9); A — обратимый вектор, такой, что для некоторой локальной левой единицы L выполняется равенство $A \circ R' = L$. Для любого натурального числа k выполняется условие взаимной коммутативности с операцией возведения в степень: $(R' \circ G \circ A)^k = R' \circ G^k \circ A$.

В работе [13] предлагается следующая процедура формирования открытого ключа в виде пары четырехмерных векторов (Y, Z) , использующая два указанных типа маскирующих операций.

1. Выбрать случайный необратимый вектор G , локальный порядок которого равен простому числу q .

2. Генерируя случайные пары значений (d, h) и используя формулу (9), вычислить две локальные правые единицы R_1 и R_2 , являющиеся обратимыми элементами рассматриваемой КНАА.

3. Генерируя случайные пары значений (d, h) и используя формулу (8), вычислить случайную локальную левую единицу L , являющуюся обратимым элементом алгебры.

4. Вычислить четырехмерный вектор A из уравнения $A \circ R_2 = L$.

5. Сгенерировать случайное натуральное число $x < q$ и вычислить открытый ключ в виде пары четырехмерных векторов Y и Z :

$$Y = R_1 \circ G^x; Z = R_2 \circ G \circ A.$$

Личным секретным ключом подписанта является четверка значений x, G, R_1 и A . Процедура вычисления ЭЦП к электронному документу M включает следующие шаги.

1. Сгенерировать случайное натуральное число $k < q$.

2. Вычислить фиксатор в виде четырехмерного вектора $W = R_1 \circ G^k \circ A$.

3. Вычислить первый элемент ЭЦП как значение хэш-функции $e = F_h(M, W)$, где F_h — некоторая хэш-функция, являющаяся частью протокола ЭЦП.

4. Вычислить второй элемент ЭЦП в виде двоичного числа s :

$$s = k + ex \text{ mod } q.$$

Проверка подлинности ЭЦП (e, s) к документу M выполняется с использованием открытого ключа (Y, Z) по следующему алгоритму.

1. Вычислить значение вектора $\tilde{W} = Y^{-e} \circ Z^s$.

2. Присоединив к документу вектор \tilde{W} , вычислить значение хэш-функции $\tilde{e} = F_h(M, \tilde{W})$.

3. Сравнить значения \tilde{e} и e . Если $\tilde{e} = e$, то подпись (e, s) принимается как подлинная ЭЦП. Если $\tilde{e} \neq e$, то подпись (e, s) отклоняется как ложная ЭЦП.

Протоколы слепой ЭЦП

В некоторых специальных информационных технологиях, например в системах электронных денег [21], одним из важных требований является обеспечение неотслеживаемости (анонимности) пользователей. Для решения этой задачи в работе [22] впервые был предложен механизм формирования ЭЦП «вслепую», реализуемый с помощью протоколов слепой подписи. В протоколах данного типа участвуют два субъекта: 1) клиент, формирующий электронный документ и желающий получить подлинную подпись другого лица к этому документу, и 2) подписант, вычисляющий некоторые параметры (элементы слепой подписи) и передающий их значения клиенту, из которых последний вычисляет подлинную ЭЦП. Проверка подлинности подписи, полученной клиентом, осуществляется по тому же алгоритму, как и проверка обычной подписи.

При этом по данным, использованным в ходе протокола слепой подписи, клиент не может получить информацию о личном секретном ключе подписанта, а подписант не может однозначно установить связь некоторого выполненного протокола слепой подписи с некоторым электронным документом и приложенной к нему подлинной ЭЦП (предполагается, что подписант многократно выполнял подписывание документов «вслепую»). Последний момент называется требованием обеспечения неотслеживаемости пользователей, которое предъявляется к протоколам слепой подписи.

Первый протокол слепой ЭЦП был реализован на основе схемы подписи RSA [23], основанной на вычислительной сложности задачи факторизации. В дальнейшем были разработаны протоколы слепой ЭЦП, основанные на вычислительной сложности ЗДЛ [14]. В обоих случаях анонимность клиента обеспечивается механизмом внесения в слепую подпись одного или двух случайных ослепляющих множителей. После получения слепой подписи от подписанта клиент удаляет ослепляющие множители, в результате чего получает подлинную подпись. Протоколы слепой подписи могут быть реализованы на основе ряда известных схем ЭЦП, например, на основе RSA [23], схемы Шнора [20], ГОСТ Р 34.10–94 и ГОСТ Р 34.10–2001 [24, 25].

Протоколы слепой ЭЦП, основанные на вычислительной сложности СЗДЛ

Использование КНАА с множеством глобальных левых единиц

Протокол слепой ЭЦП, основанный на схеме ЭЦП из статьи [19], описывается следующим образом.

1. Подписант генерирует случайное число $k < q$ и вычисляет вектор-фиксатор. Затем он направляет вектор-фиксатор $\bar{\mathbf{V}}$ клиенту, имеющему намерение сформировать подлинную ЭЦП подписанта к некоторому документу M .

2. Клиент выбирает два случайных равновероятных натуральных числа $\tau < q$ и $\varepsilon < q$, используя которые вычисляет левый \mathbf{Y}^τ и правый \mathbf{Z}^ε ослепляющие множители. Затем он модифицирует вектор-фиксатор $\bar{\mathbf{V}}$ по формуле $\mathbf{V} = \mathbf{Y}^\tau \circ \bar{\mathbf{V}} \circ \mathbf{Z}^\varepsilon$ и вычисляет первый элемент подписи $v = F_h(M, \mathbf{V})$.

3. По значению v клиент формирует первый элемент слепой подписи в виде натурального числа $\bar{v} = v + \tau \bmod q$ и передает его подписанту.

4. Подписант вычисляет второй элемент \bar{s} слепой подписи по формуле $\bar{s} = k + \bar{v}x \bmod q$, где x — значение личного секретного ключа подписанта. Затем значение \bar{s} передается клиенту.

5. Получив второй элемент слепой подписи, клиент вычисляет значение $s = \bar{s} + \varepsilon \bmod q$, которое является вторым элементом s подлинной подписи.

На выходе этого протокола клиент получает подлинную подпись (v, s) к документу, который был недоступен подписанту в ходе протокола. Доказательство корректности работы протокола состоит в том, чтобы доказать, что полученная клиентом ЭЦП проходит процедуру проверки подлинности подписи, как правильная подлинная подпись. Действительно, подавая на вход проверочной процедуры подпись (v, s) , получим следующее доказательство:

$$\begin{aligned} \mathbf{V}' &= \mathbf{Y}^{-v} \circ \mathbf{T} \circ \mathbf{Z}^s = \mathbf{Y}^{-(\bar{v}-\tau)} \circ \mathbf{T} \circ \mathbf{Z}^{\bar{s}+\varepsilon} = \\ &= \mathbf{Y}^\tau \circ \mathbf{Y}^{-\bar{v}} \circ \mathbf{T} \circ \mathbf{Z}^{\bar{s}} \circ \mathbf{Z}^\varepsilon = \mathbf{Y}^\tau \circ \bar{\mathbf{V}} \circ \mathbf{Z}^\varepsilon \Rightarrow \\ &\Rightarrow v' = F_h(M, \mathbf{V}') = F_h(M, \mathbf{V}) = v. \end{aligned}$$

Поскольку выполняется условие $v' = v$, то подпись (v, s) принимается как подлинная ЭЦП к документу M .

Аналогичным способом другие известные схемы ЭЦП, основанные на вычислительной трудности СЗДЛ и использующие в качестве алгебраического носителя КНАА с большим множеством односторонних единиц, также могут применяться для разработки протоколов слепой ЭЦП.

Использование КНАА с глобальной двухсторонней единицей

В протоколе слепой ЭЦП, основанном на схеме ЭЦП из статьи [13], выполняются следующие шаги.

1. Сформировав случайное равновероятное натуральное число $k < q$, подписант вычисляет вектор-фиксатор $\bar{\mathbf{W}} = \mathbf{R}_1 \circ \mathbf{G}^k \circ \mathbf{A}$. Затем он передает вектор-фиксатор $\bar{\mathbf{W}}$ клиенту, имеющему намерение получить подлинную ЭЦП подписанта к документу M .

2. Клиент выбирает два случайных равновероятных натуральных числа $0 < \tau < q$ и $0 < \varepsilon < q$, используя которые вычисляет левый \mathbf{Y}^τ и правый \mathbf{Z}^ε ослепляющие множители. Затем он модифицирует вектор-фиксатор по формуле $\mathbf{W} = \mathbf{Y}^\tau \circ \bar{\mathbf{W}} \circ \mathbf{Z}^\varepsilon$ и вычисляет первый элемент ЭЦП в виде значения хэш-функции $e = F_h(M, \mathbf{W})$.

3. Первый элемент слепой ЭЦП \bar{e} вычисляется клиентом по формуле $\bar{e} = e + \tau \bmod q$ и направляется подписанту.

4. Подписант находит значение второго элемента \bar{s} слепой подписи по формуле $\bar{s} = k + \bar{e}x \bmod q$, где x — личный секретный ключ подписанта. Значение \bar{s} передается клиенту.

5. Используя второй элемент \bar{s} слепой подписи, клиент вычисляет значение $s = \bar{s} + \varepsilon \bmod q$, которое является вторым элементом подлинной подписи.

Полученная клиентом подпись (e, s) к документу M проходит процедуру проверки подписи как подлинная ЭЦП. Действительно, при использовании открытого ключа (\mathbf{Y}, \mathbf{Z}) в ходе осуществления процедуры проверки подлинности ЭЦП имеем следующее доказательство:

$$\begin{aligned} \tilde{\mathbf{W}} &= \mathbf{Y}^{-e} \circ \mathbf{Z}^s = \mathbf{Y}^{-\bar{e}+\tau} \circ \mathbf{Z}^{\bar{s}+\varepsilon} = \\ &= \mathbf{Y}^\tau \circ \mathbf{Y}^{-\bar{e}} \circ \mathbf{Z}^{\bar{s}} \circ \mathbf{Z}^\varepsilon = \mathbf{Y}^\tau \circ \bar{\mathbf{W}} \circ \mathbf{Z}^\varepsilon \Rightarrow \\ &\Rightarrow \tilde{e} = F_h(M, \tilde{\mathbf{W}}) = F_h(M, \mathbf{W}) = e. \end{aligned}$$

В соответствии с проверочной процедурой при выполнении условия $\tilde{e} = e$ подпись (e, s) принимается как подлинная, т. е. описанный протокол слепой подписи является корректным.

Заключение

Предложены реализации двух протоколов слепой подписи, основанных на вычислительной трудности СЗДЛ. В качестве базовых схем ЭЦП использованы схемы ЭЦП, алгебраическими носителями которых являются четырехмерные КНАА двух разных типов. Предложенные протоколы представляют интерес как кандидаты на постквантовые протоколы слепой ЭЦП. Особенностью построения протоколов слепой подписи, основанных на вычислительной трудности СЗДЛ, является то, что один из вносимых ослепляющих множителей должен играть роль правого операнда, а второй — роль левого операнда. Данная специфика связана с тем, что алгебраическими носителями таких протоколов являются некоммутативные алгебры.

Финансовая поддержка

Работа выполнена при частичной финансовой поддержке РФФИ (грант № 18-07-00932-а).

Литература

1. **Shor P. W.** Polynomial-time algorithms for prime factorization and discrete logarithms on quantum computer. *SIAM Journal of Computing*, 1997, vol. 26, pp. 1484–1509.
2. **Smolin J. A., Smith G., Vargo A.** Oversimplifying quantum factoring. *Nature*, 2013, vol. 499, no. 7457, pp. 163–165.
3. **Yan S. Y.** *Quantum Computational Number Theory*. Springer, 2015. 252 p.
4. **Yan S. Y.** *Quantum Attacks on Public-Key Cryptosystems*. Springer, 2014. 207 p.
5. Federal Register. Announcing Request for Nominations for Public-Key Post-Quantum Cryptographic Algorithms. <https://www.gpo.gov/fdsys/pkg/FR-2016-12-20/pdf/2016-30615.pdf> (дата обращения: 13.11.2019).
6. *Post-Quantum Cryptography. 9th International Conference, PQCrypto 2018*, Fort Lauderdale, FL, USA, April 9–11, 2018, Proceedings. *Lecture Notes in Computer Science (LNCS)*, Springer, 2018, vol. 10786. 529 p. doi:10.1007/978-3-319-79063-3
7. *Post-Quantum Cryptography. 10th International Conference, PQCrypto 2019*, Chongqing, China, May 8–10, 2019. *Lecture Notes in Computer Science (LNCS)*, Springer, 2019, vol. 11505. doi:10.1007/978-3-030-25510-7
8. **Moldovyan N. A., Moldovyan A. A.** Finite non-commutative associative algebras as carriers of hidden discrete logarithm problem. *Вестник ЮУрГУ. Серия: Математическое моделирование и программирование*, 2019, т. 12, № 1, с. 66–81.
9. **Moldovyan D. N.** Non-commutative finite groups as primitive of public-key cryptoschemes. *Quasigroups and Related Systems*, 2010, vol. 18, pp. 165–176.
10. **Moldovyan N. A.** Unified method for defining finite associative algebras of arbitrary even dimensions. *Quasigroups and Related Systems*, 2018, vol. 26, no. 2, pp. 263–270.
11. **Молдовян Н. А., Абросимов И. К.** Схема постквантовой электронной цифровой подписи на основе усиленной формы скрытой задачи дискретного логарифмирования. *Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления*, 2019, т. 15, вып. 2, с. 212–220. <https://doi.org/10.21638/11702/spbu10.2019.205>
12. **Moldovyan N. A.** Finite non-commutative associative algebras for setting the hidden discrete logarithm problem and post-quantum cryptoschemes on its base. *Buletinul Academiei de Stiinte a Republicii Moldova. Matematica*, 2019, no. 1 (89), pp. 71–78.
13. **Молдовян А. А., Молдовян Д. Н.** Постквантовая схема ЭЦП на основе скрытой задачи дискретного логарифмирования в четырехмерной конечной алгебре. *Вопросы защиты информации*, 2019, № 2, с. 18–22.
14. **Camensisch J. L., Piveteau J.-M., Stadler M. A.** Blind signatures based on the discrete logarithm problem. *Advances in Cryptology — EUROCRYPT '94, Workshop on the Theory and Application of Cryptographic Techniques*, Perugia, Italy, May 9–12, 1994. Proceedings. Springer, 1995, vol. 950. *Lecture Notes in Computer Science (LNCS)*, pp. 428–432.
15. **Pointcheval D., Stern J.** Security arguments for digital signatures and blind signatures. *Journal of Cryptology*, 2000, vol. 13, no. 3, pp. 361–396.
16. **Kuzmin A. S., Markov V. T., Mikhalev A. A., Mikhalev A. V., Nechaev A. A.** Cryptographic algorithms on groups and algebras. *Journal of Mathematical Sciences*, 2017, vol. 223, no. 5, pp. 629–641.
17. **Кузьмин А. С., Марков В. Т., Михалев А. А., Михалев А. В., Нечаев А. А.** Криптографические алгоритмы на группах и алгебрах. *Фундаментальная и прикладная математика*, 2015, т. 20, № 1, с. 205–222.
18. **Глухов М. М.** К анализу некоторых систем открытого распределения ключей, основанных на неабелевых группах. *Математические вопросы криптографии*, 2010, т. 1, № 4, с. 5–22.
19. **Moldovyan D. N., Moldovyan A. A., Sklavos N.** Post-quantum signature schemes for efficient hardware implementation. *Proceedings of 10th IFIP International Conference on New Technologies, Mobility & Security (NTMS'19)*, Canary Islands, Spain, June 24–26, 2019, pp. 1–5.
20. **Schnorr C. P.** Efficient signature generation by smart cards. *Journal of Cryptology*, 1991, vol. 4, pp. 161–174.
21. **Chaum D.** Blind signatures for untraceable payments. *Advances in Cryptology: Proc. of CRYPTO'82*, Plenum Press, 1983, pp. 199–203.
22. **Chaum D.** Security without identification: Transaction systems to make big brother obsolete. *Communication of the ACM*, Oct. 1985, vol. 28, no. 10, pp. 1030–1044.
23. **Rivest R. L., Shamir A., Adleman L. M.** A method for obtaining digital signatures and public key cryptosystems. *Communications of the ACM*, 1978, vol. 21, pp. 120–126.
24. **Молдовян Н. А.** Протоколы слепой коллективной подписи на основе стандартов цифровой подписи. *Вопросы защиты информации*, 2010, № 1, с. 2–6.
25. **Moldovyan N. A.** Blind signature protocols from digital signature standards. *Int. Journal of Network Security*, 2011, vol. 13, no. 1, pp. 22–30.

UDC 003.26

doi:10.31799/1684-8853-2020-3-71-78

Blind signature protocols based on hidden discrete logarithm problemD. N. Moldovyan^a, PhD, Tech., Research Fellow, orcid.org/0000-0001-5039-7198, mdn.spectr@mail.ruA. A. Moldovyan^a, Dr. Sc., Tech., Professor, orcid.org/0000-0001-5480-6016D. Yu. Gurianov^b, PhD, Tech., Associate Professor, orcid.org/0000-0002-2923-4965^aSaint-Petersburg Institute for Informatics and Automation of the RAS, 39, 14 Line, V. O., 199178, Saint-Petersburg, Russian Federation^bAdmiral Makarov State University of Maritime and Inland Shipping, 5/7, Dvinskaya St., 198035, Saint-Petersburg, Russian Federation

Introduction: The progress in the development of quantum computing has raised the problem of constructing post-quantum public-key cryptographic algorithms and protocols, i. e. cryptoschemes resistant to quantum attacks. Based on the hidden discrete logarithm problem, some practical post-quantum digital signature schemes have been developed. The next step could be the development of post-quantum blind signature protocols. **Purpose:** To develop blind signature protocols based on the computational difficulty of the hidden discrete logarithm problem. **Method:** The use of blinding factors introduced by the client during the blind signature protocol when the parameters necessary for the blind signature formation are passed to the signer. **Results:** It has been proposed to use blinding multipliers of two different types: left-sided and right-sided ones. With them, you can develop blind signature protocols on the base of schemes with a verification equation defined in non-commutative algebraic structures. New blind signature protocols have been developed, based on the computational difficulty of the hidden discrete logarithm problem. As the algebraic carrier for the developed protocols, finite non-commutative associative algebras of two types are used: 1) those with a global two-sided unit, and 2) those with a large set of global left units. **Practical relevance:** The proposed protocols have a high performance and can be successfully implemented either in software or in hardware.

Keywords — post-quantum crypto schemes, computer security, information protection, digital signature, blind signature, discrete logarithm problem, non-commutative associative algebras, computationally difficult problem.

For citation: Moldovyan D. N., Moldovyan A. A., Gurianov D. Yu. Blind signature protocols based on hidden discrete logarithm problem. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2020, no. 3, pp. 71–78 (In Russian). doi:10.31799/1684-8853-2020-3-71-78

References

- Shor P. W. Polynomial-time algorithms for prime factorization and discrete logarithms on quantum computer. *SIAM Journal of Computing*, 1997, vol. 26, pp. 1484–1509.
- Smolin J. A., Smith G., Vargo A. Oversimplifying quantum factoring. *Nature*, 2013, vol. 499, no. 7457, pp. 163–165.
- Yan S. Y. *Quantum Computational Number Theory*. Springer, 2015. 252 p.
- Yan S. Y. *Quantum Attacks on Public-Key Cryptosystems*. Springer, 2014. 207 p.
- Federal Register. Announcing Request for Nominations for Public-Key Post-Quantum Cryptographic Algorithms*. Available at: <https://www.gpo.gov/fdsys/pkg/FR-2016-12-20/pdf/2016-30615.pdf> (accessed 13 November 2019).
- Post-Quantum Cryptography. 9th International Conference, PQCrypto 2018*, Fort Lauderdale, FL, USA, April 9–11, 2018, Proceedings. *Lecture Notes in Computer Science (LNCS)*, Springer, 2018, vol. 10786. 529 p. doi:10.1007/978-3-319-79063-3
- Post-Quantum Cryptography. 10th International Conference, PQCrypto 2019*, Chongqing, China, May 8–10, 2019. *Lecture Notes in Computer Science (LNCS)*, Springer, 2019, vol. 11505. doi:10.1007/978-3-030-25510-7
- Moldovyan N. A., Moldovyan A. A. Finite non-commutative associative algebras as carriers of hidden discrete logarithm problem. *Bulletin of the South Ural State University. Series "Mathematical Modelling, Programming & Computer Software"*, 2019, vol. 12, no. 1, pp. 66–81.
- Moldovyan D. N. Non-commutative finite groups as primitive of public-key cryptoschemes. *Quasigroups and Related Systems*, 2010, vol. 18, pp. 165–176.
- Moldovyan N. A. Unified method for defining finite associative algebras of arbitrary even dimensions. *Quasigroups and Related Systems*, 2018, vol. 26, no. 2, pp. 263–270.
- Moldovyan N. A., Abrosimov I. K. Post-quantum electronic digital signature scheme based on the enhanced form of the hidden discrete logarithm problem. *Vestnik of Saint Petersburg University. Applied Mathematics. Computer Science. Control Processes*, 2019, vol. 15, iss. 2, pp. 212–220 (In Russian). <https://doi.org/10.21638/11702/spbu10.2019.205>
- Moldovyan N. A. Finite non-commutative associative algebras for setting the hidden discrete logarithm problem and post-quantum cryptoschemes on its base. *Buletinul Academiei de Stiinta a Republicii Moldova. Matematica*, 2019, no. 1 (89), pp. 71–78.
- Moldovyan A. A., Moldovyan D. N. Post-quantum signature algorithms based on the hidden discrete logarithm problem in four-dimensional finite algebra. *Voprosy zashchity informatsii*, 2019, no. 2, pp. 18–22 (In Russian).
- Camenisch J. L., Piveteau J.-M., Stadler M. A. Blind signatures based on the discrete logarithm problem. *Advances in Cryptology — EUROCRYPT '94, Workshop on the Theory and Application of Cryptographic Techniques*, Perugia, Italy, May 9–12, 1994. Proceedings. Springer, 1995, vol. 950. *Lecture Notes in Computer Science (LNCS)*, pp. 428–432.
- Pointcheval D., Stern J. Security arguments for digital signatures and blind signatures. *Journal of Cryptology*, 2000, vol. 13, no. 3, pp. 361–396.
- Kuzmin A. S., Markov V. T., Mikhalev A. A., Mikhalev A. V., Nechaev A. A. Cryptographic algorithms on groups and algebras. *Journal of Mathematical Sciences*, 2017, vol. 223, no. 5, pp. 629–641.
- Kuzmin A. S., Markov V. T., Mikhalev A. A., Mikhalev A. V., Nechaev A. A. Cryptographic algorithms on groups and algebras. *Fundamental and Applied Mathematics*, 2015, vol. 20, no. 1, pp. 205–222 (In Russian).
- Glukhov M. M. On analysis of some public key distribution systems based on non-abelian groups. *Mathematical Aspects of Cryptography*, 2010, vol. 1, no. 4, pp. 5–22 (In Russian).
- Moldovyan D. N., Moldovyan A. A., Sklavos N. Post-quantum signature schemes for efficient hardware implementation. *Proceedings of 10th IFIP International Conference on New Technologies, Mobility & Security (NTMS'19)*, Canary Islands, Spain, June 24–26, 2019, pp. 1–5.
- Schnorr C. P. Efficient signature generation by smart cards. *Journal of Cryptology*, 1991, vol. 4, pp. 161–174.
- Chaum D. Blind Signatures for Untraceable Payments. *Advances in Cryptology: Proc. of CRYPTO'82*, Plenum Press, 1983, pp. 199–203.
- Chaum D. Security without identification: Transaction systems to make big brother obsolete. *Communication of the ACM*, Oct. 1985, vol. 28, no. 10, pp. 1030–1044.
- Rivest R. L., Shamir A., Adleman L. M. A method for obtaining digital signatures and public key cryptosystems. *Communications of the ACM*, 1978, vol. 21, pp. 120–126.
- Moldovyan N. A. Blind collective signature protocols based on digital signature standards. *Voprosy zashchity informatsii*, 2010, vol. 1, pp. 2–6 (In Russian).
- Moldovyan N. A., Moldovyan D. N. Blind signature protocols from digital signature standards. *Int. Journal of Network Security*, 2011, vol. 13, no. 1, pp. 22–30.

Lower bound for average delay in unblocked random access algorithm with orthogonal preambles

A. A. Burkov^a, Assistant Professor, orcid.org/0000-0002-0920-585X

S. V. Shneer^b, PhD, Phys.-Math., Associate Professor, orcid.org/0000-0001-6750-6995

A. M. Turlikov^a, Dr. Sc., Tech., Professor, orcid.org/0000-0001-7132-094X, turlikov@k36.org

^aSaint-Petersburg State University of Aerospace Instrumentation, 67, B. Morskaya St., 190000, Saint-Petersburg, Russian Federation

^bSchool of Mathematical and Computer Sciences, Heriot–Watt University, Edinburgh, EH14 4AS, United Kingdom

Introduction: Currently, the first versions of 5G communication standard networks are being deployed and discussions are underway on further development of cellular networks and transition to the 6G standard. Once the Internet of Things is fully developed, it will function within the framework of Massive Machine-Type Communications. Its performance will need to satisfy a number of requirements: very high energy efficiency, relatively low delay and fairly reliable communication. Due to the nature of the traffic, it is impossible to develop a channel resource sharing policy and thus random multiple access procedures will be employed. A class of unblocked algorithms with orthogonal preambles may be used to increase the efficiency of random access.

Purpose: To calculate the lower bound of the average delay for the class of unblocked random multiple access algorithms using orthogonal preambles. **Results:** We propose a model of a system with a potentially unlimited number of users who utilize random unblocked access to transmit data over a common communication channel using orthogonal preambles. For an infinite number of preambles, we prove that up to an input arrival rate of 0.5671, the system is stable. We obtain a closed-form expression for the average delay in a stable system depending on the intensity of the input arrival rate. We also demonstrate that this expression is the lower bound for the average delay in a system with a finite number of preambles. We perform a simulation in the case of a finite number of preambles which shows that with an increase in the number of preambles, the input arrival rate at which the system is stable approaches 0.5671, and the average delay tends to the lower bound. With an input arrival rate not exceeding 0.3, about 100 orthogonal preambles are sufficient to reach the lower bound. **Practical relevance:** The obtained bound allows us to estimate the lower average delay in the class of algorithms under consideration. Its application allows us to determine the possibility of using such a class of algorithms at the stage of designing random multiple access systems from the point of view of limitations on the average delay.

Keywords – random multiple access, maximum arrival rate, average delay, preambles, mMTC, IoT.

For citation: Burkov A. A., Shneer S. V., Turlikov A. M. Lower bound for average delay in unblocked random access algorithm with orthogonal preambles. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2020, no. 3, pp. 79–85. doi:10.31799/1684-8853-2020-3-79-85

Introduction

Currently, the first versions of 5G communication standard networks are being deployed and new versions will be developed in the near future [1–3]. Within the framework of the fifth generation networks, three main types of communication will be distinguished: enhanced Mobile Broadband, Ultra-Reliable Low Latency Communication, and Massive Machine-Type Communications (mMTC) [4]. In further development of cellular networks, and the transition to the 6G standard, these key scenarios are expected to remain relevant, but the level of requirements for the main characteristics will increase [1, 3]. The functionality of the currently popular idea of the Internet of Things (IoT) is expected to be within the framework of the mMTC scenario [3, 5]. For this type of communication there are a number of requirements for performance characteristics: very high energy efficiency, relatively low latency and fairly reliable communication [1]. The task of meeting these requirements is difficult, due

to the features of IoT traffic. A huge number of devices generate data periodically and unpredictably. Due to the number of devices in the network and the characteristics of the formation of data traffic, it is difficult to develop an effective channel resource allocation policy [4, 5]. Therefore, in the framework of this scenario, it is assumed that random multiple access procedures are used when devices access the shared resource uncoordinated, as well as without the need for a connection establishment and communication with the resource manager [5].

The 5G standard is advancing as the primary performance tool for IoT communications, thanks to the development of 5G New Radio. However, the revolutionary forecast for 5G networks was reduced exclusively to work on mmWave, and that would not allow for the operation of IoT systems taking into account the increasing number of devices and services [5]. When applying mmWave communication to cellular systems, an important problem that can negate its advantages is multiple access, namely because of the high energy consumption and high cost

of equipment [6]. However, this is only the first prerequisite for the development of IoT technology and future versions of cellular standards will require new solutions in this direction.

In this paper, we are going to consider the use of orthogonal preambles for data transmission in mMTC scenarios. A similar approach to data transmission using orthogonal preambles is considered in [7]. The algorithm considered there involves work in two stages. At the first stage, users transmit preambles, then they receive information about which slot the user who sent a particular preamble will utilize. The base station then allocates preamble resources for data transmissions. At the second stage, the user transmits data to the allocated time-frequency resource in accordance with the transmitted preamble and the received information. In [7], the authors note that with a limited number of preambles, the system can quickly become unstable, therefore, it considers a variant of the lossy algorithm. That is, if several users transmit under the preamble in the allocated resource, then the message is lost. We note further the following features. When working with losses, the system is always stable. When data is divided into two types and the operation of the system into two distinct stages, the user has to constantly monitor the channel output and the stages in the system, even if he does not have a message to transmit. At the command of the message, the user will have to wait for the next first stage of the system, postponing his transfer. This algorithm can be attributed to the class of blocked algorithms. The term blocked and non-blocked algorithms was first introduced in the works of B. S. Tsybakov (see, for example, [8]). The features of these types of algorithms are also discussed in [9, 10]. A blocked algorithm runs in sessions. The user can transmit only in the first slot of the session, if the user appeared in the system during the session; then the user blocks his transmission until the start of the next session. Therefore, this algorithm is called blocked. An algorithm related to this type in the framework of the mMTC scenario was also considered in [11], in which, unlike in [7], there is no separation of messages into two types and no losses of messages. One of the disadvantages of this class of algorithms is the constant listening to the channel, even in the absence of a message ready for transmission by the user. As one of mMTC traffic characteristics is a rare message from the user, the constant monitoring of the channel adversely affects the energy efficiency of the system. For transmitting devices, with a limitation on complexity and energy efficiency, a significant role is played by delays between the appearance of a message from a user and its transmission to a communication channel. Therefore,

more efficient algorithms are those in which, when a user has a message, it is transmitted in the next slot without any wait. Algorithms with this property form a class of non-blocked algorithms. With this approach, the user will listen to the channel only after the first transmission of the message that appears with him and until it is transmitted, and not continuously, as in blocked algorithms.

In this paper, we will consider an unblocked random multiple access algorithm using orthogonal preambles. The data transfer mode can be divided into two types. The first type is the first transmission of a message that occurs in a nearby slot. In the event of a collision, the preamble information will be used by the base station to schedule the time-sharing mode. The second type, in the event of a collision during the first transmission, is the message that is resent in accordance with the schedule received from the base station. For this algorithm, assuming an infinite number of preambles, we obtain a lower bound for the average number of users, the delay and the maximum input arrival rate at which the system is stable. We will also demonstrate the performance of the algorithm with a finite number of preambles for some implementation of the system, with lossless operation and a different number of preambles. We will further provide a comparison of the simulation results and the obtained lower bound.

System model

We will make a number of assumptions on the operation of the algorithm which will facilitate our analysis of the system characteristics.

Assumption 1. Messages for all users have the same length. The transmission time of one message is taken as a unit of time. The entire transmission time in the channel is divided into slots that correspond to the transmission time of one message. User messages consist of the service part (randomly selected preamble from some set) and data. Users know exactly the moments of breaking into slots and can transmit only at the beginning of the next slot (synchronous system).

Assumption 2. There are three possible events in the slot:

“Empty” (*E*), if no one was transmitting in the slot;

“Success” (*S*), if one user transmitted in the slot. In this case, the base station correctly receives both the preamble and the message itself;

“Conflict” (*C*), if more than one user was transmitted in one slot. That message cannot be correctly received by the base station. Preamble numbers are accepted correctly, even if several users choose one preamble.

Assumption 3. The base station transmits information about the event that occurred in the channel and additional information necessary for the algorithm to work (for example, a list of preambles involved in the conflict).

Assumption 4. An infinite number of users is assumed. The number of new messages that appear in the system in one slot is distributed according to the Poisson law with the parameter λ , where is the intensity of the input arrival rate.

Assumption 5. There is an infinite number of orthogonal preambles.

According to assumption 4, each message appearing in the system will be considered a new user who leaves the system after the transmission of this message.

Algorithm for users.

1. The user, when it has a message, selects one of the preambles in a random fashion. And it transmits a message consisting of the selected preamble and data at the beginning of the next slot.

2. The user receives information from the base station and acts in accordance with it:

“Success”: the user deletes the message because it was sent successfully;

“Conflict”: the user delays the transmission of its message to the slot allocated under its preamble in accordance with the schedule received from the base station.

The algorithm of the base station for an infinite number of preambles.

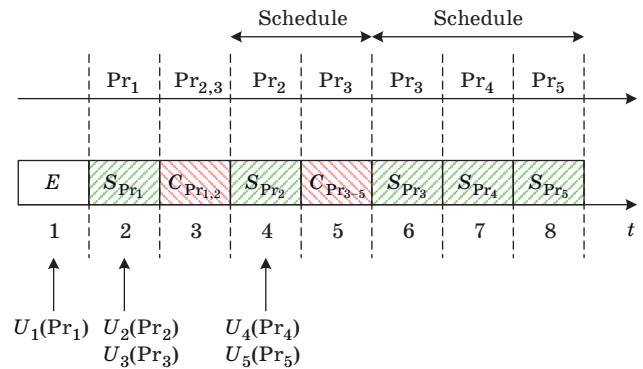
The base station receives information from the channel and operates depending on the events in the channel:

“Empty”: the base station informs all users of the “Empty” event;

“Success”: the base station informs all users about the “Success” event along with the number of the received preamble. If the system was in the time-sharing mode, the base station removes this preamble from the schedule;

“Conflict”: the base station reports a “Conflict” along with the numbers of the preambles involved in it. The base station makes or adjusts the schedule based on the detected preambles and reports it to users.

An example of the algorithm operation is presented in Fig. 1. According to assumptions 1 and 5, all users appearing in the system select unique orthogonal preambles. Assume that there are 5 users (U_1, U_2, U_3, U_4, U_5). At the beginning of the second slot, the first user (U_1), which has a message, sends it to the channel and the event that occurs is “Success”. In the second slot, messages appear for users U_2 and U_3 , which are transmitted in the third slot and the event “Conflict” occurs. The base station correctly decodes the numbers of the preambles involved in the conflict and schedules the next two slots. The



■ Fig. 1. An example of an unblocking random multiple access algorithm with orthogonal preambles

second user successfully sends the message according to the schedule. Users U_4 and U_5 appear in the fourth slot. In the fifth slot, they transmit messages, thereby causing a conflict and interrupting the current schedule. The base station decodes the numbers of the preambles involved in the conflict, adjusts the schedule, and informs its users. In the sixth through eighth slots, no new messages appear, and users (U_{3-5}), successfully transmit their messages according to the schedule.

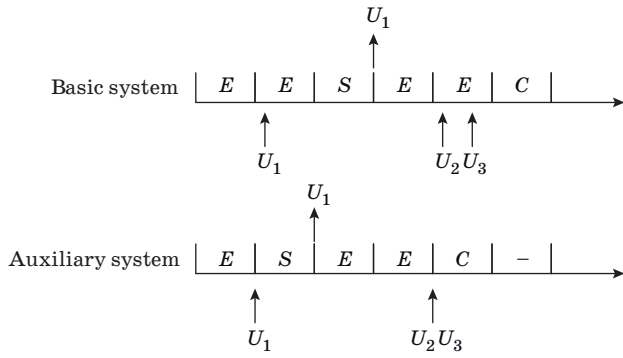
Remark 1. In real systems, there cannot be an infinite number of preambles, and assumption 5 is not valid for such systems. In the case of a limited number of preambles, it is necessary to additionally determine the rule of operation of the algorithm in the case when a “Conflict” occurs in one frequency-time resource, and several users choose the same preamble.

System analysis

In the system under consideration, messages can appear during the slot and affect the system in the next slot. So if the system is in the time-sharing mode, the message that appears during the slot will cause a conflict in the next slot. To take into account such events when describing the number of active users in the system using a recurrence equation, one needs to use a two-dimensional Markov chain (for example, [12]). The use of such a model makes an exact analysis of the system impossible. Therefore, for our analysis, we will use an auxiliary system in which messages can appear only at the beginning of the slot and therefore can be transmitted immediately. An example showing the difference between the basic and auxiliary systems is shown in Fig. 2.

By comparing the basic and auxiliary systems, one can prove the following statement.

Statement 1. Let $\bar{N}(\lambda)$ be the average number of users in the main system, and let $\bar{Q}(\lambda)$ be the aver-



■ Fig. 2. The difference between the basic and auxiliary systems

age number of users in the auxiliary system, where λ is the input stream intensity. The following equality holds:

$$\bar{N}(\lambda) = \bar{Q}(\lambda) + \lambda. \quad (1)$$

We perform further calculations for the auxiliary system. Denote by Q_k by the number of users in the slot with number k . Then the number of users in the next slot satisfies the following recursive equation describing the operation of the system:

$$Q_{k+1} = Q_k + V_k - I(V_k = 1) \times \\ \times I(Q_k = 0) - I(V_k = 0)I(Q_k > 0), \quad (2)$$

where V_k is the number of users appearing in the slot with number k , and $I(\text{statement})$ is the indicator function, which takes values:

$$I\{\text{statement}\} = \begin{cases} 1, & \text{true} \\ 0, & \text{false} \end{cases}$$

Using expression (2), one can easily analyze the stability of the system. Suppose that initially $Q_0 = q$ where q is a large value. Then

$$E[Q_1 - Q_0 | Q_0 = q] = \lambda - e^{-\lambda}. \quad (3)$$

Stability is ensured when the value of the expression (3) is negative (see [13]). This is equivalent to $\lambda < e^{-\lambda}$, a solution to this inequality can be found using the Lambert function (W) [14], and we obtain $\lambda < W(1) \approx 0.5671$.

We will assume that the system works indefinitely. If the system is stable, then Q_k and Q_{k+1} have the same distribution, and, in particular, the same moments. We take the mathematical expectation of the left and right sides of formula (2) and get

$$E[Q_{k+1}] = \\ = E[Q_k] + \lambda - \lambda e^{-\lambda} \pi_0 - e^{-\lambda} (1 - \pi_0),$$

where by π_0 we denote the stationary probability that the system is empty. Using that $E[Q_{k+1}] = E[Q_k]$, we can write

$$\lambda = \lambda e^{-\lambda} \pi_0 + e^{-\lambda} (1 - \pi_0), \quad (4)$$

which implies

$$\pi_0 = \frac{e^{-\lambda} - \lambda}{e^{-\lambda} (1 - \lambda)}. \quad (5)$$

Squaring the right- and the left-hand sides of formula (2) and taking the mathematical expectation, we obtain the following expression:

$$E[(Q_{k+1})^2] = E[(Q_k)^2] + (\lambda^2 + \lambda) + \\ + (\lambda e^{-\lambda} \pi_0 - e^{-\lambda} (1 - \pi_0)) + \\ + 2E[Q_k] \lambda - 2E[Q_k] e^{-\lambda} - 2\lambda e^{-\lambda} \pi_0. \quad (6)$$

Using $E[(Q_{k+1})^2] = E[(Q_k)^2]$ and using formula (6), we obtain

$$0 = \lambda^2 + 2\lambda - 2E[Q_k](e^{-\lambda} - \lambda) - 2\lambda e^{-\lambda} \pi_0.$$

Next, using expression (5) we obtain the equation

$$2E[Q_k](e^{-\lambda} - \lambda) = \lambda^2 + 2\lambda - 2\lambda \frac{e^{-\lambda} - \lambda}{(1 - \lambda)},$$

which implies

$$E[Q_k] = \frac{\lambda^2 + 2\lambda - 2\lambda \frac{e^{-\lambda} - \lambda}{(1 - \lambda)}}{2(e^{-\lambda} - \lambda)} = \\ = \frac{\lambda^2(1 - \lambda) + 2\lambda(1 - e^{-\lambda})}{2(e^{-\lambda} - \lambda)(1 - \lambda)}. \quad (7)$$

Expression (7) determines the average number of users in the auxiliary system at $k \rightarrow \infty$, with a fixed $\lambda < 0.5671$. Let us denote the right-hand side of expression (7) by $\bar{Q}(\lambda)$. Using statement 1, the average number of users in the main system is defined as:

$$\bar{N}(\lambda) = \frac{\lambda^2(1 - \lambda) + 2\lambda(1 - e^{-\lambda})}{2(e^{-\lambda} - \lambda)(1 - \lambda)} + \lambda = \\ = \frac{2\lambda - \lambda^2(1 + 2e^{-\lambda} - \lambda)}{2(e^{-\lambda} - \lambda)(1 - \lambda)}. \quad (8)$$

Using Little's theorem [15] and expression (8), we can determine the average delay \bar{D} in the main system as

$$\bar{D} = \frac{E[N_k]}{\lambda} + 0.5 = \frac{2 - \lambda(1 + 2e^{-\lambda} - \lambda)}{2(e^{-\lambda} - \lambda)(1 - \lambda)} + 0.5. \quad (9)$$

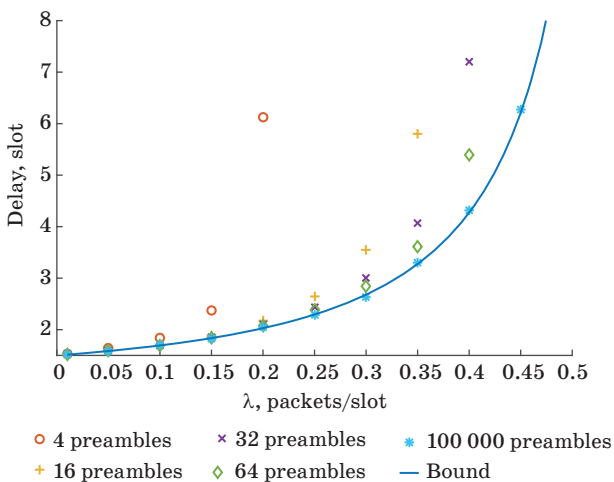
Remark 2. The addition of 0.5 in formula (9) is determined by the fact that the synchronous system is considered (see assumption 1) and we thus need to take into account the average waiting time for the next slot.

Numerical results

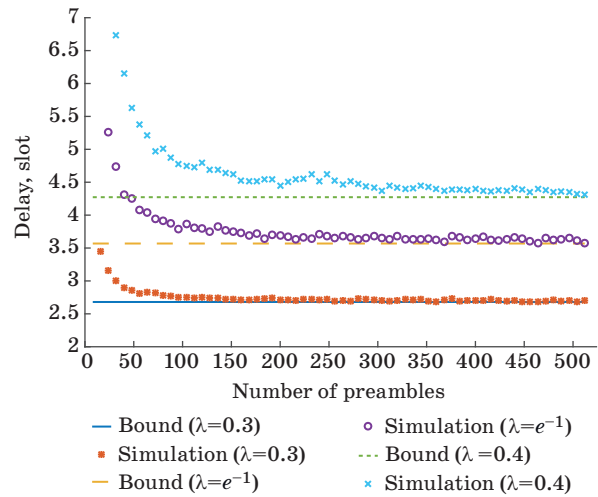
In this section we present simulation results for the system under consideration for a different number of preambles. Taking into account Remark 1, for a finite number of preambles, we modify the algorithm as follows. If a user finds itself in a conflict situation and at the same time the base station detects only one preamble, then the users participating in such a conflict will retransmit their message with probability $p = \min\left(\frac{L}{N_a}, 1\right)$, where L is the

number of orthogonal preambles and N_a is the number of users who have a message ready for transmission at the time of the start another slot. It is worth noting that in real systems, knowledge of the value N_a may not be available, but there are works [16, 17] in which methods for its evaluation are proposed. Such a modification of the algorithm for a finite number of preambles is logical, but it may not achieve the best performance. However, its influence on the performance of the system decreases with an increase in the number of preambles.

Figure 3 illustrates the average delay in the system for different values of the number of preambles



■ *Fig. 3.* The dependence of the average delay on the intensity of the input arrival rate for a different number of preambles



■ *Fig. 4.* Values of an average delay from the number of preambles for input arrival rate equal to 0.3, e^{-1} and 0.4

ambles. The bound calculated by formula (9) is also shown. One can see from the graph that with a small number of preambles, the delay values go to infinity earlier than the maximum input arrival rate obtained previously (0.5671) at which the system is stable. This is due to the fact that with a finite number of preambles this intensity is lower and tends to the obtained value with an increase in the number of preambles. One can also see that with an increase in the number of preambles, the average delay values approach the lower bound.

Figure 4 shows the dependence of the average delay in the system for fixed values of the input arrival rate (0.3, e^{-1} and 0.4) on the number of orthogonal preambles. For instance, for an input stream intensity of 0.3, about 100 preambles are enough to reach the obtained bound.

Conclusion

We considered an unblocked random access algorithm using orthogonal preambles. When a message appears, users immediately send it. Messages consist of data and orthogonal preambles. The base station uses information about orthogonal preambles in the event of a conflict to construct a schedule that is reported to users. We introduced an auxiliary system under the assumption of an infinite number of preambles and carried out its rigorous analysis. We showed that the system with a large number of preambles is stable for arrival intensities up to the value $\lambda = 0.5671$. We also obtained lower bounds for the average number of users and the average delay in the system. We further presented results of modeling with a finite number of preambles with some additional modification of the algorithm. A simulation has shown that with an increase in

the number of preambles, the maximum intensity of the input arrival rate approaches the calculated one, and the delay approaches the obtained bound. The results of our simulation showed that at an input arrival rate of 0.3, about 100 orthogonal preambles are sufficient to reach the lower bound. When using a finite number of preambles, there are various ways of implementing the algorithm in case of conflicts with the intersection of preambles for

different users. A logical continuation of the work is the search for an algorithm that provides an average delay closest to the received lower bound.

Acknowledgment

The research of A. Burkov was supported by RFBR, project number 19-31-27001.

References

1. Letaief K. B., Chen W., Shi Y., Zhang J., & Zhang Y. J. A. The roadmap to 6G: AI empowered wireless networks. *IEEE Communications Magazine*, 2019, vol. 57, no. 8, pp. 84–90. doi:10.1109/MCOM.2019.1900271
2. Piran M. J., Suh D. Y. Learning-Driven Wireless Communications, towards 6G. *2019 International Conference on Computing, Electronics & Communications Engineering (iCCECE)*, IEEE, 2019, pp. 219–224. doi:10.1109/iCCECE46942.2019.8941882
3. Mahmood N. H., Alves H., López O. A., Shehab M., Osorio D. P. M., & Latva-aho M. Six key enablers for machine type communication in 6G. *arXiv preprint arXiv:1903.05406*, 2019.
4. Popovski P., Trillingsgaard K. F., Simeone O., & Durisi G. 5G wireless network slicing for eMBB, URLLC, and mMTC: A communication-theoretic view. *IEEE Access*, 2018, vol. 6, pp. 55765–55779. doi:10.1109/ACCESS.2018.2872781
5. Clazzer F., Munari A., Liva G., Lazaro F., Stefanovic C., & Popovski P. From 5G to 6G: Has the time for modern random access come? *arXiv preprint arXiv:1903.03063*, 2019.
6. Zhu L., Xiao Z., Xia X. G., & Wu D. O. Millimeter-wave communications with non-orthogonal multiple access for B5G/6G. *IEEE Access*, 2019, vol. 7, pp. 116123–116132. doi:10.1109/ACCESS.2019.2935169
7. Choi J. On throughput of compressive random access for one short message delivery in IoT. *IEEE Internet of Things Journal*, 2020, vol. 7, no. 4, pp. 3499–3508. doi:10.1109/JIOT.2020.2972519
8. Tsybakov B. Survey of USSR contributions to random multiple-access communications. *IEEE Transactions on Information Theory*, 1985, vol. 31, no. 2, pp. 143–165. doi:10.1109/TIT.1985.1057023
9. Rom R., Sidi M. *Multiple Access Protocols: Performance and Analysis*. Springer Science & Business Media, 2012. 178 p. doi:10.1007/978-1-4612-3402-9
10. Chlebus B. S. Randomized Communication in Radio Networks. *arXiv preprint arXiv:1801.00074*, 2017.
11. Burkov A., Frolov A., Turlikov A. Contention-based protocol with time division collision resolution. *2018 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, IEEE, 2018, pp. 1–4. doi:10.1109/ICUMT.2018.8631217
12. Burkov A. A., Saveliev A. A., Turlikov A. M. Upper and lower bound for non-blocking random multiple access algorithm with time division mode. *2019 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF)*, IEEE, 2019, pp. 1–7. doi:10.1109/WECONF.2019.8840590
13. Meyn S. P., Tweedie R. L. *Markov Chains and Stochastic Stability*. Springer Science & Business Media, 2012. 559 p. doi:10.1007/978-1-4471-3267-7
14. Corless R. M., Gonnet G. H., Hare D. E., Jeffrey D. J., & Knuth D. E. On the Lambert W Function. *Advances in Computational Mathematics*, 1996, vol. 5, pp. 329–359. doi:10.1007/BF02124750
15. Brémaud P. *Probability Theory and Stochastic Processes*. Springer, Cham, 2020. 717 p. doi:10.1007/978-3-030-40183-2
16. Jeon S. W., Jin H. Online Estimation and Adaptation for Random Access with Successive Interference Cancellation. *arXiv preprint arXiv:2001.03248*, 2020.
17. Galinina O., Turlikov A., Andreev S., & Koucheryavy Y. Stabilizing multi-channel slotted aloha for machine-type communications. *2013 IEEE International Symposium on Information Theory*, IEEE, 2013, pp. 2119–2123. doi:10.1109/ISIT.2013.6620600

УДК 004.728.3.057.4

doi:10.31799/1684-8853-2020-3-79-85

Нижняя граница для средней задержки в неблокированном алгоритме случайного доступа с ортогональными преамбуламиА. А. Бурков^а, ассистент, orcid.org/0000-0002-0920-585XВ. В. Шнеер^б, канд. физ.-мат. наук, доцент, orcid.org/0000-0001-6750-6995А. М. Тюрликов^а, доктор техн. наук, профессор, orcid.org/0000-0001-7132-094X, turlikov@k36.org^аСанкт-Петербургский государственный университет аэрокосмического приборостроения, Б. Морская ул., 67, Санкт-Петербург, 190000, РФ^бШкола математических и компьютерных наук, Университет Хериота—Уатта, Эдинбург, EH14 4AS, Соединенное Королевство

Введение: в настоящее время разворачиваются первые версии сетей стандарта связи 5G и ведутся рассуждения о дальнейшем развитии сотовых сетей и переходе к стандарту 6G. Функционирование развивающейся идеи интернета вещей планируется в рамках сценария массовой межмашинной связи, к которому имеется ряд требований: очень высокая энергоэффективность, относительно невысокая задержка и достаточно надежная связь. Предполагается использование процедур случайного множественного доступа, так как ввиду особенностей трафика невозможно разработать политику разделения ресурсов канала. Для повышения эффективности случайного доступа можно применять класс неблокированных алгоритмов, использующих ортогональные преамбулы. **Цель:** расчет нижней границы средней задержки для класса неблокированных алгоритмов случайного множественного доступа с использованием ортогональных преамбул. **Результаты:** предложена модель системы с потенциально неограниченным числом абонентов, которые используют случайный неблокированный доступ для передачи данных по общему каналу связи с применением ортогональных преамбул. Для бесконечного числа преамбул доказано, что до интенсивности входного потока, равной 0,5671, система работает стабильно. Получено замкнутое выражение для расчета средней задержки в стабильной системе в зависимости от интенсивности входного потока. Обосновано, что это выражение является нижней границей для средней задержки в системе с конечным числом преамбул. Для конечного числа преамбул выполнено имитационное моделирование. Результаты моделирования показали, что при увеличении числа преамбул интенсивность входного потока, при которой система работает стабильно, приближается к 0,5671, а средняя задержка стремится к нижней границе. При интенсивности входного потока, не превышающей 0,3, достаточно порядка 100 ортогональных преамбул для достижения нижней границы. **Практическая значимость:** полученная граница дает возможность оценить снизу среднюю задержку в описанном классе алгоритмов. Ее использование позволяет определить возможность применения рассмотренного класса алгоритмов с точки зрения ограничений по средней задержке на стадии проектирования систем случайного множественного доступа.

Ключевые слова — случайный множественный доступ, предельная интенсивность входного потока, средняя задержка, преамбулы, mMTC, IoT.

Для цитирования: Burkov A. A., Shneer S. V., Turlikov A. M. Lower bound for average delay in unblocked random access algorithm with orthogonal preambles. *Информационно-управляющие системы*, 2020, № 3, с. 79–85. doi:10.31799/1684-8853-2020-3-79-85

For citation: Burkov A. A., Shneer S. V., Turlikov A. M. Lower bound for average delay in unblocked random access algorithm with orthogonal preambles. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2020, no. 3, pp. 79–85. doi:10.31799/1684-8853-2020-3-79-85

АРСЕНЬЕВ
Владимир
Николаевич



Профессор кафедры бортовых информационных и измерительных комплексов Военно-космической академии им. А. Ф. Можайского, Санкт-Петербург. В 1978 году окончил Военно-инженерную Краснознаменную академию им. А. Ф. Можайского по специальности «Системы управления, электропроверочное и специальное оборудование летательных аппаратов». В 1993 году защитил диссертацию на соискание ученой степени доктора технических наук. Является автором 120 научных публикаций и двух патентов на изобретения. Область научных интересов — методы анализа и синтеза систем управления летательных аппаратов.
Эл. адрес: vladar56@mail.ru

ВОЛКОВ
Владимир
Юрьевич



Профессор АО «НПП «Радар ммс», кафедры радиотехнических систем Санкт-Петербургского государственного университета аэрокосмического приборостроения, кафедры радиотехнических систем Санкт-Петербургского электротехнического университета. В 1972 году окончил Ленинградский институт авиационного приборостроения по специальности «Радиоэлектронные устройства». В 1993 году защитил диссертацию на соискание ученой степени кандидата технических наук. Является автором более 200 научных публикаций. Область научных интересов — методы распознавания и селекции объектов на цифровых изображениях, обработка сигналов и изображений в условиях априорной неопределенности, радиолокационные системы.
Эл. адрес: vl_volk@mail.ru

ГУРЬЯНОВ
Денис
Юрьевич



Доцент кафедры комплексного обеспечения информационной безопасности Государственного университета морского и речного флота им. адмирала С. О. Макарова, Санкт-Петербург. В 2007 году окончил Санкт-Петербургский государственный университет водных коммуникаций по специальности «Комплексное обеспечение информационной безопасности автоматизированных систем». В 2011 году защитил диссертацию на соискание ученой степени кандидата технических наук. Является автором 12 научных публикаций и двух учебных пособий. Область научных интересов — информационная безопасность, криптография, практическая реализация криптографических протоколов.
Эл. адрес: rightx@gmail.com

БУРКОВ
Артем
Андреевич



Ассистент кафедры инфокоммуникационных систем Санкт-Петербургского государственного университета аэрокосмического приборостроения. В 2017 году окончил с отличием Санкт-Петербургский государственный университет аэрокосмического приборостроения по специальности «Инфокоммуникационные технологии и системы связи». Является автором десяти научных публикаций. Область научных интересов — беспроводные системы передачи данных; алгоритмы случайного множественного доступа; системы с гибридной решающей обратной связью; цифровая обработка изображений; сжатие данных.
Эл. адрес: a.burkov@k36.org

ГЕЙДАРОВ
Полад
Шахмалы оглы



Доцент, ведущий научный сотрудник Института систем управления НАН Азербайджана, Баку. В 1996 году окончил Азербайджанский политехнический институт им. Н. Туси по специальности «Инженер системотехник». В 2006 году защитил диссертацию на соискание ученой степени кандидата технических наук. Является автором 45 научных публикаций. Область научных интересов — распознавания образов, нейросетевые технологии и модели, оптико-информационные системы, ГИС-системы, электронное правительство, электронная наука.
Эл. адрес: plbaku2010@gmail.com

ДАО
Минх Ху



Магистрант факультета информационных технологий Института науки и техники Данагского университета, Вьетнам. В 2019 году окончил Институт науки и техники Данагского университета по специальности «Информационные технологии». Область научных интересов — беспроводная сеть, машинное обучение, SDN.
Эл. адрес: minh.daohuu@gmail.com

ДОРОНИНА
Юлия
Валентиновна



Профессор кафедры информационных систем Севастопольского государственного университета; старший научный сотрудник научно-исследовательской лаборатории ЧВВМУ им П. С. Нахимова, Севастополь.

В 1993 году окончила Севастопольский государственный технический университет по специальности «Вычислительные машины, комплексы, системы и сети». В 2014 году защитила диссертацию на соискание ученой степени доктора технических наук.

Является автором более 100 научных публикаций.

Область научных интересов — системный анализ, каскадно-иерархические модели, синтез систем с переменной структурой, квалиметрия моделей Маркова, реинжиниринг информационных систем.

Эл. адрес: juvado@yandex.ru

ЛЕ
Дык Чан



Лектор факультета информационных технологий Института науки и техники Данагского университета, Вьетнам.

В 2014 году окончил Санкт-Петербургский государственный университет телекоммуникаций им. профессора М. А. Бонч-Бруевича по специальности «Многоканальные телекоммуникационные системы».

В 2018 году защитил диссертацию на соискание ученой степени кандидата технических наук. Является автором 15 научных публикаций.

Область научных интересов — беспроводная сеть, VANET, FANET, SDN.

Эл. адрес: letranduc@dut.udn.vn

МОЛДОВЯН
Александр
Андреевич



Профессор, главный научный сотрудник лаборатории кибербезопасности и постквантовых криптосистем Санкт-Петербургского института информатики и автоматизации РАН.

В 1974 году окончил Ленинградский электротехнический институт им. В. И. Ульянова (Ленина) по специальности «Автоматизированные системы управления». В 2005 году защитил диссертацию на соискание ученой степени доктора технических наук.

Является автором более 200 научных публикаций и 60 патентов на изобретения.

Область научных интересов — компьютерная безопасность, защита информации, криптография, протоколы электронной цифровой подписи.

Эл. адрес: maa1305@yandex.ru

МОЛДОВЯН
Дмитрий
Николаевич



Научный сотрудник лаборатории кибербезопасности и постквантовых криптосистем Санкт-Петербургского института информатики и автоматизации РАН.

В 2009 году окончил Ленинградский электротехнический институт им. В. И. Ульянова (Ленина) по специальности «Компьютерная безопасность».

В 2012 году защитил диссертацию на соискание ученой степени кандидата технических наук. Является автором 79 научных публикаций и шести патентов на изобретения.

Область научных интересов — информационная безопасность, защита информации, криптосистемы с открытым ключом, постквантовая криптография, конечные некоммутативные алгебры.

Эл. адрес: mdn.spectr@mail.ru

НГУЭН
Куен Ле Тхи



Лектор факультета информационных технологий Института науки и техники Данагского университета, Вьетнам.

В 2014 году окончила Институт науки и техники Данагского университета по специальности «Информационные технологии», в 2018 году — магистратуру в Университете Квангвун (Kwangwoon), Корея.

Является автором пяти научных публикаций.

Область научных интересов — NFV, SDN, информационная безопасность.

Эл. адрес: ntlequyen@dut.udn.vn

ПЯТАЕВ
Алексей
Сергеевич



Аспирант Сибирского государственного университета науки и технологий им. академика М. Ф. Решетнёва, Красноярск.

В 2013 году окончил магистратуру Сибирского федерального университета по специальности «Информатика и вычислительная техника».

Область научных интересов — анализ изображений, интеллектуальные информационные системы.

Эл. адрес: pyataev.alex@gmail.com

СКАТКОВ
Александр
Владимирович



Профессор кафедры информационных технологий и компьютерных систем Севастопольского государственного университета, эксперт РАН, заслуженный работник образования Автономной Республики Крым.

В 1969 году окончил Севастопольский приборостроительный институт по специальности «Математические и счетно-решающие приборы и устройства».

В 1983 году защитил диссертацию на соискание ученой степени доктора технических наук.

Является автором более 300 научных публикаций.

Область научных интересов — системный анализ, исследование операций, прикладная математика, моделирование, численные методы, распределенные вычислительные системы, автоматизированные системы управления.
Эл. адрес: kvf.sevntu@gmail.com

ТЮРЛИКОВ
Андрей
Михайлович



Профессор, директор Института информационных систем и защиты информации, заведующий кафедрой инфокоммуникационных систем Санкт-Петербургского университета аэрокосмического приборостроения.

В 1980 году окончил Ленинградский институт авиационного приборостроения по специальности «Информационные системы управления».

В 2011 году защитил диссертацию на соискание ученой степени доктора технических наук.

Является автором более 80 научных публикаций.

Область научных интересов — многообъектные системы связи, системы дистанционного обучения, протоколы передачи данных в реальном масштабе времени, алгоритмы сжатия видеoinформации.
Эл. адрес: turlikov@vu.spb.ru

ФАВОРСКАЯ
Маргарита
Николаевна



Профессор, заведующая кафедрой информатики и вычислительной техники Сибирского государственного университета науки и технологий им. академика М. Ф. Решетнёва, Красноярск.

В 1980 году окончила Рыбинский авиационный технологический институт по специальности «Конструирование и производство электронно-вычислительной аппаратуры».

В 2011 году защитила диссертацию на соискание ученой степени доктора технических наук.

Является автором около 160 научных публикаций.

Область научных интересов — распознавание образов, цифровая обработка изображений, кластерный анализ, интеллектуальные технологии обработки данных и др.
Эл. адрес: favorskaya@sibsau.ru

ХОМОНЕНКО
Анатолий
Дмитриевич



Профессор, заведующий кафедрой информационных и вычислительных систем Петербургского государственного университета путей сообщения Императора Александра I.

В 1975 году окончил Военный инженерный Краснознаменный институт им. А. Ф. Можайского по специальности «Электронно-вычислительная техника».

В 1991 году защитил диссертацию на соискание ученой степени доктора технических наук.

Является автором более 190 научных публикаций и 25 авторских свидетельств на изобретения.

Область научных интересов — численные методы теории массового обслуживания, программирование, компьютерная безопасность, базы данных и информационные системы, системы искусственного интеллекта.
Эл. адрес: khomon@mail.ru

ШНЕЕР
Всеволод
Владиславович



Доцент школы математических и компьютерных наук университета Хериота—Уатта (Heriot—Watt University), Великобритания.

В 2003 окончил с отличием механико-математический факультет Новосибирского государственного университета по специальности «Математика».

В 2006 году защитил диссертацию на соискание ученой степени PhD в университете Хериота—Уатта (Heriot—Watt University).

Является автором 35 научных публикаций.

Область научных интересов — прикладная вероятность.
Эл. адрес: v.shneer@hw.ac.uk

ЯДРЕНКИН
Андрей
Александрович



Доцент, начальник кафедры бортовых информационных и измерительных комплексов Военно-космической академии им. А. Ф. Можайского, Санкт-Петербург.

В 1995 году окончил Военную инженерно-космическую академию им. А. Ф. Можайского по специальности «Оптико-электронные приборы и системы».

В 2003 году защитил диссертацию на соискание ученой степени кандидата технических наук.

Является автором 50 научных публикаций.

Область научных интересов — методы анализа и синтеза систем управления летательных аппаратов.
Эл. адрес: andrei_nikita@mail.ru