ISSN 1684–8853

ИНФОРМАЦИОННО-УПРАВЛЯЮЩИЕ СИСТЕМЫ

НАУЧНЫЙ ЖУРНАЛ

1(80)/2016

ISSN 1684-8853

SCIENTIFIC JOURNAL

1(80)/2016

INFORMATSIONNO-UPRAVLIAIUSHCHIE SISTEMY (INFORMATION AND CONTROL SYSTEMS)

REFEREED EDITION

Founder
«Information and Control Systems», Ltd.
Publisher
Saint-Petersburg State University
of Aerospace Instrumentation

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

The Journal was registered in the Ministry of Press, Broadcasting and Mass Media of the Russian Federation. Registration Certificate JD № 77-12412 from April, 19, 2002.
Re-registration in the Federal Service for Supervision in the Sphere of Telecom, Information Technologies and Mass Communications (ROSKOMNADZOR) due to change of the founder: «Information and Control Systems», Ltd., JD № FS77-49181 from March, 30, 2012.

Submitted for publication 11.01.16. Passed for printing 20.02.16. Format 60×84$_{1/8}$.
Offset paper. Phototype SchoolBookC. Offset printing.

Layout original is made at the Editorial and Publishing Center, SUAI.
67, B. Morskaia, 190000, St. Petersburg, Russia
Printed from slides at the Editorial and Publishing Center, SUAI.
67, B. Morskaia, 190000, St. Petersburg, Russia

The journal is distributed by subscription. Subscription can be made in the Editorial and publishing center, SUAI as well as in any post office based on «Rospechat» catalogue:
№ 48060 — annual subscript, № 15385 — semiannual subscript.

ISSN 1684-8853

НАУЧНЫЙ ЖУРНАЛ

1(80)/2016

ИНФОРМАЦИОННО-УПРАВЛЯЮЩИЕ СИСТЕМЫ

РЕЦЕНЗИРУЕМОЕ ИЗДАНИЕ

Учредитель
ООО «Информационно-управляющие системы»

Издатель
Санкт-Петербургский государственный университет аэрокосмического приборостроения

Адрес редакции: 190000, Санкт-Петербург,
Б. Морская ул., д. 67, ГУАП, РИЦ
Тел.: (812) 494-70-02, e-mail: 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.

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

Сдано в набор 11.01.16. Подписано в печать 20.02.16. Формат 60×84 1/8.
Бумага офсетная. Гарнитура SchoolBookC. Печать офсетная.
Усл. печ. л. 13,7. Уч.-изд. л. 17,2. Тираж 1000 экз (1-й завод 150 экз). Заказ 44.
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.

Журнал распространяется по подписке. Подписку можно оформить через редакцию, а также в любом отделении связи по каталогу «Роспечать»:
№ 48060 — годовой индекс, № 15385 — полугодовой индекс.

УДК 519.614
doi:10.15217/issn1684-8853.2016.1.2

МАТРИЦЫ МЕРСЕННА И АДАМАРА

Н. А. Балонин[a], *доктор техн. наук, профессор*
М. Б. Сергеев[a], *доктор техн. наук, профессор*
[a]*Санкт-Петербургский государственный университет аэрокосмического приборостроения, Санкт-Петербург, РФ*

Цель: *показать соответствие чисел Мерсенна, Ферма и прочих числовых последовательностей малоуровневым матрицам локального максимума детерминанта, гарантирующее как существование матриц, так и взаимное соответствие матричных портретов видам чисел: простых, пар простых чисел, степеней простых чисел.* ***Методы:*** *поиск матриц глобального и локального максимумов детерминанта ведется итерационной вычислительной процедурой, ориентированной на минимизацию максимального абсолютного значения элементов ортогональной матрицы.* ***Результаты:*** *разработана теория взаимного соответствия чисел и экстремальных матриц, упрощающая поиск неизвестных матриц обращением к классификации матриц по типам чисел. Предложено расширительное толкование гипотезы Адамара адекватными ей гипотезами о существовании семейств малоуровневых квазиортогональных матриц. Приведено доказательство существования матриц Мерсенна и, следствием, доказательство существования матриц Адамара. На основе арифметики конечных полей Галуа построены алгоритмы вычисления матриц Мерсенна, согласованные по результатам с оптимизационными процедурами повышения детерминанта и дополняемые ими.* ***Практическая значимость:*** *малоуровневые матрицы локального максимума детерминанта ортогональны и имеют непосредственное практическое значение для задач помехоустойчивого кодирования, сжатия и маскирования видеоинформации.*

Ключевые слова *— ортогональные матрицы, матрицы Мерсенна, матрицы Адамара, гипотеза Адамара, конечные поля Галуа, циклические матрицы, негациклические матрицы, бициклические матрицы, вычислительные алгоритмы.*

Введение

Отношения чисел и математических объектов иной природы, таких как функции и матрицы, естественны, поскольку математика — единый предмет, разделенный на области. Выход на границы областей, как правило, лежит вне основного интереса отдельных научных направлений, и они развиваются автономно. Вместе с тем междисциплинарные исследования приносят свои плоды. Пример — гипотеза Таниямы. Любая эллиптическая кривая, заданная над полем рациональных чисел, характеризуется своими параметрами. Математический объект другой природы, модулярная форма, также дает некоторую последовательность чисел. В сентябре 1955 г. Ютака Танияма высказал предположение о том, что эллиптические кривые являются модулярами, т. е. нет такой эллиптической кривой, для которой не нашлась бы адекватная ей по набору параметров модулярная форма. Гипотезой заинтересовались, когда в 1985 г. Герхар Фрэй предположил, что она является обобщением Великой теоремы Ферма, поскольку любой контрпример к Великой теореме Ферма приводил в итоге к немодулярной эллиптической кривой. Гипотезу доказали, и эта история служит хорошей иллюстрацией продуктивности исследований в пограничных областях наук, когда устанавливаются связи между объектами совершенно различной, казалось бы, математической природы.

Интерес настоящей статьи связан с системами ортогональных векторов, далекими от числовых последовательностей, объектов не геометрической природы. Ортогональные базисы описываются ортогональными матрицами или *квазиортогональными* матрицами — масштабированными ортогональными матрицами с максимальным элементом, равным по модулю единице.

Наше видение состоит в том, что с числовыми последовательностями соотносятся не все такие матрицы, а только экстремальные, у которых детерминант максимален. То, что экстремальные квазиортогональные матрицы имеют порядки, соответствующие элементам числовой последовательности $4t$, где t — натуральное число, заметил еще Адамар [1]. Он высказал предположение, сходное с предположением Таниямы, о том, что экстремальные матрицы подобны «модулярам» для четных чисел вида $4t$. Гипотеза раскрывает разнообразие систем чисел и ортогональных базисов, которое в пределах даже такого сравнительно узкого предположения способно дать почву для столетних изысканий матриц Адамара.

Предположение, высказанное нами и подкрепленное примерами матриц в работе [2], состоит в том, что семейства экстремальных матриц существуют не только на четных порядках $4t$ и $4t-2$, но и на нечетных порядках $4t-1$ и $4t-3$.

Приведенные числовые последовательности распадаются на вложенные в них последовательности простых чисел p, степеней простых чисел p^m, где m — натуральное число, пар близких простых чисел p и $p+2$, чисел Мерсенна 2^k-1, где k — натуральное число, чисел Ферма $2^{2^k}+1$, где k — неотрицательное целое число, и др. Согласно предположению, на соответствующие подсемейства распадаются также и матрицы, подробно рассмотренные в работе [2].

Опираясь на отмеченное разнообразие последовательностей чисел, можно отойти от универсальных алгоритмов и кардинально ускорить поиск экстремальных матриц.

Интерпретация теоремы Эйлера — Ферма

В качестве иллюстрации тесной связи чисел и ортогональных матриц приведем нашу интерпретацию теоремы Эйлера — Ферма.

Одна из знаменитых теорем теории чисел связана с заменой *в левой части* уравнения для триады чисел Пифагора (пифагоровы тройки) квадратичной зависимости, типичной для уравнения круга, на линейную зависимость с получением $p = x^2 + y^2$. Жирар и Ферма заметили, что любое натуральное число представляется суммой не более чем четырех квадратов целых чисел. Ни одно число вида $4t-1$ не представимо в виде суммы двух квадратов. В 1749 г. Эйлер после семи лет работы и почти через сто лет после смерти Ферма доказал теорему о простых числах, согласно которой разложение числа p на сумму двух квадратов всегда возможно для чисел $4t-3$.

Форма представления ортогональных массивов (матриц Адамара) бициклом является специфической иллюстрацией теоремы Эйлера — Ферма для чисел $p = n/4 = x^2 + y^2$ (буквальная визуализация представления матрицы *двумя квадратами*). В популярной характеристике бинарных матриц в форме SDS($2p$; r, s; λ) разности $r = p-x$ и $s = p-y$ описывают число 1 или –1 матриц бицикла (просто проверяется по портретам матриц на рис. 1), $\lambda = p-x-y$ — четвертый параметр, равный числу сходных элементов в двух соседних строках.

Позднее результат развил Лагранж — теоремой о сумме четырех квадратов. Теорема утверждает, что всякое натуральное число можно представить в виде суммы четырех квадратов целых чисел. Она является основой поиска матриц Адамара в форме четырехблочного массива Гетхальса — Зейделя. С нашей точки зрения, это структурный избыток, поскольку достаточно бицикла и двойной каймы. Этой форме соответствует разложение матрицы Мерсенна (на единицу меньшего порядка) в виде бицикла и одинарной каймы.

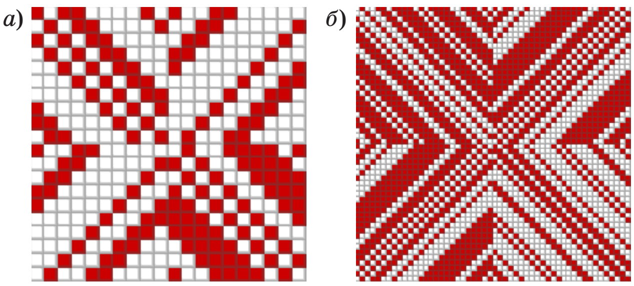
Важную для теоремы лемму о том, что произведение сумм четырех квадратов есть сумма четырех квадратов, доказал Эйлер, однако саму теорему доказал Лагранж в 1770 г. Доказательство теоремы представляет собой алгоритм, позволяющий находить разложение числа N с помощью $O(N^2\log_2\{N\})$ арифметических операций.

Матрицы Ферма и золотого сечения

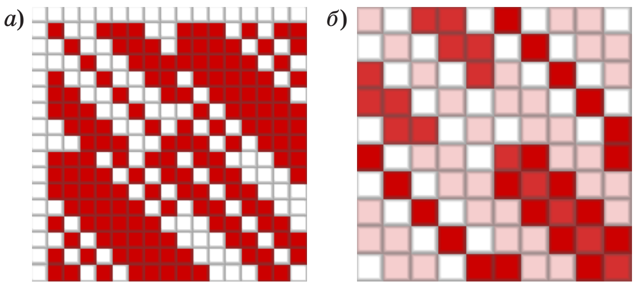
Матрицы Ферма [2, 3] порядка 17 и золотого сечения [4, 5] порядка 10 приведены на рис. 2.

Простые числа Ферма F_k являются классическими объектами теории чисел, известно пять таких чисел: 3, 5, 17, 257, 65 537. В 1796 г. Карл Фридрих Гаусс обнаружил неожиданную связь между ними и геометрическими фигурами, вписав в круг правильный семнадцатиугольник и доказав более общее положение, что если число сторон правильного многоугольника равно простому числу Ферма, то его можно построить при помощи циркуля и линейки.

Поскольку между числами и матрицами есть соответствие, то эта связь должна иметь продолжение у матриц Ферма. Матрицы Ферма — матрицы, построенные расширением регулярных матриц Адамара (суммы элементов которых по строкам и по столбцам равны) каймой. Это матрицы ортогонализуемые параметрически без изменения структуры выбором значений элементов каймы и отрицательных элементов. Первые три целочисленные матрицы Ферма порядков 3, 5, 17 — матрицы глобального максимума детерминанта $F_{k-1}/(2F_k-1)^{1/2}B$, где B — ограничение на детерминант сверху, данное Гвидо Барба. Проверка оставшихся порядков 257, 65 537 не проведена, но предположение (гипотеза), что они отвечают экстремумам, в слабом варианте — локальным, логично вытекает из знаменитой теоремы Гаусса.



■ *Рис. 1.* Бициклические матрицы порядков $n = 20$ и 52: $p = 5$ и 13



■ *Рис. 2.* Матрицы Ферма порядка 17 и золотого сечения порядка 10

Другой пример не менее примечателен — в работе [5] описана матрица золотого сечения порядка 10, ее бициклическая форма построена на паре последовательностей $[g, 1, -g, -g, 1]$, $[-1, 1, g, g, 1]$ с модулями элементов 1 и g. Условие ортогональности столбцов матрицы — уравнение золотого сечения $g^2 + g - 1 = 0$. В решении фигурирует иррациональный корень $g = 0{,}618...$, известный в теории чисел Фибоначчи.

Экстремальные квазиортогональные матрицы

Для правильного восприятия материала приведем некоторые определения.

Определение 1. Ортогональной матрицей называется квадратная матрица $\mathbf{A}$ порядка n такая, что $\mathbf{A}^{\mathrm{T}}\mathbf{A} = \mathbf{I}$, где $\mathbf{I}$ — единичная матрица.

Определение 2. Квазиортогональной матрицей называется квадратная матрица $\mathbf{A}$ порядка n с ограниченными по модулю элементами $|a_{ij}| \le 1$ такая, что $\mathbf{A}^{\mathrm{T}}\mathbf{A} = \omega(n)\mathbf{I}$, где $\mathbf{I}$ — единичная матрица; $\omega(n)$ — весовая функция.

Определение 3. *Локальный максимум* $|\det(\mathbf{A})|$ квазиортогональной матрицы достигнут, если любое достаточно малое по параметрам изменение матрицы не нарушает вида уравнения связи $\mathbf{A}^{\mathrm{T}}\mathbf{A} = \omega(n)\mathbf{I}$ при свободно заданном значении веса $\omega(n)$, но приводит к уменьшению модуля детерминанта.

Таким образом, *допустимыми* являются любые изменения параметров варьируемой матрицы, не нарушающие условие ортогональности ее столбцов. Весовая функция в задачах поиска условного максимума $|\det(\mathbf{A})|$ при ограничении вида $\mathbf{A}^{\mathrm{T}}\mathbf{A} = \omega(n)\mathbf{I}$ заранее не задана, и сама является предметом поиска. В этом состоит важное их отличие от задачи поиска матриц Адамара с жестко заданным заранее весом $\omega(n) = n$.

Квазиортогональные матрицы имеют следующие принципы деления их на семейства и подсемейства по аналогии с семействами и подсемействами чисел. Крупное семейство матриц порядков, равных числам некоторой достаточно общей числовой последовательности, отличается от прочих матриц количеством различных элементов (уровней). Например, матрицы Адамара с элементами 1 и –1 являются двухуровневыми матрицами, причем значения уровней не зависят от их порядков, функции уровней — константы.

Для экстремальных двухуровневых матриц один из уровней равен 1 (или –1), иначе значение детерминанта матрицы можно повысить элементарным масштабированием. Следовательно, варьируемая функция уровня семейства двухуровневых матриц — всего одна. К таким матрицам примыкают матрицы четных порядков, удвоение уровней наблюдается здесь лишь в силу изменения знака при блоках в блочных конструкциях.

Подсемейства квазиортогональных матриц строятся на порядках, вложенных в основные числовые последовательности. Элементы подсемейств различаются между собой структурами. Простейшими структурами являются циклические, бициклические и негациклические матрицы. Если регулярная структура не реализуема, появляются более сложные структуры, содержащие в своем составе циклические, обратные циклические, негациклические и др. блоки составных матриц, а также кайму.

Стоит отметить также возможность выделения некоторой универсальной структуры для всех подсемейств семейства матриц, разрешимой для любого вложения. Например, для матриц Адамара такой является структура из четырех блоков. Отделять универсальные структуры от частных полезно, они различимы на всех порядках $4t$, $4t-1$, $4t-2$, $4t-3$ $(4t+1)$ основных семейств.

Семейство квазиортогональных матриц порядков $4t-3$ дисперсное, значения порядков вложенных матриц нарастают в нем неаддитивно, поскольку по своей конструкции матрицы являются специфичными производными от основных матриц. В данном семействе есть оригинальные подсемейства с нарастающими по величине пропусками порядков и особыми точками, где матриц не существует или их существование ставится под сомнение. Свойство числа 9 последовательности $4t-3$ распадаться на пару множителей 3 последовательности $4t-1$ находит свое отражение в блочной структуре матрицы Якобсталя (основе матрицы Белевича из семейства порядков $4t-2$). Для разделения сугубо различных между собой подсемейств удобно пользоваться разными обозначениями $4t-3$ (матрицы Зейделя) и $4t+1$ (матрицы Ферма) этих порядков.

Семейство двухуровневых матриц Адамара порядков $4t$ характеризуется глобальным максимумом детерминанта. Как уже оказалось, для матриц соседних семейств важно не то, что максимум глобальный — он может быть и локальным, а то, что число уровней минимально — два.

Осознав это, можно уйти от исследования матриц с глобальным максимумом детерминанта по причинам, рассмотренным в работе [2]. Матрицы, отличающиеся *глобальным максимумом детерминанта* на уравнении связи $\mathbf{A}^{\mathrm{T}}\mathbf{A} = \omega(n)\mathbf{I}$, соответствуют своему семейству чисел. На нечетных порядках $4t-1$ или $4t-3$ количество уровней не постоянно, оно растет почти линейно как $(n+1)/2$. На порядках 3 и 5 имеются 2 и 3 уровня соответственно. На последующих порядках количество уровней отличается не более чем на 1 от оценочного. На четных порядках усложнения матриц не наблюдаются, для $4t-2$ они являются трехуровневыми (иногда количество уровней больше). Особые точки на $4t-2$ — предмет отдельных исследований.

Порядок 13 специальный — это критическая точка; для него и далее *для всех нечетных порядков* количество уровней квазиортогональных матриц глобального максимума детерминанта значительно превышает приведенную выше линейную оценку. Переход от матриц с абсолютным условным экстремумом к матрицам локального максимума детерминанта принципиален.

Малое число уровней гарантирует лишь «слабый оптимум», но именно оно соответствует обширным семействам чисел. Отказ от поисков матриц глобального условного экстремума, характерного только для четных порядков, позволяет установить достаточно прочную связь между числами и квазиортогональными матрицами.

Неравенства Адамара

Неравенство Адамара 1. $|\det(\mathbf{A})| \le N_1 \times N_2 \times \ldots \times N_n$, где N_i — квадратичная норма столбца.

Соответствующая теорема доказана Адамаром [1]. На множестве квазиортогональных матриц неравенство Адамара сводится к следующему.

Неравенство Адамара 2. $|\det(\mathbf{A})| \le n^{n/2}$.

Из $\mathbf{A}^{\mathrm{T}}\mathbf{A} = \omega(n)\mathbf{I}$ следует, что $|\det(\mathbf{A})|^2 = |\omega(n)\mathbf{I}| = \omega(n)^n$, максимальное значение $\omega(n) = n$ достижимо при равенстве 1 модулей всех элементов каждого столбца. Равенство модулей и ортогональность столбцов совокупно достижимы лишь на порядках 1, 2 и далее на $4t$. На остальных порядках $|\det(\mathbf{A})| < n^{n/2}$.

Определение 4. Матрица Адамара — это квазиортогональная матрица $\mathbf{H}$ порядка n с элементами 1 и –1, для которой $\omega(n) = n$.

Значение $|\det(\mathbf{H})| = n^{n/2}$, и, соответственно, для нее достигается верхняя граница неравенства Адамара. Матрицы Адамара — это матрицы глобального максимума детерминанта, как условного, с учетом уравнения связи $\mathbf{H}^{\mathrm{T}}\mathbf{H} = n\mathbf{I}$, так и безусловного, верхняя граница справедлива и для неортогональных по столбцам матриц.

Гипотеза Адамара состоит в том, что соответствующие матрицы существуют на порядках 1, 2 и $4t$.

Определение матриц Адамара через равенство, фиксирующее ортогональность их столбцов, и возможные уровни 1 и –1, не эквивалентно их определению через абсолютный максимум детерминанта. Оптимизационная постановка шире, это обстоятельство важно отметить в связи с другими нецелочисленными двухуровневыми матрицами.

Определение через равенство приводит к определению матриц основного семейства через совокупность всех его подсемейств. Так как подсемейства неисчерпаемы, как и вложенные в базисные порядки разнообразные подсистемы чисел, определение неконструктивно — оно не способствует собиранию подсемейств в единое семейство в рамках основной и общей для них черты — иметь максимум детерминанта.

Матрицы Мерсенна

Квазиортогональные матрицы вложенных в $4t$ порядков $n = 2^k$ выделены еще Сильвестром ввиду их орнаментальных свойств — способности создавать сложный фрактальный узор совокупно с исключительно простым правилом инверсии (транспонированием с масштабированием).

Адамар нашел пару матриц с аналогичными свойствами на порядках 12 и 20, выходящих за пределы последовательности, и предложил расширить семейство до $4t$, но не указал путь его пополнения.

Аналогично можно построить квазиортогональные матрицы порядков $n = 2^k - 1$, вложенных в последовательность $4t-1$ и обладающих локальным максимумом детерминанта. Последовательность 2^k-1 называется последовательностью чисел Мерсенна, поэтому сопровождающие их матрицы названы матрицами Мерсенна [6].

Возможны два определения матриц порядков $4t-1$.

Определение 5. Матрица Мерсенна — это квазиортогональная матрица $\mathbf{M}$ порядка n с элементами $a = 1$ и $-b$, где $b = \dfrac{t}{t+\sqrt{t}}$ для $n = 2^k - 1$.

Эти матрицы существуют, алгоритмы их вычисления описаны в работах [2, 6, 7]. Заметим, что у матриц Мерсенна каждый второй порядок отличается тем, что один их уровень иррационален. Как и в случае с матрицами Адамара, матрицы Мерсенна найдены на порядках 11 и 19 [8], что позволило сформулировать гипотезу о существовании матриц с нужной функцией уровня на порядках $4t-1$ [9].

Определение 6. Матрица Мерсенна — это двухуровневая квазиортогональная матрица порядка $4t-1$, имеющая локальный максимум $|\det(\mathbf{M})|$ на уравнении $\mathbf{M}^{\mathrm{T}}\mathbf{M} = \omega(n)\mathbf{I}$ при свободной правой части $\omega(n)$, с уровнем $b = \dfrac{t}{t+\sqrt{t}}$, согласованным с уровнем матриц чисел Мерсенна 2^k-1.

В отличие от матриц Адамара, отсутствие которых на соседних нечетных порядках необходимо еще доказать, какие-либо другие порядки, кроме $4t-1$, не включены, поскольку в них не содержится последовательность чисел Мерсенна.

Оптимизация при свободной правой части $\omega(n)$ существенно отличается от оптимизации при зажатом нормировании. Не получить решение в таких условиях невозможно. Не все решения могут быть признаны матрицами Мерсенна, а только те, которые согласованы по функции уровня с матрицами Мерсенна в узком их смысле.

Функция уровня является основой классификации семейств квазиортогональных матриц и составляет дополнительный критерий отбора входящих в семейство матриц. Такой подход предлагается распространить на все семейства квазиортогональных матриц, связанных с вложенными числовыми последовательностями.

Задачи разрешимые и неразрешимые

В теории и практике разрешимых и неразрешимых задач хорошо известен пример системы линейных алгебраических уравнений (СЛАУ) $\mathbf{Ax} = \mathbf{b}$, где $\mathbf{b}$ — вектор правой части. Система считается разрешимой, если существует вектор $\mathbf{x}$, при котором соблюдается тождество левой и правой частей этого уравнения.

Неразрешимая система называется *несовместной*.

Неразрешимую задачу можно формально «решить». Большое значение для систематизации решений неразрешимых задач матричной алгебры имеют работы Пенроуза, предложившего определения псевдорешения и псевдообратной матрицы, основанные на свойстве экстремальности.

Для решения приведенной задачи необходимо изменить ее формулировку на оптимизационную. Рассмотрим невязку левой и правой части линейного уравнения $\xi = \mathbf{Ax} - \mathbf{b}$. Поставим цель оптимизировать невязку в смысле минимума квадрата ее нормы $\xi^{\mathrm{T}}\xi = \|\mathbf{Ax} - \mathbf{b}\|^2$.

Все обычные решения СЛАУ определяются вектором нулевых невязок, самым малым из возможных значений квадратичной нормы. Они являются также решениями оптимизационной задачи.

В отличие от исходной задачи точного решения СЛАУ, оптимизационная совместна при любом векторе правой части, поскольку при нежестких условиях на область поиска решения всегда есть некоторый минимум.

Следует отметить, что переход к оптимизационной трактовке переводит задачу решения формально неразрешимых задач к заведомо разрешимым.

Пенроуз предусмотрел минимизацию нормы самого решения как дополнительное условие, если основная оптимизация дает не одно решение, а множество. Так, менее ста лет назад в линейной алгебре появилось новое понятие: псевдорешение $\mathbf{x} = \mathbf{A}^{+}\mathbf{b}$, где $\mathbf{A}^{+}$ — псевдообратная матрица Пенроуза [10, 11].

Другой, более близкой рассматриваемой в данной статье задаче, является задача на разрешимость квадратичного уравнения с одной переменной $\mathbf{x}^2 = \mathbf{b}$. Для $\mathbf{b} = 2$ задача не разрешима, например, в целых числах или отношениях целых чисел. Вместе с тем геометрический объект, который уравнение описывает, — прямоугольный треугольник с катетами единичной длины — существует.

Для поиска такого «решения» нужна смена парадигмы. От равенства переходят к задаче минимизации невязки $\xi = |\mathbf{x}^2 - \mathbf{b}|$, выражая x все более точно через значения частных сумм ряда. Иными словами, осмысленное решение формально неразрешимой задачи получено аккуратной аксиоматикой, приведшей к возникновению понятия иррационального числа.

Квадратичное ограничение $\mathbf{M}^{\mathrm{T}}\mathbf{M} = \omega(n)\mathbf{I}$ — матричное обобщение уравнения $\mathbf{x}^2 = \mathbf{b}$. Поиск целочисленного решения при зажатой правой части, как это делается при ограничении $\mathbf{H}^{\mathrm{T}}\mathbf{H} = n\mathbf{I}$, приводит к той же проблеме, что и в элементарном частном случае. Поступим иначе — будем искать экстремальные по детерминанту квазиортогональные матрицы определенных порядков $4t-1$, не зажимая условие в правой части, т. е. не назначая $\omega(n)$. Эта функция задается не априорно, а узнается апостериорно. После оптимизации из имеющихся экстремальных решений должно быть выбрано одно, согласованное с решением для числовой последовательности Мерсенна.

Это иной путь решения, повторяющий в некоторых деталях логику определения псевдообратных матриц по Пенроузу.

Смена парадигмы

Новый подход приводит к широкой программе исследований, в которой, в силу наличия различных числовых последовательностей и связей чисел с матрицами, возникают предположения и утверждения.

Мы утверждаем, что:

— числам Мерсенна соответствуют специальным образом определенные матрицы Мерсенна, числам Ферма — матрицы Ферма и т. п.;

— вложенность числовой последовательности чисел Мерсенна 2^k-1 в последовательности нечетных чисел $4t-1$ обусловливает существование обобщенных матриц Мерсенна, соответствующих более широкой области чисел. У матриц, как и у чисел, существует наследование признаков принадлежности к более широким семействам;

— семейства квазиортогональных матриц различаются инвариантами, в которые входят *значения их элементов* (уровни): все матрицы Мерсенна и обобщенные матрицы Мерсенна порядков $4t-1$ имеют два возможных значения модулей уровней $a = 1$ и $b = \dfrac{t}{t+\sqrt{t}}$. Остальные семейства квазиортогональных матриц также отличаются своими функциями уровней;

— несуществование квазиортогональных матриц на порядках, соответствующих некоторым числам определенной последовательности, означает их неэквивалентность друг другу по матрицам, и там, где для возникновения неэквивалентности нет основания, справедливы теоремы, аналогичные теореме Таниямы.

В последовательности чисел $4t-1$ нет пропусков. Это означает, что существуют все без исключения обобщенные матрицы Мерсенна. Заметим, что гипотезы о существовании всех матриц Мерсенна нечетных порядков $4t-1$ и матриц Адамара четных порядков $4t$ сходны. Следуя логике теоремы Таниямы, можно говорить о сопоставлении матриц Адамара и Мерсенна между собой по их числовым инвариантам (порядкам).

Иными словами, если существуют матрицы Мерсенна, то существуют и матрицы Адамара, и наоборот.

Тождество SBIBD для матриц Мерсенна и Адамара

Для бинарных матриц порядков v количество элементов одного знака (пусть $a = 1$) в строке принято описывать параметром k. Параметр λ равен числу элементов одного знака, встречающихся попарно в двух соседних строках или столбцах. Сочетание параметров $\{v, k, \lambda\}$, которое принято называть *симметричным блочным дизайном* и обозначать устоявшейся аббревиатурой SBIBD, предложено как средство поиска междисциплинарных связей.

Один и тот же SBIBD могут иметь различные между собой математические объекты, столь же далекие, как числа и матрицы. Понятие «блок» в него пришло от графических задач с блоками, а симметрия напоминает о симметрии таких схем. Пример: матроид Фано — объект конечномерной геометрии, имеющий общий SBIBD с матрицей Адамара порядка $n = 8$.

Все матрицы Адамара характеризует один и тот же SBIBD $\{4t-1, 2t, t\}$, называемый *адамаровым дизайном* [12].

Если существует один объект какой-либо одной математической природы (блок, схема или таблица чисел) с SBIBD $\{v, k, \lambda\}$, то существуют и другие объекты, описываемые тем же самым SBIBD. Этим свойством пользуются для нахождения особенно сложных матриц Адамара или сходных с ними трехуровневых конференц-матриц Белевича. Так была найдена матрица Белевича порядка $n = 46$ конструкции Матона [13].

В адамаровом дизайне $\{4t-1, 2t, t\}$ понятно все, кроме размера $v = 4t-1$, не имеющего прямого отношения к порядку матрицы.

Заметим, что числа положительных элементов k и попарных соответствий элементов λ рассчитываются на всю матрицу. Наиболее часто происхождение несогласованного с порядком матрицы числа $v = 4t-1$ объясняется тем, что оно описывает порядок основы (core) нормализованной матрицы Адамара без ее каймы в предположении, что она не несет информационной нагрузки.

Числа $4t-1$ являются порядками обобщенных матриц Мерсенна, которые имеют два уровня (бинарны) и могут быть описаны SBIBD. Размер дизайна соответствует *мерсеннову дизайну*, поскольку в нем фигурирует порядок матриц Мерсенна. Покажем, что приведенный выше дизайн является общим для матриц Адамара и Мерсенна.

Параметры k, λ связаны с коэффициентами квадратичного условия связи. Произведение двух соседних строк матрицы содержит λ произведений aa, $2(k-\lambda)$ произведений ab ($k-\lambda$ элементов a каждой из строк умножено на b), и остающиеся $n-2k+\lambda$ произведений b^2. Согласно $\mathbf{A}^{\mathrm{T}}\mathbf{A} = \omega(n)\mathbf{I}$, оно равно нулю, что дает уравнение $(n-2k+\lambda) \times b^2 - 2(k-\lambda)ab + \lambda a^2 = 0$, которое будем называть *характеристическим*. Оно в сжатой форме выражает условие ортогональности бинарной по уровням матрицы.

После подстановки значений $\{n, k, \lambda\} = \{4t-1, 2t, t\}$ характеристическое уравнение сводится к $(t-1)b^2 - 2tba + ta^2 = 0$. Положительный корень полинома левой части $b = \dfrac{t}{t+\sqrt{t}}$ является модульным уровнем матриц Мерсенна.

Следовательно, SBIBD — прямое параметрическое описание матриц Мерсенна порядков $4t-1$, отвечающее характеристическому уравнению для искомого варьируемого уровня. Если матрицы порядков $4t-1$ с иррациональными уровнями, называемые матрицами Мерсенна, существуют, то существуют и целочисленные матрицы Адамара. И наоборот: если существуют матрицы Адамара, то существуют и матрицы Мерсенна.

Существование матриц Мерсенна

Заметим, что в задачах определения существования решения не требуется уметь находить оптимальные матрицы. Важно указать на неразрешимое противоречие, которое возникает, если решения задачи нет.

Переход от целочисленных задач к задачам с иррациональными элементами описан выше, остается доказать существование матриц Мерсенна, причем в облегченной постановке, когда они рассматриваются как результат оптимизации. Задача на оптимум не меняется для порядков, входящих в последовательность $4t-1$, она одинаково формулируется для любого t, и нет причин выделять одну матрицу среди других, опираясь только на различие в значениях t. Хотя искомых матриц бесконечно много, для нахождения не самих матриц, а всего лишь функции модуля уровня в виде формул $b = b(n)$ или $b = b(t)$, не обязательно искать все такие матрицы.

Пусть матрица порядка $n = 4t - 1$ имеет уровни $(a = 1, -b)$ при $b \le 1$.

Значение $b = b(t)$ удовлетворяет квадратичному условию связи $\mathbf{M}^{\mathrm{T}}\mathbf{M} = \omega(n)\mathbf{I}$, т. е. это корень полинома второго порядка $a_2b^2 + a_2ba + a_0a^2 = 0$. Для идентификации зависимости трех коэффициентов полинома a_2, a_2, a_0 от порядка матрицы (а значит, и искомой формулы для корня) нам достаточно найти несколько таких матриц на порядках, все равно каких, поскольку они не отличаются и ничем не предпочтительны. В табл. 1 приведены найденные нами для порядков, равных первым числам Мерсенна $n = 2^k - 1$, параметры матриц и полиномы, которым они удовлетворяют согласно условию $\mathbf{M}^{\mathrm{T}}\mathbf{M} = \omega(n)\mathbf{I}$, отвечающих при свободно заданной правой части $\omega(n)$ локальному максимуму детерминанта. Они найдены не из уравнений, а строго в оптимизационной постановке, включая многие порядки $n = 4t - 1$ (в табл. 1 содержится только часть обработанного нами материала, для иллюстрации).

Таблицу можно продолжить, заметив, что для $n = 4t - 1$ все описанные ею полиномы являются частными случаями полинома $(t-1)b^2 - 2tba + ta^2 = 0$.

Отсюда, при уровне $a = 1$, имеем формулу для $b = \dfrac{t}{t+\sqrt{t}}$, график этой зависимости приведен на рис. 3.

Зависимость коэффициентов b матриц Мерсенна от порядка n не содержит особых точек, свидетельствующих об отсутствии матриц. Так как функция уровня монотонна и *не имеет особых точек* на области ее определения $n = 4t - 1$, отсюда следует вывод, что такая оптимизационная задача совместна всегда. В противном случае мы пришли бы к противоречию: уровень предвычислен, включая даже возможный тут иррациональный уровень, а матрицы нет. Это свидетельствовало бы о неоднородности такой задачи, а она, как нам известно, одинакова для всех порядков.

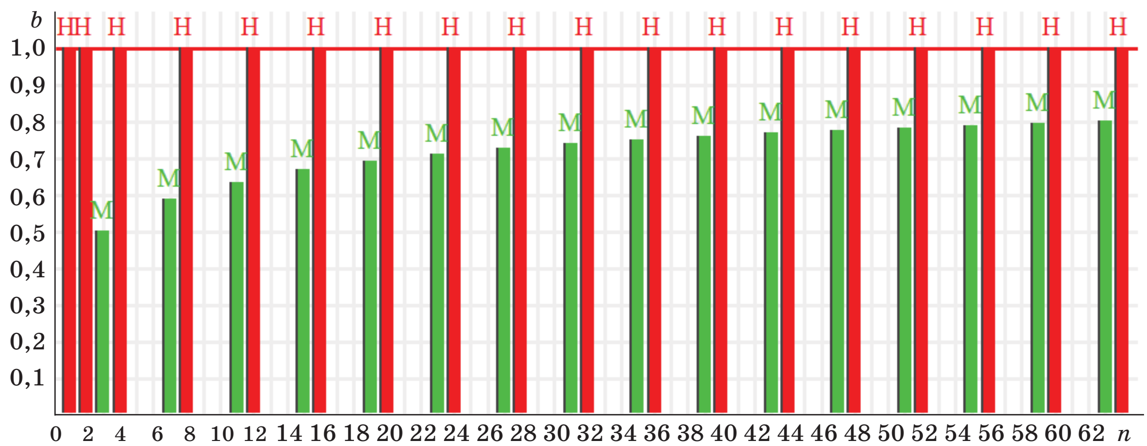
Итак, матрицы Мерсенна существуют для всей оси порядков вида $n = 4t - 1$.

■ ***Таблица 1.*** Полиномы связи и значения уровней матриц Мерсенна

k	Порядок $n = 2^k - 1$	Полином	Уровни
1	1	$b = a$	$b = a$
2	3	$2b - a = 0$	$b = a/2$
3	7	$b^2 - 4ab + 2a^2 = 0$	$b = \left(2 - \sqrt{2}\right)a$
4	15	$3b^2 - 8ab + 4a^2 = 0$	$b = 2a/3$
5	31	$7b^2 - 16ab + 8a^2 = 0$	$b = \left(8 - 2\sqrt{2}\right)a/7$
6	63	$15b^2 - 32ab + 16a^2 = 0$	$b = 4a/5$
7	127	$31b^2 - 64ab + 32a^2 = 0$	$b = \left(32 - 4\sqrt{2}\right)a/31$
8	255	$63b^2 - 128ab + 64a^2 = 0$	$b = 8a/9$

Выводы из существования матриц Мерсенна

Выводы касаются не только порядков матриц Адамара, но и всей числовой оси — всех последовательностей чисел и всех квазиортогональных матриц. Выделенные числами экстремальные



■ ***Рис. 3.*** Зависимость уровней матриц Мерсенна (**М**) и Адамара (**Н**) от порядка n

■ *Таблица 2.* Значения уровней и функции веса семейств матриц

Порядок матрицы n	Матрица	Возможные значения элементов матрицы	Функция веса $\omega(n)$
$4t$	Адамара	1, –1	n
$2t$, $4t$	Белевича	1, –1, 0	$n-1$
t, $2t$, $3t$, $4t$	Себерри (взвешенная)	1, –1, 0	$n-k$
$4t-1$	Мерсенна	1, $-b$, где $b=\dfrac{t}{t+\sqrt{t}}$	$((n+1)+(n-1)b^2)/2=2t+(2t-1)b^2$
$4t-2$	Эйлера*	1, $-b$, где $b=\dfrac{t}{t+\sqrt{2t}}$	$((n+2)+(n-2)b^2)/2=2t+(2t-2)b^2$
$4t-3$	Зейделя	1, $-b$, d, где $b=1-2d$, $d=\dfrac{1}{1+\sqrt{n}}$	$(n-1)(1+b^2)/2+d^2=2(t-1)(1+b^2)+d^2$
$4t-3$	Ферма	1, $-b$, s, где $q=n-1=4u^2$, $p=q+\sqrt{q}$, $b=\dfrac{2n-p}{p}=1-\dfrac{2u-1}{2u+1}\times\dfrac{1}{u}$, $s=\dfrac{\sqrt{nq-2\sqrt{q}}}{p}=\dfrac{\sqrt{nu-1}}{2u+1}\times\dfrac{1}{\sqrt{u}}$	$1+4u^2s^2=k+(q-k)b^2+s^2$, где $k=\dfrac{q-\sqrt{q}}{2}=2u^2-u$

* Для матрицы Эйлера четного порядка указаны значения двух блоков ее бицикла.

ортогональные базисы существуют, им отвечают как целочисленные, так и иррациональные матрицы.

К матрицам Мерсенна примыкают, например, матрицы Эйлера [2], дополняющие порядки $4k$ и $4k-1$ четными порядками $4k-2$, не кратными 4.

Порядки матриц Ферма $2^{2^k}+1$ погружены в более общее семейство чисел $2^{2k}+1$, на которые соответствующие матрицы можно распространить. Они вложены в $4t+1$ не прямо, а опосредованно. Более общее их описание состоит в том, что они соответствуют произведениям близких пар чисел $4t-1$ и $4t-3$ или (с точность до 1) взвешенным квадратам чисел, т. е. $4t^2+1$. Порядки идут неравномерно 3, 5, 17, 35, 65, ..., среди них встречаются все числа Ферма 3, 5, 17, ... Среди меньших на 1 чисел им соответствуют порядки регулярных матриц Адамара, для которых все суммы столбцов и строк равны друг другу.

Матрицы Ферма, Адамара, Мерсенна, Эйлера (соседствующие с ними матрицы Белевича и взвешенные матрицы) и Зейделя, описанные в работе [2], отвечают основным числовым системам для порядков $n=4t+1$, $n=4t$, $n=4t-1$, $n=4t-2$, $n=4t-3$ соответственно. Матрицы Зейделя порядков $n=4t-3$ сопровождают матрицы Белевича порядков $n=4t-2$, это не экстремальные матрицы, а матрицы седловых точек. Они существуют не всегда, но если нет матрицы Зейделя, то нет и матрицы Белевича, и наоборот. Матриц Зейделя нет, если их порядок $n=4t-3$ не выразить суммой двух целых чисел. Это повторение взаимоотношений матриц Адамара порядков $n=4t$ и Мерсенна порядков $n=4t-1$.

Основная ветвь матриц описывается SBIBD $\{4t-1,\ 2t,\ t\}$, включающим в себя непосредственно параметры матриц Мерсенна (табл. 2).

Примеры подсемейств матриц Мерсенна

Получение подсемейств матриц Мерсенна опирается на различие числовых последовательностей, входящих в $n=4t-1$. Эти нечетные числа, включая числа Мерсенна, распадаются на простые (или степени простых чисел) и прочие. Для вычисления матриц соответствующих порядков в первом случае естественно привлекать конечные поля Галуа. Второй случай сложнее, поскольку конечное поле не может быть использовано. Особняком стоят порядки, равные произведениям пар простых чисел $n=p(p+2)$. Близким соседством к степеням простого числа 2^m отличаются и числа Мерсенна $n=2^m-1$, поэтому для них существует процедура с использованием поля Галуа GF(2), хотя сами по себе числа Мерсенна могут быть простыми и составными.

Сказывается выделенность числовых последовательностей Мерсенна и Ферма, базовых для всех квазиортогональных матриц.

Расчеты в поле GF(2). Начнем примеры с этого случая. Матрицы, отвечающие порядкам чисел Мерсенна $n=2^m-1$, отличает самая простая структура — циклическая. Циклические матрицы находят, рассматривая их первую строку как выход динамической системы $\mathbf{x}_k = \mathbf{F}\mathbf{x}_{k-1}$ с компонентами вектора состояния в поле GF(2), $\mathbf{F}$ — матрица фробениусовой формы порядка m.

В имеющейся на этот счет обширной литературе свойственно подчеркивать необходимость адекватного выбора вектора начального состояния и параметров матрицы фробениусовой формы. От них, на самом деле, мало что зависит. Основной параметр здесь — порядок динамической системы m, а влияние каких-либо других чисел на вычислительный расчет означает их внутреннюю связь с числами Мерсенна, которой нет. Поэтому почти любой выбранный случайным образом вектор начального состояния и параметры динамической системы, определенные в поле GF(2), генерируют матрицу.

Итерационные процессы такого сорта хорошо известны в практике нахождения собственных векторов, теории цепей Маркова и т. п., когда вектор состояния стремится от некоторого почти произвольного начального состояния к главному собственному вектору матрицы системы. Конечное поле вносит свои коррективы, но $n = 2^m - 1$ — максимальная длина генерируемой последовательности, после которой эволюции динамической системы утрачивают свое разнообразие. По истечении n тактов дискретная система может только воспроизвести тот же самый процесс, появляется период.

Заменяя в выходном сигнале 0 на $-b$, получаем первую строку циклической матрицы, генерируемую сдвиговым регистром. Заметим, что динамическая система тоже может быть реализована на сдвиговом регистре, и этот метод усиленно рекламируется, хотя он малопродуктивен для порядков вне основной зависимости. Кроме того, в литературе этот генератор не ассоциируют с матрицами Мерсенна, хотя он генерирует именно их.

Дополняя циклическую матрицу каймой, получим матрицу Адамара порядка Сильвестра $n = 2^m$, на ней и концентрируется внимание. Это иной метод придания ортогональности циклическому массиву, не связанный с адаптацией положительного или отрицательного уровня.

Отсюда родом твиттер Мерсенна, описывающий блок в виде полосы верхних элементов матрицы Мерсенна шириною m, используемый для бинарного кодирования символов (кодов столбцов этой полосы).

Расчеты в поле GF(p). Выше была описана несложная, но большая динамическая система порядка m, теперь поле большое, и наступает черед упростить порядок системы до первого. Выход такой динамической системы первого порядка — показательная функция $x^{\lambda k}$ (экспонента, для краткости); от основания x и параметра λ мало что зависит (противоположное означало бы связи внутри числовой системы, они почти произвольны).

Значения этой экспоненты непосредственно описывают индексы отрицательных элементов $-b$ первой строки циклической матрицы Мерсенна.

Для большинства случаев подходит выбор $x = 2$ и $\lambda = 1$. Чтобы не выбирать с начальным условием, часто переходят к иному сорту комбинаторики, вычисляя степенные функции $k^2 \bmod p$. Те значения k, которые совпадут с любыми такими квадратами (мы ищем пересечение множества индексов с множествами квадратов индексов), называются квадратичными вычетами. Им присваивается значение *символа Лежандра*, равное 1. Остальным индексам соответствуют отрицательные символы Лежандра — в нашем случае выгодно считать их равными $-b$. Стартовому значению $k = 0$ отвечает 1. Трудность состоит в том, что k сравнивается не с собственным квадратом, а со всеми полученными квадратами, возникает перебор. Так что метод экспонент, описанный выше, более прост. Не надо подбирать, сразу имеем нужный нам индекс.

Для генерации матриц Мерсенна, поскольку эти матрицы пока малоизвестны, метод этот никогда не применялся, он описывается нами впервые.

Расчет для пары простых чисел p, $p+2$. Пóля Галуа размера $n = p(p+2)$ нет. Можно вычислить p значений $x^k \bmod(n)$ и $q - p$ значений $y^k \bmod(n)$ с основаниями $x = y \bmod(p)$, $x = 0 \bmod(p+2)$, где y — примитивный элемент групп GF(p) и GF($p+2$), $q = (n-1)/2$. Примитивный элемент — это любой элемент, показательная функция от которого обходит всю группу, такие элементы связаны с размерами группы и подбираются несложно, можно случайным поиском. Этот расчет отличается от предыдущего только тем, что индексы отрицательных элементов $-b$ первой строки циклической матрицы Мерсенна генерируются совокупно двумя динамическими системами первого порядка.

Метод этот никогда не применялся для генерации матриц Мерсенна по той же причине, что и предыдущий, он описывается нами впервые.

Расчеты в поле GF(p^m). В данном случае также рассчитываются две функции, содержащие сумму и разность двух «экспонент» $x^{\lambda_1 k} + x^{\lambda_2 k}$, $x^{\lambda_1 k} - x^{\lambda_2 k}$, аналогов тригонометрических функций, косинуса и синуса (при равенстве, с точностью до знака, показателей, рассчитываемых в поле GF(p)). Бинарной последовательности или последовательности индексов в таком поле нет. Все элементы его имеют размер m, это векторы. Значения «парных символов Лежандра» рассчитаем сопоставлениями $x^{\lambda 0 k}$ с двумя указанными функциями. В этой удивительной математике мало что зависит от второстепенных параметров, при выборе $\lambda_0 = \lambda_1 = 1$, $\lambda_2 = 0$ «экспонента» генератора упрощается до x^k, она сопоставляется на предмет пересечения с $x^k + 1$, $x^k - 1$.

Если экспонента пересекает функцию, то ставится 1, если нет, то –1, амплитудой отрицательного элемента заниматься пока рано. Отсюда получим две последовательности, для которых построим циклические матрицы **A**, **B** размера $(n-1)/2$, $n = p^m$.

С их помощью складывается бициклическая форма (бицикл) $\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^{\mathrm{T}} & -\mathbf{A}^{\mathrm{T}} \end{pmatrix}$, охватываемая каймой с элементами из –1 и 1 напротив соответствующих матриц, первый элемент каймы 1 (рис. 4).

На завершающей стадии расчета отрицательный элемент –1 замещается на уровень матрицы Мерсенна $-b$. У этого метода есть альтернатива, когда вместо матрицы Мерсенна рассчитывается негациклическая матрица Белевича четного порядка $n = p^m + 1$. В данном случае нужна только одна последовательность (рис. 5). Она генерируется суммой экспоненты $\alpha = x^{\lambda_1 k + n/2}$ со своей степенью, т. е. функцией $\alpha + \alpha^{n-1}$. Сопоставляемая с нею на предмет пересечения вторая функция построена на квадрате основания $\beta = x^{2\lambda_2 k}$, в таком случае аддитивной добавкой в виде степени можно пренебречь, она не меняет расчет. Показатели $\lambda_1 = n-1$ (нечетное число) и $\lambda_2 = n$ (четное число) можно упростить, взяв первый параметр равным 1. Вторую функцию рассчитываем для меньшего количества $(n-1)/2$ точек, остальные будут повторениями в силу ее четного показателя. Если сумма пересекает вторую функцию, то ставится –1, если нет, то 1, первое число последовательности равно 0.

Далее матрица разбивается на четыре блока разделением четных и нечетных строк и столбцов, верхние два блока **A**, **B** после добавления к **A** единичной матрицы образуют бицикл четного порядка — матрицу Адамара. После нормирования от нее отнимают кайму, начинается стадия вытеснения –1 уровнем матрицы Мерсенна $-b$.

Этот метод расчета базируется на работах [14, 15], не соотнесенных, впрочем, с расчетом матриц Мерсенна, т. е. этот метод расчета матриц Мерсенна также нов. Упоминается он потому, что иллюстрирует сложный вид, который могут принимать матрицы подсемейств, основанных на разных системах чисел.

Нет никакого поля. Случай порядков $n = 4t - 1$ максимально общий, в него входят все предыдущие порядки. Сложность формы матрицы Мерсенна ограничена — это бициклическая матрица Мерсенна с одной каймой. Следовательно, все разнообразие числовой системы не повышает сложность структуры в сравнении с максимально сложным случаем расчета в полях Галуа. Алгоритм поиска универсален и нами освещен отдельно в работе [16]. Его общность и безотносительность к виду числовой системы, в которую входят порядки искомых матриц, подтверждают высказанные в работе [17] положения.

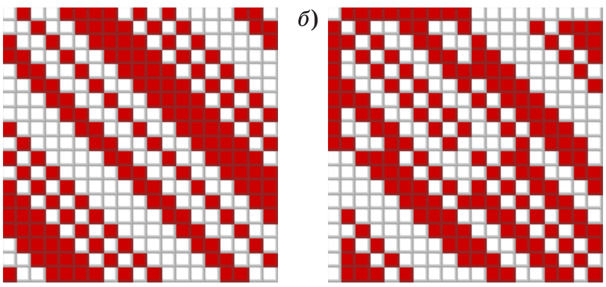
Произведение матриц Мерсенна. В силу тождества SBIBD матриц Мерсенна **M** порядка q и Адамара **H** порядка n имеет место тождество $\mathbf{H} = \begin{pmatrix} 1 & \mathbf{e}^{\mathrm{T}} \\ \mathbf{e} & \mathbf{C} \end{pmatrix}$, где **e** — вектор единичных элементов; $\mathbf{C} = -\mathrm{sign}(\mathbf{M})$. Свойством округлять отрицательные элементы матрицы Мерсенна до –1 обладает также *произведение Скарпи* [18] (Scarpic product)

$$\mathbf{H}_{nq} = \mathbf{C} \times \mathbf{C} =$$

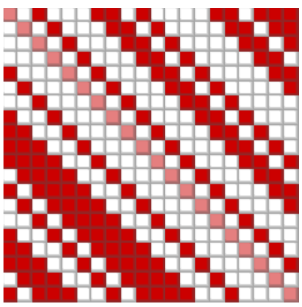
$$= \begin{pmatrix} \begin{pmatrix} 1 & c_{11}\mathbf{e}^{\mathrm{T}} \\ c_{11}\mathbf{e} & \mathbf{C} \end{pmatrix} & \begin{pmatrix} 1 & c_{12}\mathbf{e}^{\mathrm{T}} \\ c_{12}\mathbf{e} & \mathbf{C} \end{pmatrix} & \cdots & \begin{pmatrix} 1 & c_{1q}\mathbf{e}^{\mathrm{T}} \\ c_{1q}\mathbf{e} & \mathbf{C} \end{pmatrix} \\ \begin{pmatrix} 1 & c_{21}\mathbf{e}^{\mathrm{T}} \\ c_{21}\mathbf{e} & \mathbf{C} \end{pmatrix} & \begin{pmatrix} 1 & c_{22}\mathbf{e}^{\mathrm{T}} \\ c_{22}\mathbf{e} & \mathbf{TC} \end{pmatrix} & \cdots & \begin{pmatrix} 1 & c_{2q}\mathbf{e}^{\mathrm{T}} \\ c_{2q}\mathbf{e} & \mathbf{T}^{q-1}\mathbf{C} \end{pmatrix} \\ \cdots & \cdots & \ddots & \cdots \\ \begin{pmatrix} 1 & c_{q1}\mathbf{e}^{\mathrm{T}} \\ c_{q1}\mathbf{e} & \mathbf{C} \end{pmatrix} & \begin{pmatrix} 1 & c_{q2}\mathbf{e}^{\mathrm{T}} \\ c_{q2}\mathbf{e} & \mathbf{T}^{q-1}\mathbf{C} \end{pmatrix} & \cdots & \begin{pmatrix} 1 & c_{qq}\mathbf{e}^{\mathrm{T}} \\ c_{qq}\mathbf{e} & \mathbf{T}^{(q-1)(q-1)}\mathbf{C} \end{pmatrix} \end{pmatrix},$$

где **T** — матрица циклического сдвига в степени произведения индексов $(i-1)(j-1)$. Конструкция

а) б)



■ ***Рис. 4.*** Циклическая и бициклическая матрицы Мерсенна порядка 19



■ ***Рис. 5.*** Негациклическая матрица порядка 20 симметричная, сдвигать можно как строки, так и столбцы. Это аналог кронекерова произведения матриц Адамара, ориентированный на вложение друг в друга матриц Мерсенна, знаки элементов **C** оказывают влияние только на кайму.

Итоговую матрицу Адамара порядка nq, q — простое число, можно, в свою очередь, представить в виде матрицы Мерсенна размера $nq-1$ и продолжить квадрирование. Матричная степенная функция дает связанные между собой пары матриц Мерсенна и Адамара новых порядков по отношению к матрицам, вычисляемым выше. Произведение Скарпи известно применительно к матрицам Адамара [18], первое применение к матрицам Мерсенна с поднятием негативного уровня до –1 описано в работе [19].

Некоторые рекомендации по исследованиям. Несмотря на обилие литературы по полям Галуа, а может быть, ввиду этого, найти конкретное описание операции, служащей для вычисления показательных функций, сложно. Правило сложения и вычитания векторов такое же, как и в обычной векторной алгебре. Операция умножения в полиномиальной арифметике полей Галуа основана на *свертке*, которая дает примерно вдвое (менее на 1, чем вдвое) больше коэффициентов, чем нужно элементу поля.

Нередуцируемый полином, с помощью которого образуется поле $\mathrm{GF}(p^m)$, по сути, представляет собой обратную связь, с помощью которой к m коэффициентам младшей части полинома произведения прибавляются старшие.

Чаще всего это $x^m = sx^d + r$, при $d = 1$ ко всем m младшим коэффициентам свертки прибавляются все старшие с множителями s, r за исключением крайних членов, к младшему коэффициенту произведения не дотягивается ветвь с весом s, к старшему коэффициенту — ветвь с весом r.

При $d > 1$ в этом алгоритме возникают поправки, описываемые следующей вычислительной схемой для векторов $\mathbf{C} = \mathbf{AB}$: на начальной стадии $\mathbf{C} = \mathrm{conv}(\mathbf{A}, \mathbf{B})$, это обычное произведение коэффициентов полиномов, поправки формируем в цикле для индекса $0 \le i \le m$:

$c = \mathbf{C}[i]$;
if $(i < m-1)$ $\{c += r * \mathbf{C}[m + i]$;
if $(i < d-1)$ $c += r * s * \mathbf{C}[2 * m - d + i]\}$;
if $(i > d-1)$ $\{c += s * \mathbf{C}[m + i - d]$;
if $(i < 2 * d - 1)$ $c += s * s * \mathbf{C}[2 * (m - d) + i]\}\}$;
$\mathbf{C}[i] = c\%p$.

Схема полиномиального умножения напоминает нейронную, вспомогательный полином дает веса обратных связей, желательно, чтобы большую часть ветвей ликвидировали нулевые коэффициенты. Потребность в s-ветви отпадает совсем для $\mathrm{GF}(p^2)$, p — нечетное простое, произведение пары чисел $(a, b)(c, d) = (ac - rbd), (ad + bc)$. Это поле сходно с полем комплексных чисел, где $-r$ — аналог –1 (квадрат невычета). Параметр $r = 1$ (при $p + 1$ кратно 4) или $r = 3$ (при $p + 1$ кратно 6), в остальных случаях $r = 2$. Для более сложных полей параметры полинома приведены в табл. 3.

■ ***Таблица 3.*** Параметры неприводимого полинома

p	m	d	s	r	p	m	d	s	r	p	m	d	s	r
2	2	1	1	1	7	3	1	0	2	19	3	1	0	2
2	3	1	1	1	7	4	1	1	3	19	4	1	1	1
2	4	1	1	1	7	5	1	1	3	19	5	1	1	3
2	5	2	1	1	7	6	1	0	3	23	3	1	1	4
2	6	1	1	1	7	7	1	1	1	23	4	1	1	3
2	7	1	1	1	7	8	1	1	1	23	5	1	1	2
2	8	5	1	1	7	9	1	0	2	29	3	1	1	1
3	3	1	1	1	7	10	1	2	1	29	4	1	1	1
3	4	1	1	1	11	3	1	1	3	31	3	1	0	3
3	5	1	1	1	11	4	1	1	6	31	4	1	1	1
3	6	1	1	1	11	5	1	1	1	37	3	1	0	2
3	7	2	1	2	11	6	1	1	1	37	4	1	1	1
3	8	2	1	1	13	3	1	0	2	41	3	1	0	2
3	9	5	1	1	13	4	1	0	2	41	4	1	1	7
5	3	1	1	1	13	5	1	1	1	47	3	1	1	1
5	4	1	1	1	13	6	1	0	2	47	4	1	1	1
5	5	1	1	1	17	3	1	1	2	53	3	1	1	4
5	6	1	1	3	17	4	1	0	3	53	4	1	1	2
5	7	1	1	2	17	5	1	1	6	57	3	1	1	8
5	8	1	0	2	17	6	1	1	4	57	4	1	1	5

Заключение

Описав конкретные методы нахождения матриц Мерсенна, теперь позволим себе несколько фраз об общей философии рассматриваемого нами подхода. Достаточно очевидно, что матрицы Адамара — это фрактальный объект (его сравнивают с треугольником Серпинского), так же, как и все прочие такие матрицы. Мы показали, что матрицы могут быть иррациональными. Такой подход преимущественен с точки зрения меньших сомнений в существовании искомых матриц. Нет матриц — нет и соответствующих им чисел, а они есть, такая привязка дает существенные гарантии и позволяет по-новому взглянуть на старую проблему поиска целочисленных решений. Гипотеза Адамара, как и прочие сходные с ней гипотезы, уступает в своей сложности Великой теореме Ферма. Их проверка не должна быть столь же замысловатой.

В конечном поле показательная функция заметает собой все пространство и играет роль спирали, пронизывающей каждую его точку без пересечения, что свойственно моделям детерминированного хаоса. Таковы решения некоторых нелинейных дифференциальных уравнений в бесконечномерном трехмерном пространстве. 3D нужно, чтобы интегральной кривой было место, где разместиться без самопересечений. Интегральная кривая (странный аттрактор) прошивает любую точку заполняемого ею объема один раз. В конечных полях происходит нечто аналогичное. «Аттракторы» динамических систем с пространством состояния в конечном поле образованы степенями примитивного элемента мультипликативной циклической группы (или подгруппы), выступающего в качестве начального условия показательной функции. Системы просты, сложность их поведения содержится в логике модульной или полиномиальной арифметики.

Не забываем и об оптимизируемом таким движением квадратичном критерии — детерминанте. Задача в любом случае — квадратичная. Квазиортогональные матрицы глобального максимума детерминанта нечетного порядка увеличиваются в размерах путем бифуркации их уровней, причем есть критическая точка — порядок 13. Матрицы Адамара — островные области, где ни количество уровней, ни сами эти уровни не меняются с ростом порядка.

Вся числовая система связана с малоуровневыми квазиортогональными матрицами. Матрицы высоких порядков (см. рис. 3 и табл. 3) при увеличении их размера все менее отличаются друг от друга и от матриц Адамара. Варьируемый отрицательный уровень позволяет матрице Мерсенна удерживать ортогональность столбцов, причем с ростом порядка элемент $-b$ стремится к -1.

Даже если читатель не согласен с рядом выдвинутых здесь положений или затрудняется их проверить самостоятельно, все же, согласитесь, это полезные примеры познания чисел и матриц в их взаимосвязи, имеющие перспективы в теории и практике вычисления экстремальных матриц ортогональных базисов.

Благодарности

Наше исследование циклических, бициклических, негациклических структур экстремальных матриц (называемых ради большей краткости *критскими* [20]) в том виде, как оно изложено, невозможно было бы без действенной помощи и совета профессоров Дженифер Себерри (Jennifer Seberry) и Драгомира Джоковича (Dragomir Đoković).

Дж. Себбери и Д. Джокович своими интересными работами и творчеством внесли большой вклад в создание алгоритмов поиска экстремальных квазиортогональных матриц с использованием полей Галуа.

В отношении доказательства гипотезы Адамара Дж. Себерри поддерживает ту часть утверждения, которая констатирует эквивалентность матриц Мерсенна и Адамара. Согласно более осторожному подходу, если нет матрицы Адамара, то нет и матрицы Мерсенна в том первом ее определении, где она задана равенством и сопоставляется с матрицей Адамара по системе совместных числовых инвариантов [21].

Литература

1. **Hadamard J.** Résolution d'une Question Relative aux Déterminants// Bulletin des Sciences Mathématiques. 1893. Vol. 17. P. 240–246.
2. **Балонин Н. А., Сергеев М. Б.** Матрицы локального максимума детерминанта // Информационно-управляющие системы. 2014. № 1(68). С. 2–15.
3. **Балонин Н. А., Сергеев М. Б., Мироновский Л. А.** Вычисление матриц Адамара — Ферма // Информационно-управляющие системы. 2012. № 6(61). С. 90–93.
4. **Балонин Н. А., Сергеев М. Б.** Матрица золотого сечения G10 // Информационно-управляющие системы. 2013. № 6(67). С. 2–5.
5. **Balonin N. A., Vostrikov A. A., Sergeev M. B.** Two-Circulant Golden Ratio Matrices // Информационно-управляющие системы. 2014. № 5(72). С. 5–11.
6. **Балонин Н. А., Сергеев М. Б., Мироновский Л. А.** Вычисление матриц Адамара — Мерсенна // Информационно-управляющие системы. 2012. № 5(60). С. 92–94.
7. **Балонин Ю. Н.** Программный комплекс MMatrix-2 и найденные им М-матрицы // Вестник компьютерных и информационных технологий. 2013. № 10(112). С. 58–64.
8. **Балонин Н. А.** О существовании матриц Мерсенна 11-го и 19-го порядков // Информационно-управляющие системы. 2013. № 2(63). С. 89–90.
9. **Сергеев А. М.** Обобщенные матрицы Мерсенна и гипотеза Балонина // Автоматика и вычислительная техника. 2014. № 4. С. 35–43.
10. **Bellman R.** Introduction to Matrix Analysis. — Philadelphia: SIAM, 1997. — 395 p.
11. **Воеводин В. В., Кузнецов Ю. А.** Матрицы и вычисления. — М.: Наука. Гл. ред. физ.-мат. лит., 1984. — 320 с.
12. Handbook of Combinatorial Designs. Second Edition (Discrete Mathematics and its Applications). 2nd Ed. / Charles J. Colbourn, Jeffrey H. Dinitz Ed. — Chapman and Hall, 2006. — 1000 p.
13. **Balonin N. A., Seberry J.** A Review and New Symmetric Conference Matrices // Информационно-управляющие системы. 2014. № 4(71). С. 2–7.

14. **Balonin N. A., Djokovic D. Z.** Negaperiodic Golay Pairs and Hadamard Matrices // Информационно-управляющие системы. 2015. № 5(78). С. 2–17. doi:10.15217/issn1684-8853.2015.5.2
15. **Балонин Н. А., Джокович Д. Ж.** Симметрия двуциклических матриц Адамара и периодические пары Голея // Информационно-управляющие системы. 2015. № 3(76). С. 2–16. doi:10.15217/issn1684-8853.2015.3.16
16. **Балонин Н. А., Сергеев М. Б.** О значении матриц начального приближения в алгоритме поиска обобщенных взвешенных матриц глобального и локального максимума детерминанта // Информационно-управляющие системы. 2015. № 6(79). С. 2–9. doi:10.15217/issn1684-8853.2015.6.2
17. **Балонин Н. А., Сергеев М. Б.** К вопросу существования матриц Мерсенна и Адамара // Информационно-управляющие системы. 2013. № 5(66). С. 2–8.
18. **Scarpis U.** Sui Determinanti di Valore Massimo// Rendiconti della R. Istituto Lombardo di Scienze e Lettere. 1898. Vol. 31. P. 1441–1446.
19. **Балонин Н. А., Балонин Ю. Н., Сергеев М. Б.** Вычисление матриц Мерсенна и Адамара методом Скарпи // Вестник информационных технологий, механики и оптики. 2014. № 3. С. 104–112.
20. **Balonin N. A., Seberry J.** Remarks on Extremal and Maximum Determinant Matrices with Real Entries ≤ 1 // Информационно-управляющие системы. 2014. № 5(71). С. 2–4.
21. **Seberry J., Balonin N. A.** Equivalence of the Existence of Hadamard Matrices and Cretan($4t$–1,2)-Mersenne Matrices. http://arxiv.org/abs/1501.07012v1. (дата обращения: 20.12.2015).

UDC 519.614
doi:10.15217/issn1684-8853.2016.1.2

Mersenne and Hadamard Matrices

Balonin N. A.[a], Dr. Sc., Tech., Professor, korbendfs@mail.ru
Sergeev M. B.[a], Dr. Sc., Tech., Professor, mbse@mail.ru
[a]Saint-Petersburg State University of Aerospace Instrumentation, 67, B. Morskaia St., 190000, Saint-Petersburg, Russian Federation

Purpose: The goal of the paper is to show the correspondence of Mersenne numbers, Fermat numbers and numbers of other numerical sequences to multilevel matrices of local maximum determinant, which would guarantee both the existence of the matrices and the mutual correspondence of the matrix portraits to various kinds of numbers such as prime numbers, prime pairs or prime powers. **Methods:** The search for global or local maximum determinant matrices is performed by an iterative computational procedure focused on the minimization of the maximum absolute values of the elements of an orthogonal matrix. **Results:** A theory of mutual correspondence between numbers and extremal matrices has been developed, which simplifies the search for unknown matrices, classifying the matrices by types of numbers. We propose a broader interpretation of Hadamard conjecture by adequate hypotheses about the existence of multilevel quasiorthogonal matrices. A proof is given for the existence of Mersenne matrices and, as a consequence, of Hadamard matrices, too. For the calculation of Mersenne matrices, we propose algorithms based on Galois finite field arithmetics. These algorithms are concordant by results with the optimization procedures of increasing the determinant, being complemented by them. **Practical relevance:** Multilevel local maximum determinant matrices are orthogonal and have a direct practical value for the problems of error-correcting coding, video compression and masking.

Keywords — Orthogonal Matrices, Mersenne Matrices, Hadamard Matrices, Hadamard Conjecture, Galois Finite Fields, Cyclic Matrices, Negacyclic Matrices, Bicirculant Matrices, Numerical Methods.

References

1. Hadamard J. Résolution d'une Question Relative aux Déterminants. *Bulletin des Sciences Mathématiques*, 1893, vol. 17, pp. 240–246 (In French).
2. Balonin N. A., Sergeev M. B. Local Maximum Determinant Matrices *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2014, no. 1(68), pp. 2–15 (In Russian).
3. Balonin N. A., Sergeev M. B., Mironovsky L. A. Calculation of Hadamard Fermat Matrices. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2012, no. 6(61), pp. 90–93 (In Russian).
4. Balonin N. A., Sergeev M. B. Matrix of Golden Ratio G10. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2013, no. 6(67), pp. 2–5 (In Russian).
5. Balonin N. A., Vostrikov A. A., Sergeev M. B. Two-Circulant Golden Ratio Matrices. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2014, no. 5(72), pp. 5–11.
6. Balonin N. A., Sergeev M. B., Mironovsky L. A. Calculation of Hadamard–Mersenne Matrices. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2012, no. 5(60), pp. 92–94 (In Russian).
7. Balonin Yu. N. The Software Complex MMatrix-2 and Searched Minimax Matrices. *Vestnik komp'iuternykh i informatsionnykh tekhnologii* [Herald of Computer and Information Technologies], 2013, no. 10(112), pp. 58–64 (In Russian).
8. Balonin N. A. Existence of Mersenne Matrices of 11th and 19th Orders. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2013, no. 2(63), pp. 89–90 (In Russian).
9. Sergeev A. M. Generalized Mersenne Matrices and Balonin's Conjecture. *Avtomatika i vychislitel'naia tekhnika*, 2014, vol. 48, no. 4, pp. 214–220 (In Russian).
10. Bellman R. *Introduction to Matrix Analysis*. Philadelphia, SIAM, 1997. 395 p.
11. Voevodin V. V., Kuznetsov Iu. A. *Matritsy i vychisleniia* [Matrices and Calculations]. Moscow, Nauka Publ., 1984, 320 p. (In Russian).
12. *Handbook of Combinatorial Designs. Second Edition* (*Discrete Mathematics and its Applications*). 2nd Ed. (Charles J. Colbourn, Jeffrey H. Dinitz Ed.). Chapman and Hall/CRC, 2006. 1000 p.
13. Balonin N. A., Seberry J. A Review and New Symmetric Conference Matrices. *Informatsionno-upravliaiushchie*

sistemy [Information and Control Systems], 2014, no. 4(71), pp. 2–7.
14. Balonin N. A., Djokovic D. Z. Negaperiodic Golay Pairs and Hadamard Matrices. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2015, no. 5(78), pp. 2–17. doi:10.15217/issn1684-8853.2015.5.2
15. Balonin N. A., Djokovic D. Z. Symmetry of Two Circulant Hadamard Matrices and Periodic Golay Pairs. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2015, no. 3(76), pp. 2–16 (In Russian). doi:10.15217/issn1684-8853.2015.3.16
16. Balonin N. A., Sergeev M. B. Initial Approximation Matrices in Search for Generalized Weighted Matrices of Global or Local Maximum Determinant. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2015, no. 6, pp. 2–9 (In Russian). doi:10.15217/issn1684-8853.2015.6.2
17. Balonin N. A., Sergeev M. B. On the Issue of Existence of Hadamard and Mersenne Matrices. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2013, no. 5(66), pp. 2–8 (In Russian).
18. Scarpis U. Sui Determinanti di Valore Massimo. *Rendiconti della R. Istituto Lombardo di Scienze e Lettere*, 1898, vol. 31, pp. 1441–1446 (In Italian).
19. Balonin N. A., Balonin Yu. N., Sergeev M. B. Mersenne and Hadamard Matrices Calculation by Scarpis Method. *Vestnik informatsionnykh tekhnologii, mekhaniki i optiki*, 2014, no. 3, pp. 104–112 (In Russian).
20. Balonin N. A., Seberry J. Remarks on Extremal and Maximum Determinant Matrices with Moduli of Real Entries ≤ 1. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2014, no. 5(71), pp. 2–4.
21. Seberry J., Balonin N. A. *Equivalence of the Existence of Hadamard Matrices and Cretan(4t–1,2)-Mersenne Matrices*. Available at: http://arxiv.org/abs/1501.07012v1 (accessed 20 December 2015).

УДК 004.932.2
doi:10.15217/issn1684-8853.2016.1.16

ДЕТЕКТИРОВАНИЕ ДЫМА НА ОТКРЫТЫХ ПРОСТРАНСТВАХ В СЛОЖНЫХ ПОГОДНЫХ УСЛОВИЯХ С ПРИМЕНЕНИЕМ ПРОСТРАНСТВЕННО-ВРЕМЕННЫХ ЛОКАЛЬНЫХ БИНАРНЫХ ШАБЛОНОВ

М. Н. Фаворская[а]*, доктор техн. наук, профессор*
А. В. Пятаева[а]*, аспирант*
[а]*Сибирский государственный аэрокосмический университет им. академика М. Ф. Решетнёва, Красноярск, РФ*

Постановка проблемы: *раннее обнаружение дыма на открытых пространствах с использованием видеопоследовательностей является одной из важнейших задач в современных системах видеонаблюдения. Видеоинформация, полученная в результате видеосъемки в реальных условиях, может содержать объекты с динамическим поведением, помехи оборудования или линии передачи, а также артефакты, связанные с погодными условиями (например, дождем или снегом, недостаточным освещением в утренние и вечерние часы). В связи с этим возникает необходимость окончательной проверки областей-кандидатов после того, как области, похожие на дым, сегментированы в видеопоследовательности.* ***Результаты:*** *предложен алгоритм обработки изображений, содержащих артефакты, связанные с атмосферными явлениями, такими как морось и дымка, а также недостаточным освещением. Исследован метод пространственно-временных локальных бинарных, локальных тернарных и расширенных бинарных шаблонов детектирования густого и прозрачного дыма при наличии следующих артефактов: шума типа «соль-перец» со значениями до 10 дБ; аддитивного белого шума Гаусса, имитирующего атмосферные осадки; размытия изображений и слабой освещенности с применением алгоритма Ретинекс. Для классификации областей дыма использован гистограммный подход как один из наиболее простых и быстрых способов анализа изображений. В качестве меры различия гистограмм двух сопоставляемых изображений при построении соответствующего решающего правила использовалось расстояние Кульбака — Лейблера.* ***Практическая значимость:*** *разработанный метод верификации дыма на видеопоследовательностях с использованием пространственно-временных локальных бинарных шаблонов и 3D расширенных локальных бинарных шаблонов обеспечивает точность верификации 96–99 % для густого и 86–94 % для прозрачного дыма в зависимости от артефактов и шумов.*

Ключевые слова *— детектирование дыма, шум, погодные условия, пространственно-временные локальные бинарные шаблоны.*

Введение

В системах раннего обнаружения возгораний используются, в основном, противопожарные датчики, измеряющие концентрацию продуктов сгорания в воздухе. Такие системы успешно применяются для детектирования дыма в помещении, однако не дают результатов при использовании на открытых пространствах, так как концентрация продуктов сгорания быстро уменьшается из-за воздействия естественных метеорологических условий [1]. Для детектирования дыма на открытых пространствах используются оптические (иногда инфракрасные) системы видеонаблюдения. К преимуществам таких систем относятся: широкая распространенность (городские условия или специально установленные противопожарные вышки в лесных массивах), небольшая стоимость и высокое быстродействие. Одними из эффективных методов детектирования дыма на видеоизображении являются методы анализа динамических текстур, обладающих пространственной и временной регулярностью [2].

В работе [3] рассмотрен метод детектирования дыма на видеоизображении с применением одного из методов анализа динамических текстур — семейства локальных бинарных шаблонов. Рассматриваются такие виды шаблонов, как базовые бинарные шаблоны, локальные тернарные шаблоны, расширенные бинарные шаблоны. Показана эффективность распознавания густого и прозрачного дыма по видеопоследовательности. Особенностью работы является использование темпоральных шаблонов, основанных на анализе трех последовательных кадров видеопоследовательности.

При детектировании объектов на видеоизображении сложные погодные условия играют существенную роль. При этом используют различные приемы, помогающие избежать негативного влияния таких факторов, как дождь, снег, сумерки, ночь, яркое солнце, морось. Для анализа изображений с низкой освещенностью предлагаются различные методики, например, составление матрицы изменения яркости изображения [4], ручная настройка параметров системы детектирования стоп-линий на перекрестках [5], применение различных фильтров, в частности фильтра энергетических текстурных характеристик Лавса [6]. В настоящей работе предложен независимый от типа оборудования алгоритм детектирования дыма в сложных погодных условиях, а также при наличии шума.

Влияние слабой освещенности и сложных погодных условий

Рассмотрим алгоритмическую компенсацию слабой освещенности сложных погодных условий съемки при детектировании дыма по видеопоследовательностям.

Слабая освещенность

Время суток существенно влияет на особенности отображения искомого объекта на видеопоследовательности. Для эффективной работы алгоритма классификации динамических текстур необходимо рассмотреть все аспекты, от которых зависит корректность их определения. Уровень освещенности объектов на изображении является одним из основных таких аспектов, так как правильно различать предметы, например в сумерках при смене времени суток, бывает сложно даже человеку. В статье [5] показано влияние времени суток на качество работы системы видеонаблюдения по обнаружению стоп-линий на регулируемых перекрестках. При этом количество ложных срабатываний зависит от полосы движения, времени суток, солнечной или пасмурной погоды. Среднее количество ложных срабатываний варьировалось от 2 до 36 %. Для распознавания лиц при изменяющемся освещении предлагается алгоритм, основанный на представлении исходного изображения в виде матрицы изменения яркостей, что позволяет добиться устойчивости относительно изменений условий освещенности [4].

Для получения изображений, имитирующих низкую освещенность, использован алгоритм Ретинекс (Retinex) на базе локального сглаживания [7], представленный уравнением

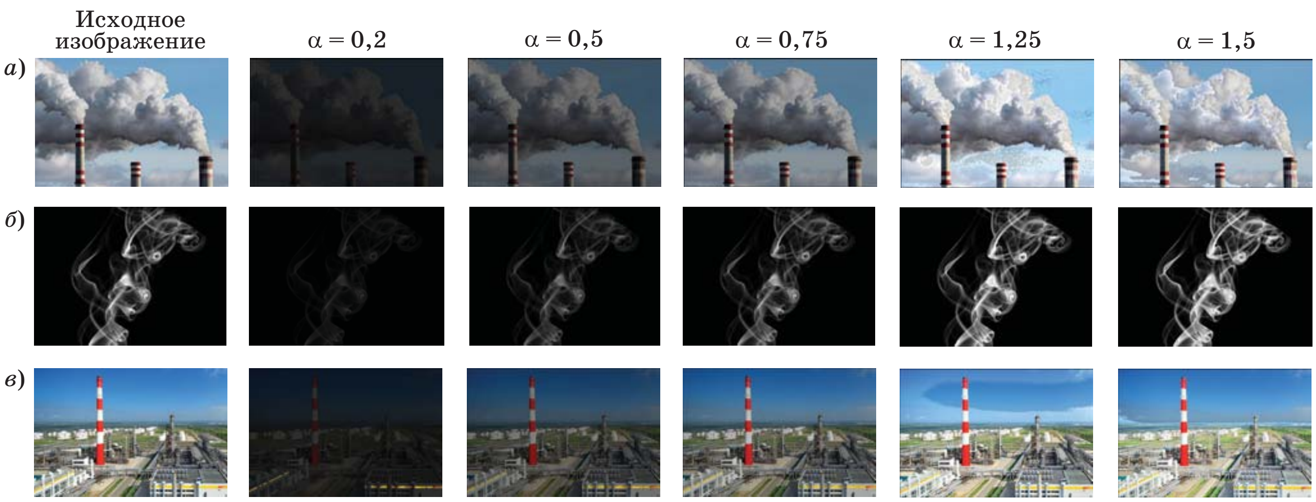
$$R_k(x, y) = \alpha S + (1-\alpha)(\log S_k(x, y) - \\ -\log\left[S_k(x, y) * F(x, y)\right], k = 1\ldots N, \quad (1)$$

где $R_k(x, y)$ — выходное изображение в k-м цветовом канале; N — количество цветовых каналов; α — масштабирующий коэффициент; $S_k(x, y)$ — исходное изображение в k-м цветовом канале; $F(x, y)$ — фильтр Гаусса; знак «*» — свертка изображений.

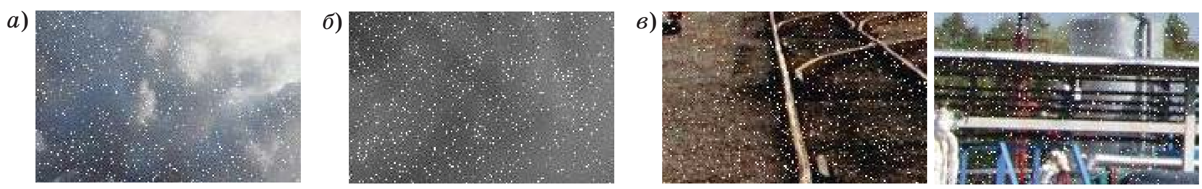
Кадры из видеопоследовательностей густого и прозрачного дыма и без дыма для различных значений параметра α показаны на рис. 1, *а–в*.

Атмосферные осадки (дождь/снег)

Для имитации атмосферных осадков на изображение накладывался аддитивный белый гауссов шум. Для этого к цветовому значению в каждом из каналов пиксела прибавлялась нормально распределенная случайная величина. При программной реализации нормально распределенные случайные величины были получены с помощью стандартного генератора случайных чисел по центральной предельной теореме. В случае



■ ***Рис. 1.*** Применение алгоритма Ретинекс к изображениям с густым дымом (*а*); с прозрачным дымом (*б*); без дыма (*в*)



■ ***Рис. 2.*** Изображения с аддитивным белым гауссовым шумом: густой дым (*а*); прозрачный дым (*б*); без дыма (*в*) если при прибавлении случайной величины цветовое значение канала выходило за предельный диапазон [0...255], то цветовое значение такого пиксела оставалось неизменным. На рис. 2, *а–в* показаны примеры воздействия аддитивного белого гауссова шума на изображения, в котором среднеквадратическое отклонение $\sigma = 2000$, среднее значение $\mu = 0$.

Шум типа «соль-перец»

Импульсный шум типа «соль-перец» (рис. 3, *а* и *б*) возникает из-за проблем с оборудованием, на котором проводится видеосъемка. Эта модель шума также может быть связана с ошибками при передаче изображений.

Отношение сигнал/шум N_{bb} [дБ] рассчитывается по формуле

$$N_{bb} = 10\,\log_{10} A_n / A_s\,,$$

где A_n — мощность шума; A_s — мощность полезного сигнала. В работе учитывался шум, достигающий 5,2; 7,01 и 10 дБ.

Размытие

В качестве фильтра размытия использован фильтр Гаусса. Фильтр размытия Гаусса основывается на матрице свертки, которая заполняется нормализованными коэффициентами, полученными с помощью функции Гаусса двух переменных [8]:

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}, \tag{2}$$

где σ — коэффициент среднеквадратического отклонения, определяющий степень размытия; x, y — координаты центрального элемента матрицы. Размерность матрицы вычисляется по формуле

$$dim = n \times n,\ n = 2R + 1, \tag{3}$$

где R — «радиус» матрицы свертки, т. е. выбранный радиус окрестности обрабатываемого пиксела. Степень размытия изображения зависит от двух параметров — R и σ.

С помощью нормализованной матрицы свертки для каждого пиксела изображения вычисляется среднее взвешенное значение соседних пикселов

$$g(x, y) = \sum_{i=-R}^{R} \sum_{j=-R}^{R} f(x+i, y+j) G^*(i+R, j+r), \tag{4}$$

где $g(x, y)$ — выходное изображение; $f(x, y)$ — исходное изображение; $G^*(i, j)$ — нормализованные коэффициенты матрицы свертки:

$$G^*(i, j) = G(i, j) \frac{1}{\sum_{i=0}^{2R} \sum_{j=0}^{2R} G(i, j)}. \tag{5}$$

В данной работе нормализованная матрица свертки, рассчитанная по формулам (2)–(5) с параметрами $\sigma = 0{,}84089642$ и $R = 2$, выглядит следующим образом:

$$\begin{bmatrix} 0{,}000789 & 0{,}006581 & 0{,}013347 & 0{,}006581 & 0{,}000789 \\ 0{,}006581 & 0{,}054901 & 0{,}111345 & 0{,}054901 & 0{,}006581 \\ 0{,}013347 & 0{,}111345 & 0{,}225821 & 0{,}111345 & 0{,}013347 \\ 0{,}006581 & 0{,}054901 & 0{,}111345 & 0{,}054901 & 0{,}006581 \\ 0{,}000789 & 0{,}006581 & 0{,}013347 & 0{,}006581 & 0{,}000789 \end{bmatrix}. \tag{6}$$

Размытие в сравнении с исходным изображением для видеоизображений показано на рис. 4, *а* и *б*.

Компенсация слабой освещенности и погодных условий фильтром Лапласа

Для повышения резкости изображения использован фильтр Лапласа. Дискретный лапласиан определяется как сумма вторых производных, а вычисляется как линейная аппроксимация суммы перепадов на соседних пикселах относительно центрального пиксела:

$$\begin{aligned} &\left[f(x+1, y) + f(x-1, y) - 2f(x, y)\right] + \\ &+ \left[f(x, y+1) + f(x, y-1) - 2f(x, y)\right] = \\ &= f(x+1, y) + f(x-1, y) + f(x, y+1) + \\ &\quad + f(x, y-1) - 4f(x, y). \end{aligned} \tag{7}$$

Рис. 3. Шум типа «соль-перец»: изображения с дымом (*а*) и без дыма (*б*)

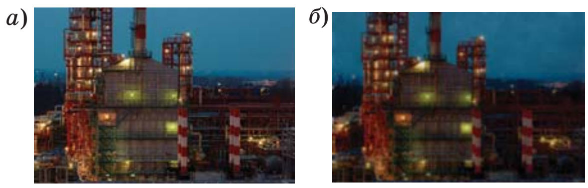
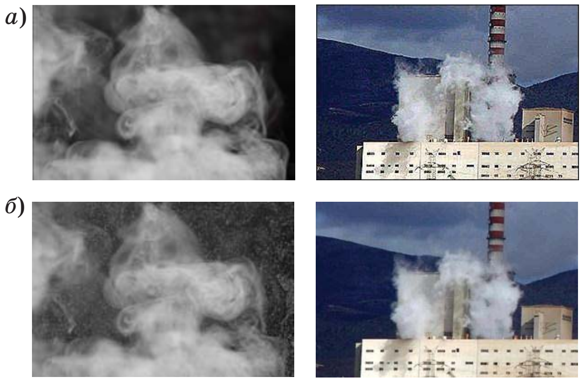


Рис. 4. Применение фильтра Гаусса: *а* — исходное изображение; *б* — размытие



■ *Рис. 5.* Иллюстрация использования фильтра Лапласа: исходные (*а*) и фильтрованные (*б*) изображения

Таким образом, маска фильтра, используемого для реализации дискретного лапласиана, задается матрицей

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}. \quad (8)$$

Для повышения резкости изображения с помощью лапласиана необходимо вычесть из исходного изображения лапласиан. Итоговая маска задается матрицей

$$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}. \quad (9)$$

Изображения до преобразования резкости и после него представлены на рис. 5, *а* и *б*.

Рассмотренные алгоритмы показывают целесообразность предварительной обработки изображений, что положительно сказывается на результатах работы алгоритма детектирования дыма.

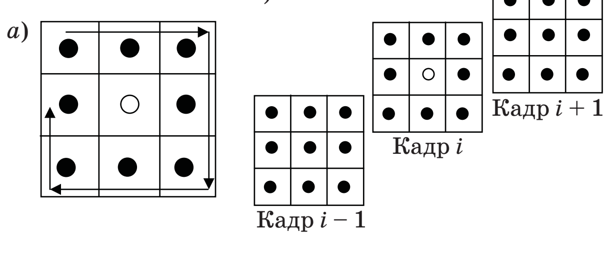
Алгоритм автоматического детектирования дыма

В работе предложен алгоритм, основанный на анализе текстуры с помощью локальных бинарных шаблонов (Local Binary Patterns — LBP), локальных тернарных шаблонов (Local Ternary Patterns — LTP) и расширенных локальных бинарных шаблонов (Extended Local Binary Patterns — ELBP). Существуют две основные версии 3D локальных бинарных шаблонов: пространственно-временные локальные бинарные шаблоны (Spatio-Temporal Local Binary Patterns — STLBP) и объемные локальные бинарные шаблоны (Volume Local Binary Patterns — VLBP).

Объемные шаблоны — шаблоны с тремя ортогональными планами — анализируют информацию из трех ортогональных проекций XY, XT и YT, где T — это время [9]. Оператор LBP в этом случае объединяет информацию в этих трех направлениях. Однако информация в некоторых пикселах может повторяться и тем самым учитываться дважды. Пространственно-временные шаблоны анализируют информацию соседних кадров относительно центрального пиксела. STLBP являются одним из способов исследования динамических свойств текстур в видеопоследовательности. 2D- и 3D-шаблоны в пространственно-временной области показаны на рис. 6, *а* и *б*.

Основная идея алгоритма состоит в исследовании применимости различных методов пространственно-временных локальных бинарных шаблонов при детектировании дыма для изображений, подверженных воздействию описанных ранее шумов, а также при различных уровнях освещенности.

Базовый оператор LBP представляет собой описание окрестности пиксела радиуса R. Количество пикселов окрестности n зависит от R и определяется формулой $n = (2R + 1)^2 - 1$. Оператор LBP, который применяется к окрестности пиксела изображения, использует n пикселов окрестности, принимая яркость центрального пиксела в качестве порогового значения. Пикселы, которые имеют значения яркости больше, чем значение яркости центрального пиксела (или равные ему), принимают значения «1»; те пикселы, которые имеют меньшие значения относительно центрального пиксела, принимают значения «0». В результате получается n-разрядный бинарный код, который описывает окрестность пиксела. В локальных тернарных шаблонах для уменьшения влияния шума и решения проблемы неотличимости двух пикселов с близкими значениями яркости от пикселов со значительной разницей яркости вводится тернарный код. В тернарном коде вместо порогового значения, основанного только на значении яркости пиксела, вводится константа t (определяемая экспериментально), позволяющая учитывать влияние шума на построение кода шаблона [10].



■ *Рис. 6.* Виды LBP: *а* — базовый LBP; *б* — 3D STLBP: ● — центральный пиксел

Алгоритм детектирования дыма включает следующие шаги [11].

1. Преобразование изображения из RGB цветового пространства в цветовое пространство YUV.

2. Преобразование изображений с помощью выбранного метода зашумления.

3. Выбор способа расчета шаблона (вычисление шаблона в каждой точке, шаблоны с прореживанием).

4. Вычисление STLBP для выбранного метода:

$$LBP(P,R) = \sum_{n=0}^{P-1} s\left(I_n - I_c\right) \cdot 2^n,$$

где P — количество пикселов в окрестности; R — радиус окрестности; $s(x) = 1$, если $x \geq 0$ и $s(x) = 0$ в противном случае; I_n и I_c — значения яркости текущего и центрального пикселов, представляющие собой значение яркости Y из цветовой схемы YUV [12];

$$LTP(P,R) = \sum_{n=0}^{P-1} s^*\left(I_n - I_c\right) \cdot 3^n,$$

где $s^*(x, t)$ — пороговая функция, а t — предопределенное пороговое значение, в данной работе $t = 5$ [13],

$$s^*(x,t) = \begin{cases} 1 \text{ при } x \geq t \\ 0 \text{ при } |x| < t \\ -1 \text{ при } x \leq -t \end{cases};$$

$$ELBP(P, R) = LBP(P, R);$$

$$STLBP_{i,j}(P, R) =$$
$$= LBP_{i-1,j}(P, R) + LBP_{i,j}(P, R) + LBP_{i+1,j}(P, R),$$

где i — номер текущего кадра, для которого происходит вычисление шаблона.

Для тернарных (STLTP) и расширенных (STELBP) шаблонов 3D-вариант рассчитывается аналогичным образом по следующему алгоритму.

1. Предварительная обработка полученных шаблонов (построение UP LTP, LOW LTP, подсчет количества равномерных шаблонов для ELBP).

2. Применение гистограммного подхода для классификации результата.

3. Выполнение шагов 1–2 в цикле для обработки всех кадров видеопоследовательности.

4. Сравнение эффективности полученных результатов.

5. Рекомендации для каждого типа погодных условий для анализируемого типа дыма.

В работе использовано два способа расчета шаблонов: шаблоны с прореживанием и вычисление шаблона в каждой точке. При использовании шаблонов с прореживанием исходное изображение разбивается на блоки для расчета локального шаблона исходя из значения радиуса R. Изображение разбивается на квадраты размера $dim = (2R + 1)^2$ пикселов, начиная от верхнего левого угла. При $R = 1$ размер квадрата равен 3×3 пикселов, при $R = 2$ — 5×5 и т. д. Для каждого квадрата вычисляется свой локальный шаблон, таким образом, шаблоны не пересекаются.

При расчете шаблона в каждом пикселе изображение не разбивается на непересекающиеся блоки, бинарная строка локального шаблона рассчитывается для каждого пиксела, происходит пересечение шаблонов. Вычисление паттерна локального шаблона проводится одинаково как в случае работы с прореживанием, так и при работе без него.

Построение вектора признаков дыма

Для классификации областей дыма в работе использован гистограммный подход как один из наиболее простых и быстрых способов. Классический LBP не требует предварительной обработки, в то время как для равномерных и тернарных шаблонов требуется предварительная обработка перед построением гистограммы. Предварительная обработка тернарного n-разрядного кода происходит следующим образом. Полученный код разбивается на два бинарных n-разрядных кода: UP LTP — отрицательные значения заменяются 1, LOW LTP — отрицательные значения заменяются 0. Далее по каждому из полученных наборов бинарных n-разрядных кодов строится своя гистограмма.

Для расширенных бинарных шаблонов предварительная обработка заключается в нахождении равномерных шаблонов (не более трех переходов в шаблоне) и всех их циклических сдвигов. Равномерные шаблоны определяют такие важные особенности изображения, как пятна, концы линий, углы, грани [14]. Для обеспечения инвариантности к вращению используются только равномерные шаблоны и все их циклические сдвиги [15]. Каждый равномерный шаблон является сдвигом одной из комбинаций, этой комбинации присваивается номер, этот номер является инвариантным к вращению. Если минимальное число переходов больше трех, то исходный бинарный вектор рассматривается особо: ему сопоставляется число, равное максимальному возможному целому для данной размерности, увеличенному на 1. В итоге получается список десятичных чисел, каждое из которых соответствует своему бинарному вектору. Конечная гистограмма ELBP строится классическим способом по итоговому набору десятичных чисел.

Построение гистограмм по сформированному набору n-разрядных бинарных кодов происходит двумя способами.

1. Классический способ — каждая бинарная строка преобразуется в десятичное число, под-

считывается количество одинаковых чисел, определяющих положение и высоту столбцов гистограммы.

2. Альтернативный способ (сумма единиц для каждого разряда LBP) — вычисляется количество единиц в каждом разряде бинарного кода, тем самым определяется высота столбца гистограммы.

Построение решающего правила

Для классификации применяются меры сравнения расстояний, таких как χ^2-расстояние, расстояние пересечения гистограмм, расстояние Кульбака — Лейблера, расстояние на основе статистик. В данном случае была выбрана мера Кульбака — Лейблера как наиболее часто рекомендуемая для сравнения гистограмм.

Расстояние Кульбака — Лейблера, которое служит обобщением меры энтропии, введенной Шенноном, является относительной мерой энтропии, которая адаптирована для измерения расстояний между гистограммами. Данное расстояние применяется для анализа вероятности возникновения кодовых значений. Вначале строится гистограмма кодовых значений изображения (в данном случае — кодовых значений локальных бинарных шаблонов). Затем полученные гистограммы тестовых изображений нормализуются, и вычисляется расстояние Кульбака — Лейблера:

$$D_{K,L}(f,g)=\sum_{k=1}^{P(P-1)+3} f_k \ln\frac{f_k}{g_k},$$

где f и g — гистограммы первого и второго изображения; P — число точек в окрестности LBP; k — номер столбца.

Данное расстояние используется в качестве меры различия гистограмм двух сопоставляемых изображений при построении соответствующего решающего правила.

Эксперименты

Для проведения экспериментальных исследований анализировались данные из базы данных динамических текстур DynTex [16]. Использовано три видеопоследовательности густого дыма, одна видеопоследовательность с прозрачным дымом и шесть видеопоследовательностей без дыма, из которых вручную были извлечены фрагменты дыма или фрагменты с отсутствием дыма.

В экспериментальном исследовании шаблоны для густого дыма рассчитываются в каждой точке, так как результат таких расчетов лучше в среднем на 2–5 %, чем для вычисления шаблонов с прореживанием, кроме того, для густого дыма приведены результаты только STLBP, STLTP и STELPB (только 3D-вариант). Для прозрачного дыма дает приемлемый результат только метод расширенных бинарных шаблонов в 2D-варианте.

Слабая освещенность (Ретинекс α = 0,2). В табл. 1–3 приведены лучшие результаты работы алгоритма детектирования прозрачного и густого дыма в условиях низкой освещенности при значении параметра $\alpha = 0{,}2$.

Для густого дыма при низкой освещенности [$\alpha = 0{,}2$ в выражении (1)] расширенные локальные бинарные шаблоны и тернарные локальные шаблоны показали лучший результат. При этом не имеет значения величина радиуса, при котором происходит вычисление шаблона. Однако расчет при единичном радиусе с восемью точками окрестности происходит приблизительно в два раза быстрее. Изменение параметра α до значения 0,5 увеличило точность детектирования дыма на 1 % при уменьшении количества ошибок на порядок. При 75 % освещенности от начальной по сравнению со значением $\alpha = 0{,}5$ тернарные и расширенные шаблоны показали лучшие результаты. Также эти типы шаблонов устойчивы и при увеличенной яркости относительно базовой.

Для прозрачного дыма единственным способом расчетов шаблонов, обеспечивающим приемлемую эффективность, являются расширенные бинарные шаблоны в двумерном случае. При этом лучшее значение TR = 0,711 в условиях низкой освещенности и 0,88 при увеличении параметра α до 0,5. Точность распознавания возрастает по мере возрастания освещенности и является максимальной при значении параметра $\alpha = 1$ в неизменном изображении. При увеличении яркости ($\alpha = 1{,}25$ и $\alpha = 1{,}5$) точность алгоритма падает до значения TR = 0,628, что связано с особенностями преобразования изображения с помощью алгоритма Ретинекс.

Атмосферные осадки. Лучшие результаты расчетов для детектирования дыма при гауссовом шуме, в котором $\sigma = 2000$, $\mu = 0$ (см. рис. 2), приведены в табл. 4.

Для густого дыма устойчивость к аддитивному белому гауссовому шуму показали тернарные шаблоны, что объясняется их особенностью — устойчивостью к шуму тернарного кода. Расширенные бинарные шаблоны также дали значения TR на уровне 0,985. При этом не имеет значения радиус расчета такого шаблона для опознавания густого дыма. Однако отсутствие дыма детектируется в этом случае лучшим образом при радиусе два пиксела и восьми точках окрестности. Прозрачный дым детектируется со значениями точности распознавания 0,843 для дыма и 0,527 при отсутствии дыма.

■ ***Таблица 1.*** Ретинекс $\alpha = 0{,}2$ для бинарных (LBP) и тернарных (LTP) шаблонов

Видеофрагмент	LBP			LTP		
	TR	FRR	FAR	TR	FRR	FAR
Гистограммы «сумма единиц для каждого разряда LBP»						
Густой дым	0,907	0,127	0,093	0,99	0,077	0,01
Отсутствие дыма	0,866	0,099	0,134	0,909	0,089	0,091
Гистограммы классические						
Густой дым	0,975	0,053	0,025	0,98	0,102	0,02
Отсутствие дыма	0,941	0,056	0,059	0,881	0,986	0,119

TR — правильное распознавание; FRR — ошибка I рода; FAR — ошибка II рода.

■ ***Таблица 2.*** Ретинекс $\alpha = 0{,}2$ для пространственно-временных расширенных бинарных шаблонов (STELBP)

Видеофрагмент	$R = 1$; 8 точек окрестности			$R = 2$					
				8 точек окрестности			12 точек окрестности		
	TR	FRR	FAR	TR	FRR	FAR	TR	FRR	FAR
Густой дым	0,98	0,086	0,02	0,985	0,012	0,015	0,99	0,08	0,01
Отсутствие дыма	0,9	0,032	0,1	0,989	0,035	0,011	0,906	0,07	0,09

■ ***Таблица 3.*** Ретинекс $\alpha = 0{,}2$ для видеопоследовательности с прозрачным дымом, расширенные бинарные шаблоны (ELBP)

Видеофрагмент	2D-вариант; $R = 1$; 8 точек окрестности					
	В каждой точке			С прореживанием		
	TR	FRR	FAR	TR	FRR	FAR
Прозрачный дым	0,711	0,422	0,289	0,627	0,859	0,373
Отсутствие дыма	0,545	0,232	0,455	0,021	0,141	0,979

■ ***Таблица 4.*** Результаты детектирования дыма для шума Гаусса

Видеофрагмент	**LBP**						**LTP**		
	–			TR	FRR	FAR	TR	FRR	FAR
	Гистограммы «сумма единиц для каждого разряда LBP»								
Густой дым				0,917	0,16	0,083	0,995	0,043	0,005
Отсутствие дыма	–			0,824	0,015	0,176	0,949	0,023	0,051
							Гистограммы классические		
Густой дым	–			–	–	–	0,985	0,023	0,015
Отсутствие дыма	–			–	–	–	0,976	0,029	0,024
	STELBP								
	$R = 1$; 8 точек окрестности			$R = 2$					
				8 точек окрестности			12 точек окрестности		
	TR	FRR	FAR	TR	FRR	FAR	TR	FRR	FAR
Густой дым	0,985	0,048	0,015	0,985	0,012	0,015	0,985	0,054	0,015
Отсутствие дыма	0,945	0,002	0,055	0,989	0,052	0,011	0,938	0,002	0,062
	ELBP; 2D-вариант; $R = 1$; 8 точек окрестности								
	–			В каждой точке			С прореживанием		
	–			TR	FRR	FAR	TR	FRR	FAR
Прозрачный дым	–			0,843	0,411	0,157	0,843	0,42	0,157
Отсутствие дыма	–			0,527	0,026	0,473	0,515	0,099	0,485

Шум типа «соль-перец». Результаты работы алгоритма детектирования дыма для зашумленных изображений с различными значениями шума приведены в табл. 5 и 6.

При густом дыме устойчивыми к шуму типа «соль-перец» являются тернарные шаблоны в случае классических гистограмм, что обусловлено особенностями построения бинарного кода для таких шаблонов, также шум до 10 дБ выдерживают и расширенные шаблоны радиуса 2. При этом для расширенных шаблонов предпочтительно рассчитывать шаблон для радиуса два пиксела и восьми точек окрестности.

Прозрачный дым в условиях шума «соль-перец» можно детектировать только в 2D-случае с использованием тернарных локальных шаблонов, при этом, если уровень шума возрастает до 10 дБ, эффективность этого метода падает до значений TR 0,723 для дыма и до 0,688 при отсутствии дыма.

Размытие. Точность детектирования дыма на изображениях, подвергнутых фильтру Гаусса [выражения (2)–(6)], в случае STELBP снижалась на 1–2 % для густого дыма и на 5–10 % для прозрачного дыма. Это означает, что, прежде чем приступить к детектированию, требуется провести дополнительные методы борьбы с размытием изображений.

Фильтр Лапласа. В табл. 7 показаны лучшие результаты детектирования густого дыма на видеопоследовательности, к которой для повышения резкости применен фильтр Лапласа [выражения (7)–(9)].

■ ***Таблица 5.*** Результаты детектирования дыма с шумом типа «соль-перец», расширенные бинарные шаблоны (STELBP)

Видеофрагмент	Значение шума, дБ	$R = 2$					
		8 точек окрестности			12 точек окрестности		
		TR	FRR	FAR	TR	FRR	FAR
Густой дым	5,2	0,917	0,16	0,083	0,995	0,043	0,005
Отсутствие дыма		0,824	0,015	0,176	0,949	0,023	0,051
Густой дым	7,01	0,99	0,02	0,01	0,99	0,013	0,01
Отсутствие дыма		0,986	0,012	0,014	0,978	0,012	0,022
Густой дым	10	0,971	0,038	0,029	0,951	0,1	0,005
Отсутствие дыма		0,961	0,04	0,039	0,89	0,05	0,051

■ ***Таблица 6.*** Результаты детектирования дыма с шумом типа «соль-перец», тернарные шаблоны

Видеофрагмент	$R = 1$			$R = 2$					
	5,2 дБ			7,01 дБ			10 дБ		
	TR	FRR	FAR	TR	FRR	FAR	TR	FRR	FAR
	STLTP, альтернативные гистограммы								
Густой дым	0,92	0,135	0,08	0,912	0,17	0,088	0,903	0,02	0,097
Отсутствие дыма	0,88	0,114	0,12	0,876	0,02	0,124	0,813	0,09	0,187
	LTP, классические гистограммы, 2D-вариант								
Прозрачный дым	0,904	0,107	0,096	0,90	0,115	0,096	0,723	0,305	0,277
Отсутствие дыма	0,819	0,112	0,181	0,881	0,114	0,119	0,688	0,205	0,312

■ ***Таблица 7.*** Результаты детектирования густого дыма с применением фильтра Лапласа, расширенные бинарные шаблоны (STELBP)

Видеофрагмент	$R = 1$			$R = 2$					
	8 точек окрестности			8 точек окрестности			12 точек окрестности		
	TR	FRR	FAR	TR	FRR	FAR	TR	FRR	FAR
Густой дым	0,985	0,003	0,015	0,98	0,003	0,01	0,985	0,005	0,015
Отсутствие дыма	1	0,001	0	1	0,002	0	0,997	0,004	0,003

В большинстве случаев фильтр Лапласа обеспечивает незначительное улучшение точности детектирования на 1–2 %. В случае с прозрачным дымом точность детектирования снижалась до 78 %.

Заключение

Анализ результатов при детектировании густого дыма показывает, что тернарные шаблоны обладают преимуществом перед бинарными за счет снижения влияния шума на изображение. Пространственно-временные локальные тернарные шаблоны (STLTP) и пространственно-временные расширенные локальные бинарные шаблоны (STETBP) являются устойчивыми к шуму, STELBP обеспечивают точность детектирования густого дыма 98–99 %, тернарные шаблоны показывают значение точности 90–98 %. Для прозрачного дыма наиболее эффективными являются ELBP в двумерном случае.

В условиях слабой освещенности эффективность детектирования густого дыма падает на 1–2 %, что позволяет говорить о том, что выбранный тип шаблонов является достаточно устойчивым к изменению условий освещенности. При зашумлении изображений аддитивным белым шумом Гаусса эффективность работы тернарных шаблонов не изменяется, а эффективность расширенных бинарных шаблонов падает на 1,5 % для густого дыма. Для шума типа «соль-перец» расширенные бинарные шаблоны выдерживают (без падения точности) шум до 7,01 дБ, при увеличении шума до 10 дБ происходит падение точности на 3 % для густого дыма. Тернарные шаблоны работают хуже на 6,8–8,7 % при увеличении шума до 5–10 дБ для густого дыма. При детектировании прозрачного дыма ELBP выдерживают падение освещенности в два раза, точность детектирования при этом 86–88 %, при шуме типа «соль-перец» для прозрачного дыма точность составляет 90 % при шуме 7 дБ.

Эксперименты с фильтром размытия показали, что для густого и прозрачного дыма требуются дополнительные методы борьбы с размытием, фильтр Лапласа обеспечивает незначительное увеличение точности детектирования. Эффективность работы STELBP объясняется тем, что расширенные шаблоны учитывают специфические особенности изображения, такие как грани, углы, линии, поскольку и густой, и прозрачный дым сглаживают резкие грани изображения. STLTP обладают устойчивостью к шуму, а следовательно, обеспечивают высокое качество детектирования.

Литература

1. **Пятаева А. В., Фаворская М. Н.** Применение локальных бинарных шаблонов при раннем обнаружении дыма на открытых пространствах // Цифровая обработка сигналов и ее применение: тр. 17-й Междунар. конф., Москва, 25–27 марта 2015 г. М., 2015. Т. 2. С. 531–535.
2. **Gonçalves W. N., Machado B. B., Bruno O. M.** A Complex Network Approach for Dynamic Texture Recognition//Neurocomputing. 2015. N 153. P. 211–220.
3. **Пятаева А. В.** Обнаружение областей задымления на видеопоследовательности с применением локальных бинарных шаблонов // Вестник Сибирского государственного аэрокосмического университета. 2014. № 5(57). С. 108–114.
4. **Гончаров А. В., Каракищенко А. Н.** Влияние освещенности на качество распознавания фронтальных лиц // Изв. ЮФУ. Технические науки. 2008. № 4. С. 88–92.
5. **Chitturi M. V., Medina J. C., Benekohal R. F.** Effect of Shadows and Time of Day on Performance of Video Detection Systems at Signalized Intersections // Transportation Research Part C: Emerging Technologies. 2010. N 18(2). P. 176–186.
6. **Костин К. А., Аксёнов С. В.** Анализ влияния изменения освещенности на фильтрацию изображений с помощью фильтров Лавса // Молодежь и современные информационные технологии: тр. XII Междунар. науч.-практ. конф. студентов, аспирантов и молодых ученых, Томск, 11–14 ноября 2014 г. Томск, 2014. С. 90–91.
7. **Монич Ю. И., Старовойтов В. В.** Преобразование цветных изображений на базе технологии Ретинекс // Искусственный интеллект. 2007. № 3. С. 256–261.
8. **Zhang Z.**, et al. Blurring-Invariant Riemannian Metrics for Comparing Signals and Images/ Z. Zhang, E. Klassen, A. Srivastava, P. Turaga, R. Chellappa // 6th Intern. Conf. on Computer Vision. 5–7 March 2011. Vilamoura, Algarve, Portugal, 2011. P. 1770–1775.
9. **Zhao G., Pietikäinen M.** Dynamic Texture Recognition Using Volume Local Binary Patterns // Dynamical Vision. LNCS. 2007. Vol. 4358. P. 165–177.
10. **Mohamed A. A., Yampolskiy R. V.** Adaptive Extended Local Ternary Pattern (AELTP) for Recognizing Avatar Faces // 11th Intern. Conf. on Machine Learning and Applications, 12–15 December 2012, Boca Raton, Florida, USA. P. 57–62.
11. **Favorskaya M., Pyataeva A., Popov A.** Verification of Smoke Detection in Video Sequences Based on Spatio-Temporal Local Binary Patterns//Procedia Computer Science. 2015. N 60. P. 671–680.
12. **Ko B. C., Park J. O., Nam J. Y.** Spatiotemporal Bag-of-Features for Early Wildfire Smoke Detection // Image and Vision Computing. 2013. N 31(10). P. 786–795.

13. **Zhao G., Pietikainen M.** Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions // Transactions on Pattern Analysis and Machine Intelligence. 2007. N 29(6). P. 915–928.
14. **Feiniu Y.** Rotation and Scale Invariant Local Binary Pattern Based on High Order Directional Derivatives for Texture Classification // Digital Signal Processing. 2014. N 26. P. 142–152.
15. **Zho H., Wang R., Wang C.** A Novel Extended Local-Binary-Pattern Operator for Texture Analysis // Information Sciences. 2008. N 178(22). P. 4314–4325.
16. **Renaud P., Fazekas S., Huiskes M. J.** DynTex: A Comprehensive Database of Dynamic Textures // Pattern Recognition Letters. 2010. N 31(12). P. 1627–1632.

UDC 004.932.2
doi:10.15217/issn1684-8853.2016.1.16

Smoke Detection on Video Sequences based on Spatio-Temporal Local Binary Patterns in Outdoor Spaces under Complicated Weather Conditions

Favorskaya M. N.[a], Dr. Tech. Sc., Professor, favorskaya@sibsau.ru
Pyataeva A. V.[a], Post-Graduate Student, anna4u@list.ru
[a]Siberian State Aerospace University named after academician M. F. Reshetnev, 31, Krasnoyarsky Rabochy St., 660037, Krasnoyarsk, Russian Federation

Purpose: Early smoke detection in outdoor scenes using video sequences is one of the important tasks in modern surveillance systems. Visual information as a result of real-time video shooting may include objects with dynamic behavior, noise of the hardware or transmission lines, as well as artefacts affected by weather conditions (for example, rain or snow, poor luminance in the morning or evening). Therefore, smoke-like regions segmented into video sequences should be finally verified. **Results:** An algorithm is proposed to process images with atmospheric artefacts like drizzle or fog, or images with poor luminance. A method has been studied which involves texture analysis based on spatio-temporal local binary patterns, local ternary patterns, and extended binary patterns to detect dense or transparent smoke with the following artefacts: salt-pepper noise up to 10 dB, additive white Gauss noise simulating atmospheric precipitates, image blurring and poor luminance, using Laplace filter. To classify smoke regions, histograms were applied as one of the simplest and fastest methods of image analysis. As a measure of histogram difference between two images, Kullback–Leibler divergence was used in order to provide a decision rule. **Practical relevance:** The developed method of smoke verification on video sequences using spatio-temporal local binary patterns and 3D extended local binary patterns provides 96–99 % of accuracy for dense smoke and 86–94 % of accuracy for transparent smoke, depending on the artefacts and noise.
Keywords — Smoke Detection, Noise, Weather Conditions, Spatio-Temporal Local Binary Patterns.

References

1. Pyataeva A. V., Favorskaya M. N. Early Smote Detection in Outdoor Scenes by Using Local Binary Patterns. *Trudy 17-i Mezdunarodnoy konferencii "Tsifrovaia obrabotka signalov i ee primenenie"* [The 17th Intern. Conf. "Digital Signal Processing and its Applications"], Moscow, 2015, pp. 531–535 (In Russian).
2. Gonçalves W. N., Machado B. B., Bruno O. M. A Complex Network Approach for Dynamic Texture Recognition. *Neurocomputing,* 2015, no. 153, pp. 211–220.
3. Pyataeva A. V. Video-Based Smoke Detection Using Local Binary Patterns. *Vestnik Sibirskogo gosudarstvennogo aerokosmicheskogo universiteta imeni akademika M. F. Reshetneva (Vestnik SibGAU)*, 2014, no. 5(57), pp. 108–114 (In Russian).
4. Goncharov A. V., Karakishchenko A. N. Influence of Illumination on a Quality of Frontal Face Recognition. *Izvestiya SFedU. Tekhnicheskie nauki*, 2008, no. 4, pp. 88–92 (In Russian).
5. Chitturi M. V., Medina J. C., Benekohal R. F. Effect of Shadows and Time of Day on Performance of Video Detection Systems at Signalized Intersections. *Transportation Research Part C: Emerging Technologies*, 2010, no. 18(2), pp. 176–186.
6. Kostin K. A., Aksenov S. V. Analysis of the Impact of Luminance Changes on Image Filtering Using Laws Filters. *Trudy XII Mezhdunarodnoi nauchno-prakticheskoi konferentsii studentov, aspirantov i molodykh uchenykh "Molodezh' i sovremennye informatsionnye tekhnologii"* [Proc. of XII Intern. Scientific-practical Conf. of Students, PhD Students, and Young Scientists "Young People and Modern Information Technologies"], Tomsk, 2014, pp. 90–91 (In Russian).
7. Monich Y. I., Starovoitov V. V. Transformation of Color Images using Retinex Technology. *Iskusstvennyi intellekt*, 2007, no. 3, pp. 256–261 (In Russian).
8. Zhang Z., Klassen E., Srivastava A., Turaga P., Chellappa R. Blurring-Invariant Riemannian Metrics for Comparing Signals and Images. *6th Intern. Conf. on Computer Vision*, 2011, pp. 1770–1775.
9. Zhao G., Pietikäinen M. Dynamic Texture Recognition Using Volume Local Binary Patterns. *Dynamical Vision*, LNCS, 2007, vol. 4358, pp. 165–177.
10. Mohamed A. A., Yampolskiy R. V. Adaptive Extended Local Ternary Pattern (AELTP) for Recognizing Avatar Faces. *11th Intern. Conf. on Machine Learning and Applications*, 2012, pp. 57–62.
11. Favorskaya M., Pyataeva A., Popov A. Verification of Smoke Detection in Video Sequences Based on Spatio-Temporal Local Binary Patterns. *Procedia Computer Science*, 2015, no. 60, pp. 671–680.
12. Ko B. C., Park J. O., Nam J. Y. Spatiotemporal Bag-of-Features for Early Wildfire Smoke Detection. *Image and Vision Computing*, 2013, no. 31(10), pp. 786–795.
13. Zhao G., Pietikainen M. Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions. *Transactions on Pattern Analysis and Machine Intelligence*, 2007, no. 29(6), pp. 915–928.
14. Feiniu Y. Rotation and Scale Invariant Local Binary Pattern Based on High Order Directional Derivatives for Texture Classification. *Digital Signal Processing*, 2014, no. 26, pp. 142–152.
15. Zho H., Wang R., Wang C. A Novel Extended Local-Binary-Pattern Operator for Texture Analysis. *Information Sciences*, 2008, no. 178(22), pp. 4314–4325.
16. Renaud P., Fazekas S., Huiskes M. J. DynTex: A Comprehensive Database of Dynamic Textures. *Pattern Recognition Letters*, 2010, no. 31(12), pp. 1627–1632.

УДК 681.3
doi:10.15217/issn1684-8853.2016.1.26

АНАЛИЗ ПРОБЛЕМЫ ПОВЫШЕНИЯ РАДИАЦИОННОЙ СТОЙКОСТИ ИНФОРМАЦИОННО-УПРАВЛЯЮЩИХ СИСТЕМ НА ЭТАПЕ ФУНКЦИОНАЛЬНО-ЛОГИЧЕСКОГО ПРОЕКТИРОВАНИЯ

И. В. Егоров[a], *аспирант*
В. Ф. Мелехин[a], *доктор техн. наук, профессор*
[a]*Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, РФ*

Постановка проблемы: *в условиях радиации в элементах вычислительных устройств возникают мягкие отказы, интенсивность которых значительно превышает интенсивность невосстанавливаемых отказов. Меры повышения радиационной стойкости за счет технологии изготовления интегральных схем и схемотехники библиотечных элементов недостаточны. В целях повышения радиационной стойкости актуальна организация периодического восстановления искаженной информации в резервированных структурах устройств. Это требует решения ряда задач, связанных с синтезом и анализом при функционально-логическом проектировании систем.* ***Цель:*** *аналитический обзор опубликованных результатов по рассматриваемой проблеме, выработка концептуальных положений подхода к решению проблемы повышения радиационной стойкости систем на этапе функционально-логического проектирования, постановка задач дальнейших исследований.* ***Результаты:*** *анализ показал, что мягкие информационные отказы, возникающие при радиационных воздействиях, оказывают иное влияние на работу систем по сравнению с невосстанавливаемыми отказами и, как следствие, изменяют структурную значимость элементов разных типов с точки зрения надежности. Определен подход к организации вычислительных процессов и построению резервированных структур выделенных типов блоков системы, основанный на использовании резерва времени для периодического восстановления информации, искаженной под воздействием потока мягких отказов. Сформулированы задачи дальнейших исследований, направленных на разработку методики проектирования информационно-управляющих систем с повышенной радиационной стойкостью.*

Ключевые слова *— радиационные эффекты в полупроводниковых структурах, мягкие отказы, структурное резервирование, восстановление информации в резервированных структурах, восстанавливаемые вычислительные системы.*

Введение

Из работ [1–3], посвященных анализу эффектов в полупроводниковых структурах при действии радиации, известно, что наиболее часто частица высокой энергии, попадая в МОП-транзистор, вызывает ионизацию подзатворной области полупроводника. При этом если транзистор закрыт, в подзатворной области возникает заряд неосновных носителей определенной величины, вызывающий проскакивание соответствующего импульса тока, а на выходе вентиля, в схему которого входит транзистор, появляется кратковременный импульс напряжения (ложный сигнал). Он может возникнуть в вентиле при состоянии как логического нуля, так и логической единицы. Длительность импульса зависит от величины заряда неосновных носителей, а он — от энергии частицы и технологических параметров вентиля. Обычно длительность импульса находится в диапазоне до 1–2 нс. Воздействие таких импульсов на цифровую схему зависит от задержек переключения вентиля и задержек в связях. В свою очередь эти параметры зависят от проектной нормы изготовления интегральных схем (ИС). При микронной и субмикронной технологии производства ИС эти импульсы были неопасны вследствие инерционности элементов. При современной нанотехнологии производства ИС с проектной нормой меньше 0,1 мкм такие наведенные ложные импульсы сравнимы с полезными импульсными сигналами и приводят к искажению информации в СБИС. Рассматриваемые в работе вопросы связаны с особенностями поведения ИС, выполненных по современным технологиям микроэлектроники.

Наиболее опасны рассматриваемые эффекты для транзисторов и вентилей, входящих в состав бистабильной ячейки триггера либо в запоминающий элемент блока памяти. Тогда эффект воздействия частицы приводит к искажению бита информации, которое принято называть «мягким отказом» (soft error). Если вентиль, подвергшийся воздействию, входит в состав логического элемента комбинационной схемы, то возникший ложный сигнал будет распространяться по цепи элементов и может достигнуть триггера и изменить его состояние. Это тоже приводит к сбою, но вероятность возникновения сбоя в этом случае существенно меньше, чем при непосредственном воздействии на триггер. Особенности распространения ложных сигналов по сети элементов требуют отдельного рассмотрения. Группой сотрудников Санкт-Петербургского политехнического университета, в которую входят авторы, выполнен ряд исследований по проблеме повышения радиационной стойкости информационно-управляющих систем (ИУС).

Цель данной работы — обобщить опубликованные результаты и обосновать постановку задач дальнейших исследований, решение которых позволит создать методику проектирования (маршрут функционально-логического проектирования) ИУС со структурной, информационной и временной избыточностью, обладающих повышенной радиационной стойкостью. Публикации по новым результатам решения поставленных задач готовятся. В настоящей статье дано общее концептуальное представление о проводимых исследованиях, объединяющее последующие публикации, посвященные частным результатам.

Анализ опубликованных результатов по проблеме повышения радиационной стойкости ИУС на этапе функционально-логического проектирования

В работе [4] выполнен аналитический обзор методов повышения радиационной стойкости, описанных в литературе. Эти методы относятся к технологии изготовления ИС и построению элементов библиотеки, на базе которой производится проектирование схем устройств.

Эффективной в плане повышения радиационной стойкости при изготовлении ИС является технология SOI (кремний на изоляторе), которая создавалась в целях уменьшения паразитных емкостей и, как следствие, повышения быстродействия элементов. Оказалось, что технология SOI позволяет также уменьшить интенсивность потока мягких отказов до 10 раз. Недостатками метода являются значительное удорожание изготовления ИС и количественный характер достигаемого эффекта. Уменьшение толщины кремния (полупроводника) в подзатворном слое транзистора уменьшает величину наводимого в нем заряда неосновных носителей воздействием частицы высокой энергии, но не исключает этот эффект. При этом уменьшается длительность «ложного» импульса, но по мере уменьшения проектной нормы изготовления ИС эффект повышения радиационной стойкости будет снижаться.

На уровне построения библиотеки элементов для повышения радиационной стойкости предложены схемы триггеров с увеличенным числом транзисторов (например, DICE-cells), что, как утверждают авторы этих ячеек, тоже позволяет уменьшить интенсивность мягких отказов в несколько раз. Анализ [4] переключения такого триггера под действием наведенного «ложного» импульса показывает, что и в этом случае эффект имеет количественный характер: для переключения требуется большая длительность импульса.

Таким образом, из проведенного анализа [4] следует, что предложенные методы повышения радиационной стойкости не снимают проблемы. Наряду с ними актуально рассмотреть дополнительные методы повышения радиационной стойкости ИУС на уровне функциональной организации системы и организации вычислительных процессов.

В работе [5] введено понятие восстанавливаемых вычислительных систем при действии потока мягких отказов. В этих системах со структурной и информационной избыточностью наряду с основным вычислительным процессом исполнения прикладной программы организованы вспомогательные процессы периодического контроля и исправления информации в узлах с информационным отказом. Показано, что такая организация позволяет в условиях потока мягких отказов в несколько раз повысить время работоспособного состояния системы со структурным резервированием.

Применительно к задаче повышения надежности все блоки структуры системы делятся на два типа [5]: автоматы с памятью, реализуемой на триггерах (блоки типа 1), и блоки памяти с линейно-адресной организацией (блоки типа 2). В автоматах с памятью для блокирования распространения отказов используется троирование блока и мажорирование сигналов от экземпляров блока. Варианты структурной организации невосстанавливаемых систем со структурным резервированием исследованы в работе [6]. Обоснован способ так называемой «доменной» организации, обеспечивающий независимость блокирования распространения отказов в доменах, а также масштабируемость структурного резервирования. В работе [5] рассматривается структура блоков типа 1, построенная в соответствии со способом доменной организации. Как будет показано в следующем разделе, для проектирования восстанавливаемых систем целесообразна модификация подхода к структурному резервированию блоков типа автомата с памятью, подверженных потоку мягких отказов, по отношению к невосстанавливаемым системам.

В работе [7] предложена математическая модель для анализа надежности функциональных узлов цифровых СБИС со структурным резервированием и периодическим восстановлением работоспособного состояния. Исследуется поведение узлов типа конечного автомата с памятью, в которых используется структурное резервирование по способу, разработанному для невосстанавливаемых систем. Показано, что при периодах восстановления, меньших среднего интервала возникновения мягкого отказа, интервал работоспособного состояния блока восстанавливаемой системы увеличивается в несколько раз.

Этот результат доказывает, что организация периодического восстановления информации, искаженной под действием мягких отказов, весьма эффективна. Кроме того, результат наглядно демонстрирует сильную зависимость показателей надежности от временной избыточности блоков, т. е. от того, какую часть времени блок используется для реализации основного вычислительного процесса, а какую — для вспомогательного процесса контроля и восстановления информации. В связи с этим для анализа надежности восстанавливаемых ИУС, подверженных потоку мягких отказов, необходимо учитывать не только структуру системы, отражающую состав и связи блоков, но и организацию вычислительных процессов, в которой существенную роль играют системное и прикладное программное обеспечение (ПО).

Изменения в подходе к проблеме повышения надежности ИУС при потоке мягких отказов, вызванных действием радиации

Прежде всего, ответим на вопрос: «Почему проблему обеспечения работоспособности системы при воздействии потока мягких отказов можно выделить в самостоятельную проблему повышения надежности системы?»

Как уже отмечалось, при воздействии радиации на современные ИС мягкие отказы возникают существенно чаще (в сотни раз), чем невосстанавливаемые отказы элементов. Для вычислительного процесса мягкий отказ, проявляющийся в искажении запомненной в текущий момент информации, — это такой же отказ, как и невосстанавливаемый отказ элемента, который также проявляется в искажении информации. Поэтому для блокирования распространения мягкого отказа и обеспечения выполнения вычислительного процесса применяются те же методы повышения надежности, что и в случае невосстанавливаемых отказов элементов. Кратность резервирования при введении структурной избыточности ограничена увеличением сложности реализации системы. Из этого следует, что при действии радиации отказ резервированной системы с большой вероятностью произойдет из-за мягких отказов. Таким образом, можно поставить задачу повышения надежности системы при действии потока мягких отказов на таком интервале времени, когда невосстанавливаемые отказы элементов маловероятны. Получив эффективные способы обеспечения работоспособности системы в условиях только мягких отказов, далее можно рассматривать и более общую задачу с учетом невосстанавливаемых отказов.

Еще одним аргументом актуальности рассмотрения работы системы при потоке мягких отказов является возможность сбора статистики о мягких отказах в блоке и прогнозирование критического увеличения потока отказов для своевременного выхода действующей системы из рабочего режима с целью предотвратить аварию. Возможность восстановления состояния элемента после мягкого отказа во много раз увеличивает интервал работоспособного состояния резервированной системы. Поэтому, регистрируя каждый случай выявленного мягкого отказа в блоке, можно собирать и обрабатывать статистику отказов. При отсутствии периодических восстановлений информации в блоках такая возможность отсутствовала.

Ограничиваясь рассмотрением только потока мягких отказов, уточним понятие «отказ». Как и в классической теории надежности вычислительных систем, отказ связан с нарушением работоспособности на нижнем уровне организации вычислительного процесса, т. е. с аппаратурой. Отказ проявляется в искажении информации, представленной сигналами в аппаратуре.

В классической теории надежности справедливо считается, что отказ любого логического элемента, будь то элемент комбинационной схемы или памяти, вызывает искажение информации, представленной хранением или распространением сигналов по сети. Поэтому применительно к показателям надежности речь об информации не идет, а рассматривается исправность элементов, объединенных в сеть.

При возникновении мягких отказов под воздействием радиации ситуация иная. Если воздействие оказано на логический элемент комбинационной схемы, то на выходе может возникнуть ложный импульс, что влечет изменение соответствующего бита информации. Но этот ложный сигнал существует всего 1–2 нс, после чего опять восстановится тот же сигнал, который был до воздействия. Таким образом, можно считать, что логические элементы комбинационных схем не подвержены мягким отказам. Однако их влияние необходимо учитывать при оценке надежности памяти, так как они являются источником потока кратковременных ложных импульсов, поступающих на вход запоминающего элемента. Воздействие радиации непосредственно на элементы памяти вызывает поток мягких отказов, так как искажает хранимую информацию, т. е. изменяет состояния системы.

Итак, мягкие отказы возникают только в элементах памяти. В элементах комбинационных схем через 1–2 нс наступает самовосстановление. Это существенно влияет на подход к структурному резервированию блоков 1-го типа и требует отдельного рассмотрения. В частности, отказ от троирования мажоритаров существенно упрощает связи между блоками и снимает ограничения, связанные с доменной организацией структур.

Отличия построения блоков 2-го типа (блоков памяти) в восстанавливаемых системах обусловлены следующими дополнительными требованиями к их функциональности.

— Для обнаружения отказов и периодического восстановления информации необходимо организовать в фоновом режиме в дополнение к основному процессу выполнения прикладной программы процесс циклического чтения всех ячеек, контроля, регистрации и исправления ошибок. Взаимодействие основного и фонового процессов следует осуществлять на аппаратном уровне с использованием способа построения двухпортовой памяти.

— Период восстановления информации при мягких отказах существенно влияет на показатели надежности. Поэтому при необходимости в целях повышения надежности накопитель памяти можно разделить на несколько банков, в каждом из которых вспомогательный процесс чтения контроля и восстановления информации должен протекать параллельно.

— В блоках памяти целесообразно предусмотреть средства регистрации ошибок и формирования сообщений, передаваемых на верхний уровень организации.

Это позволит определить текущий уровень деградации системы по частоте возникновения сбоев и заблаговременно вывести систему из эксплуатации.

Интенсивность отказов блоков со структурным резервированием, подверженных потоку «мягких» отказов, пропорциональна периоду восстановления информации [7]. Это очень сильная зависимость. Поэтому временной ресурс блоков, не используемый основным процессом, необходимо употреблять для контроля и восстановления. В связи с этим для анализа надежности восстанавливаемой системы необходимо рассматривать не только структуру (состав блоков и связи), что соответствует статическому состоянию системы, но и процессы в системе, отражающие ее динамическое поведение. Чем меньше период восстановления, тем устойчивее система к воздействию «мягких» отказов. Но уменьшению периода восстановления препятствует тот факт, что блок, требующий восстановления, может быть в текущий момент времени задействован основным вычислительным процессом. Следовательно, при проектировании отказоустойчивых систем и при анализе их надежности необходимо иметь информацию о том, в какие моменты требуется доступ вычислительного процесса к аппаратным ресурсам.

Это выходит за рамки традиционных подходов к анализу надежности и классических моделей надежности, оперирующих лишь информацией о возникновении случайных событий — отказов аппаратных элементов системы.

Возникает задача интеграции сведений о протекании основного вычислительного процесса и встроенных средствах периодического восстановления в процедуру анализа надежности. Попытка моделирования программных компонентов по аналогии с аппаратными блоками в традиционных моделях надежности не позволяет добиться положительных результатов, поскольку:

— поведение программы определяется детерминированным алгоритмом, а не описывается законами распределения случайных событий, используемых в классических моделях надежности;

— в предлагаемом подходе ПО необходимо анализировать именно с точки зрения использования им аппаратных ресурсов, а не с позиции источника возникновения/распространения отказов, как это делается в большинстве моделей надежности.

Моделирование процессов периодического восстановления также связано с рядом проблем. Основная их причина в том, что события восстановления происходят периодически (обусловлены детерминированным процессом), что нарушает требование использования потока случайных событий, характерное для большинства классических моделей надежности. По этой причине применение таких моделей может привести к возникновению значительной погрешности в оценке. Данное ограничение затрудняет построение адекватной модели надежности. Но если аппаратная составляющая системы спроектирована в виде сети восстанавливаемых функциональных блоков, то оценка влияния ПО на надежность может быть сведена к определению периодичности обращения вычислительного процесса к конкретному аппаратному ресурсу. Используя знания об алгоритме программы, для каждого блока можно оценить резерв времени, который может быть использован в целях восстановления данного блока. Этот подход позволяет учесть влияние ПО еще до построения модели надежности. Для этого нужно:

— по алгоритму программы определить, к каким функциональным ресурсам происходит обращение в конкретные моменты времени;

— связать функциональный ресурс с соответствующим аппаратным блоком в структуре, определив тем самым моменты времени, когда данный блок доступен для процесса восстановления;

— для блока задать характеристики восстановления: в простейшем случае — выбрать минимально допустимый период восстановления исходя из сведений о доступности блока.

После определения периода восстановления для каждого восстанавливаемого блока в системе можно «забыть» о влиянии ПО и далее оценивать надежность чисто аппаратной системы с учетом полученных характеристик.

Подход к проектированию восстанавливаемых ИУС с учетом требований к надежности

На основании предложенного выше подхода к оценке влияния ПО на надежность восстанавливаемых систем предлагается разделить процесс проектирования отказоустойчивой системы на следующие этапы.

1. Оценка допустимых периодов восстановления в резервированных функциональных блоках.

2. Анализ надежности каждого функционального блока с учетом оцененных периодов восстановления. Для корректной оценки на этом этапе необходимо учитывать протекание детерминированных процессов восстановления в блоке, поэтому традиционные методы оценки надежности могут оказаться неприменимыми либо приводить к большим погрешностям в результатах. Для решения этой проблемы при оценке надежности блока предлагается использовать имитационные модели вместо традиционных аналитических. Имитационная модель позволяет более точно отразить специфику протекания детерминированных процессов, что на текущем этапе может привести к более точной оценке. Результатом данного этапа является оцененная вероятность отказа (вероятностная функция работоспособности) каждого восстанавливаемого блока. Так как внутренние отказы блока не распространяются в сети, на следующем этапе можно абстрагироваться от структуры блока и представлять его в виде «черного ящика» с предрассчитанной вероятностной функцией работоспособности.

3. Оценка надежности системы на основании оценок надежности блоков с учетом всех аппаратных элементов: самих функциональных блоков, а также источников питания, интерфейсных модулей, оперативной памяти, ПЗУ, процессора и т. д.

Последний этап может быть осуществлен традиционными методами анализа надежности: логико-вероятностными методами, использованием марковской модели, структурной схемой надежности и т. д. — либо с помощью имитационных моделей.

Целями проведения последнего этапа анализа являются:

— определение общих характеристик надежности системы: среднего времени наработки до отказа, вероятностной функции работоспособности;

— определение структурного вклада отдельных элементов системы в ее общую надежность.

Если по результатам анализа требования к надежности не удовлетворяются, то:

— либо осуществляется реорганизация вычислительного процесса в целях уменьшения периодов восстановления блоков;

— либо применяются методы структурного или информационного резервирования к блокам, оказывающим наибольшее влияние на надежность;

— либо реализуются более трудозатратные способы повышения надежности: резервирование на более низком уровне структурного разбиения, изменение технологии производства и т. д.

Перепроектирование осуществляется до тех пор, пока требования к надежности не будут удовлетворены.

Заключение

В результате проведенного анализа установлено, что обеспечение защиты от «мягких» отказов позволяет значительно увеличить надежность вычислительных систем, эксплуатируемых в условиях повышенной радиации. Выявлено, что эффективной является реализация механизмов защиты как на аппаратном уровне, так и на уровне организации вычислительного процесса путем выделения резерва времени, используемого для периодического восстановления искаженной информации.

Предложен подход к проектированию отказоустойчивых систем, основанный на анализе надежности отдельных функциональных блоков с учетом встроенных в них механизмов самовосстановления, по результатам которого производится оценка надежности всей системы и выбор структуры, обеспечивающей выполнение требований к производительности системы и к характеристикам надежности.

Выявлен ряд задач, требующих дополнительных исследований:

— разработка имитационной модели для оценки показателей надежности восстанавливаемых ИУС и определение условий применения аналитических и имитационной модели в процессе проектирования;

— исследование вариантов функциональной организации блоков типа автомата с памятью со структурным резервированием и периодическим восстановлением для ИУС с повышенной радиационной стойкостью;

— анализ вариантов функциональной организации блоков памяти с информационной избыточностью и периодическим восстановлением для ИУС с повышенной радиационной стойкостью;

— создание методики проектирования ИУС с повышенной радиационной стойкостью и разработка инструментальных средств для их синтеза.

Литература

1. **Edmonds L. D., Barnes C. E., Scheick L. Z.** An Introduction to Space Radiation Effects on Microelectronics. — JPL publication 00-06, 2000. — 83 p. http://snebulos.mit.edu/projects/reference/NASA-Generic/JPL-00-06.pdf (дата обращения: 05.12.2015).
2. **Gaillard R.** Single Event Effects Mechanisms and Classification // Frontiers in Electronic Testing. 2011. Vol. 41. P. 27–54.
3. **Amusan O. A.**, et al. Single Event Upsets in Deep-Submicrometer Technologies due to Charge Sharing / O. A. Amusan, L. W. Massengill, M. P. Baze, A. L. Sternberg, A. F. Witulski, B. L. Bhuva, J. D. Black // IEEE Transactions on Device and Materials Reliability. 2008. Vol. 8. N 3. P. 582–589.
4. **Максименко С. Л., Мелехин В. Ф., Филиппов А. С.** Анализ проблемы построения радиационно-стойких информационно-управляющих систем // Информационно-управляющие системы. 2012. № 2(57). С. 18–25.
5. **Максименко С. Л., Мамутова О. В., Филиппов А. С., Мелехин В. Ф.** Методология проектирования восстанавливаемых встраиваемых вычислительных систем // Университетский научный журнал. 2014. № 8. С. 144–153.
6. **Глухих М. И.** Разработка методов синтеза информационно-управляющих систем специального назначения со структурным резервированием: автореф. дис. ... канд. техн. наук. — СПб.: СПбПУ, 2006. — 173 с.
7. **Максименко С. Л., Мелехин В. Ф.** Анализ надежности функциональных узлов цифровых СБИС со структурным резервированием и периодическим восстановлением работоспособного состояния // Информационно-управляющие системы. 2013. № 2(63). С. 18–23.

UDC 681.3
doi:10.15217/issn1684-8853.2016.1.26

Analysis of Radiation Resistance Improvement Issue for Information and Control Systems at the Stage of Functional and Logical Design

Egorov I. V.[a], Post-Graduate Student, iegorov@kspt.icc.spbstu.ru
Melekhin V. F.[a], Dr. Sc., Tech., Professor, melekhin@kspt.ftk.spbstu.ru
[a]Peter the Great St.Petersburg Polytechnic University, 29, Polytechnicheskaya St., 195251, Saint-Petersburg, Russian Federation

Introduction: Under radiation, "soft failures" can occur in elements of computing devices. Their intensity is much higher than the intensity of nonrecoverable failures. To improve the radiation resistance, it is not enough to change the IC manufacturing technology or the circuitry of the standard libraries. The distorted information in the redundant structures of the devices should be repeatedly restored. To provide that, we have to solve a number of synthesis and analysis problems during the functional and logical design of the systems. **Purpose:** The paper gives an analytical overview of the published results on the problem, proposes a conceptual approach to radiation resistance improvement at the stage of functional and logical design, and poses problems for further research. **Results:** The analysis showed that soft information failures under radiation have a different impact on the system compared to nonrecoverable failures. Consequently, they change the structural importance of elements of different types in terms of reliability. An approach is proposed to the organization of computing processes and to the construction of redundant structures of allocated types of system blocks, based on using a time reserve for periodic recovery of the information distorted by the impact of a soft failure stream. The further research should be focused on developing a design technique for information and control systems with improved radiation resistance.
Keywords — Radiation Effects in Semiconductors, Soft Failures, Structural Reservation, Recovery of Information in Redundant Structures, Recoverable Computing Systems.

References

1. Edmonds L. D., Barnes C. E., Scheick L. Z. *An Introduction to Space Radiation Effects on Microelectronics*. JPL publication, 2000, no. 00-06. 83 p. Available at: http://snebulos.mit.edu/projects/reference/NASA-Generic/JPL-00-06.pdf (accessed 05 December 2015).
2. Gaillard R. Single Event Effects Mechanisms and Classification. *Frontiers in Electronic Testing*, 2011, vol. 41, pp. 27–54.
3. Amusan O. A., Massengill L. W., Baze M. P., Sternberg A. L., Witulski A. F., Bhuva B. L, Black J. D. Single Event Upsets in Deep-Submicrometer Technologies due to Charge Sharing. *IEEE Transactions on Device and Materials Reliability*, 2008, vol. 8, no. 3, pp. 582–589.
4. Maximenko S. L., Melekhin V. F., Filippov A. S. Analysis of the Problem of Radiation-Tolerant Information and Control-Systems Implementation. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2012, no. 2(57), pp. 18–25 (In Russian).
5. Maximenko S. L., Filippov A. S., Melekhin V. F., Mamoutova O. V. Design Methodology for Embedded Systems with Built-in Self-Recovery. *Universitetskii nauchnyi zhurnal*, 2014, no. 8, pp. 144–153 (In Russian).
6. Glukhikh M. I. *Razrabotka metodov sinteza informatsionno-upravliaiushchikh sistem spetsial'nogo naznacheniia so strukturnym rezervirovaniem*. Dis. kand. tehn. nauk. [Development of Methods of Synthesis of Special Purpose Control Information Systems with Structural Reservation. PhD tech. sci. diss.]. Saint-Petersburg, Sankt-Peterburgskii politekhnicheskii universitet Publ., 2006. 173 p. (In Russian).
7. Maximenko S. L., Melekhin V. F. Analysis of Reliability of Functional Nodes of Digital VLSI Circuits with Structural Redundancy and Periodic Operational State Recovery. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2013, no. 2(63), pp. 18–23 (In Russian).

УДК 004.89+004.94
doi:10.15217/issn1684-8853.2016.1.32

КОМПЛЕКС ПРОГРАММ ДЛЯ НАВИГАЦИИ МОБИЛЬНЫХ УСТРОЙСТВ ВНУТРИ ПОМЕЩЕНИЙ С ПОМОЩЬЮ НЕЙРОННЫХ СЕТЕЙ

П. А. Новиков[а]*, аспирант*
А. Д. Хомоненко[а]*, доктор техн. наук, профессор*
Е. Л. Яковлев[б]*, начальник учебной лаборатории*
[а]*Петербургский государственный университет путей сообщения Императора Александра I, Санкт-Петербург, РФ*
[б]*Военно-космическая академия им. А. Ф. Можайского, Санкт-Петербург, РФ*

Введение: *навигацию внутри помещений (на транспорте, в общественных заведениях и др.) затруднительно проводить с помощью существующих технологий. Рассматриваются компоненты, общая характеристика построения и подходы к навигации посредством мобильных устройств внутри помещений.* ***Цель:*** *реализация подхода к навигации на основе отпечатков радиосигналов точек доступа Wi-Fi и создание мобильного приложения, предназначенного для навигации мобильных устройств внутри помещений.* ***Результаты:*** *проведен сравнительный анализ современных подходов к навигации. Предложен механизм хранения и обработки радиоотпечатков с помощью искусственных нейронных сетей. При создании комплекса программ навигации мобильных устройств с помощью нейронной сети применены модули открытой библиотеки Fast Artificial Neural Network Library. Описаны используемый алгоритм обучения — упругого распространения ошибки RProp, диаграмма классов и технология работы с комплексом программ навигации. Приведены результаты сравнительного анализа основных алгоритмов обучения нейронной сети, используемой для навигации мобильных приложений внутри помещений, с позиции соотношения точности и трудоемкости в зависимости от числа нейронов скрытого слоя, обозначены направления дальнейших исследований.* ***Практическая значимость:*** *разработан программный комплекс для мобильных устройств, способный осуществлять навигацию внутри помещений на базе существующей инфраструктуры в виде Wi-Fi-сетей.*

Ключевые слова *— навигация внутри помещений, мобильные устройства, Wi-Fi, нейронная сеть, мобильное приложение, отпечатки радиосигналов.*

Введение

Сегодняшний мир не может существовать без систем точной навигации. Спутниковые навигационные системы GPS и ГЛОНАСС [1] широко используются во множестве сфер деятельности, таких как навигация транспорта, строительство, геодезия, сотовая связь и пр. GPS-приемники потребительского уровня установлены практически во все современные телефоны. На открытой местности подобные приемники способны обеспечить точность определения местоположения в районе 1–5 м. Они могут быть применены внутри больших торговых комплексов, складских помещений или систем типа «умный дом», в которых различные домашние системы (отопления, освещения, кондиционирования и пр.) могут централизованно управляться в автоматическом режиме.

Наглядным аргументом необходимости навигации внутри помещений могут служить задачи навигации в транспортных системах, например в зданиях аэропортов и вокзалов [2]:

— для посетителей: определение местоположения; поиск стоек регистрации, касс, камер хранения, кафе, парковки, такси и пр.; прокладка маршрута с учетом этажности зданий; поиск коллег внутри аэропорта (вокзала);

— для аэропортов, вокзалов: обширная аналитика перемещений посетителей и персонала; сервис, основанный на знании положения посетителя (location-based service); геоконтекстная реклама (location-based advertising) товаров и услуг; рекламные акции с учетом местоположения людей.

Однако такие приемники не удовлетворяют существующей потребности выполнения навигации в закрытых помещениях. В подобных условиях GPS либо вообще не работает, либо предоставляет данные о местоположении с очень большой погрешностью около 100–150 м.

В настоящее время активно проводятся работы по совершенствованию и практическому использованию методов навигации мобильных устройств в помещениях. Рассмотрим кратко современные подходы к решению указанной задачи и предложения по обоснованию выбора алгоритмов обучения и подбору параметров нейронных сетей (НС).

Характеристика подходов к навигации мобильных устройств

Прежде всего, отметим, что современные мобильные устройства содержат множество различных датчиков и приемников. К основным из них относятся:

1) GPS;

2) Wi-Fi, работает с радиосигналом на частотах 2,4 и 5 Ггц;

3) Bluetooth, работает на тех же частотах;

4) акселерометр и гироскоп — инерционные датчики, измеряющие линейные и угловые ускорения;

5) магнитомер — датчик, измеряющий напряженность магнитных полей;

6) датчики влажности, освещенности, приближения; барометр и пр.

Датчики 1–5 широко представлены во многих современных мобильных устройствах, которые работают под управлением операционных систем iOS и Android. Эти системы предоставляют программный доступ к датчикам через собственные программные интерфейсы (API), и любые приложения могут получать данные с этих датчиков (с некоторыми несущественными ограничениями).

В настоящее время существует несколько способов навигации с применением мобильных устройств на основе следующих технологий.

1. GPS (Global Positioning System) и ГЛОНАСС (Глобальная навигационная спутниковая система) [1, 3] — спутниковая навигация. Хорошо подходит для определения местоположения на открытых пространствах. Из-за необходимости быть в поле видимости как минимум трех спутников плохо подходит для закрытых помещений, так как они сильно ухудшают спутниковый сигнал. Первоначальный поиск спутников может занимать порядка нескольких минут.

2. AGPS (Assisted GPS) — навигация с использованием радиосигналов от вышек сотовой связи в комбинации с GPS. Ускоряет первоначальное определение координат за счет обработки данных об обслуживающих мобильное устройство сотовых вышках.

3. Навигация с использованием радиомаячков, работающих на технологии Bluetooth Low Energy (iBeacons) [3]. Относительно молодая технология, толчком к развитию которой послужил последний стандарт технологии Bluetooth 4.0(+) с низким энергопотреблением. Маячок представляет собой микросхему с радиомодулем, который с заданной частотой посылает радиопакеты с информацией о себе. Приемник, зная карту расположения маячков и силу сигнала до ближайших из них, вычисляет свое относительное местоположение.

4. Навигация на основе отпечатков (fingerprints) радиосигналов точек доступа Wi-Fi [4]. Подобные системы навигации носят, скорее, исследовательский характер, так как из-за относительной новизны данного подхода еще не выработаны проверенные способы осуществления точной навигации.

5. Инерциальная система навигации [5] по данным с инерциальных датчиков устройства, на основе которых строится модель положения устройства в пространстве.

Приведем сравнительную характеристику наиболее распространенных технологий навигации с применением мобильных устройств [2] (табл. 1).

Можно отметить работы, в которых развиваются идеи рассматриваемых нами технологий и приводятся сравнительно новые способы навигационных систем. Статья [6] посвящена интеграции Wi-Fi и инерциальной навигационной системе. В работе [7] описаны мобильные навигационные сервисы и применение технологии OpenCellID для определения местоположения мобильных устройств. Сравнительно новый подход к позиционированию мобильных устройств в помещениях на основе двумерных штрихкодов предложен в работе [8]. В статье [9] представлена модель контекстно-зависимых вычислений (контекстно-зависимый браузер), основанная на сетевой близости. При этом мобильный телефон используется как сенсор близости, и геопозиционная информация заменена сетевой близостью. Приведен алгоритм вычисления траекторий в мобильных сетях на базе информации о сетевой близости.

■ ***Таблица 1.*** Сравнительная характеристика технологий

Система	Достоинства	Недостатки
GPS	Средняя точность (5 м) Удобство использования Хорошая совместимость	Невозможность работы внутри помещения
GSM	Удобство использования Хорошая совместимость	Низкая точность (100 м)
Wi-Fi	Средняя точность (5 м) Работа внутри помещений	Необходимость развертывания Wi-Fi-сети Ограниченная совместимость (на iOS ввиду запрета компании Apple) Высокое энергопотребление
iBeacon	Высокая точность (1–2 м) Удобство использования Хорошая совместимость (iOS, Android) Низкое энергопотребление	Необходимость развертывания BLE-сети

О навигации мобильных устройств с помощью нейронных сетей

Работа системы навигации с помощью отпечатков радиосигналов от точек доступа Wi-Fi состоит из двух частей:

1) составление базы данных известных отпечатков радиосигналов с сопоставленными им известными координатами;

2) получение новых координат из базы данных по новым отпечаткам радиосигналов.

В качестве механизма хранения и обработки радиоотпечатков рассматривается модель искусственных нейронных сетей (ИНС). При этом нами используется ИНС со скрытым слоем (рис. 1), которая в настоящее время получила широкое применение.

Рассматриваемый подход интересен по следующим причинам [10]:

— задача определения местоположения на основе ранее заданных радиоотпечатков может быть рассмотрена как задача классификации многомерных данных;

— основная вычислительная нагрузка этого метода приходится на процесс обучения, который выполняется однократно и может быть проведен на независимом вычислительном комплексе;

— данные об окружающих Wi-Fi-точках можно дополнять данными от других источников, например от Bluetooth LE маячков.

Возможности современных ИНС являются предметом активных исследований. Подход к использованию ИНС для решения названной задачи рассмотрен в нескольких публикациях, например [11–14]. В отличие от указанных работ, в настоящей статье предлагается решать задачу навигации мобильных устройств на основе ИНС и программного комплекса, реализуемого в виде *мобильного приложения*.

Характеристика программного комплекса навигации

Разработанный программный комплекс для навигации по Wi-Fi-сигналам с использованием мобильного телефона состоит из двух компонентов:

1) мобильного приложения, которое:

а) собирает данные об окружающих Wi-Fi-точках доступа;

б) использует обученную НС для определения текущего местоположения мобильного устройства на основе новых данных о Wi-Fi-сигналах;

2) настольного приложения, которое:

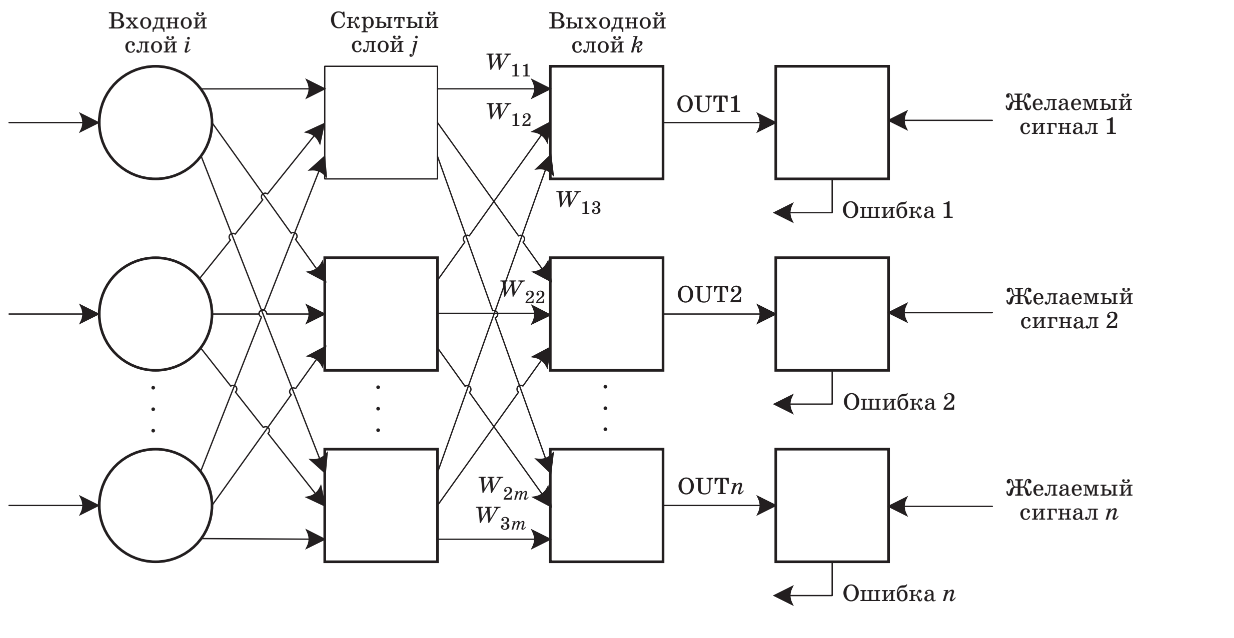
а) подготавливает данные о радиоотпечатках для последующего обучения НС;

б) обучает и тестирует НС.

В качестве реализации НС в комплексе используется многослойная ИНС, представленная свободной библиотекой FANN (Fast Artificial Neural Network Library). Основные преимущества этой библиотеки заключаются в ее открытости, наличии нескольких алгоритмов обучения и простоте использования.

Для обучения ИНС используется алгоритм упругого распространения ошибки (Resilient Propagation — RProp) [15]. Такой выбор обусловлен тем, что алгоритм RProp превосходит алгоритм обратного распространения ошибки (Backprop) по следующим показателям:

— меньшее количество эпох обучения;



■ ***Рис. 1.*** Нейронная сеть со скрытым слоем

— большая надежность работы алгоритма при различных начальных параметрах.

В процессе обучения алгоритм RProp использует знаки частных производных для подстройки весовых коэффициентов. Для каждого весового коэффициента w_{ij} вводится отдельное значение-модификатор Δ_{ij}, и только оно служит для определения размера изменения связанного весового коэффициента. Это значение-модификатор изменяется в течение всего процесса обучения, и его изменение определяется по правилу [15]

$$\Delta_{ij}^{(t)} = \begin{cases} \eta^{+}\Delta_{ij}^{(t-1)}, \text{если } \dfrac{\partial E^{(t-1)}}{\partial w_{ij}} \dfrac{\partial E^{(t)}}{\partial w_{ij}} > 0; \\ \eta^{-}\Delta_{ij}^{(t-1)}, \text{если } \dfrac{\partial E^{(t-1)}}{\partial w_{ij}} \dfrac{\partial E^{(t)}}{\partial w_{ij}} < 0, \end{cases}$$

где $0 < \eta^{-} < 1 < \eta^{+}$; $\dfrac{\partial E^{(t)}}{\partial w_{ij}}$ — частная производная активационной функции по весовому коэффициенту w_{ij} в момент времени t.

Каждый раз, когда частная производная соответствующего весового коэффициента w_{ij} меняет знак (это означает, что предыдущее изменение коэффициента было слишком большим, и алгоритм перешагнул через локальный минимум), значение-модификатор Δ_{ij} уменьшается на коэффициент η^{-}. Если производная сохранила свой знак, то значение-модификатор Δ_{ij} немного увеличивается для ускорения сходимости в пологих областях.

После того как обновлены значения-модификаторы для всех весовых коэффициентов w_{ij}, изменение самих коэффициентов происходит по следующему правилу: если производная положительна (при увеличивающейся ошибке), вес уменьшается на значение-модификатор; если производная отрицательна, значение-модификатор добавляется к значению веса:

$$\Delta w_{ij}^{(t)} = \begin{cases} -\Delta_{ij}^{(t)}, \text{если } \dfrac{\partial E^{(t)}}{\partial w_{ij}} > 0; \\ +\Delta_{ij}^{(t)}, \text{если } \dfrac{\partial E^{(t)}}{\partial w_{ij}} < 0; \\ 0, \text{иначе}, \end{cases}$$

$$w_{ij}^{(t+1)} = w_{ij}^{(t)} + \Delta w_{ij}^{(t)}.$$

В работе [15] авторы предлагают и обосновывают использование следующих начальных параметров:

— начальное весовое значение-модификатор $\Delta_0 = 0{,}1$. Для рассматриваемого алгоритма выбор начального значения этого параметра не критичен и не влияет на скорость сходимости алгоритма;

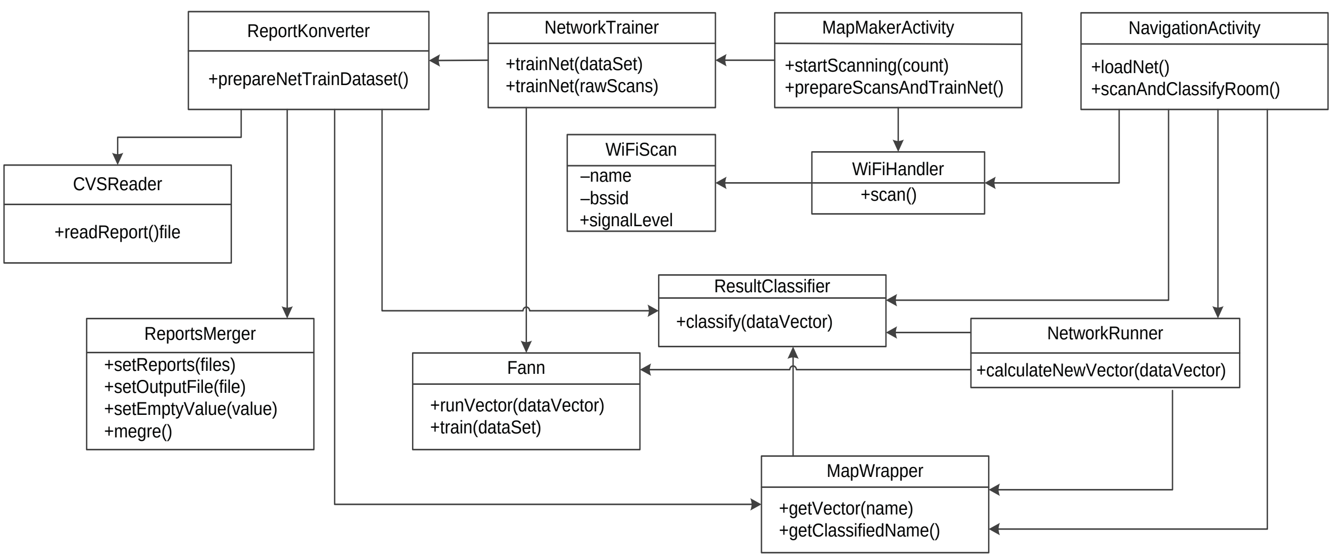
— величина значения-модификатора ограничивается пределами $\Delta_{\min} = 1\text{e}^{-6}$ и $\Delta_{\max} = 50$;

— коэффициенты $\eta^{-} = 0{,}5$ и $\eta^{+} = 1{,}2$.

Текущая реализация программного комплекса навигации представляет собой одномерный классификатор, который по новым радиоотпечаткам способен определить помещение, в котором находится мобильное устройство. На рис. 2 показана упрощенная диаграмма классов программного комплекса.

Взаимодействие с пользователем происходит с помощью двух экранов, представленных классами MapMakerActivity и NavigationActivity.

Класс MapMakerActivity отвечает за создание карты для дальнейшей навигации. С помощью класса WiFiHandler производится сканирование



■ ***Рис. 2.*** Диаграмма классов программного комплекса

Wi-Fi-точек доступа. Далее полученный набор данных передается в класс ReportsConverter, который отвечает за преобразование исходных данных в обучающий набор в формате библиотеки FANN. Полученный обучающий набор передается в класс Fann, в котором происходит обучение ИНС.

Класс NavigationActivity отвечает за классификацию новых сканированных данных с помощью обученной ранее ИНС.

Работа с программным комплексом навигации

Работа с созданным программным комплексом навигации реализуется с использованием настольного и мобильного программного обеспечения и включает следующие этапы.

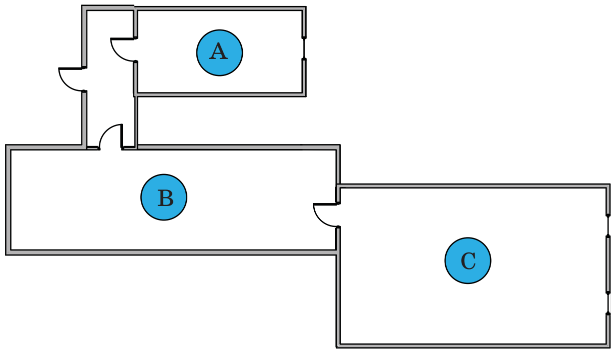
1. Сбор данных о радиоотпечатках в исследуемых помещениях. На выходе этого этапа получается набор файлов, каждый из которых содержит совокупность векторов (матрицу) измерений сигналов Wi-Fi-точек (табл. 2).

2. Объединение всех исходных файлов в один. В результате получается один общий файл с измерениями во всех исследуемых помещениях (табл. 3).

Пустые значения (0) заменяются на –100 (очень слабый сигнал).

Кроме того, создается файл “names.txt”, в котором перечислены имена всех сетей в том порядке, в котором они располагаются в файле обучения (для этапа 4).

3. Обозначается карта сопоставления названий помещений (рис. 3) с их формальным числовым представлением: A [0]; B [0,5] и C [1].



■ *Рис. 3.* Схема помещений

4. На основе содержимого файлов этапов 2 и 3 создается файл с данными для обучения сети:

```
3  15  1

-86.0 -100.0 -54.0 -69.0 -100.0 -100.0 -88.0 -100.0 ↲
-100.0 -100.0 -100.0 -100.0 -100.0 -100.0 -100.0
0.0

-89.0 -83.0 -56.0 -66.0 -100.0 -100.0 -91.0 -100.0 ↲
-100.0 -100.0 -100.0 -100.0 -100.0 -100.0 -100.0
0.5

-90.0 -71.0 -51.0 -67.0 -100.0 -100.0 -88.0 -100.0 ↲
-100.0 -100.0 -100.0 -100.0 -100.0 -100.0 -100.0
1.0
```

В первой строке указываются количество обучающих векторов, число нейронов входного слоя и число нейронов выходного слоя.

В следующих парах строк указываются значения входного вектора и значения, которым НС должна обучиться для этого вектора.

■ ***Таблица 2.*** Набор радиоотпечатков в одном помещении

№ измерения	Сила измеренного сигнала, дБм, до точки						
	Net1	Net2	Net3	Net4	Net5	Net6	Net7
1	–86	0	–54	–69	0	0	–88
2	–89	–83	–56	–66	0	0	–91
…	…	…	…	…	…	…	…

Значение 0 показывает, что точка была недоступна в этот момент времени.

■ ***Таблица 3.*** Объединенный набор радиоотпечатков нескольких помещений

№ измерения	Net1	Net2	Net3	Net4	Net5	Net6	Net7
Помещение A							
1	–86	–100	–54	–69	–100	–100	–88
2	–89	–83	–56	–66	–100	–100	–91
Помещение B							
1	…	…	…	…	…	…	…

Результатом этого этапа является файл "network", представляющий собой обученную НС в формате библиотеки FANN. Имя сети выбирается произвольно.

Следующие этапы описывают процесс работы программного приложения, запущенного на мобильном устройстве.

5. Мобильное приложение запускает библиотеку FANN и передает ей файл с обученной НС.

6. Мобильное приложение начинает сканирование окружающих Wi-Fi-точек доступа. После завершения очередной итерации сканирования приложение получает массив доступных в текущее время точек доступа, к которым устройство может попробовать подключиться. Каждый объект этого массива, кроме прочего, содержит следующие два поля: «Название» и «Сила сигнала до заданной точки».

С помощью массива имен из файла "names.txt", полученного на втором этапе, и массива текущих видимых точек доступа строится радиоотпечаток, который, по сути, является простым массивом чисел. Создание радиоотпечатка включает следующие шаги:

1) создать массив чисел (float[]) размером, определяемым числом элементов массива имен;

2) пройтись по массиву имен в цикле:

а) если в массиве точек доступа есть точка с текущим именем, то добавить в радиоотпечаток силу сигнала до этой точки по текущему индексу итерации;

б) иначе, добавить по текущему индексу итерации значение по умолчанию для отсутствующей точки, которое было принято на шаге 2 (–100).

На выходе итерации сканирования получается радиоотпечаток в виде вектора значений (–100, –50, –75, –100, ...).

7. Полученный вектор передается в НС, которая возвращает результирующий вектор (float[]). Так как НС была обучена на основе одномерных данных, то нас интересует только первый элемент результирующего вектора.

8. Полученное значение сравнивается с данными из карты помещений. В используемой версии программного комплекса в карте ищется наиболее близкое значение к текущему, полученному из сети, значению. После его нахождения на экран выводится название текущего помещения, которому соответствует измеренный радиоотпечаток.

Одной из основных причин выбора алгоритма RProp обучения НС в программном комплексе стал ограниченный набор реализованных алгоритмов в библиотеке FANN. Для того чтобы в дальнейшем добиться большей скорости обучения в комплексе, далее проводится моделирование и сравнение других известных алгоритмов обучения на существующем наборе данных.

Сравнение алгоритмов обучения нейронной сети

Сравнение алгоритмов обучения НС проведено в широко распространенной среде MatLab. Данные для обучения НС взяты из файла "train.txt", используемого при проверке работоспособности программного комплекса навигации. Для проверки характеристик обучения НС исходный файл целесообразно разбить на два набора (обучения и проверки).

Сравнительный анализ основных алгоритмов обучения НС, используемой для навигации мобильных приложений внутри помещений, проведем по аналогии с работами [10, 16] — с позиции соотношения точности и трудоемкости в зависимости от числа нейронов скрытого слоя. Для анализа были выбраны часто используемые алгоритмы: алгоритм обучения Левенберга — Маркарта [17]; метод шкалированных связанных градиентов [18]; метод байесовской регуляризации, представляющий собой модернизированный алгоритм Левенберга — Маркарта [19]. Результаты сравнения алгоритмов обучения приведены в табл. 4.

На основе анализа табл. 4 можно сделать вывод, что оптимальное число нейронов скрытого слоя для данного набора исходных данных находится в диапазоне 15÷25. Наибольшую точность показал алгоритм байесовской регуляризации, но он значительно более трудоемок по сравнению с другими алгоритмами. Оптимальным по соотношению точности и сходимости оказался алгоритм Левенберга — Маркарта.

Отметим общую для всех алгоритмов низкую точность, что связано с исходными данными. Вероятно, следует уменьшить размерность входного вектора (использовались данные от 15 WiFi-

■ ***Таблица 4.*** Характеристики трудоемкости и точности алгоритмов обучения НС

Число нейронов	trainlm		trainscg		trainbr	
	MSE	Epochs(s)	MSE	Epochs(s)	MSE	Epochs(s)
5	0,11	12(0,1)	0,08	15(0,1)	0,03	592(25)
10	0,06	16(0,1)	0,09	20(0,2)	0,02	700(20)
15	0,07	9(0,1)	0,05	13(0,2)	0,013	800(23)
20	0,05	11(0,2)	0,13	14(0,2)	0,015	269(48)
30	0,13	6(0,2)	0,19	23(0,3)	0,018	889(120)
50	0,23	13(0,2)	0,21	30(0,5)	0,02	483(187)

trainlm — алгоритм обучения Левенберга — Маркарта; trainscg — метод шкалированных связанных градиентов; trainbr — метод байесовской регуляризации; MSE — средняя квадратическая ошибка; Epochs(s) — число циклов обучения, s в скобках — время в секундах.

точек, большая часть которых недоступна) и увеличить общее количество измерений. В то же время в исходных данных набор векторов измерений соответствовал всего трем разным помещениям, поэтому точность получилась достаточной.

Заключение

В качестве рекомендации к применению программного комплекса отметим важность использования mac-адреса точки доступа вместо ее имени, так как оно не является уникальным. Возникшие коллизии могут нарушить работу всего комплекса.

Достоинством программного комплекса является то, что он способен работать на существующей инфраструктуре Wi-Fi-сетей, развернутых в различных помещениях, вроде жилых домов и торговых комплексов.

К определенным недостаткам можно отнести особенность рассматриваемого подхода, заключающуюся в том, что перед его использованием необходимо вручную составить карту радиоотпечатков.

Дальнейшие исследования, на наш взгляд, целесообразно продолжить в следующих направлениях:

— реализация многомерного классификатора для преобразования радиоотпечатков в многомерные координаты (x, y, высота);

— проведение исследований влияния других алгоритмов обучения ИНС (Quickprop, Batch, Incremental) на скорость обучения и точность навигации;

— разработка метода снижения размерности входного вектора радиоотпечатков;

— повышение качества навигации при комбинировании с инерционной системой навигации;

— добавление обработки сигналов от Bluetooth LE маячков к сигналам от Wi-Fi-точек доступа;

— определение оптимального числа и, возможно, размещения Wi-Fi-точек внутри каждого помещения (к примеру, по аналогии с подходом, предлагаемым в работе [20]).

Литература

1. **Соловьев Ю. А.** Системы спутниковой навигации. — М.: ЭКО ТРЕНДЗ, 2000. — 267 с.
2. http://www.indoorsnavi.com/ (дата обращения: 20.10.2015).
3. *Навигация в помещениях с iBeacon и ИНС.* http://habrahabr.ru/post/245325/ (дата обращения: 28.10.2015).
4. **Kukolj D., Vuckovic M., Pletl S.** Indoor Location Fingerprinting Based on Data Reduction: Proc. 2011 Intern. Conf. on Broadband and Wireless Computing, Communication and Applications (BWCCA 2011), Barcelona, Oct. 26–28, 2011. P. 327–332.
5. **Бутаков Н. А.** Применимость инерциальной навигации в мобильных устройствах//Intern. Journal of Open Information Technologies. 2014. Vol. 2. N 5. P. 24–32.
6. **Evennou F., Marx F.** Advanced Integration of WiFi and Inertial Navigation Systems for Indoor Mobile Positioning//EURASIP Journal on Applied Signal Processing. 2006. Vol. 2006. P. 1–11. doi:10.1155/ASP/2006/86706
7. **Дворкина Н. Б., Намиот Д. Е., Дворкин Б. А.** Мобильные навигационные сервисы и применение технологии OpenCellID для определения местоположения//Геоматика. 2010. № 2. С. 80–87.
8. **Абдрахманова А. М., Намиот Д. Е.** Использование двумерных штрихкодов для создания системы позиционирования и навигации в помещении// Прикладная информатика. 2013. № 1(43). С. 31–39.
9. **Намиот Д. Е., Шнепс-Шнеппе М. А.** Анализ траекторий в мобильных сетях на базе информации о сетевой близости // Автоматика и вычислительная техника. 2013. № 3. С. 48–60.
10. **Novikov P. A., Khomonenko A. D., Yakovlev E. L.** Justification of the Choice Learning Algorithms of Neural Networks for Indoor Mobile Positioning: Proc. of the 11th Central and Eastern European Software Engineering Conf. in Russia. Moscow, Oct. 22–23 (24), 2015 (in the press).
11. **Mok E., Cheung Bernard K. S.** An Improved Neural Network Training Algorithm for Wi-Fi Fingerprinting Positioning//ISPRS Intern. Journal Geo-Inf. 2013. N 2. P. 854–868.
12. **Hamid Mehmood, Nitin K.** Tripathi and Taravudh Tipdecho. Indoor Positioning System Using Artificial Neural Network//Journal of Computer Science. 2010. N 6(10). P. 1219–1225.
13. **Gogolak L., Pletl S., Kukolj D.** Neural Network-based Indoor Localization in WSN Environments//Acta Polytechnica Hungarica. 2013. Vol. 10. N 6. P. 221–235.
14. **Fang S. H., Lin T. N.** Indoor Location System Based on Discriminant-Adaptive Neural Network in IEEE 802.11 Environments//IEEE Trans. Neural Networks. 2008. Vol. 19. N 11. P. 1973–1978.
15. **Riedmiller M., Braun H.** A Direct Adaptive Method for Faster Backpropagation Learning: Proc. of the IEEE Intern. Conf. on Neural Networks. 1993. P. 586–591.
16. **Хомоненко А. Д., Яковлев Е. Л.** Нейросетевая аппроксимация характеристик многоканальных немарковских систем массового обслуживания // Тр. СПИИРАН. 2015. № 4(41). С. 81–93.
17. **Hagan M. T., Menhaj M.** Training Feedforward Networks with the Marquardt Algorithm//IEEE

Transactions on Neural Networks. 1994. Vol. 5. N 6. P. 989–993.
18. **Moller M. F.** A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning//Neural Networks. 1993. Vol. 6. P. 525–533.
19. **Foresee F. D., Hagan M. T.** Gauss-Newton Approximation to Bayesian Regularization: Proc. of the 1997 Intern. Joint Conf. on Neural Networks. 1997. P. 1930–1935.
20. **Блаунштейн Н. Ш., Сергеев М. Б.** Иерархия размещения фемтосот/пикосот/макросот в городской среде с плотным расположением абонентов, находящихся в помещениях и вне их // Информационно-управляющие системы. 2013. № 4(65). С. 36–47.

UDC 004.89+004.94
doi:10.15217/issn1684-8853.2016.1.32

Software for Mobile Indoor Navigation using Neural Networks

Novikov P. A.[a],Post-Graduate Student, imortan@gmail.com
Khomonenko A. D.[a], Dr. Sc., Tech., Professor, khomon@mail.ru
Yakovlev E. L.[b], Head of the Training Laboratory, evgen-1932@yandex.ru
[a]Petersburg State Transport University, 9, Moskovskiy Pr., 190031 Saint-Petersburg, Russian Federation
[b]A. F. Mozhaiskii Military Space Academy, 13, Zhdanovskaia St., 197198, Saint-Petersburg, Russian Federation

Introduction: The paper discusses indoor mobile device navigation (on transport, in public institutions, etc.), giving comparative characteristics of conventional and modern technologies. **Purpose:** The navigation should be provided with an approach based on Wi-Fi access point radio fingerprints. One more goal is developing a mobile app for indoor navigation. **Results:** A mechanism is proposed to store and process radio fingerprints using artificial neural networks. For the development of neural-network navigation mobile software, we used modules from the public Fast Artificial Neural Network Library. The paper discusses the training algorithm we used (resilient propagation of an error), the class diagram and the navigation software technology. The basic neural network training algorithms for mobile indoor navigation have been compared in the aspect of accuracy/complexity ratio depending on the number of neurons in the hidden layer. Directions for further research have been designated. **Practical relevance:** Mobile software has been developed which is capable of indoor navigation on the base of the existing Wi-Fi network infrastructure.
Keywords — Indoor Navigation, Neural Network, Wi-Fi, Mobile Application, Radio Fingerprints.

References

1. Solov'ev Iu. A. *Sistemy sputnikovoi navigatsii* [Satellite Navigation Systems]. Moscow, EKO TRENDZ Publ., 2000. 267 p. (In Russian).
2. Available at: http://www.indoorsnavi.com/ (accessed 20 October 2013).
3. *Navigatsiia v pomeshcheniiakh s iBeacon i INS* [Navigation in Areas with iBeacon and ANN]. Available at: http://habrahabr.ru/post/245325/ (accessed 28 October 2015).
4. Kukolj D., Vuckovic M., Pletl S. Indoor Location Fingerprinting Based on Data Reduction. *Proc. 2011 Intern. Conf. on Broadband and Wireless Computing, Communication and Applications (BWCCA 2011)*, Barcelona, October 26–28, 2011, pp. 327–332.
5. Butakov N. A. Applicability of Inertial Navigation in Mobile Devices. *International Journal of Open Information Technologies*, 2014, vol. 2, no. 5, pp. 24–32 (In Russian).
6. Evennou F., Marx F. Advanced Integration of WiFi and Inertial Navigation Systems for Indoor Mobile Positioning. *EURASIP Journal on Applied Signal Processing*, vol. 2006, pp. 1–11. doi:10.1155/ASP/2006/86706
7. Dworkina N. B., Namiot D. E., Dvorkin B. A. Mobile Navigation Services and the use of Technology to Determine the Location OpenCellID. *Geomatika* [Geomatics], 2010, no. 2, pp. 80–87 (In Russian).
8. Abdrakhmanova A. M., Namiot D. E. Using Two-Dimensional Bar Code to Create a System of Positioning and Navigation Inside. *Prikladnaia informatika* [Applied Informatics], 2013, no. 1(43), pp. 31–39 (In Russian).
9. Namiot D. E., Shneps-Shneppe M. A. Analysis of the Trajectories in Mobile Networks on the Basis of Information about Network Proximity. *Avtomatika i vychislitel'naia tekhnika* [Automation and Computer Science], 2013, no. 3, pp. 48–60 (In Russian).
10. Novikov P. A., Khomonenko A. D., Yakovlev E. L. Justification of the Choice Learning Algorithms of Neural Networks for Indoor Mobile Positioning. *Proc. of the 11th Central and Eastern European Software Engineering Conf. in Russia*, Moscow, October 22–23 (24), 2015 (in the press).
11. Mok E., Cheung Bernard K. S. An Improved Neural Network Training Algorithm for Wi-Fi Fingerprinting Positioning. *ISPRS Intern. Journal Geo-Inf*, 2013, no. 2, pp. 854–868.
12. Hamid Mehmood, Nitin K. Tripathi and Taravudh Tipdecho. Indoor Positioning System Using Artificial Neural Network. *Journal of Computer Science*, 2010, no. 6(10), pp. 1219–1225.
13. Gogolak L., Pletl S., Kukolj D. Neural Network-based Indoor Localization in WSN Environments. *Acta Polytechnica Hungarica*, 2013, vol. 10, no. 6, pp. 221–235.
14. Fang S. H., Lin T. N. Indoor Location System Based on Discriminant-Adaptive Neural Network in IEEE 802.11 Environments. *IEEE Trans. Neural Networks*, 2008, vol. 19, no. 11, pp. 1973–1978.
15. Riedmiller M., Braun H. A Direct Adaptive Method for Faster Backpropagation Learning. *Proc. of the IEEE Intern. Conf. on Neural Networks*, 1993, pp. 586–591.
16. Khomonenko A. D., Yakovlev E. L. Neural Network Approximation of Characteristics of Multi-Channel non-Markovian Queuing Systems. *Trudy SPIIRAN*, 2015, no. 4(41), pp. 81–93 (In Russian).
17. Hagan M. T., Menhaj M. Training Feedforward Networks with the Marquardt Algorithm. *IEEE Transactions on Neural Networks*, 1994, vol. 5, no. 6, pp. 989–993.
18. Moller M. F. A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning. *Neural Networks*, 1993, vol. 6, pp. 525–533.
19. Foresee F. D., Hagan M. T. Gauss-Newton Approximation to Bayesian Regularization. *Proc. of the 1997 Intern. Joint Conf. on Neural Networks*, 1997, pp. 1930–1935.
20. Blaunshteyn N. S., Sergeev M. B. A Hierarchy of Deployment of Femto/Pico/Macrocells in Urban Environment with Dense Distribution of Outdoor and Indoor Subscribers. *Informatsionno-aupravliaiushchie sistemy* [Information and Control Systems], 2013, no. 4(65), pp. 36–47 (In Russian).

УДК 629.051
doi:10.15217/issn1684-8853.2016.1.40

РЕАЛИЗАЦИЯ ПОСТРОЕНИЯ МАРШРУТНЫХ ТРАЕКТОРИЙ ДЛЯ ОТОБРАЖЕНИЯ НА БОРТОВЫХ МНОГОФУНКЦИОНАЛЬНЫХ ИНДИКАТОРАХ

В. В. Маркелов[а]*, ведущий инженер-программист*
М. О. Костишин[а, б]*, ассистент, старший инженер*
И. О. Жаринов[а, б]*, заведующий кафедрой, руководитель учебно-научного центра*
В. А. Нечаев[а, б]*, доцент, главный конструктор*
[а]*АО «ОКБ «Электроавтоматика», Санкт-Петербург, РФ*
[б]*Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург, РФ*

Введение: *для обеспечения эффективного выполнения экипажем воздушного судна заданного плана полета предусматривается отображение траектории маршрута полета на бортовых многофункциональных индикаторах, что требует использования в бортовых комплексах специализированного математического аппарата, обеспечивающего преобразование данных плана полета и их визуализацию в виде прогнозируемой траектории движения воздушного судна в горизонтальной плоскости.* ***Цель:*** *разработка принципов и алгоритмов построения маршрутных траекторий полета летательного аппарата в горизонтальной плоскости, основанных на унификации используемых маршрутных траекторий и реализуемых в бортовых вычислительных комплексах в условиях ограничения потребляемых вычислительных ресурсов.* ***Результаты:*** *сформулированы алгоритмы формирования и расчета массивов данных маршрутных траекторий для построения их отображения на многофункциональных индикаторах. Траектория полета отображается в виде последовательной прорисовки определенных графических примитивов в соответствии с массивом данных для отображения маршрута. Массив данных сформирован программным обеспечением бортового комплекса на основе информации, представленной в полетном задании и соответствующей базе данных Jeppesen или аналогичной. Формирование массива осуществлено приведением заданных траекторий маршрута к форматам трех типовых описанных траекторий. При этом каждая из типовых траекторий имеет стандартное описание и алгоритм расчета параметров, составляющих массив данных для отображения.* ***Практическая значимость:*** *полученные алгоритмы построения маршрутных траекторий унифицируют алгоритмы формирования информации для отображения на индикаторах навигационной обстановки и оптимизируют набор расчетных данных для управления полетом по траектории в горизонтальной плоскости.*

Ключевые слова *— траектория маршрута летательного аппарата, навигационная обстановка, зональная навигация.*

Введение

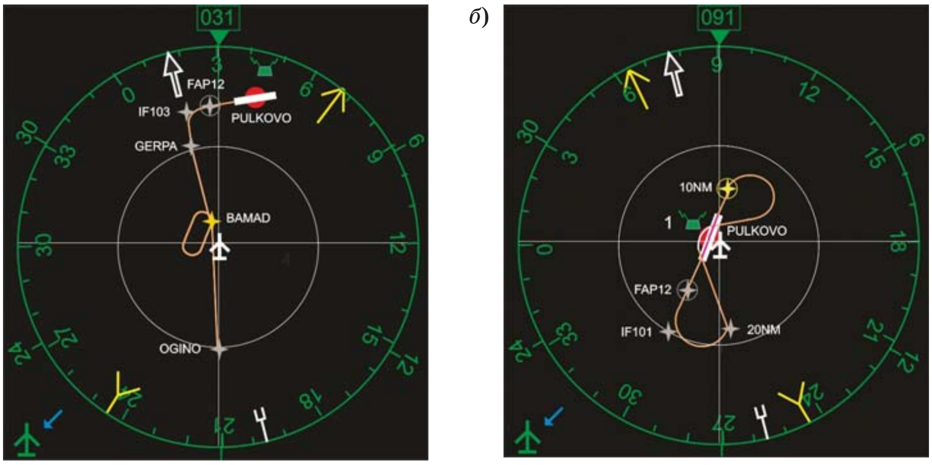
Для обеспечения выполнения полета по заданному в полетном задании (плане полета) маршруту предусматривается отображение данного маршрута на индикаторах навигационной обстановки в кабине экипажа.

Современные требования по точности навигации при полете по маршруту, в том числе по правилам зональной навигации с вероятностью выдерживания заданной траектории 0,95 не более 1,85 км (навигационные характеристики *RNP* не хуже *RNP*1), обусловливают повышенные требования и к точности отображения маршрута на бортовых индикаторах, включая отображение участков перехода между фиксированными траекториями маршрута, которые должны максимально соответствовать фактической реализации данных переходов, обеспечиваемой бортовой системой управления [1–11]. Выполнение данных требований реализуется при использовании современных многофункциональных индикаторов и детализации маршрута на отдельные конечные траектории.

В общем случае представление маршрута состоит из набора заданных контрольных точек маршрута и набора образующих маршрут траекторий, определенных в геодезической системе координат. Непосредственно отображение маршрута на индикаторе обеспечивается преобразованием контрольных точек и траекторий из геодезической в экранную систему координат с учетом выбранного окна индикатора и масштаба изображения [12–17].

Исходные данные для построения маршрута включают в себя массив навигационной информации, определенный полетным заданием (планом полета) и содержащий последовательный перечень контрольных точек и описание траекторий полета между ними. В зависимости от поставленных задач описание контрольных точек и траекторий может соответствовать стандартам ARINC-424 или собственной структуре базы данных [18–21].

Бортовой комплекс преобразует заданную навигационную информацию в массив образующих маршрут траекторий, каждая из которых формируется из характерных расчетных точек, соединенных друг с другом. Координаты характерных точек траектории и тип соединительной линии определяются расчетным путем в зависимости *а*) *б*)



■ ***Рис. 1.*** Отображение участков маршрута на многофункциональном индикаторе: *а* — полет по маршруту; *б* — повторный заход на посадку

от заданных параметров соответствующего участка маршрута. Принципы и алгоритмы преобразования не являются регламентированными и определяются целевым назначением воздушного судна и техническими характеристиками бортового комплекса [1, 2, 22].

Вариант отображения траекторий полета на бортовом многофункциональном индикаторе представлен на рис. 1.

В качестве одного из вариантов алгоритма преобразования навигационной информации, определенной полетным заданием, представлен типовой вариант формирования массива данных для отображения маршрута (участков маршрута) на бортовых индикаторах в целях обеспечения полетов по правилам зональной навигации или аналогичным.

Массив данных траектории

Массив данных каждой траектории включает в себя последовательный набор графических примитивов дуга — отрезок, определенных в геодезической системе координат. В общем случае каждая траектория состоит из набора дуг, построенных из характерных расчетных точек и соединенных отрезками.

Схема построения типовой траектории представлена на рис. 2. Для ее отображения используется ряд следующих рассчитываемых параметров, составляющих два или три участка типа дуга — прямая:

— геодезические координаты текущего местоположения ($B0$, $L0$);

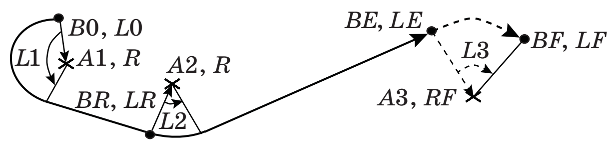
— радиус разворота (R);

— азимут центра первого разворота ($A1$);

— угол дуги первого разворота ($L1$) (знаковый);

— геодезические координаты точки конечного разворота (BR, LR);

— азимут центра конечного разворота ($A2$);



■ ***Рис. 2.*** Схема построения траектории

— угол дуги конечного разворота ($L2$) (знаковый);

— геодезические координаты точки окончания траектории (BE, LE);

— радиус дуги (RF) (только для полета по дуге);

— азимут центра дуги ($A3$) (только для полета по дуге) (знаковый);

— угол полета по дуге ($L3$) (только для полета по дуге).

При построении маршрута последовательно отображаются все траектории, из которых маршрут состоит. При этом точка окончания траектории (BE, LE или BF, LF — для полета по дуге) принимается в качестве точки текущего местоположения ($B0$, $L0$) для следующей траектории.

Величина радиуса разворота (R), используемая при построении траектории, определяется скоростью полета и углом крена и вычисляется по формуле

$$R = V_{\text{ист}}^2 / (9{,}81 \cdot \mathrm{tg}(\gamma)),$$

где $V_{\text{ист}}$ — истинная скорость (заданная на данной траектории); γ — значение номинального угла крена на текущей траектории.

Для определения окончания траектории с упреждением разворота используется значение расчетного радиуса разворота для полета по следующей траектории (RE), определяемое данной формулой, в качестве исходных данных которой являются заданные параметры, относящиеся к следующей траектории.

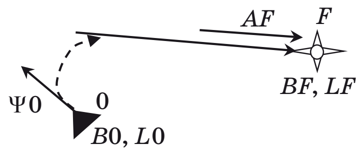
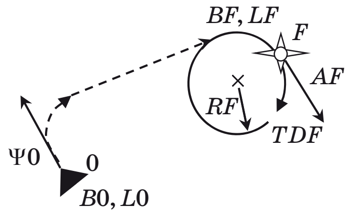
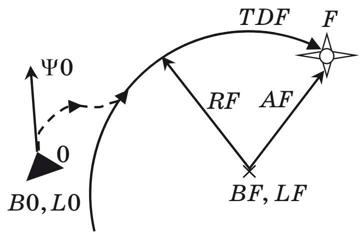
Траектории полета в горизонтальной плоскости

Траектории полета в горизонтальной плоскости могут быть представлены тремя типовыми траекториями. Варианты реализуемых типовых траекторий и исходные данные для их расчета представлены в табл. 1.

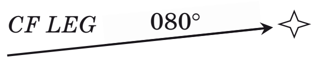
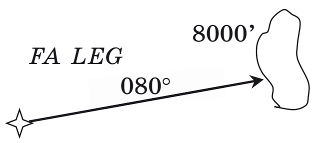
При построении маршрута, состоящего из последовательности траекторий, в качестве текущего путевого угла принимается расчетный азимут окончания предыдущей траектории.

Представление стандартных типов траекторий, в том числе траекторий зональной навигации [1, 2, 18, 19], в виде типовых приведено в табл. 2. Там же представлено преобразование

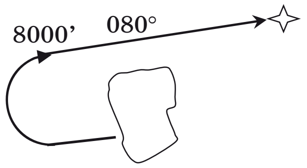
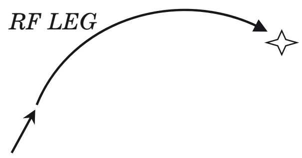
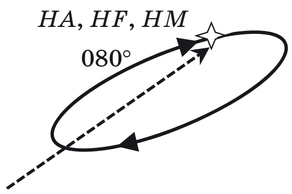
■ ***Таблица 1.*** Типовые траектории в горизонтальной плоскости

Траектория	Схема	Исходные данные
Линия пути		*B*0, *L*0 — координаты текущего местоположения Ψ0 — текущий путевой угол *BF*, *LF* — координаты исполняемой навигационной точки *AF* — конечный азимут ортодромии
Прямо на точку (на касательную к окружности, построенной в заданном направлении)		*B*0, *L*0 — координаты текущего местоположения Ψ0 — текущий путевой угол *BF*, *LF* — координаты исполняемой навигационной точки *AF* — заданное направление подхода *RF* — радиус окружности *TDF* — направление разворота
Полет по дуге		*B*0, *L*0 — координаты текущего местоположения Ψ0 — текущий путевой угол *BF*, *LF* — координаты центра дуги *AF* — заданный исполняемый азимут навигационной точки *RF* — радиус дуги *TDF* — направление разворота

■ ***Таблица 2.*** Представление траекторий в типовом формате

Описание траектории	Исходные данные траектории	Исходные данные типовой траектории
Линия пути между двумя точками 	*B*1, *L*1 — координаты первой точки *B*2, *L*2 — координаты второй точки	*Линия пути* $BF, LF = B2, L2$ $AF = F_{об}(B2, L2, B1, L1) + \pi$
Полет с выдерживанием ЗПУ* на точку 	*B*, *L* — координаты точки ЗПУ	*Линия пути* $BF, LF = B, L$ $AF =$ ЗПУ
Полет прямо на точку 	*B*, *L* — координаты точки	*Прямо на точку* $BF, LF = B, L$ $AF = RF = 0$ $TDF = 0$
Полет с выдерживанием ЗПУ от точки до высоты 	*B*, *L* — координаты точки ЗПУ Δ*H* — необходимый набор высоты	*Линия пути* $BF, LF = F_{пр}(B, L, \text{ЗПУ}, H/grd)$, где *grd* — заданный градиент набора высоты $AF =$ ЗПУ

■ *Окончание табл. 2*

Описание траектории	Исходные данные траектории	Исходные данные типовой траектории
Выход на точку с ЗПУ на заданном удалении 	B, L — координаты точки ЗПУ $D_{уп}$ — дальность упреждения	*1-й участок прямо на точку* BF, $LF = F_{пр}(B, L, ЗПУ + \pi, D_{уп})$ AF = ЗПУ RF = радиус разворота $TDF = 0$ *2-й участок линии пути* BF, $LF = B$, L AF = ЗПУ
Полет по дуге заданного радиуса до точки 	B, L — координаты точки $B_ц$, $L_ц$ — координаты центра дуги $R_{дуги}$ — радиус дуги НР — направление разворота	*Полет по дуге* BF, $LF = B_ц$, $L_ц$ $AF = F_{об}(B_ц, L_ц, B, L)$ $RF = R_{дуги}$ TDF = НР
Полет в зоне ожидания 	B, L — координаты точки зоны ожидания ЗПУ S — длина участка приближения НР — направление разворота	*1-й участок прямо на точку* BF, $LF = F_{пр}(B, L, ЗПУ + \pi, S)$ AF = ЗПУ RF = радиус разворота TDF = НР *2-й участок линии пути* BF, $LF = B$, L AF = ЗПУ

* ЗПУ — заданный путевой угол.

исходных данных для расчета траекторий по типовым методикам.

Алгоритмы расчета параметров трех типовых траекторий: линия пути, прямо на точку и полет по дуге — приведены далее.

Функция типа $F_{пр}(B, L, A, D)$ представляет собой решение прямой геодезической задачи в пространстве и обеспечивает расчет геодезических координат навигационной точки по известным координатам заданной точки, азимуту и дальности. Функция типа $F_{об}(B1, L1, B2, L2)$ представляет собой решение обратной геодезической задачи на эллипсоиде и обеспечивает расчет азимута от первой заданной координатами навигационной точки ко второй.

Исходные данные для расчета параметров типовых траекторий

В качестве исходных данных для расчета траекторий принимаются:

— текущий путевой угол ($\Psi 0$);

— азимут текущего местоположения от исполняемой навигационной точки или центра дуги — для траектории полета по дуге ($AF0$);

— азимут на исполняемую навигационную точку или центр дуги — для траектории полета по дуге от текущего местоположения ($A0F$);

— расстояние до исполняемой навигационной точки или центра дуги ($D0F$).

Для траектории линии пути дополнительно:

— конечный азимут ортодромии (AF).

Для путевого участка комбинированного способа или для выхода на точку касания окружности в заданном направлении дополнительно:

— заданное направление подхода (AF);

— радиус окружности конечного разворота (RF);

— направление разворота по окружности (TDF) — при необходимости.

Для траектории полета по дуге дополнительно:

— радиус дуги (RF);

— заданный азимут исполняемой навигационной точки от центра дуги (AF);

— направление разворота (TDF).

Все развороты на участке выхода на траекторию осуществляются с расчетным радиусом (R).

Расчет параметров траектории линии пути

Участок выхода на траекторию полета по линии пути — ортодромии — включает в себя дугу первого разворота от текущего местоположения до достижения угла приближения и дугу конечного разворота выхода на ортодромию от точки начала конечного разворота, соединенные прямой. Радиусы дуг равны расчетному радиусу разворота. Параметры участка выхода на линию пути представлены на рис. 3.

Боковое отклонение от ортодромии

$$Z = D0F \cdot \sin(AF - AF0).$$

Азимут ортодромии

$$\Psi F = AF + \pi + (A0F - AF0) \text{ в диапазоне } [-\pi; \pi].$$

Направление первого разворота при угле приближения, равном 45°:

— при $|Z| \geq R \cdot (1 - \cos(\pi/4))$

$$TD = \mathrm{sign}(\sin(\mathrm{sign}(Z) \cdot \pi/4 - (\Psi 0 - \Psi F)));$$

— при $|Z| < R \cdot (1 - \cos(\pi/4))$

$$TD = \mathrm{sign}(\sin(\mathrm{sign}(Z) \cdot \arccos(1 - |Z|/R) - (\Psi 0 - \Psi F))).$$

Здесь и далее при определении направления разворотов: при 1 — разворот по часовой стрелке, при минус 1 — против.

В случае иной величины угла приближения значение $\pi/4$ в приведенных формулах заменяется на соответствующее значение.

Угол приближения представляет собой угол выхода на дугу окружности, касательную к ортодромии, и отсчитываемый от направления ортодромии.

Угол приближения:

— при $TD \cdot Z > R \cdot (1 + \cos(\Psi 0 - \Psi F) - 2\cos\pi/4)$

$$\Phi = TD \cdot \pi/4;$$

— при $TD \cdot Z \geq R \cdot (\cos(\Psi 0 - \Psi F) - 1)$ и $TD \cdot Z \leq R \cdot (1 + \cos(\Psi 0 - \Psi F) - 2\cos\pi/4)$

$$\Phi = TD \cdot \arccos(0{,}5 \cdot (1 + \cos(\Psi 0 - \Psi F) - TD \cdot Z/R));$$

— при $TD \cdot Z < R \cdot (\cos(\Psi 0 - \Psi F) - 1)$

$$\Phi = -TD \cdot \pi/4.$$

Путевой угол приближения

$$\Psi D = \Psi F + \Phi \text{ в диапазоне } [-\pi; \pi].$$

Азимут и дальность точки начала конечного разворота от текущего местоположения (AR, DR), определяющие координаты точки начала конечного разворота (BR, LR):

$$AR = \Psi F + \mathrm{arctg}((Z - ZR) / SR);$$

$$DR^2 = ((Z - ZR)^2 + SR^2),$$

где (ZR, SR) — координаты точки начала конечного разворота в ортодромической системе координат:

$$ZR = \mathrm{sign}(\Phi) \cdot R \cdot (1 - \cos(\Phi));$$

$$SR = TD \cdot R/\sin(\Phi)(1 - \cos(\Psi 0 - \Psi F - \Phi)) + (Z - ZR)/\mathrm{tg}(\Phi); \text{ при } \Phi = 0 \; SR = 0.$$

Азимут центра первого разворота

$$A1 = \Psi 0 + \pi/2 \cdot TD.$$

Угол дуги первого разворота

$$L1 = TD \cdot L10,$$

где $L10 = TD \cdot (\Psi D - \Psi 0)$ в диапазоне $[0; 2\pi]$.

Азимут центра конечного разворота

$$A2 = \Psi D - \pi/2 \cdot \mathrm{sign}(\Phi).$$

Угол дуги конечного разворота

$$L2 = (-\Phi) \cdot R.$$

Суммарная длина прямолинейного участка ортодромии

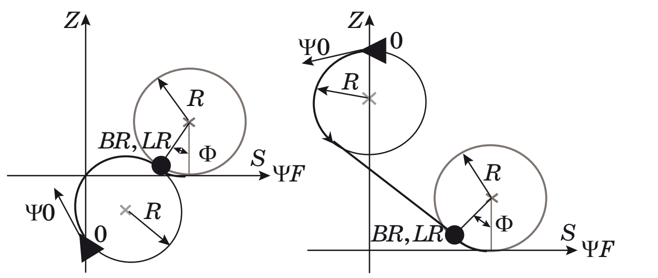
$$S0 = -D0F \cdot \cos(AF0 - AF) - (SR + \mathrm{sign}(\Phi) \cdot R \cdot \sin(\Phi)).$$

Таким образом, определен необходимый набор параметров, составляющих массив данных траектории линии пути для ее отображения.

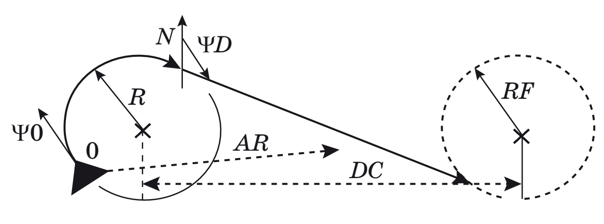
Расчет параметров траектории прямо на точку

Участок выхода на траекторию включает в себя дугу первого разворота от текущего местоположения до достижения угла приближения с радиусом, равным расчетному радиусу разворота. Параметры участка выхода на траекторию представлены на рис. 4.

Направление первого разворота определяется из условия обеспечения меньшего угла разворо-



■ *Рис. 3.* Параметры выхода на траекторию полета по линии пути



■ *Рис. 4.* Параметры траектории выхода прямо на точку та, потребного для выхода на азимут приближения к заданной навигационной точке или точке касания окружности.

Направление разворота рассчитывается по формуле

$$TD = \mathrm{sign}(\sin(AR - \Psi 0 - TDF \cdot \arcsin(RF/DR))),$$

где дальность и азимут к центру окружности конечного разворота (DR, AR):

$$DR^2 = D0F^2 + RF^2 +$$

$$+ 2 \cdot TDF \cdot D0F \cdot RF \cdot \sin(AF - AF0);$$

$$AR = A0F - TDF \cdot \arcsin(RF \cdot \cos(AF - AF0)/DR);$$

TDF — направление конечного разворота, определяемое для путевого участка комбинированного способа или для выхода на точку касания окружности в заданном направлении. Направление конечного разворота подлежит расчету при отсутствии его однозначного задания и определяется по формуле

$$TDF = \mathrm{sign}\,(R \cdot \mathrm{sign}(\sin(\Psi 0 - AF))(1 + \cos(A0F -$$

$$- AF0 + AF - \Psi 0)) + D0F\sin(AF0 - AF)).$$

Угол приближения есть путевой угол выхода на исполняемую навигационную точку либо, для путевого участка комбинированного способа полета или для выхода на точку касания окружности, путевой угол выхода на касательную к окружности конечного разворота. Угол приближения

$$\Psi D = AC + \arcsin((TD \cdot R - TDF \cdot RF)/DC)$$

в диапазоне $[-\pi; \pi]$,

где расстояние между центрами окружностей первого и конечного разворотов (DC) и угол между ними (AC):

$$DC^2 = DR^2 + R^2 + TD \cdot 2 \cdot DR \cdot R \cdot \sin(\Psi 0 - AR);$$

$$AC = AR - TD \cdot \arcsin(R/DC \cdot \cos(\Psi 0 - AR)).$$

Азимут центра первого разворота

$$A1 = \Psi 0 + \pi/2 \cdot TD.$$

Угол дуги первого разворота

$$L1 = TD \cdot L10,$$

$$L10 = TD \cdot (\Psi D - \Psi 0) \text{ в диапазоне } [0; 2\pi].$$

Суммарная длина прямолинейного участка траектории

$$S0^2 = DC^2 - TDF \cdot (R - TD \cdot TDF \cdot RF)^2.$$

Вычисленные параметры определяют в полном объеме массив данных траектории прямо на точку.

Расчет параметров траектории полета по дуге

Участок выхода на траекторию полета по дуге включает в себя дугу первого разворота от текущего местоположения до достижения угла приближения и дугу конечного разворота выхода на заданную исполняемую дугу от точки начала конечного разворота, соединенные прямой. Радиусы дуг при разворотах равны расчетному радиусу разворота.

Параметры траектории выхода на дугу представлены на рис. 5.

Направление первого разворота определяется из условия обеспечения меньшего угла разворота, потребного для выхода на траекторию приближения к дуге. Направление разворота:

— при $D0F^2 \geq (RF^2 + 2 \cdot RF \cdot R)$

$$TD = \mathrm{sign}(\sin(A0F - \Psi 0));$$

— при $D0F \geq RF$ и $D0F^2 < (RF^2 + 2 \cdot RF \cdot R)$

$$TD = \mathrm{sign}(TDF \cdot \cos(A0F - \Psi 0 -$$

$$- TDF \cdot \arccos((D0F^2 - RF^2)/2/R/D0F - RF/D0F)));$$

— при $D0F < RF$ и $D0F^2 > (RF^2 - 2 \cdot RF \cdot R)$ и $D0F > (2 \cdot R - RF)$

$$TD = -\,\mathrm{sign}(TDF \cdot \cos(A0F - \Psi 0 -$$

$$- TDF \cdot \arccos((D0F^2 - RF^2)/2/R/D0F + RF/D0F)));$$

— при $D0F^2 \leq (RF^2 - 2 \cdot RF \cdot R)$ и $D0F > (2 \cdot R - RF)$

$$TD = -\,\mathrm{sign}(\sin(A0F - \Psi 0));$$

— при $D0F \leq (2 \cdot R - RF)$

$$TD = TDF.$$

Угол приближения есть путевой угол выхода на дугу окружности, касательную к заданной исполняемой дуге. Угол приближения:

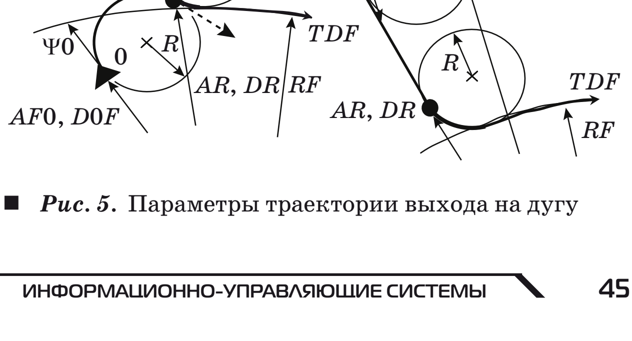
— при $DC \geq (RF + R)$

$$\Psi D = AC + TD \cdot \arcsin(R/DC) \text{ в диапазоне } [-\pi; \pi];$$

— при $DC < (RF + R)$ и $DC > (RF - R)$

$$\Psi D = AC + TD \cdot \pi/2 - TDF \cdot \arccos((DC^2 + 4R^2 -$$

$$- (RF + TDF \cdot TD \cdot R)^2)/4/DC/R) \text{ в диапазоне } [-\pi; \pi];$$



■ ***Рис. 5.*** Параметры траектории выхода на дугу

— при $DC \le (RF - R)$

$$\Psi D = AC - TD \cdot \arcsin(R/DC) + \pi \text{ в диапазоне } [-\pi; \pi],$$

где дальность от центра первого разворота до центра заданной дуги (DC):

$$DC^2 = D0F^2 + R^2 + 2 \cdot TD \cdot D0F \cdot R \cdot \sin(\Psi 0 - A0F).$$

Азимут от центра первого разворота до центра заданной дуги:

— при $D0F^2 > R^2 + DC^2$

$$AC = \Psi 0 +$$

$$+ TD \cdot (\pi/2 - \arcsin(D0F \cdot \cos(\Psi 0 - A0F)/DC));$$

— при $D0F^2 \le R^2 + DC^2$

$$AC = \Psi 0 -$$

$$- TD \cdot (\pi/2 - \arcsin(D0F \cdot \cos(\Psi 0 - A0F)/DC)).$$

Дальность и азимут точки начала конечного разворота от центра дуги (DR, AR), определяющие координаты точки начала конечного разворота (BR, LR):

— при $DC \ge (RF + R)$

$$DR^2 = RF^2 + 2 \cdot RF \cdot R;$$

$$AR = AF0 - A0F + \Psi D;$$

— при $DC < (RF + R)$ и $DC > (RF - R)$

$$DR^2 = DC^2 + R^2 + TD \cdot 2 \cdot DC \cdot R \cdot \sin(AC - \Psi D);$$

$$AR = AF0 - A0F + AC +$$

$$+ TDF \cdot \arccos((DC^2 + DR^2 - R^2)/2/DC/DR);$$

— при $DC \le (RF - R)$

$$DR^2 = RF^2 - 2 \cdot RF \cdot R;$$

$$AR = AF0 - A0F + \Psi D + \pi.$$

Азимут центра первого разворота

$$A1 = \Psi 0 + \pi/2 \cdot TD.$$

Угол дуги первого разворота

$$L1 = TD \cdot L10,$$

где $L10 = TD \cdot (\Psi D - \Psi 0)$ в диапазоне $[0; 2\pi]$.

Угол дуги конечного разворота:

— при $DR \ge RF$

$$L2 = -TDF \cdot \arccos((R^2 +$$

$$+ (RF + R)^2 - DR^2)/2/R/(RF + R));$$

— при $DR < RF$

$$L2 = TDF \cdot \arccos(-(R^2 +$$

$$+ (RF - R)^2 - DR^2)/2/R/(RF - R)).$$

Азимут центра конечного разворота

$$A2 = \Psi D + \pi/2 \cdot \mathrm{sign}(L2).$$

Вычисленные параметры однозначно определяют массив данных траектории полета по дуге и ее отображение.

Окончание траекторий

Азимут окончания траектории полета по линии пути

$$\Psi E = AF \text{ в диапазоне } [-\pi; \pi].$$

Азимут окончания траектории полета прямо на точку

$$\Psi E = \Psi D + A0F - AF0 + \pi \text{ в диапазоне } [-\pi; \pi].$$

Окончание траектории полета по линии пути, а также полета прямо на точку, если траектория не является участком комбинированного способа или выходом на точку касания окружности, определяется признаком прохождения исполняемой навигационной точки, который может принимать значения:

— пролет исполняемой точки;

— достижение упреждения разворота.

Упреждение разворота предусматривает сдвиг точки окончания траектории от исполняемой навигационной точки на линейное упреждение разворота (SE) в целях обеспечения выхода на следующую траекторию полета в один разворот.

Если следующая траектория — линия пути, в качестве расчетного параметра принимается азимут на точку следующей траектории от исполняемой точки (AN).

Если следующая траектория — полет по дуге, расчетные параметры:

— азимут на центр следующей дуги от исполняемой точки (AN);

— дальность до центра следующей дуги от исполняемой точки (DN);

— радиус следующей дуги (RN);

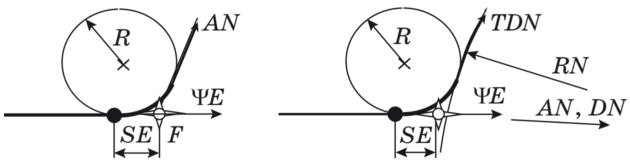
— направление разворота по следующей дуге (TDN).

Развороты на участке выхода на следующую траекторию осуществляются с расчетным радиусом разворота (RE).

Параметры линейного упреждения разворота представлены на рис. 6.

Если следующая траектория — линия пути, упреждение разворота

$$SE = RE \cdot |\mathrm{tg}((\Psi E - AN)/2)|.$$



■ *Рис. 6.* Параметры линейного упреждения разворота

Если следующая траектория — полет по дуге, упреждение разворота при $|RN + TDR \cdot RE| > |TDN \cdot RE - TDR \cdot DN \cdot \sin(\Psi E - AN)|$

$$SE2 = TDR \cdot \mathrm{sqrt}((RN + TDR \cdot RE)^2 - (TDN \cdot RE - TDR \cdot DN \cdot \sin(\Psi E - AN))^2) - DN \cdot \cos(\Psi E - AN),$$

где $TDR = \mathrm{sign}(\cos(\Psi E - AN))$.

При невыполнении данного условия или при $SE < 0$ $SE = 0$.

Во избежание некорректного перехода между соседними траекториями величина линейного упреждения разворота ограничивается некоторой заданной величиной, в частности радиусом разворота.

Азимут точки окончания траектории от исполняемой навигационной точки, определяющий координаты точки окончания траектории (BL, LE):

$$AE = \Psi E + \pi.$$

В случае путевого участка комбинированного способа полета по траектории или выхода на точку касания заданной окружности траектория оканчивается в точке касания окружности конечного разворота на исполняемую точку. Дальность и азимут точки окончания траектории (DE, AE), отсчитываемые от исполняемой навигационной точки, определяющие координаты точки окончания траектории (BE, LE):

$$DE = RF \cdot \mathrm{sqrt}(2 \cdot (1 + \cos(A0F - AF0 + \Psi D - AF));$$

$$AE = \arccos(DE/RF/2).$$

Траектория полета по дуге строится от точки выхода на дугу до исполняемой навигационной точки. Координаты точки выхода на дугу (BE, LE) определяются расчетным азимутом точки выхода на дугу от центра и дальностью, равной радиусу дуги. Азимут точки выхода на дугу от центра дуги

$$A0 = \Psi D + L2 - \pi/2 \cdot TDF.$$

Азимут центра разворота по заданной дуге до исполняемой точки

$$A3 = A0 + \pi.$$

Угол разворота по заданной дуге

$$L3 = TDF \cdot L30,$$

где $L30 = TDF \cdot (AF - A0)$ в диапазоне $[0; 2\pi]$.

Азимут окончания траектории полета по дуге

$$\Psi E = AF + \pi/2 \cdot TDF \text{ в диапазоне } [-\pi; \pi].$$

Прерывание траектории

Прерывание траектории полета по линии пути определяется условием

$$S0 - SE < 0.$$

Прерывание траектории полета прямо на точку определяется условием

$$|TD \cdot R - TDF \cdot RF| > DC \text{ или } S0 - SE < 0.$$

Прерывание траектории полета по дуге определяется условием

$$|L3| > (2\pi - \pi/12).$$

Выполнение условий прерывания означает невозможность выхода на траекторию линии пути до достижения точки окончания траектории. При этом данная траектория игнорируется, и ее расчет не производится.

Заключение

Построение маршрута полета летательного аппарата в горизонтальной плоскости для отображения на бортовых индикаторах обеспечивается преобразованием заданных траекторий, составляющих маршрут, определенный полетным заданием (планом полета), в фиксированный массив данных траекторий. Массив данных содержит определенный набор расчетных точек в геодезической системе координат, соединенных друг с другом либо дугой, либо отрезком.

При этом любая заданная траектория полета воздушного судна (в том числе траектория зональной навигации) может быть представлена в виде одной или нескольких определенных типовых траекторий: линия пути, прямо на точку или полет по дуге. Данное представление унифицирует алгоритмы формирования информации для отображения на индикаторах навигационной обстановки и оптимизирует набор расчетных данных для управления полетом по траектории в горизонтальной плоскости.

Литература

1. Производство полетов воздушных судов. Т. 1: Правила производства полетов. Doc 8168. — Монреаль: ИКАО, 2006. http://www.icao.int (дата обращения: 12.05.2015).
2. Производство полетов воздушных судов. Т. 2: Построение схем визуальных полетов и полетов по приборам. Doc 8168. — Монреаль: ИКАО, 2006. http://www.icao.int (дата обращения: 12.05.2015).
3. Minimum Aviation System Performance Standards: Required Navigation Performance for Area Navigation. RTCA DO-236B. (EUROCAE ED-75). http://www.rtca.org (дата обращения: 12.05.2015).
4. Global Positioning System. Standard Positioning Service. Performance Standard. Department of Defense USA. Washington. DC 20301-6000, Feb. 2007. http://www.gps.gov (дата обращения: 07.06.2014).

5. Aeronautical Information Services Manual. Doc 8126-AN/872. http://www.icao.int (дата обращения: 12.05.2015).
6. **Вовк В. И., Липин А. В., Сарайский Ю. Н.** Зональная навигация. — СПб.: Академия ГА, 2004. — 127 с.
7. Организация воздушного движения. Doc 4444-RAC/501/12. — Монреаль: ИКАО, 2007. http://www.icao.int (дата обращения: 12.05.2015).
8. **Гриневич А. С., Музалев А. А., Ройзензон А. Л.** Руководство по построению аэродромных схем и определению безопасных высот пролета препятствий. Изд. 3-е, доп. — СПб.: Министерство транспорта РФ, 1998. — 300 с.
9. **Щепилов Ю. Н.** Построение аэродромных схем: учеб. пособие. — СПб.: СПбГУ ГА, 2013. — 120 с.
10. Performance-based Navigation (PNB) Manual. Vol. I: Concept and Implemtation Guidance. Doc 9613. — Montreal: ICAO, 2008. http://www.icao.int (дата обращения: 12.05.2015).
11. Performance-based Navigation (PNB) Manual. Vol. II: Implementing RNAV and RNP. Doc 9613. — Montreal: ICAO, 2008. http://www.icao.int (дата обращения: 12.05.2015).
12. Cocpit Display System Interfaces to User Systems. ARINCS Specification 661. http://www.aviation-ia.com (дата обращения: 24.11.2014).
13. **Парамонов П. П.** и др. Реализация структуры данных, используемых при формировании индикационного кадра в бортовых системах картографической информации/ П. П. Парамонов, П. В. Коновалов, И. О. Жаринов, Ю. А. Кирсанова, С. Б. Уткин // Научно-технический вестник информационных технологий, механики и оптики. 2013. № 2. С. 165–167.
14. **Жаринов И. О., Жаринов О. О.** Бортовые средства отображения информации на плоских жидкокристаллических панелях: учеб. пособие. — СПб.: ГУАП, 2005. — 144 с.
15. **Костишин М. О.** и др. Оценка точности визуализации местоположения объекта в геоинформационных системах и системах индикации навигационных комплексов пилотируемых летательных аппаратов/ М. О. Костишин, И. О. Жаринов, О. О. Жаринов, В. А. Нечаев, В. Д. Суслов // Научно-технический вестник информационных технологий, механики и оптики. 2014. № 1. С. 87–93.
16. **Парамонов П. П., Ильченко Ю. А., Жаринов И. О., Тарасов П. Ю.** Структурный анализ и синтез графических изображений на экранах современных средств бортовой индикации на плоских жидкокристаллических панелях // Авиакосмическое приборостроение. 2004. № 5. С. 50–57.
17. **Парамонов П. П.** и др. Принцип формирования и отображения массива геоинформационных данных на экран средств бортовой индикации/ П. П. Парамонов, М. О. Костишин, И. О. Жаринов, В. А. Нечаев, С. А. Сударчиков // Научно-технический вестник информационных технологий, механики и оптики. 2013. № 6. С. 136–142.
18. Navigation System Data Base. ARINC Specification 424. http://www.aviation-ia.com (дата обращения: 24.11.2014).
19. Guidance Material for Design of Terminal Procedures for Area Navigation (DME/DME, B-GNSS, Baro-VNAV & RNP-RNAV). — Eurocontrol, March 2003. — 217 с.
20. Сборники аэронавигационной информации. ЦАИ ГА России. http://www.caiga.ru (дата обращения: 21.04.2014).
21. **Липин А. В., Ключников Ю. И.** Применение зональной навигации при обслуживании воздушного движения. Зональная навигация. — СПб.: Университет ГА, 2011. — 79 с.
22. **Харин Е. Г.** Комплексная обработка информации навигационных систем летательных аппаратов. Опыт многолетнего практического применения. — М.: Изд-во МАИ, 2002. — 264 с.

UDC 629.051
doi:10.15217/issn1684-8853.2016.1.40

Forming Route Trajectories for Aiborne Multi-Function Displays

Markelov V. V.[a], Senior Software Engineer, vvmarkel@gmail.com
Kostishin M. O.[a, b], Post-Graduate Student, Engineer, maksim@kostishin.com
Zharinov I. O.[a, b], Dr. Sc., Tech., Chef of Department, Chef of Learning-Scientist Center, igor_rabota@pisem.net
Nechaev V. A.[a, b], Assistance of Professor, Mail Constructor, nil-12@mail.ru
[a]Saint-Petersburg Scientific Design Bureau «Electroavtomatika», 40, Marshala Govorova St., 198095, Saint-Petersburg, Russian Federation
[b]Saint-Petersburg National Research University of Information Technologies, Mechanics and Optics, 49, Kronverksky Pr., 197101, Saint-Petersburg, Russian Federation

Introduction: To ensure that the aircraft crew efficiently follows the specified flight plan, the flight route trajectory is mapped onto airborne multi-function displays. This requires specialized mathematical tools which would transform the flight plan data and visualize it as a predictable trajectory of the aircraft in the horizontal plane. **Purpose:** The goal is to develop algorithms for building aircraft route trajectories in the horizontal plane, based on the unification of common trajectories. The algorithms should be implemented on airborne computer systems under limited computing resources. **Results:** Algorithms have been built for the formation and calculation of route trajectories and their visual output. A flight trajectory is drawn on an airborne multi-function display as a series of graphic primitives in accordance with the route data array. The array is generated by the on-board software, using the information provided in the flight mission in Jeppesen database format or similar. When building the array, the given route is reduced to the formats of three standard predefined trajectories. Each of the three has a standard description and an algorithm for calculating the parameters which form an array of the data to display. **Practical relevance:** The obtained algorithms which form the route trajectories unify the algorithms which generate the information to display and optimize the calculated data in order to control the flight path in the horizontal plane.

Keywords — Aircraft Route Trajectory, Navigation Conditions, Area Navigation.

Reference

1. *Proizvodstvo poletov vozdushnykh sudov. T. 1. Pravila proizvodstva poletov* [Procedures for Air Navigation Services — Aircraft Operations. Vol. 1 (PANS-OPS)]. Doc 8168. Available at: http://www.icao.int (accessed 12 May 2015).
2. *Proizvodstvo poletov vozdushnykh sudov. T. 2. Postroenie skhem vizual'nykh poletov i poletov po priboram* [Procedures for Air Navigation Services — Aircraft Operations. Vol. 2 (PANS-OPS)]. Doc 8168. Available at: http://www.icao.int (accessed 12 May 2015).
3. *Minimum Aviation System Performance Standards: Required Navigation Performance for Area Navigation*. RTCA DO-236. (EUROCAE ED-75). Available at: http://www.rtca.org (accessed 12 May 2015).
4. *Global Positioning System. Standard Positioning Service. Performance Standard. Department of Defense USA*. Washington, DC 20301-6000, Feb. 2007. Available at: http://www.gps.gov (accessed: 07 June 2014).
5. *Aeronautical Information Services Manual*. Doc 8126-AN/872. Available at: http://www.icao.int (accessed 12 May 2015).
6. Vovk V. I., Lipin A. V., Sarayskij Yu. N. *Zonal'naia navigatsiia* [Area Navigation]. Saint-Petersburg, Akademia grazhdanskoi aviatsii Publ., 2004. 127 p. (In Russian).
7. *Organizatsiia vozdushnogo dvizheniia* [Rules of the Air and Air Traffic Services]. Doc 4444-RAC/501/12. ICAO. Available at: http://www.icao.int (accessed 12 May 2015).
8. Grinevich A. S., Muzalev A. A., Roynezon A. L. *Rukovodstvo po postroeniiu aerodromnykh skhem i opredeleniiu bezopasnykh vysot proleta prepiatstvii* [Development of Terminal Procedures and Minimum Obstacle Clearance Manual]. Saint-Petersburg, Ministerstvo transporta RF Publ., 1998. 300 p. (In Russian).
9. Shchepilov Yu. N. *Postroenie aerodromnykh skhem* [Development of Terminal Procedures]. Saint-Petersburg, Sankt-Peterburgskii gosudarstvennyi universitet grazhdanskoi aviatsii Publ., 2013. 120 p. (In Russian).
10. *Performance-based Navigation (PNB) Manual. Vol. I. Concept and implemtation guidance*. Doc 9613. Montreal, ICAO, 2008. Available at: http://www.icao.int (accessed 12 May 2015).
11. *Performance-based Navigation (PNB) Manual. Vol. II. Implementing RNAV and RNP*. Doc 9613. Montreal, ICAO, 2008. Available at: http://www.icao.int (accessed 12 May 2015).
12. *Cocpit Display System Interfaces to User Systems. ARINCS specification 661*. Available at: http://www.aviation-ia.com (accessed 24 November 2014).
13. Paramonov P. P., Konovalov P. V., Zharinov I. O., Kirsanova Ju. A., Utkin S. B. Implementation of Data Structure in Indication Frame Formation for onboard Geospatial Data Computer Systems. *Nauchno-tekhnicheskii vestnik informatsionnykh tekhnologii, mekhaniki i optiki*, 2013, no. 2, pp. 165–167 (In Russian).
14. Zharinov I. O., Zharinov O. O. *Bortovye sredstva otobrazheniia informatsii na ploskikh zhidkokristallicheskikh paneliakh* [Onboard Display on Flat Liquid Crystal Panels]. Saint-Petersburg, Gosudarstvennyi universitet aerokosmicheskogo priborostroeniia Publ., 2005. 144 p. (In Russian).
15. Kostishin M. O., Zharinov I. O., Zharinov O. O., Nechaev V. A., Suslov V. D. Accuracy Evaluation of the Object Location Visualization for Geo-information and Display Systems of Manned Aircraft Navigation Complexes. *Nauchno-tekhnicheskii vestnik informatsionnykh tekhnologii, mekhaniki i optiki*, 2014, no. 1, pp. 87–93 (In Russian).
16. Paramonov P. P., Il'chenko Yu. A., Zharinov I. O., Tarasov P. Yu. Structural Analysis and Synthesis of Graphic Images on the Screens of Modern on-board Display on Flat Liquid Crystal Panels. *Aviakosmicheskoe priborostroenie*, 2004, no. 5, pp. 50–57 (In Russian).
17. Paramonov P. P., Kostishin M. O., Zharinov I. O., Nechaev V. A., Sudarchikov S. A. Formation and Display Principles for an Array of Geoinformation Data by means of onboard Display Screen. *Nauchno-tekhnicheskii vestnik informatsionnykh tekhnologii, mekhaniki i optiki*, 2013, no. 6, pp. 136–142 (In Russian).
18. *Navigation System Data Base. ARINC Specification 424*. Available at: http://www.aviation-ia.com (accessed 24 November 2014).
19. *Guidance Material for Design of Terminal Procedures for Area Navigation (DME/DME, B-GNSS, Baro-VNAV & RNP-RNAV)*. Eurocontrol, March 2003. 217 p.
20. *Sborniki aeronavigatsionnoi informatsii* [Aeronautical Information Publications]. Tsentr aeronavigatsionnoi informatsii grazhdanskoi aviatsii Rossii Publ. (In Russian). http://www.caiga.ru (accessed 21 April 2014).
21. Lipin A. V., Klyuchnikov Yu. I. *Primenenie zonal'noi navigatsii pri obsluzhivanii vozdushnogo dvizheniia. Zonal'naia navigatsiia* [Use of Area Navigation in Airtraffic Service. Area Navigation]. Saint-Petersburg, Universitet grazhdanskoi aviatsii Publ., 2011. 79 p. (In Russian).
22. Kharin E. G. *Kompleksnaia obrabotka informatsii navigatsionnykh sistem letatel'nykh apparatov. Opyt mnogoletnego prakticheskogo primeneniia* [Complex Information Processing of Aircraft Navigation Systems. The Experience of Many Years of Practical Application]. Moscow, Moskovskii aviatsionnyi institut Publ., 2002. 264 p. (In Russian).

УДК 629.784
doi:10.15217/issn1684-8853.2016.1.50

ПОДХОД К РАЗРАБОТКЕ СИСТЕМЫ РАСПОЗНАВАНИЯ И ПРЕДУПРЕЖДЕНИЯ НЕШТАТНЫХ СИТУАЦИЙ ТРАНСПОРТНОГО ГРУЗОВОГО КОРАБЛЯ «ПРОГРЕСС» НА ОСНОВЕ МУЛЬТИАГЕНТНЫХ ТЕХНОЛОГИЙ И ОНТОЛОГИЙ

Юрыгина Ю. С.[а]*, руководитель проекта*
Скорюпина Е. Г.[а]*, аналитик*
Лахин О. И.[а]*, руководитель направления*
Мишурова Н. В.[б]*, инженер*
[а]*ООО «Научно-производственная компания «Разумные решения», Самара, РФ*
[б]*ОАО «Ракетно-космическая корпорация «Энергия» им. С. П. Королева, Королев, РФ*

Постановка проблемы: *при управлении современными космическими аппаратами специалистам необходимо оперативно принимать решения на основе анализа данных о текущем состоянии корабля. Выполнение данной трудоемкой задачи в реальном времени во многом зависит от уровня подготовки специалиста. Существующие решения для управления сложным техническим объектом применимы только в определенной предметной области, что препятствует использованию подобных программных комплексов в смежных областях. Основная цель разработки системы — предоставить специалистам помощь в мониторинге, анализе и своевременном принятии решений по предупреждению и парированию нештатных ситуаций.* ***Методы:*** *основой предлагаемого подхода является использование мультиагентных технологий и семантических сетей (онтологий), концептуальной модели сложного технического объекта и формирование киберфизической модели.* ***Результаты:*** *разработан универсальный подход к созданию интеллектуальной системы обеспечения безопасности при управлении транспортным грузовым кораблем «Прогресс»: устройство системы сложного технического объекта описано в онтологии, каждой подсистеме ставится в соответствие программный агент. Агенты анализируют состояние системы на основе модели штатных и нештатных состояний узлов, сигнализируют о возможности возникновения нештатной ситуации. План ликвидации нештатной ситуации формируется на основе регламентированных правил, а также знаний о процессах, протекающих в системе. Добавление новых технических средств, нештатных ситуаций, изменение предметной области не требует перепрограммирования системы.* ***Практическая значимость:*** *система, разработанная на основе предлагаемого подхода, может работать автономно или совместно с имеющимися средствами мониторинга и управления космическим аппаратом. Подход не ограничен рамками предметной области и применим в других отраслях, требующих решения аналогичных задач. Ожидается получение экономического эффекта за счет снижения рисков отказа сложного технического объекта и повышения показателей надежности.*

Ключевые слова *— предметная область, мультиагентные технологии, онтология, киберфизическая модель, нештатные ситуации, транспортный грузовой корабль «Прогресс».*

Введение

Управление транспортным грузовым кораблем (ТГК) «Прогресс» и обеспечение его безопасности и надежности является сложной задачей, которую должны решать специалисты главной оперативной группы управления (ГОГУ). Анализ в реальном времени большого объема разнородной информации, отражающей текущее состояние изделия, и оперативное принятие решения без средств автоматизации являются трудоемкими и зависят от уровня подготовки специалиста.

Анализ программных продуктов на мировом рынке показывает, что комплексный универсальный подход к решению задач по управлению сложным техническим объектом на текущий момент не разработан. Частные решения применимы только для определенной предметной области, что не позволяет внедрять и использовать их в смежных отраслях. Это свидетельствует о необходимости разработки универсального метода решения поставленной задачи.

В данной статье рассматривается подход к разработке прототипа интеллектуальной системы поддержки процессов обеспечения безопасности и надежности ТГК «Прогресс». В основе предлагаемого подхода лежит использование мультиагентных технологий, предметных онтологий и киберфизической модели ТГК «Прогресс». Подход не только предлагает комплексное решение для создания рассматриваемой системы, но и открывает возможности его применения в аналогичных задачах других предметных областей.

Постановка проблемы

Сложность управления полетами современных космических аппаратов состоит в необходимости своевременно предупреждать, обнаруживать и ликвидировать нештатные ситуации (НШС) на борту. Под нештатной ситуацией понимается совокупность обстоятельств, обусловленных действием возмущающих факторов, представляющих угрозу безопасности экипажа, космического аппарата, выполнению задач полета.

Предпосылками для разработки системы стали проблемы и трудности, с которыми постоянно сталкиваются специалисты ГОГУ Центра управления полетами:

— сложность устройства объекта управления, физических и логических связей между компонентами;

— большой объем разнородной информации в кадрах телеметрии, отражающей качество работы систем/узлов/элементов конструкции космического аппарата;

— необходимость оперативного принятия решений в условиях ограниченного времени;

— определение гарантированного запаса времени, которым располагает экипаж и (или) специалист ГОГУ для парирования НШС;

— необходимость комплексного анализа знаний об объекте управления: физических, структурных, логических, функциональных, пространственных связей для принятия решений по выходу из НШС.

Основной целью разработки системы является сокращение времени на анализ состояния космического аппарата, мгновенное сигнализирование о возможности возникновения НШС, подготовка рекомендаций по дальнейшей работе космического аппарата с использованием новых методов и средств автоматизации.

Обзор существующих решений

Среди решений, представленных на рынке, можно выделить систему ProcessVue Alarm Management Suite [1] компании MAC Solutions (Великобритания), которая осуществляет мониторинг факторов, влияющих на безопасность производства, сигнализирует об отклонениях от штатного режима работы. Система включает модули, занимающиеся сбором и хранением данных об аварийных сигналах и событиях (ProcessVue SOE), отображением и анализом текущей ситуации (ProcessVue Analyser), обеспечением управления и отслеживания сигналов тревоги предприятия (ProcessVue Guardian). Это позволяет гибко реагировать на изменение производственных факторов, своевременно выявлять ситуации, в которых может возникнуть угроза жизни сотрудников и существованию самого предприятия. Однако для расширения программного продукта компании потребовалось привлекать дополнительные ресурсы, что указывает на сложность масштабируемости. Более того, внедрение данного компонента в распределенные системы управления на предприятии значительно увеличивает потребление электроэнергии, что может привести к отказу системы при вычислительных нагрузках в нештатном режиме работы.

Система DynAMo Alarm Suite [2] компании Honeywell (США) применяется на промышленных предприятиях, позволяя своевременно идентифицировать НШС и простои производства, информировать пользователей системы о возникших проблемах. Осуществляемый мониторинг системы управления позволяет определить причины возникновения НШС, осуществить прогноз последствий и выдать первоначальные рекомендации на основе инструкций, имеющихся в базе данных. Недостатком системы является подробное документирование событий, возникающих в системе, что вызывает необходимость хранить большое количество избыточных данных.

Продукт Alarm Management System [3] компании Dräger (Германия) позволяет проводить комплексный мониторинг системы подачи медицинских газов. Общее состояние объекта фиксируется датчиками и становится доступным для анализа в сети. В случае обнаружения аварийной ситуации на главный пункт в автономном режиме передаются сигналы об опасности. Система состоит из интеллектуальных модулей, которые обмениваются информацией по единой шине данных. Однако в случае возникновения непредвиденной ситуации шина данных является потенциально узким местом системы.

Системы поддержки процессов безопасности находят применение не только в организации производства, но также и в медицинских учреждениях для наблюдения за состоянием пациентов. Система Nellcor™ SatSeconds [4] компании Covidien (Ирландия) позволяет уменьшить количество ложных срабатываний сигнализации оборудования отделений реанимации, например мониторов пульсоксиметрии в случае обнаружения отклонений от нормы или сбоев в работе. Принцип работы состоит в расчете и анализе длительности показаний датчиков. Применяемый подход не является универсальным, поскольку способ расчета для каждого нового устройства необходимо пересматривать.

В Санкт-Петербургском научно-инженерном центре [5] разрабатывается интеллектуальная система, осуществляющая мониторинг и контроль состояния космодромов. В задачи системы входит определение состояния оборудования по показаниям датчиков, прогноз изменения значений параметров для выявления опасных тенденций, предпосылок развития НШС. В случае обнаружения отклонений в техническом состоянии объекта формируются и выдаются рекомендации для принятия решений. В настоящее время система осуществляет комплексный мониторинг ракетно-космической техники и наземной космической инфраструктуры космодрома Плесецк. Существенный недостаток данной системы — невозможность применять ее в других предметных областях.

Предлагаемый в данной статье подход отличается от рассмотренных ранее, во-первых, способом представления модели сложного технического объекта. Рассмотренные решения используют хранилища с описанием условий возникновения ситуаций, типов, характера их протекания. В предлагаемом подходе модель сложного технического объекта описывается при помощи семантической сети (онтологии). Это обеспечивает гибкость и легкость масштабирования системы, не требует перепрограммирования при внесении изменений: добавлении новых технических средств, новых типов НШС, изменении предметной области.

Следующее отличие состоит в методе построения рекомендаций при возникновении НШС. Многие системы работают на основе регламентированных инструкций, внесенных в базу данных. Некоторые системы сохраняют для дальнейшего использования алгоритм решения, информацию о событиях, произошедших на контролируемом объекте. Это ограничивает область решаемых задач кругом ситуаций, имеющихся в базе системы. Данное ограничение устраняет использование мультиагентных технологий, когда на основе имеющихся в системе знаний агент может получить новые знания и построить алгоритм выхода из непредвиденной ситуации. Взаимодействие между агентами децентрализовано, что свидетельствует об отсутствии узких мест в системе.

Предлагаемый подход основан на универсальных способах управления сложным техническим объектом и не зависит от предметной области. Это позволяет применять разработанные методы в других отраслях, где возникают аналогичные задачи.

Описание предлагаемой системы

Основная цель системы — предоставить специалистам помощь в мониторинге, анализе и своевременном принятии решений по предупреждению и парированию НШС [6]. Для этого система должна решать следующие задачи [7]:

— обработка телеметрической информации с оценкой текущего состояния космического аппарата;

— своевременное распознавание отклонений от штатного состояния;

— наблюдение тенденции к возникновению НШС и распознавание НШС;

— формирование рекомендаций по парированию НШС;

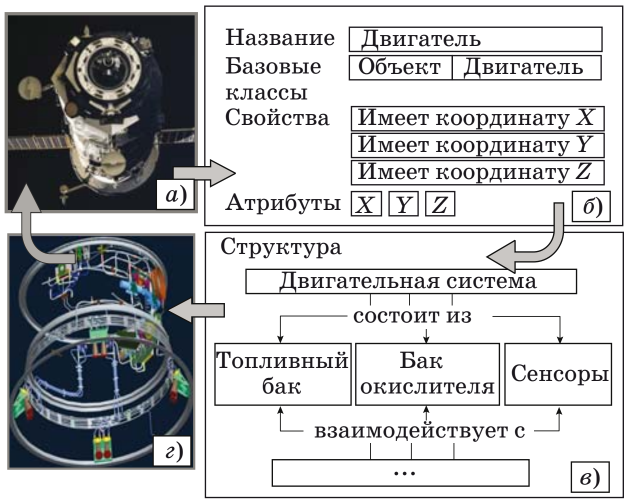
— прогнозирование последствий НШС.

Система может работать автономно либо совместно с текущими средствами мониторинга и управления изделием. При работе в автономном режиме все необходимые входные данные должны вводиться или загружаться пользователем вручную. В случае интегрированной работы с текущими средствами мониторинга и управления изделием часть данных может автоматически загружаться в режиме реального времени при помощи сервисов интеграции либо запрашиваться у пользователя.

В основе взаимодействия системы с пользователем лежит киберфизическая модель ТГК «Прогресс», позволяющая оперативно получать необходимую информацию об изделии и его составных системах и узлах. Киберфизическая модель является визуальным отражением концептуальной модели изделия. Система поддерживает модель изделия в 2D- и 3D-представлении с отображением функциональных связей между узлами. При выборе интересующего узла специалист может посмотреть текущие характеристики, расчетные параметры: показатель надежности работы узла, риск отказа, статистику НШС и др. Это позволит следить за общим состоянием системы, своевременно фиксировать и устранять неисправности.

Схема, отражающая базовые принципы системы, представлена на рис. 1.

Для конкретного космического аппарата строится онтология, содержащая необходимые классы, далее с помощью онтологии строится концептуальная модель космического аппарата. На основе концептуальной модели формируется киберфизическая модель, которая позволяет визуа-



■ *Рис. 1.* Базовые принципы организации системы: *а* — «Прогресс-М»; *б* — онтологическое описание класса «Двигатель_ДПО», используемого в информационной модели космического аппарата; *в* — физическая и иерархическая структуры двигателя; *г* — киберфизическая модель космического аппарата лизировать процессы, протекающие внутри космического аппарата, взаимодействие внутренних компонентов между собой и с внешней средой.

Предлагаемый подход

В основе предлагаемого подхода лежит использование мультиагентных технологий и онтологий [8]. Каждой подсистеме или узлу объекта ставится в соответствие программный агент. Такой агент представляет собой автономную программу, которая может реагировать на события, принимать решения и взаимодействовать с другими агентами или пользователями согласно встроенному внутреннему циклу управления. Для хранения знаний агенты используют онтологию, которая отражает концептуальную модель изделия. В концептуальную модель изделия входят следующие части:

— иерархическая структура, отражающая состав узлов изделия;

— физическая структура, отражающая физические и пространственные отношения между узлами;

— модель штатных состояний узлов;

— модель НШС;

— правила выполнения операций по ликвидации НШС;

— функциональная структура, отражающая функциональные зависимости между узлами;

— модель физических процессов.

Схематично архитектура агента представлена на рис. 2.

Агент обладает поведением, которое состоит из набора действий и событий. При возникновении события агент выполняет связанное с данным событием действие, работая со сценой. В данном контексте сцена — это отражение концептуальной модели в текущий момент времени.

1. Для *анализа и оценки текущего состояния узла* его агент обрабатывает любое изменение в сцене и выполняет следующие действия:

а) определяет текущие состояния узла — сопоставляет текущие характеристики узла с описанием его состояний в онтологии, в случае нахождения соответствия агент узла фиксирует его состояние;

б) рассчитывает степень обладания состоянием — рассчитывает значение функции, которая отражает степень нахождения данного узла в анализируемом состоянии в зависимости от характеристик и показателей узла в текущий момент времени;

в) рассчитывает показатель надежности на основании степени нахождения узла в штатном состоянии: чем больше степень нахождения узла в штатном состоянии, тем больше показатель надежности узла;

г) своевременно распознает и фиксирует отклонения от штатного состояния — анализирует, находится ли узел в штатном состоянии, в случае отклонения агент узла инициирует соответствующее событие.

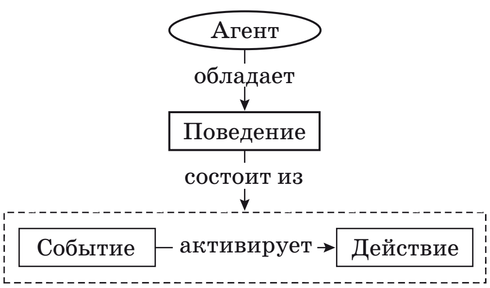
2. Для *наблюдения тенденции к возникновению НШС и распознавания нештатных ситуаций* агент узла фиксирует событие по отклонению от штатного состояния, далее анализирует, с какими НШС может быть связано зафиксированное отклонение. Для каждой связанной НШС агент узла рассчитывает тенденцию к ее возникновению с помощью функции, отражающей степень приближения к НШС в зависимости от характеристик и показателей узла в текущий момент времени.

3. Для выполнения задачи *формирования рекомендаций по устранению* НШС агент сначала фиксирует событие НШС узла, далее формирует план рекомендаций на основании указанных в онтологии правил ликвидации зафиксированной НШС.

4. После фиксации НШС в узле выполняется *прогнозирование последствий данной НШС* при помощи функциональной структуры и модели физических процессов [9].

Обработанные данные предоставляются специалисту для дальнейшего принятия решений и анализа текущей ситуации.

Основное преимущество предлагаемого подхода заключается в возможности гибко управлять моделью предметной области: расширять новыми знаниями концептуальную модель и добавлять новое поведение агентов узлов. Таким образом, специалисты предметной области смогут постоянно расширять знания о системе и адаптировать поведение агентов для решения различных задач в рамках обеспечения безопасности и повышения показателей надежности изделия.



■ ***Рис. 2.*** Архитектура агента

Пример использования предлагаемого подхода

Рассмотрим основные принципы использования предлагаемого подхода на примере возникновения НШС в комбинированной двигательной установке ТГК «Прогресс» (рис. 3). Для упрощения и наглядности некоторые элементы были опущены.

Кратко опишем сценарий развития ситуации. На начальном участке полета выполняется операция по наддуву магистралей топливом. Во время операции азот, находящийся под высоким давлением в баллоне, поступает в бак с топливом. Топливо вытесняется в двигательную установку для дальнейшего использования. По окончании операции наддува двигательная установка находится в состоянии готовности. При выполнении операции может возникнуть НШС — «неоткрытие клапана». Рассмотрим применение подхода для решения рассматриваемых выше задач.

Построение концептуальной модели начинается с описания иерархической и физической структур (рис. 4, *а* и *б*).

Для узлов описывается модель штатных и нештатных состояний (рис. 5, *а* и *б*).

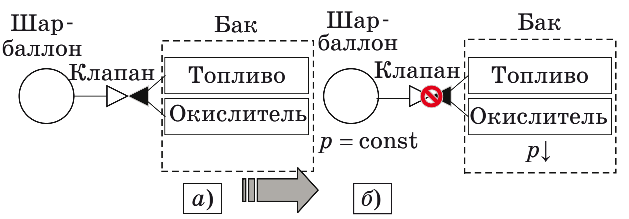
Рассмотрим описание штатного состояния для двигательной установки на начальном участке полета. Оно включает следующие составляющие:

— участок полета, определяющий временной интервал;

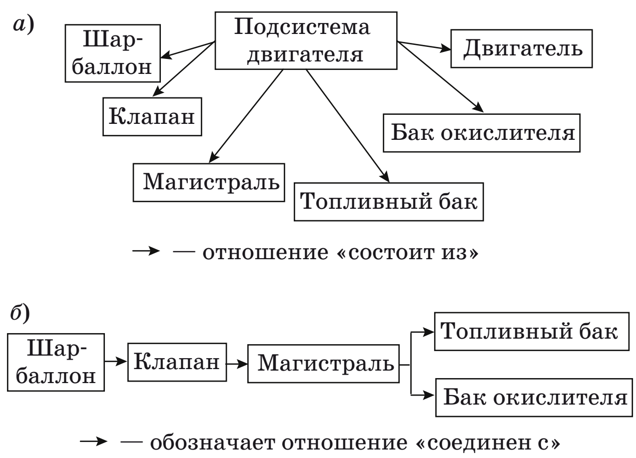
— состояние клапана, определяющее некоторый контекст ситуации;

— состояния бака наддува и бака горючего, отражающие ограничения для характеристик объектов, в рамках которых они могут изменяться.

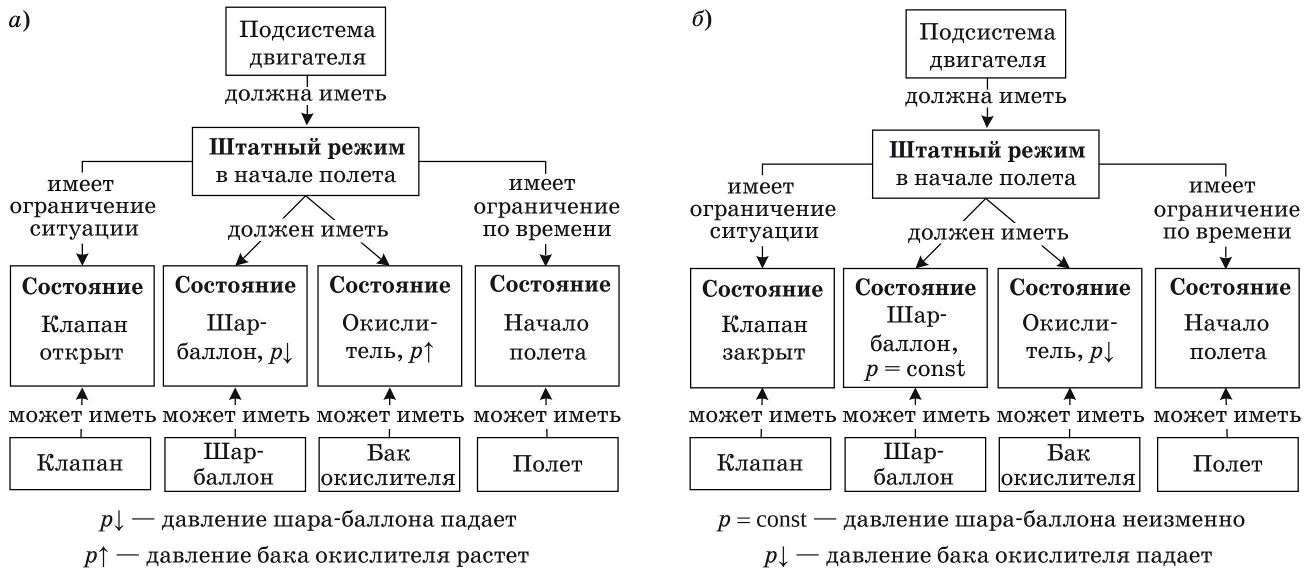
Описание нештатного состояния имеет аналогичную структуру. Отличие заключается в способе обработки описанных состояний агентом узла. Агент узла «Подсистема двигателя», анализируя рассмотренное штатное состояние, сначала проверяет состояние, определяющее участок полета. Если участок полета совпадает (т. е. «Полет» имеет состояние *Начало полета*), агент уточняет состояние задействованных узлов, если контекст ситуации совпадает (т. е. «Клапан» имеет состояние *Клапан открыт*), агент проверяет состояния, отражающие штатную работу: давление



■ ***Рис. 3.*** Фрагмент двигательной установки: *а* — до НШС; *б* — после НШС



■ ***Рис. 4.*** Иерархическая (*а*) и физическая (*б*) структура двигательной установки



■ ***Рис. 5.*** Модель штатного (*а*) и нештатного (*б*) режима работы двигательной установки в баке наддува должно падать, давление в баке окислителя должно расти. Если хотя бы одно из состояний, отражающих штатную работу, не совпадет, агент зафиксирует отклонение от данного штатного состояния.

При анализе нештатного состояния агент узла «Подсистема двигателя» аналогично проверяет состояния, определяющие:

— участок полета;

— контекст ситуации;

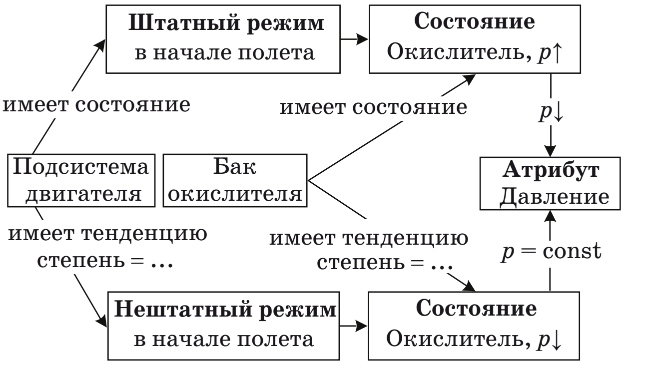
— состояния, отражающие нештатную работу: давление в баке наддува должно быть постоянным, давление в баке окислителя должно падать.

Если оба состояния, отражающие нештатную работу, совпадут, агент зафиксирует, что узел «Подсистема двигателя» имеет нештатное состояние.

Для описания состояния, отражающего нештатную работу, используются ограничения на значение атрибута. Например, для состояния *Окислитель*, $p\downarrow$ указывается, что значение атрибута *Давление* должно уменьшаться с определенной скоростью. Кроме ограничения на значения атрибута, можно указать функцию, отражающую тенденцию к возникновению нештатного состояния. Например, чем ближе значение скорости уменьшения давления к указанному в ограничении, тем больше вероятность возникновения нештатного состояния. Значение показателя тенденции к возникновению нештатного состояния может храниться в атрибуте отношения между объектом и состоянием (рис. 6).

Значение показателя тенденции к возникновению нештатного состояния может заключаться в диапазоне от 0 до 1: 1 означает, что нештатное состояние зафиксировано, 0 — отсутствие предпосылок.

Если нештатное состояние зафиксировано, агент узла «Подсистема двигателя» формирует рекомендации по его устранению. Для этого используется описанный в концептуальной модели *Алгоритм восстановления* (рис. 7). При необходимости агент узла может запрашивать дополнительные данные от пользователя, например, успешно ли выполнена операция *Попытаться открыть клапан*, и в зависимости от полученной информации формировать дальнейшие рекомендации. В случае сложных, разветвленных алгоритмов это может существенно помочь специалистам действовать строго по регламентированным инструкциям.

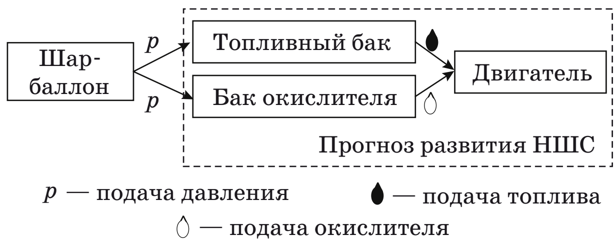


■ ***Рис. 6.*** Тенденция к нештатному состоянию

Для прогнозирования последствий НШС могут использоваться функциональная структура и модель физических процессов, являющиеся частью концептуальной модели. Анализируя физические и функциональные связи, агенты узлов определяют, какие связи будут нарушены, и делают предположения о последствиях в случае выхода из строя одного из узлов. На рис. 8 представлена функциональная структура для рассматриваемого примера, требуемая для штатной работы.

Отношения «подача давления» между узлами «Шар-баллон» и «Топливный бак», «Бак окислителя» устанавливаются автоматически при выполнении процесса «подача давления». Данный процесс должен обеспечивать штатную работу узла «Подсистема двигателя» на начальном участке полета. В процессе участвуют следующие объекты: «Шар-баллон», «Клапан», «Топливный бак»,

■ ***Рис. 7.*** Порядок восстановления



■ ***Рис. 8.*** Функциональная структура двигательной установки

«Бак окислителя». Для работы процесса имеется следующее ограничение: узел «Клапан» должен иметь состояние *Клапан открыт*. Таким образом, при отказе узла «Клапан» процесс не будет функционировать и требуемые отношения «подача давления» от узла «Шар-баллон» не будут установлены, что может вывести из штатного состояния узлы «Топливный бак» и «Бак окислителя». В свою очередь, данные узлы имеют функциональные связи «подача топлива», «подача окислителя» с узлом «Двигатель», поэтому данный узел также находится под угрозой отклонения от штатной работы.

Заключение

Применение рассмотренного подхода позволит создать прототип интеллектуальной системы поддержки процессов обеспечения безопасности и надежности ТГК «Прогресс». Использование данной системы обеспечит своевременное реагирование на возникновение тенденций и предпосылок возникновения нештатных ситуаций, поможет оперативно принимать решения по ликвидации отказов и стабилизации работы изделия.

Ожидается получение экономического эффекта за счет снижения рисков отказа изделия и повышения показателей надежности.

Систему, разработанную на основе предложенного подхода, можно использовать не только для изделия ТГК «Прогресс», но и для других изделий аэрокосмической отрасли: самолетов, пилотируемых космических кораблей, спутников [10]. Ввиду отсутствия в предлагаемом подходе непосредственной зависимости от предметной области, описанные методы можно применять в других отраслях, где используются требующие решения аналогичных задач сложные технические изделия: буровые установки и насосные станции; газорегуляторные пункты и установки; теплоэлектроцентрали; ядерные реакторы; морские суда и подводные лодки и т. д.

Литература

1. ProcessVue Suite Overview. http://www.processvue.com/en/products/processvue-suite-overview (дата обращения: 13.05.2015).
2. HoneyWell Alarm Management. https://www.honeywellprocess.com/en-US/online_campaigns/alarm-management/Pages/alarm-management.html (дата обращения: 13.05.2015).
3. Dräger Alarm Management System. http://www.draeger.com/sites/en_seeur/Pages/Applications/Alarm-Management-System.aspx?navID=1877 (дата обращения: 14.05.2015).
4. Covidien Nellcor™ SatSeconds. http://www.covidien.com/rms/clinical-solutions/alarm-management (дата обращения: 14.05.2015).
5. Научно-инженерный центр электротехнического университета. Интеллектуальные системы мониторинга и контроля состояния технически сложных объектов. http://www.nicetu.ru/reshenija-i-produkty/intellektualnye-sistemy-monitoringa-i-kontrolja-sostojanija-tekhnicheski-slozhnykh-obektov/ (дата обращения: 20.05.2015).
6. **Матюшин М. М.** и др. Автоматизированная система поддержки принятия решений в аварийных ситуациях/ М. М. Матюшин, С. И. Потоцкий, П. О. Скобелев, В. И. Потапов, О. И. Лахин // Программные продукты и системы. 2013. № 3. С. 62–69.
7. **Матюшин М. М., Вакурина Т. Г., Мишурова Н. В.** и др. Интеллектуальная система предупреждения нештатных ситуаций в бортовых системах ТГК «Прогресс» // XII Всерос. совещание по проблемам управления ВСПУ-2014, Москва, 16–19 июня 2014 г.: тр. М.: Ин-т проблем управления им. В. А. Трапезникова РАН, 2014. С. 9014–9021.
8. **Матюшин М. М.** и др. Методы и средства построения онтологий для визуализации связанных информационных объектов произвольной природы в сложных информационно-аналитических системах / М. М. Матюшин, Т. Г. Вакурина, В. В. Котеля, П. О. Скобелев, И. О. Лахин, С. С. Кожевников, Е. В. Симонова, А. И. Носкова// Информационно-управляющие системы. 2014. № 2(69). С. 9–17.
9. **Коршиков Д. Н., Лахин О. И., Носкова А. И., Юрыгина Ю. С.** Методы представления знаний для решения задач моделирования // Материалы V Междунар. науч.-техн. конф. «Открытые семантические технологии проектирования интеллектуальных систем» Open Semantic Technologies for Intelligent Systems (OSTIS-2015), Минск, 19–21 февраля 2015 г. Минск: БГУИР, 2015. С. 425–428.
10. **Лахин О. И., Полников А. С., Симонова Е. В., Скобелев П. О.** Теория сложности и проблема управления жизненным циклом изделий аэрокосмической промышленности // Информационно-управляющие системы. 2015. № 1. С. 4–12. doi:10.15217/issn1684-8853.2015.1.4

UDC 629.784
doi:10.15217/issn1684-8853.2016.1.50

Developing a System of Progress Vehicle Emergency Recognition and Prevention Based on Multi-Agent Technologies and Ontologie

Iurygina I. S.[a], Project Manager, yurygina@smartsolutions-123.ru
Skoryupina E. G.[a], Analyst, skoryupina@smartsolutions-123.ru
Lakhin O. I.[a], Project Manager, lakhin@yandex.ru
Mishurova N. V.[b], Engineer, post@rsce.ru
[a]Software Engineering Company «Smart Solutions», 17, Moscovskoe St., 443013, Samara, Russian Federation
[b]PLC S. P. Korolev Rocket & Space Corporation «Energia», 4A, Lenin St., 141070, Korolev, Moscow, Russian Federation

Purpose: Modern space flight control requires quick decision making based on data analysis, reflecting the current condition of the vehicle. Execution of this laborious task in real time largely depends on the proficiency of the specialist. The existing solutions for managing a complex technical object are limited by their knowledge domain, which prevents them from being used in related fields. The primary objective of the system development is to assist Flight Control Team specialists in monitoring, analyzing and decision making for quick prevention and elimination of emergences. **Methods:** The suggested approach is based on using multi-agent technologies and semantic networks (ontologies), the conceptual model of a complex technical object, and a cyber-physical model. **Results:** A universal approach has been developed to design an intelligent system supporting the safety and reliability of Progress Vehicle. The units of the complex technical object are described in the ontology. Each subsystem of the object has a correspondent software agent. The agents analyze the system's state using a model of normal states and emergencies, signalling when an emergency can occur. Recommendations for an emergency elimination are formed using the emergency elimination rules and the model of physical processes described in the ontology. Adding a new piece of hardware, new types of emergencies or changing the knowledge domain do not require the system to be reprogrammed. **Practical relevance:** The system developed on the basis of the proposed approach can operate autonomously or combined with the existing facilities of monitoring and managing a space vehicle. The approach is not strictly limited by a certain knowledge domain, therefore the described methods can be used in other industries where complex technical objects face similar problems. It is expected that economic benefits will be obtained by reducing failure risks and improving the object reliability.

Keywords — Knowledge Domain, Multi-Agent Technologies, Ontology, Cyber-Physical Model, Emergency, Progress Vehicle.

References

1. *ProcessVue Suite Overview*. Available at: http://www.processvue.com/en/products/processvue-suite-overview (accessed 11 May 2015).
2. *HoneyWell Alarm Management*. Available at: https://www.honeywellprocess.com/en-US/online_campaigns/alarm-management/Pages/alarm-management.html (accessed 10 May 2015).
3. *Dräger Alarm Management System*. Available at: http://www.draeger.com/sites/en_seeur/Pages/Applications/Alarm-Management-System.aspx?navID=1877 (accessed 11 May 2015).
4. *Covidien Nellcor™ SatSeconds*. Available at: http://www.covidien.com/rms/clinical-solutions/alarm-management (accessed 14 May 2015).
5. *Nauchno-inzhenernyi tsentr elektrotekhnicheskogo universiteta. Intellektual'nye sistemy monitoringa i kontrolia sostoianiia tekhnicheski slozhnykh ob"ektov* [St. Petersburg Electro-Technical University Research and Engineering Center. Intelligent Systems for Monitoring and Control of State of Complex Technical Objects]. Available at: http://www.nicetu.ru/reshenija-i-produkty/intellektualnye-sistemy-monitoringa-i-kontrolja-sostojanija-tekhnicheski-slozhnykh-obektov/ (accessed 20 May 2015).
6. Matjushin M. M., Potockij S. I., Skobelev P. O., Potapov V. I., Lahin O. I. Automated System of Decision Making Support in Emergencies. *Programmnye produkty i sistemy*, 2013, no. 3, pp. 62–69 (In Russian).
7. Matjushin M. M., Vakurina T. G., Mishurova N. V., Lahin O. I., Jurygina Ju. S., Noskova A. I. Intelligent System for Prevention of Abnormal Situations in Progress Vehicle On-Board Systems. *Trudy XII Vserossiiskogo soveshchaniia po problemam upravleniia VSPU-2014* [Proc. of the XII All-Russian Meeting on Control Sciences]. Moscow, Institut problem upravleniia im. V. A. Trapeznikova RAN Publ., 2014, pp. 9014–9021 (In Russian).
8. Matjushin M. M., Vakurina T. G., Kotelja V. V., Skobelev P. O., Lahin O. I., Kozhevnikov S. S., Simonova E. V., Noskova A. I. Methods and Software for Creation of Ontologies for Visualizing Connected Information Objects of Random Nature in Complex Information-Analytical Systems. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2014, no. 2(69), pp. 9–17 (In Russian).
9. Korshikov D. N., Lahin O. I., Noskova A. I., Jurygina Ju. S. Methods of Knowledge Presentation for Solving Modeling Tasks. *Materialy V Mezhdunarodnoi nauchno-tekhnicheskoi konferentsii "Otkrytye semanticheskie tekhnologii proektirovaniia intellektual'nykh system"* [Proc. of the V Intern. Scientific and Technical Conf. "Open Semantic Technologies for Intelligent Systems" (OSTIS-2015)], Minsk, Belorusskii gosudarstvennyi universitet informatiki i radioelektroniki Publ., 2015, pp. 425–428 (In Russian).
10. Lahin O. I., Polnikov A. S., Simonova E. V., Skobelev P. O. Complexity Theory and Challenges of Aerospace Products Lifecycle Management. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2015, no. 1, pp. 4–12. doi:10.15217/issn1684-8853.2015.1.4

УДК 614.841.332
doi:10.15217/issn1684-8853.2016.1.58

АНАЛИЗ МАТЕМАТИЧЕСКИХ МЕТОДОВ ОЦЕНКИ БЕЗОПАСНОСТИ ЭЛЕКТРИФИЦИРОВАННЫХ ПОМЕЩЕНИЙ

С. В. Соленый[a]**,** *канд. техн. наук, доцент*
О. Я. Солёная[a]**,** *канд. техн. наук, доцент*
[a]*Санкт-Петербургский государственный университет аэрокосмического приборостроения, Санкт-Петербург, РФ*

Постановка проблемы: *обширность методов оценки пожарной безопасности систем низковольтного электроснабжения, а также подходов к определению факторов опасности приводит к усложнению выбора метода и процесса такой оценки.* ***Цель:*** *унификация методов математического моделирования и оценки безопасной эксплуатации систем низковольтного электроснабжения путем классификации и анализа существующих подходов к определению уровня их пожарной безопасности.* ***Результаты:*** *анализ существующих методов оценки безопасности эксплуатации систем низковольтного электроснабжения позволил установить, что не весь имеющийся на сегодняшний день в этой области математический аппарат напрямую можно использовать для проведения оценки пожарной безопасности систем низковольтного электроснабжения. Вероятностные методы являются весьма объемными из-за необходимости проводить анализ с высокой детализацией рассматриваемых аварийных процессов. Эвристические методы не позволяют адекватно оценить результаты, так как вызывает сомнения достоверность используемой в них системы баллов. Логико-вероятностные методы не учитывают периодическое диагностирование защитно-коммутационных аппаратов в процессе их эксплуатации, что приводит к ошибкам в оценке на несколько порядков. С применением методов, построенных на использовании теории однородных марковских процессов, разработана адекватная с точки зрения достоверности математическая модель оценки безопасности электрифицированных помещений, не требующая большого инженерного ресурса при проведении оценки.* ***Практическая значимость:*** *на основе однородных случайных марковских процессов с дискретным числом состояний и непрерывным временем получена математическая модель оценки пожарной безопасности систем низковольтного электроснабжения, которая позволяет учитывать все возможные случаи нахождения элементов системы в опасном и безопасном состояниях и может применяться для анализа подобных систем любой конфигурации и сложности с точки зрения оценки их пожарной и эксплуатационной безопасности.*

Ключевые слова *— система низковольтного электроснабжения, электропроводка, математическое моделирование, контактное соединение, электрическое искрение, переходное сопротивление, изоляция, пожарная безопасность.*

Введение

В рамках Федеральной целевой программы «Пожарная безопасность в Российской Федерации на период до 2017 г.» № 1481 от 30.12.2012 г. предусмотрено проведение исследований в области разработки математических моделей и методов по определению и прогнозированию вероятности возникновения пожаров, эффективных способов их предупреждения и ликвидации последствий. Также, согласно Федеральному закону Российской Федерации № 117-ФЗ «О внесении изменений в Федеральный закон № 123-ФЗ “Технический регламент о требованиях пожарной безопасности” от 10.07.2012 г.» [1], предусмотрена разработка новых технических решений и организационных мероприятий, использование действующих и создание новых нормативных документов, применение которых позволит обеспечивать пожарную безопасность промышленных и бытовых электрифицированных объектов.

Анализ статистики пожаров в Российской Федерации по причине повреждений электрической проводки за период с 2005 по 2015 г. [2] свидетельствует о том, что если не разрабатывать новые и не совершенствовать известные методы и способы обеспечения безопасной эксплуатации систем низковольтного электроснабжения (СНЭ), то в них с доверительной вероятностью 0,95 ежегодно будет происходить до 40 000 пожаров по электротехническим причинам [3].

Следовательно, работа, направленная на создание новых и совершенствование известных математических моделей и методов оценки и прогнозирования безопасной эксплуатации СНЭ, является актуальной научной задачей, решение которой позволит не допускать случаев появления возгораний горючего материала (изоляции и др.) в СНЭ и обеспечит их пожарную безопасность на уровне действующих нормативных документов [4].

Методы оценки пожарной безопасности СНЭ

В связи с принятием законов [1, 5] особую актуальность приобретают разработка и внедрение в отечественную практику научно обоснованных рекомендаций по обеспечению пожарной безопасности СНЭ и использование методов, с помощью которых предоставляется возможность проводить оценку существующего уровня пожарной опасности объектов и сравнивать его с законодательно установленным ГОСТ значением $1 \cdot 10^{-6}$ [4].

В последнее время происходит переход от жесткого нормирования требований пожарной безопасности при проектировании СНЭ к гибкому или объектно-ориентированному. Этот подход заключается в том, что устанавливаются цели, которым должна соответствовать система безопасной эксплуатации низковольтного электроснабжения, но при этом не регламентируются проектные решения для их достижения. Таким образом, сводятся к минимуму ограничения в устройстве СНЭ, а приветствуется использование новых подходов к обеспечению ее безопасной эксплуатации, чем достигается более высокая экономическая эффективность проектных решений [6].

На сегодняшний день отсутствует единый всемирно используемый метод оценки пожарной безопасности СНЭ, который был бы принят в качестве обязательного в нормативной документации, регламентирующей данные вопросы [7]. В странах с высоким уровнем развития промышленности применяются методы оценки пожарной безопасности СНЭ на основе логических деревьев. Для объектов, представляющих повышенную опасность, применяются специально разработанные методы оценки пожарной безопасности СНЭ, утвержденные на законодательном уровне [8].

Пожарная безопасность СНЭ должна оцениваться по общим законодательно прописанным принципам, а методы ее оценки могут выбираться самостоятельно. В качестве расчетных методов допускается использовать как качественный анализ, так и количественный, включая индексные методы и полный вероятностный анализ. Выбор метода должен производиться в соответствии с целями проведения оценки пожарной безопасности СНЭ и данными о ее расположении в помещении, а также с учетом графика выполнения анализа [9].

Исходя из документов [1, 4, 5], пожарная опасность объекта защиты — это такое состояние объекта, которое характеризуется возможностью возникновения и развития пожара, а также воздействием на людей и имущество его опасных факторов. Кроме того, существует пожарный риск — это мера возможной реализации пожарной опасности объекта защиты, ее последствий для людей и материальных ценностей. Далее рассмотрим подходы к анализу и определению величин этих двух понятий, которые отличаются друг от друга тем, что при нахождении пожарного риска учитывается вероятность воспламенения (загорания) изоляции и конструктивных материалов СНЭ. Риск может быть измерен числом возгораний, происшедших на объекте в единицу времени, или вероятностью его появления в течение времени [10].

Вероятностный подход к оценке пожарной безопасности СНЭ

Количественная оценка безопасности СНЭ с использованием вероятностного подхода является фундаментом значительного числа современных методов анализа. В общем виде риск возникновения пожароопасной ситуации в СНЭ от проявления какого-либо аварийного режима (короткого замыкания, перегрузки, утечки тока на землю, электрического искрения через переходное сопротивление) можно определить следующим образом [11]:

$$R = P \cdot U, \quad (1)$$

где P — вероятность возникновения пожароопасной ситуации в СНЭ от проявления какого-либо аварийного режима в ней; U — ожидаемый ущерб от этой пожароопасной ситуации.

Таким образом, при возникновении ущерба от проявления N-го количества аварийных режимов работы СНЭ совокупный риск реализации пожароопасной ситуации определяется суммированием вероятностей всех возможных аварийных режимов работы СНЭ:

$$R = \sum_{i=1}^{N} P_i \cdot U_i. \quad (2)$$

В итоге количественная оценка риска возникновения пожароопасной ситуации в СНЭ сводится к выявлению возможных аварийных режимов ее работы и определению последствий каждого режима.

В мировой практике оценки безопасности объектов наиболее распространенными являются следующие программные продукты [12]: CRISP2 (Великобритания), FRAMEworks (США), Probabilistic Fire Simulator (Финляндия), Fire Risk Evaluator (Швеция), CESARE-RISK (Австралия). Широко известна программа FiRECAM (Канада), которая обладает подробной детализацией факторов, определяющих пожарную опасность объекта [13]. FiRECAM позволяет оценивать два основных вида риска пожароопасной ситуации: риск для жизни людей, находящихся в здании, и ожидаемый ущерб от пожара. Математический аппарат, заложенный в FiRECAM, строится на всестороннем использовании выражения (1).

Подводя итог, можно отметить, что задача количественной оценки пожарной безопасности СНЭ существующими вероятностными методами является трудоемкой в практическом исполнении. При всей привлекательности вероятностных методов оценки пожарной безопасности применительно к СНЭ, они являются весьма объемными из-за необходимости проводить анализ с высокой детализацией рассматриваемых аварийных процессов.

Эвристический подход к оценке пожарной безопасности СНЭ

Если обстоятельства (недостаточно статистических данных либо они не удовлетворяют точности математической модели) не позволяют применить строгую количественную оценку пожарной безопасности СНЭ на основе вероятностных методов, то существенное преимущество получают методы, реализующие эвристический подход. Данные методы строятся на экспертном оценивании и нахождении субъективных вероятностей тех или иных аварийных режимов работы СНЭ, а также определении набора факторов, формирующих ее пожарную опасность и степень защиты. После определения факторов происходит их оценка в баллах с постепенным подведением общей оценки — индекса пожарной опасности СНЭ [14].

Существует большое разнообразие индексных методов оценки пожарной безопасности. Это связано с тем, что они разрабатываются как методы оценки пожарной опасности для конкретных типов объектов (производств). Из-за этого характеристики, которые описывают пожарную опасность, и методы обработки выставленных баллов очень отличаются друг от друга [15]. Рассмотрим наиболее широко применяемые индексные методы — Дау Кемикал и Гретенера.

Метод Дау Кемикал используется для оценки пожарной опасности химических производств [16]. Он базируется на выделении производственных участков по роду производства и формировании для каждого из них фактора опасности материала, который характеризует выделяемую энергию при его возгорании в пределе 1–40 баллов. После чего оценивается фактор опасности для данного участка производства, зависящий от показателей первого (условия, которые способны усилить либо ослабить последствия пожара) и второго класса (условия, предопределяющие вероятность реализации пожара либо взрыва). Затем показатели по каждому классу суммируются, а их результаты перемножаются, и полученный фактор опасности находится в пределе 1–8 баллов. Далее путем произведения фактора материала на фактор опасности определяется индекс пожаровзрывоопасности объекта (FEI — fire and explosion index). При этом анализируются факторы, позволяющие снизить уровень опасности. В итоге дается заключение о возможном масштабе последствий пожара либо взрыва (размере зоны разрушений, степени материального ущерба, вероятности гибели либо травматизма персонала).

Метод Гретенера используется для определения пожарной безопасности объектов путем оценки вероятности возникновения возгорания и оценки материального ущерба по выражению [17]

$$R = A \cdot B = \frac{A \cdot P}{N \cdot S \cdot F}, \tag{3}$$

где A — вероятность возникновения возгорания; B — материальные потери; P — потенциальная пожарная опасность; N — показатель, который учитывает мероприятия по обеспечению пожарной безопасности, предусмотренные нормативными документами; S — специализированные мероприятия, направленные на повышение пожарной безопасности; F — степень огнеупорности помещения.

Таким образом, данный метод является мерой некоего баланса между вероятностью возникновения возгорания, потенциальными материальными убытками и действиями, направленными на защиту объекта.

Итак, рассмотренные индексные методы отображают эвристический подход к оценке пожарной безопасности, а также позволяют проводить это с минимальными затратами на математический аппарат, при этом адекватность их применения напрямую зависит от достоверности проставленных баллов по критериям интерпретации пожарной опасности.

Логико-вероятностные методы оценки пожарной безопасности СНЭ

Согласно ГОСТ [18], пожарная опасность СНЭ — это возможность возникновения или развития в ней или от нее пожара. Таким образом, ее пожарную опасность в целом можно характеризовать так: вероятность возникновения пожара, связанная с возгоранием узлов СНЭ, а также веществ и материалов, соприкасающихся с ней или находящихся в зоне ее радиационного излучения или в зоне воздействия электрической дуги и выпадающих из нее раскаленных частиц и газов; вероятность появления и воздействия на человека, строительные конструкции и оборудование дыма, токсичных продуктов и других опасных факторов, образующихся при горении ее деталей и узлов, находящихся в объекте [19]. Следовательно, оценка пожарной опасности СНЭ должна выполняться методами определения вероятности возникновения пожара, которые должны включать в себя вероятность возникновения возгорания от электротехнического изделия или электропроводки, а также вероятность перехода этого возгорания в пожар.

В том случае если возгорание горючего материала возможно при появлении n независимых случайных событий, то вероятность появления возгорания в течение времени t можно найти по формуле [20]

$$q(t) = \prod_{i=1}^{n} q_i(t). \tag{4}$$

Если при формировании возгораний в СНЭ учитываются только три события ($n = 3$), тогда из формулы (4) получим

$$q(t) = q_1(t)\cdot q_2(t)\cdot q_3(t), \qquad (5)$$

где $q(t)$ — вероятность возгорания горючего материала, расположенного вблизи прокладки электропроводки в течение времени t; $q_1(t)$ — вероятность повреждения электропроводки (короткое замыкание, электрическое искрение через переходное сопротивление контактного соединения, перегрузка, утечка тока на землю и др.); $q_2(t)$ — вероятность нахождения горючего материала в течение времени t вблизи прокладки электропроводки; $q_3(t)$ — вероятность отказа в срабатывании защитно-коммутационного аппарата.

В том случае если интервалы времени между повреждениями в СНЭ, отказами в срабатывании защиты и наличием горючего материала вблизи прокладки электропроводки не противоречат экспоненциальной функции распределения с параметрами λ_1, ω_1 и ω_2 соответственно, тогда вероятности, входящие в формулу (5), можно найти следующим образом:

$$q_1(t) = 1 - e^{-\lambda_1 t};\ \ q_2(t) = 1 - e^{-\omega_1 t};$$

$$q_3(t) = 1 - e^{-\omega_2 t}, \qquad (6)$$

где λ_1 — частота повреждения СНЭ; ω_1 — частота появления горючего материала вблизи СНЭ; ω_2 — частота отказа защитно-коммутационных аппаратов.

Подставив значения (6) в формулу (5), получим

$$q(t) = \left(1 - e^{-\lambda_1 t}\right)\left(1 - e^{-\omega_1 t}\right)\left(1 - e^{-\omega_2 t}\right) =$$
$$= 1 - e^{-\lambda_1 t} - e^{-\omega_1 t} - e^{-\omega_2 t} + e^{-(\lambda_1+\omega_1)t} +$$
$$+ e^{-(\omega_1+\omega_2)t} + e^{-(\lambda_1+\omega_2)t} - e^{-(\lambda_1+\omega_1+\omega_2)t}. \qquad (7)$$

Вероятность того, что в течение времени t не произойдет возгорание горючего материала в СНЭ, находится по формуле

$$R(t) = 1 - q(t) = e^{-\lambda_1 t} + e^{-\omega_1 t} + e^{-\omega_2 t} - e^{-(\lambda_1+\omega_1)t} -$$
$$- e^{-(\omega_1+\omega_2)t} - e^{-(\lambda_1+\omega_2)t} + e^{-(\lambda_1+\omega_1+\omega_2)t}. \qquad (8)$$

Среднее время до появления возгорания в СНЭ определим следующим образом:

$$T = \int_0^\infty R(t)\mathrm{d}t = \int_0^\infty \left(\begin{array}{c} e^{-\lambda_1 t} + e^{-\omega_1 t} + e^{-\omega_2 t} - \\ - e^{-(\lambda_1+\omega_1)t} - e^{-(\omega_1+\omega_2)t} - \\ - e^{-(\lambda_1+\omega_2)t} + e^{-(\lambda_1+\omega_1+\omega_2)t} \end{array} \right)\mathrm{d}t =$$
$$= \frac{1}{\lambda_1} + \frac{1}{\omega_1} + \frac{1}{\omega_2} - \frac{1}{\lambda_1+\omega_1} -$$
$$- \frac{1}{\omega_1+\omega_2} - \frac{1}{\lambda_1+\omega_2} + \frac{1}{\lambda_1+\omega_1+\omega_2}. \qquad (9)$$

Вероятность возгорания горючего материала в СНЭ можно определить по приближенной формуле

$$Q(t) = 1 - e^{-\frac{t}{T}}. \qquad (10)$$

Интенсивность возгорания горючего материала в течение времени t находим следующим образом [21]:

$$\lambda(t) = -\frac{\mathrm{d}R(t)}{\mathrm{d}t \cdot R(t)}. \qquad (11)$$

Подставив формулу (8) в (11), получим

$$\lambda(t) = \frac{\lambda_1 e^{-\lambda_1 t} + \omega_1 e^{-\omega_1 t} + \omega_2 e^{-\omega_2 t} -}{e^{-\lambda_1 t} + e^{-\omega_1 t} + e^{-\omega_2 t} -}$$
$$\frac{-(\lambda_1+\omega_1)e^{-(\lambda_1+\omega_1)t} - (\omega_1+\omega_2)e^{-(\omega_1+\omega_2)t} -}{-e^{-(\lambda_1+\omega_1)t} - e^{-(\omega_1+\omega_2)t} -}$$
$$\frac{-(\lambda_1+\omega_2)e^{-(\lambda_1+\omega_2)t} + (\lambda_1+\omega_1+\omega_2)e^{-(\lambda_1+\omega_1+\omega_2)t}}{-e^{-(\lambda_1+\omega_2)t} + e^{-(\lambda_1+\omega_1+\omega_2)t}}. \qquad (12)$$

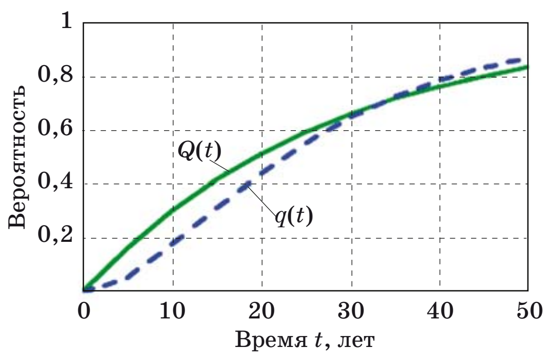
Рассмотрим случай, когда параметры процесса равны: $\lambda_1 = 0{,}049$ год$^{-1}$; $\omega_1 = 0{,}061$ год$^{-1}$; $\omega_2 = 0{,}62$ год$^{-1}$. Используя формулы (7)–(9), при $t = 1$ год получим:

— вероятность появления возгорания в течение года $q(1) = 1{,}31\cdot 10^{-3}$;

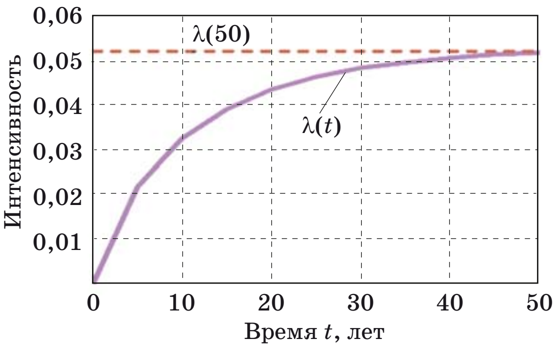
— среднее время до появления возгорания в СНЭ $T = 27$ лет;

— вероятность безаварийной работы $R(1) = 3{,}54\cdot 10^{-2}$.

По точной (7) и приближенной (10) формулам получены функции вероятностей возгорания горючего материала в СНЭ в течение времени t (рис. 1). Из графика видно, в какой степени вероятность возгорания горючего материала, полученная по точной формуле, отличается от значения вероятности возгорания, полученного по приближенной формуле. Для времени $t = 34$ года вероятности появления возгорания совпадают, а затем с течением времени результаты расходятся.



■ *Рис. 1.* Функции вероятностей появления возгорания по причине короткого замыкания в СНЭ



■ ***Рис. 2.*** Функция интенсивности возгорания горючего материала в СНЭ по причине короткого замыкания

Интенсивность возгорания горючего материала в СНЭ $\lambda(t)$ определяется по формуле (12). При $t = 1$ год находим $\lambda(1) = 3{,}49 \cdot 10^{-3}$ год$^{-1}$.

Используя формулу (12), можно построить функцию интенсивности возгорания горючего материала от короткого замыкания в СНЭ в течение времени t (рис. 2). Из графика видно, что при $t = 50$ лет интенсивность отказов системы «источник зажигания — защита — среда» уже не зависит от времени: $\lambda(50) = 0{,}052$ год$^{-1}$.

Используя полученные формулы (7)–(9) и (12), можно определить основные характеристики пожарной опасности СНЭ. Однако применение логико-вероятностных методов оценки пожарной опасности дает лишь качественный анализ пожарной безопасности, поскольку данные методы не имеют возможности учета влияния на пожарную безопасность длительности действия опасного источника зажигания, времени нахождения защитно-коммутационных аппаратов в необнаруженном отказавшем состоянии и длительности нахождения горючего материала вблизи возможного источника зажигания. Также логико-вероятностные методы не позволяют оценить влияние диагностических проверок работоспособности защитно-коммутационных аппаратов и не учитывают влияние профилактических проверок по обнаружению и устранению горючего материала вблизи возможного источника зажигания, что завышает истинный уровень пожарной безопасности на несколько порядков.

Следовательно, возникает необходимость в применении методов оценки пожарной безопасности, которые позволят учитывать вышеперечисленные факторы.

Методы, учитывающие частоту появления опасного источника и длительность его существования

В работе [22] на основе теории регулярных однородных марковских случайных процессов с дискретным числом состояний и непрерывным временем была разработана и предложена обобщенная математическая модель «источник зажигания — среда», которая позволяет описать формирование аварий и катастроф на участке угольных шахт при эксплуатации электрооборудования. В этой работе показано, что такие разные по своему характеру аварии, как взрыв, пожар, случаи поражения человека электрическим током, происходят по одной логической схеме и имеют общую математическую модель.

Число аварий, происшедших в единицу времени, определяется следующим образом:

$$H = \frac{d_1 + d_2}{\bar{d}_1 \cdot \bar{d}_2}, \tag{13}$$

где $\bar{d}_1$ — среднее время между появлением опасного состояния среды; d_1 — средняя длительность нахождения среды в опасном состоянии; $\bar{d}_2$ — средний интервал времени между появлением источника зажигания; d_2 — средняя длительность существования опасного источника.

Формула (13) справедлива для случая, когда выполняются условия

$$\bar{d}_1 \gg d_1; \;\; \bar{d}_2 \gg d_2. \tag{14}$$

Итак, использование теории марковских случайных процессов для оценки пожарной безопасности СНЭ позволит учитывать все возможные случаи нахождения элементов системы в опасном и безопасном состояниях, а также задавать нормы надежности на средства защиты, частоту появления и длительность существования горючего материала вблизи возможного повреждения СНЭ, что даст возможность не допускать случаев появления возгораний горючего материала и тем самым обеспечит безопасность эксплуатации электрооборудования, и, как следствие, снизит угрозу для жизни людей [23].

Заключение

Установлено, что применение логико-вероятностных методов для оценки пожарной опасности дает лишь качественный анализ пожарной безопасности, поскольку данные методы не учитывают влияния на пожарную безопасность длительности действия опасного источника зажигания, времени нахождения защитно-коммутационных аппаратов в необнаруженном отказавшем состоянии и наличия горючего материала вблизи возможного источника зажигания. Таким образом, логико-вероятностные методы не позволяют оценить влияние проведения диагностических проверок защитно-коммутационных аппаратов и профилактических проверок наличия горючего материала вблизи возможного источника зажигания на пожарную безопасность электрифицированного объекта. Для создания унифицированной методики оценки пожарной безопасности СНЭ целесообразно применять марковские процессы с дискретным числом состояний и непрерывным временем, поскольку математические модели, построенные с их использованием, позволяют учитывать кроме частоты появления событий, связанных с возникновением возгораний, еще и длительность их существования.

Литература

1. Федеральный закон Российской Федерации от 22 июля 2008 г. № 123-ФЗ «Технический регламент о требованиях пожарной безопасности» с изменениями согласно Федеральному закону Российской Федерации № 117-ФЗ от 10.07.2012 г.
2. Официальный сайт Министерства Российской Федерации по делам гражданской обороны, чрезвычайным ситуациям и ликвидации последствий стихийных бедствий http://www.mchs.gov.ru/folder/3788548 (дата обращения: 22.10.2015).
3. **Ковалёв А. П., Солёная О. Я.** О проблемах оценки безопасности электротехнических объектов // Электричество. 2014. № 11. С. 36–45.
4. ГОСТ 12.1.004-91. Пожарная безопасность. Общие требования. — М.: Изд-во стандартов, 1992. — 77 с.
5. Федеральный закон Российской Федерации от 21.12.1994 г. № 69-ФЗ «О пожарной безопасности» с изменениями согласно Федеральному закону Российской Федерации № 234-ФЗ от 13.07.2015 г.
6. **Молчанов В. П., Болодьян И. А., Дешевых Ю. И.** и др. Концепция объектно-ориентированного нормирования промышленных предприятий по пожарной безопасности // Пожарная безопасность. 2001. № 4. С. 94–106.
7. **Якуш С. Е., Эсманский Р. К.** Анализ пожарных рисков. Ч. I: Подходы и методы // Проблемы анализа риска. 2009. Т. 6. № 3. С. 8–27.
8. NFPA 551. Guide for the Evaluation of Fire Risk Assessments. — Quincy, MA: National Fire Protection Association, 2007. — 26 p.
9. PAS 79. Fire Risk Assessment — Guidance and a Recommended Methodology. — British Standards Institution, 2007. — 120 p.
10. ГОСТ Р 51901.1-2002 (МЭК 60300-3-9:1995). Менеджмент риска. Анализ риска технологических систем. — М.: Изд-во стандартов, 2002. — 28 с.
11. **Kumamoto H.** Satisfying Safety Goals by Probabilistic Risk Assessment. — London: Springer-Verlag, 2007. — 253 p.
12. **Hall J., Watts J.** Fire Risk Analysis // NFPA. Ch. 8. — Fire Protection Handbook, 2008. — P. 135–143.
13. **Yung D., Hadjisophocleous G., Proulx G.** A Description of the Probabilistic and Deterministic Modeling Used in FiRECAM™//Intern. Journal on Engineering Performance-Based Fire Codes. 1999. N 1. P. 18–26.
14. **Акимов В. А., Лесных В. В., Радаев Н. Н.** Основы анализа и управления риском в природной и техногенной сферах. — М.: Деловой экспресс, 2004. — 195 с.
15. **Rasbash D.,** et al. Evaluation of Fire Safety/ D. Rasbash, G. Ramachandran, B. Kandola, J. Watts, M. Law. — N. Y.: J. Wiley & Sons, 2004. — 496 p.
16. Dow's Fire and Explosion Index Hazard Classification Guide. 7th ed. — N. Y.: Dow Chemical Company; American Institute of Chemical Engineers, 2010. — 83 p.
17. **Gretener M.** Evaluation of Fire Hazard and Determining Protective Measures. — Zurich: Association of Cantonal Institutions for Fire Insurance and Fire Prevention Service for Industry and Trade, 1973. — 87 p.
18. ГОСТ 12.1.033-81. Пожарная безопасность. Термины и определения. — М.: Изд-во стандартов, 1981. — 9 с.
19. **Смелков Г. И.** Пожарная безопасность электропроводок. — М.: Кабель, 2009. — 328 с.
20. **Ковалёв П. Ф.** Разработка и исследование средств и мер обеспечения безопасности применения электрической энергии в угольных шахтах: доклад о содержании опубликованных работ и изобретений, представленных на соискание ученой степени доктора техн. наук в соответствии с решением президиума ВАК. — М.: МГИ, 1974. — 86 с.
21. **Вентцель Е. С.** Теория вероятностей: учебник для студ. вузов. 10-е изд., стер. — М.: Академия, 2005. — 576 с.
22. **Ковалёв П. Ф., Коптиков В. П., Ковалёв А. П.** О критериях оценки эффективности мер и средств обеспечения безопасности применения электрооборудования в шахтах // Безопасность труда в промышленности. 1972. № 8. С. 34–36.
23. **Ковалёв А. П., Солёная О. Я.** Оценка пожарной опасности сети 0,4/0,22 кВ в промышленных электрифицированных помещениях // Промышленная энергетика. 2013. № 11. С. 43–46.

UDC 614.841.332
doi:10.15217/issn1684-8853.2016.1.58

Analysis of Mathematical Methods of Electrified Area Safety Assessment

Solyonyj S. V.[a], PhD, Tech., Associate Professor, ssv555ssv@yandex.ru
Solenaya O. Ya.[a], PhD, Tech., Associate Professor, osolenaya@list.ru
[a]Saint-Petersburg State University of Aerospace Instrumentation, 67, B. Morskaia St., 190000, Saint-Petersburg, Russian Federation

Introduction: Low-voltage power supply systems can use so many fire safety assessment methods and so many approaches to identifying hazards, that it can be difficult to choose a method and carry out the assessment. Purpose: The goal is to unify the mathematical modeling methods and safe operation assessments for low-voltage power supply systems by classifying and analyzing the existing approaches to their fire safety assessment. **Results:** The analysis of the existing methodologies revealed that not all mathematical tools available today in this area can be directly used to assess the fire safety of low-voltage power supply systems. Probabilistic methods are very voluminous because of the need to analyze highly detailed emergency processes. Heuristic methods cannot adequately assess the results, as they use a point system whose reliability raises doubts. Logical-and-probabilistic methods do not take into account the regular checks of the protective switching devices during their operation, which leads to evaluation errors of several orders of magnitude. But the methods based on the theory of homogeneous Markov processes gave rise to a mathematical model of assessment which is adequate in terms of reliability and does not require too much engineering resources. **Practical relevance:** On the base of homogeneous random Markov processes with discrete number of states and continuous time, a mathematical model was obtained for low-voltage power supply system fire safety assessment, which allows you to consider all possible situations when the system elements stay in safe or unsafe states. This model can be used to analyze systems of any configuration and complexity in terms of their fire safety and operation security.

Keywords — Low-Voltage Power Supply System, Electric Wiring, Mathematical Modeling, Connectors, Electrical Arcing, Contact Resistance, Insulation, Fire Safety.

References

1. Federal Law of the Russian Federation from 22.07.2008 № 123-FZ "Technical Regulation on the Requirements of Fire Safety" with Change comply with the Russian Federation Federal Law № 117-FZ from 10.07.2012.
2. *Ofitsial'nyi sait Ministerstva Rossiiskoi Federatsii po delam grazhdanskoi oborony, chrezvychainym situatsiiam i likvidatsii posledstvii stikhiinykh bedstvii* [The Official Website of the Ministry of the Russian Federation for Civil Defense, Emergencies and Elimination of Consequences of Natural Disasters]. Available at: http://www.mchs.gov.ru/folder/3788548 (accessed 22 October 2015).
3. Kovalev A. P., Solenaya O. Ya. About Problems of Safety Evaluation of Electrical Facilities. *Elektrichestvo*, 2014, no. 11, pp. 36–45 (In Russian).
4. State Standard 12.1.004-91. Fire Safety. General Requirements. Moscow, Standartov Publ., 1992. 77 p. (In Russian).
5. Federal Law of the Russian Federation from 21.12.1994. № 69-FZ "About Fire Safety" with Change Comply with the Russian Federation Federal Law № 234-FZ from 13.07.2015. (In Russian).
6. Molchanov V. P., Bolodyan I. A., Deshevih Yu. I., et al. The Concept of Object-Oriented a Valuation of Industrial Enterprises on Fire Safety. *Pozharnaia bezopasnost'*, 2001, no. 4, pp. 94–106.
7. Yakush S. E., Esmansky R. K. Analysis of Fire Risks. Part I. Approaches and Methods. *Problemy analiza riska*, 2009, vol. 6, no. 3, pp. 8–27 (In Russian).
8. NFPA 551. Guide for the Evaluation of Fire Risk Assessments. Quincy, MA, National Fire Protection Association, 2007. 26 p.
9. PAS 79. Fire Risk Assessment — Guidance and a Recommended Methodology. British Standards Institution, 2007. 120 p.
10. State Standard 51901.1-2002 (IEC 60300-3-9:1995). Risk Management. Risk Analysis of Technological Systems. Moscow, Standartov Publ., 2002. 28 p. (In Russian).
11. Kumamoto H. *Satisfying safety goals by probabilistic risk assessment*. London, Springer-Verlag, 2007. 253 p.
12. Hall J., Watts J. Fire Risk Analysis. In: *NFPA*. Ch. 8. Fire Protection Handbook, 2008, pp. 135–143.
13. Yung D., Hadjisophocleous G., Proulx G. A Description of the Probabilistic and Deterministic Modeling used in FiRECAM™. *International Journal on Engineering Performance-Based Fire Codes*, 1999, no. 1, pp. 18–26.
14. Akimov V. A., Lesnih V. V., Radaev N. N. *Osnovy analiza i upravleniia riskom v prirodnoi i tekhnogennoi sferakh* [Basics of Analysis and Risk Management of Natural and Man-Made Areas]. Moscow, Delovoi ekspress Publ., 2004. 195 p. (In Russian).
15. Rasbash D., Ramachandran G., Kandola B., Watts J., Law M. *Evaluation of Fire Safety*. New York, J. Wiley & Sons, 2004. 496 p.
16. Dow's Fire and Explosion Index Hazard Classification Guide. 7th ed. New York, Dow Chemical Company, American Institute of Chemical Engineers, 2010. 83 p.
17. Gretener M. *Evaluation of Fire Hazard and Determining Protective Measures*. Zurich, Association of Cantonal Institutions for Fire Insurance and Fire Prevention Service for Industry and Trade, 1973. 87 p.
18. State Standard 12.1.033-81. Fire Safety. Terms and Definitions. Moscow, Standartov Publ., 1981. 9 p. (In Russian).
19. Smelkov G. I. *Pozharnaia bezopasnost' elektroprovodok* [Fire Safety Wirings]. Moscow, Kabel' Publ., 2009. 328 p. (In Russian).
20. Kovalev P. F. *Razrabotka i issledovanie sredstv i mer obespecheniia bezopasnosti primeneniia elektricheskoi energii v ugol'nykh shakhtakh* [Development and Research of Means and Measures to Ensure Safety of Electricity in Coal Mines]. Report on the contents of published works and inventions submitted for the degree of the doctor tehn. science in accordance with the decision of the Presidium of the Higher Attestation Commission. Moscow, MGI Publ., 1974. 86 c. (In Russian).
21. Ventzel E. S. *Teoriia veroiatnostei* [Probability Theory]. Moscow, Akademiia Publ., 10 ed., 2005. 576 p. (In Russian).
22. Kovalev P. F., Koptikov V. P., Kovalev A. P. About Criteria of Evaluating the Effectiveness of Measures and Means to Ensure the Safety of Electrical Equipment in Mines. *Bezopasnost' truda v promyshlennosti*, 1972, no. 8, pp. 34–36 (In Russian).
23. Kovalev A. P., Solenaya O. Ya. Estimation of Fire Hazard the Network 0,4 / 0,22 kV in Industrial Electrified Areas. *Promyshlennaia energetika*, 2013, no. 11, pp. 43–46 (In Russian).

УДК 004.942
doi:10.15217/issn1684-8853.2016.1.65

МОДЕЛЬ УПРАВЛЕНИЯ ДВУМЯ ПАРАЛЛЕЛЬНЫМИ FIFO-ОЧЕРЕДЯМИ, ДВИГАЮЩИМИСЯ ДРУГ ЗА ДРУГОМ В ОБЩЕЙ ПАМЯТИ

Е. А. Барковский[а], *соискатель*
А. В. Соколов[а, б], *доктор физ.-мат. наук, профессор*
[а]*Петрозаводский государственный университет, Петрозаводск, РФ*
[б]*Институт прикладных математических исследований Карельского научного центра Российской академии наук, Петрозаводск, РФ*

Постановка проблемы: *при разработке многих аппаратных и программных приложений применяют структуру данных «FIFO-очередь». В различных сетевых устройствах и встроенных операционных системах применяется несколько FIFO-очередей, расположенных в общем пространстве памяти. Также существуют архитектуры многоядерных процессоров, где каждому ядру выделено две FIFO-очереди. Программные и аппаратные решения, применяемые для реализации описанных задач, должны увеличивать надежность (снижение доли потерянных при переполнении памяти элементов очередей) работы таких устройств.* ***Цель:*** *построение и анализ математической модели процесса работы с двумя FIFO-очередями, двигающимися друг за другом по кругу, в общей памяти, где на нечетном шаге происходят операции включения элементов в одну из очередей, а на четном шаге — исключения (возможно как последовательное, так и параллельное выполнение операций).* ***Результаты:*** *построены математическая и имитационная модели этого процесса для двух очередей и проведены численные эксперименты, основывающиеся на теоретических данных. Математическая модель представлена в виде случайного блуждания по целочисленной пирамиде с отражающими экранами. Предложен алгоритм нумерации состояний, установлен вид матрицы переходных состояний полученной регулярной цепи Маркова (доказаны соответствующие утверждения), разработаны алгоритм и программа вычисления средней доли потерянных при переполнении элементов очередей. Особенностью данного исследования является специфическое выполнение операций над очередями: включение и исключение элементов происходит в зависимости от шага (были сделаны поправки для сохранения качеств однородности и регулярности цепи) и создана возможность выполнять операции параллельно.* ***Практическая значимость:*** *предложенные модель, алгоритм и программный комплекс для анализа метода движения очередей «друг за другом» могут применяться при проектировании сетевых устройств, например маршрутизаторов, микросхем, реализующих работу с несколькими FIFO-очередями, и других программных и аппаратных устройств, где потери элементов являются допустимой, но нежелательной ситуацией. С помощью разработанной модели можно, при заданных вероятностных характеристиках очередей, выбрать лучший метод представления очередей, например, из двух методов — классического последовательного циклического метода или метода «Друг за другом».*

Ключевые слова *— структуры данных, FIFO-очереди, случайные блуждания, регулярные марковские цепи.*

Введение

В большом количестве приложений используются несколько FIFO-очередей, расположенных в общем пространстве памяти. При разработке сетевых устройств и встроенных операционных систем требуются эффективные алгоритмы работы с этими структурами данных — в приложениях управляющих потоками пакетов Internet, требования на время обработки пакетов маршрутизатором могут быть очень жесткие. Так, в Cisco IOS механизм страничной виртуальной памяти не используется, вместо него имеются пулы оперативной памяти, где и происходит вся работа. Для представления FIFO-очередей применяют различные программные или аппаратные решения [1–3].

Нужно также отметить, что существуют архитектуры многоядерных процессоров без кэш-памяти. Например, в архитектуре AsAP-II каждое ядро имеет два FIFO-буфера, а в архитектуре SEAforth — два стека [4]. Очереди и стеки реализованы циклически и раздельно, при переполнении происходит потеря элементов. В наших задачах для хранения нескольких структур данных используется общая память. В ряде случаев это позволяет снизить потери элементов при переполнении.

Модель последовательного, связанного и страничного способов представления нескольких FIFO-очередей в памяти одного уровня описаны в работах [5–8]. Здесь предполагается, что на каждом шаге дискретного времени происходят некоторые операции со структурами данных (с заданными вероятностями). Так как время выполнения операций — не случайная величина, а константа, фиксированным является и шаг времени. Первоначально такие модели в виде случайного блуждания в треугольнике [9–14] были построены для решения задачи анализа процесса работы с двумя стеками, растущими навстречу друг другу, поставленной в работе [3].

Немного другой способ выполнения операций с несколькими FIFO-очередями в общей памяти предложен в работе [1]. В этом способе на нечетном шаге допускаются только операции включения элементов в одну из n очередей, а на четном шаге — исключения из очередей (в первом и втором случае операции равновероятны). Там была поставлена задача разработки математической модели для описания работы системы таких очередей. При этом не рассматривался конкретный способ представления очередей в памяти, т. е. предполагалось, что очереди могут быть неограниченной длины, что на практике невыполнимо.

Авторами [15] предложена математическая модель и решена задача оптимального разбиения общей памяти для двух FIFO-очередей в случае их последовательного циклического представления. Операции с очередями (с заданными вероятностями) выполнялись по вышеупомянутому принципу как последовательно, так и параллельно. В качестве критерия оптимальности рассмотрена минимальная средняя доля потерянных при переполнении элементов на бесконечном времени работы.

Этот критерий оптимальности стоит рассматривать, если переполнение очереди является стандартной ситуацией. Когда размер очереди превышает размер предоставленной ей памяти, все поступающие в нее элементы отбрасываются до тех пор, пока не появится свободный участок памяти (этот участок появится только после исключения элемента из очереди). Сетевые маршрутизаторы [2] работают по этой схеме. Такие потери элементов приводят к нежелательному результату. Например, в некоторых сетевых протоколах отброшенные пакеты будут посылаться снова, что приводит к замедлению работы системы — ее эффективность падает. Поэтому число таких ситуаций необходимо свести к минимуму.

В данной работе мы предлагаем математическую и имитационную модели процесса работы с двумя параллельными очередями, когда они двигаются по кругу друг за другом. Здесь общая память заранее не делится между очередями. Этот метод работы с FIFO-очередями предложен в работе [16]. Математическая модель строится в виде случайного блуждания по целочисленной пирамиде. Предварительные результаты исследования докладывались на конференции [17].

Математическая модель

Пусть в памяти размера m единиц мы работаем с двумя циклическими параллельными FIFO-очередями, которые двигаются друг за другом по кругу. Возобновление движения пустой очереди начинается с середины промежутка между очередями.

Операции, производимые с очередями, выполняются по следующей схеме: на нечетном шаге совершается операция включения в одну из очередей (включение происходит в конец очереди), на четном шаге — операция исключения из какой-либо очереди (исключение происходит из начала очереди), причем известны некоторые вероятностные характеристики операций, производимых с очередями. Можно сказать, что у нас имеется регулярный поток событий, которые происходят через фиксированный шаг дискретного времени. Следует заметить, что этот поток не является простейшим (в терминологии теории массового обслуживания), так как у него есть последействие.

Пусть p_1 и p_2 — вероятности включения элемента в первую и вторую очереди, соответственно, p_{12} — вероятность одновременного включения в обе очереди; q_1 и q_2 — вероятности исключения элемента из первой и второй очередей, соответственно, q_{12} — вероятность одновременного исключения из обеих очередей.

Поскольку построенная на основе такой постановки задачи марковская цепь не будет являться регулярной и однородной, два последовательных шага объединяем в один, а также вводим следующие вероятности операций, не изменяющих длины очередей (например, чтение): r_o — на нечетном шаге и r_e — на четном шаге, при этом $r_o \neq 0$, $r_e \neq 0$. Соответственно: $p_1 + p_2 + r_o = 1$, $q_1 + q_2 + r_e = 1$.

Тогда состояние на каждом шаге определяется наступлением одной из следующих комбинаций событий, где сумма всех вероятностей равна 1:

— включение в первую, исключение из второй очереди с вероятностью p_1q_2;

— включение во вторую, исключение из первой очереди с вероятностью p_2q_1;

— включение в первую очередь с вероятностью $p_1r_e + p_{12}q_2$;

— включение во вторую очередь с вероятностью $p_2r_e + p_{12}q_1$;

— включение параллельно в обе очереди с вероятностью $p_{12}r_e$;

— исключение из первой очереди с вероятностью $q_1r_o + p_2q_{12}$;

— исключение из второй очереди с вероятностью $q_2r_o + p_1q_{12}$;

— исключение параллельно из обеих очередей с вероятностью $q_{12}r_o$;

— выполнение над очередями сохраняющих их состояние противоположных операций с вероятностью $r_or_e + p_1q_1 + p_2q_2 + p_{12}q_{12}$.

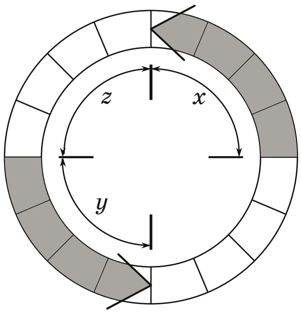
Предполагается, что в очередях хранятся данные фиксированного размера. При исключении информации из пустой очереди не происходит завершения работы.

Целью исследования является определение средней доли потерянных элементов для сравнения со средней долей потерянных элементов при последовательном циклическом способе организации параллельных очередей в случае оптимального разбиения общей памяти [15]. По закону больших чисел для регулярных цепей Маркова

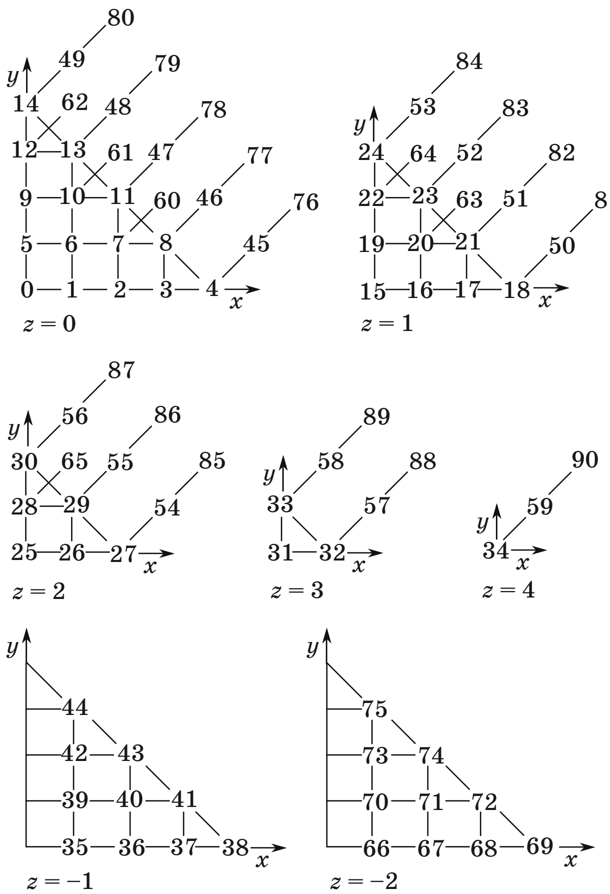
[18] это эквивалентно нахождению решения, вычисляющего значение вероятности переполнения памяти на бесконечном промежутке времени.

Обозначим через x и y текущие длины очередей, через z — расстояние между концом первой очереди и началом второй. Схема движения очередей показана на рис. 1.

В качестве математической модели рассмотрим случайное блуждание (которое конкретизировано с учетом специфики объекта исследования) по целочисленной трехмерной пирамиде



■ *Рис. 1.* Схема движения очередей



■ *Рис. 2.* Область блуждания, нумерация состояний при $m = 4$

с вершиной (0;0;0) и основанием $x + y + z = m$ (рис. 2).

Для вычисления средней доли потерянных при переполнении элементов очередей введены искусственные отражающие экраны:

1) $z = -1$ и $x + y' + z = m + 1$ ($y' = y + 1$). Блуждание переходит на экран $z = -1$, если происходит включение в первую очередь в то время, когда $z = 0$, и на экран $x + y' + z = m + 1$, если происходит включение во вторую очередь в то время, когда $x + y + z = m$. В данном случае элемент будет потерян, а очереди будут находиться на экранах до тех пор, пока поступают новые элементы. В случае исключения элемента очереди вернутся в состояния, принадлежащие пирамиде;

2) в случае когда происходит включение во вторую очередь и одновременное исключение из первой, в то время когда $x + y + z = m$, происходит потеря элемента и увеличение расстояния между концом второй очереди и началом первой. Для такой ситуации также необходимо ввести отражающий экран — $x' = x - 1$;

3) если происходит операция одновременного включения и исключения элемента из первой очереди при $z = 0$ либо операция включения и исключения из второй очереди при $x + y + z = m$, происходит потеря элемента, и очереди возвращаются в состояния, принадлежащие пирамиде. Для учета потери элемента вводим экраны $z = -2$ и $x + y' + z = m + 2$ ($y' = y + 2$) соответственно.

Определим схему переходов между состояниями. Величина *distance* означает расстояние между концом первой и началом второй очереди. Если оживает первая очередь, то $\text{distance} = \left[\dfrac{m-y-1}{2}\right]$, и $\text{distance} = \left[\dfrac{m-x-1}{2}\right]$, если оживает вторая очередь. Пусть (x, y, z) — текущее состояние процесса, тогда блуждание по пирамиде можно описать следующим образом:

$$(x, y, z) \xrightarrow[r_o r_e]{} (x, y, z);$$

$$(x, y, z) \xrightarrow[p_1 r_e]{} (x', y', z') =$$

$$= \begin{cases} (x+1, y, z-1) & x, y, z > 0 \\ (x+1, 0, 0) & y = 0, x < m \\ (1, y, \text{distance}-1) & x = 0, y < m \\ (x, y, -1) & z = 0, y \neq 0 \text{ либо } x = m \end{cases};$$

$$(x, y, z) \xrightarrow[p_2 r_e]{} (x', y', z') =$$

$$= \begin{cases} (x, y+1, z) & x, y, z > 0 \\ (x, 1, \text{distance}) & y = 0, x < m \\ (0, y, 0) & x = 0, y < m \\ (x, y+1, z) & x + y + z = m \end{cases};$$

$$(x,y,z)\xrightarrow[r_o q_1]{}(x',y',z')=$$
$$=\begin{cases}(0,y,0) & x=0 \text{ либо } x=1\\ (x-1,0,0) & y=0\\ (x-1,y,z) & \text{иначе}\end{cases};$$

$$(x,y,z)\xrightarrow[r_o q_2]{}(x',y',z')=$$
$$=\begin{cases}(x,0,0) & y=0 \text{ либо } y=1\\ (0,y-1,0) & x=0\\ (x,y-1,z+1) & \text{иначе}\end{cases};$$

$$(x,y,z)\xrightarrow[p_1 q_2]{}(x',y',z')=$$
$$=\begin{cases}(x+1,y-1,z) & x,y,z>0\\ (x+1,0,0) & x<m, y=0 \text{ либо } y=1\\ (1,y-1,\text{distance}) & x=0, y<m\\ (x,y-1,-1) & z=0, y\neq 0 \text{ либо } x=m\end{cases};$$

$$(x,y,z)\xrightarrow[p_2 q_1]{}(x',y',z')=$$
$$=\begin{cases}(x-1,y+1,z) & x,y,z>0\\ (x-1,1,\text{distance}) & y=0, x<m\\ (0,y+1,0) & y<m, x=0 \text{ либо } x=1\\ (x',y,z) & x+y+z=m\end{cases};$$

$$(x,y,z)\xrightarrow[p_1 q_1]{}(x',y',z')=$$
$$=\begin{cases}(x,y,z-1) & x,y,z>0\\ (0,y,0) & x=0\\ (x,0,0) & y=0\\ (x,y,-2) & z=0, y\neq 0 \text{ либо } x=m\end{cases};$$

$$(x,y,z)\xrightarrow[p_2 q_2]{}(x',y',z')=$$
$$=\begin{cases}(x,y,z+1) & x,y,z>0\\ (x,0,0) & y=0\\ (0,y,0) & x=0\\ (x,y+2,z+1) & x+y+z=m\end{cases};$$

$$(x,y,z)\xrightarrow[p_{12} r_e]{}(x',y',z')=$$
$$=\begin{cases}(x+1,y+1,z-1) & x,y,z>0\\ (x+1,1,\text{distance}-1) & y=0, x<m\\ (1,y+1,\text{distance}-1) & x=0, y<m\\ (x,y+1,-1) & z=0, y\neq 0 \text{ либо } x=m\\ (x+1,y+1,z-1) & x+y+z=m\end{cases};$$

$$(x,y,z)\xrightarrow[r_o q_{12}]{}(x',y',z')=$$
$$=\begin{cases}(0,y-1,0) & x=0 \text{ либо } x=1\\ (x-1,0,0) & y=0 \text{ либо } y=1\\ (x-1,y-1,z+1) & \text{иначе}\end{cases};$$

$$(x,y,z)\xrightarrow[p_1 q_{12}]{}(x',y',z')=$$
$$=\begin{cases}(x,y-1,z) & x,y,z>0\\ (0,y-1,0) & x=0\\ (x-1,0,0) & y=0\\ (x,y-1,-2) & z=0, y\neq 0 \text{ либо } x=m\end{cases};$$

$$(x,y,z)\xrightarrow[p_2 q_{12}]{}(x',y',z')=$$
$$=\begin{cases}(x-1,y,z+1) & x,y,z>0\\ (x-1,0,0) & y=0\\ (0,y-1,0) & x=0\\ (x-1,y+2,z+1) & x+y+z=m\end{cases};$$

$$(x,y,z)\xrightarrow[p_{12} q_1]{}(x',y',z')=$$
$$=\begin{cases}(x,y+1,z-1) & x,y,z>0\\ (x,1,\text{distance}-1) & y=0, x<m\\ (0,y+1,0) & x=0, y<m\\ (x,y+1,-2) & z=0, y\neq 0 \text{ либо } x=m\\ (x',y,z-1) & x+y+z=m\end{cases};$$

$$(x\ y,z)\xrightarrow[p_{12} q_2]{}(x',y',z')=$$
$$=\begin{cases}(x+1,y,z) & x,y,z>0\\ (x+1,0,0) & y=0, x<m\\ (1,y,\text{distance}) & x=0, y<m\\ (x,y,-1) & z=0, y\neq 0 \text{ либо } x=m\\ (x+1,y+2,z) & x+y+z=m\end{cases};$$

$$(x,y,z)\xrightarrow[p_{12} q_{12}]{}(x',y',z')=$$
$$=\begin{cases}(x,y,z) & x,y,z>0\\ (0,y,0) & x=0\\ (x,0,0) & y=0\\ (x,y,-2) & z=0, y\neq 0 \text{ либо } x=m\\ (x,y+2,z) & x+y+z=m\end{cases}.$$

Переходы с *первого* фиктивного экрана:
1) переходы с плоскости $z = -1$:

$$(x,y,-1)\xrightarrow[r_o r_e]{}(x,y,0);$$
$$(x,y,-1)\xrightarrow[p_1 r_e]{}(x,y,-1);$$
$$(x,y,-1)\xrightarrow[p_2 r_e]{}(x,y+1,0);$$
$$(x,y,-1)\xrightarrow[r_o q_1]{}(x-1,y,0);$$
$$(x,y,-1)\xrightarrow[r_o q_2]{}(x',y',z')=$$
$$=\begin{cases}(x,0,0) & y=0 \text{ либо } y=1\\ (x,y-1,1) & y>1\end{cases};$$

$$(x, y, -1) \xrightarrow[p_1 q_1]{} (x, y, -2);$$

$$(x, y, -1) \xrightarrow[p_2 q_2]{} (x', y', z') = \begin{cases} (x, y+2, 1) & x+y=m \\ (x, y, 1) & \text{иначе} \end{cases};$$

$$(x, y, -1) \xrightarrow[p_1 q_2]{} (x', y', z') = \begin{cases} (x, 0, -1) & y=0 \\ (x, y-1, -1) & \text{иначе} \end{cases};$$

$$(x, y, -1) \xrightarrow[p_2 q_1]{} (x', y', z') = \begin{cases} (x', y, 0) & x+y=m \\ (x-1, y+1, 0) & \text{иначе} \end{cases};$$

$$(x, y, -1) \xrightarrow[p_{12} r_e]{} (x, y+1, -1);$$

$$(x, y, -1) \xrightarrow[r_o q_{12}]{} (x', y', z') =$$
$$= \begin{cases} (x-1, 0, 0) & y=0 \text{ либо } y=1 \\ (x-1, y-1, 1) & y>1 \end{cases};$$

$$(x, y, -1) \xrightarrow[p_1 q_{12}]{} (x, y-1, -2);$$

$$(x, y, -1) \xrightarrow[p_2 q_{12}]{} (x', y', z') = \begin{cases} (x-1, y+2, 1) & x+y=m \\ (x-1, y, 1) & \text{иначе} \end{cases};$$

$$(x, y, -1) \xrightarrow[p_{12} q_1]{} (x', y', z') = \begin{cases} (x, y, -2) & x+y=m \\ (x, y+1, -2) & \text{иначе} \end{cases};$$

$$(x, y, -1) \xrightarrow[p_{12} q_2]{} (x', y', z') = \begin{cases} (x, y+2, 1) & x+y=m \\ (x, y, -1) & \text{иначе} \end{cases};$$

$$(x, y, -1) \xrightarrow[p_{12} q_{12}]{} (x, y, -2);$$

2) переходы из состояния $x + y' + z = m + 1$ ($y' = y + 1$):

$$(x, y+1, z) \xrightarrow[r_o r_e]{} (x, y, 0);$$

$$(x, y+1, z) \xrightarrow[p_1 r_e]{} (x', y', z') = \begin{cases} (0, y+1, 0) & y=m \\ (x+1, y, z-1) & \text{иначе} \end{cases};$$

$$(x, y+1, z) \xrightarrow[p_2 r_e]{} (x, y+1, z);$$

$$(x, y+1, z) \xrightarrow[r_o q_1]{} (x', y', z') =$$
$$= \begin{cases} (0, y, 0) & x=0 \text{ либо } x=1 \\ (x-1, y, z) & \text{иначе} \end{cases};$$

$$(x, y+1, z) \xrightarrow[r_o q_2]{} (x', y', z') = \begin{cases} (0, y-1, 0) & x=0 \\ (x, y-1, z+1) & \text{иначе} \end{cases};$$

$$(x, y+1, z) \xrightarrow[p_1 q_2]{} (x', y', z') =$$
$$= \begin{cases} (x+1, y-1, z) & x, y, z>0 \\ (x, y-1, -1) & z=0, x \neq 0; \\ (0, y+2, 0) & y=m \end{cases}$$

$$(x, y+1, z) \xrightarrow[p_2 q_1]{} (x', y', z') = \begin{cases} (0, y+1, 0) & x=0 \\ (x', y, z) & x \neq 0, y \neq 0 \end{cases};$$

$$(x, y+1, z) \xrightarrow[p_1 q_1]{} (x', y', z') = \begin{cases} (x, y, z-1) & x, y, z>0 \\ (x, y, -2) & z=0, x \neq 0; \\ (0, y+1, 0) & y=m \end{cases}$$

$$(x, y+1, z) \xrightarrow[p_2 q_2]{} (x, y+2, z);$$

$$(x, y+1, z) \xrightarrow[p_{12} r_e]{} (x', y', z') =$$
$$= \begin{cases} (0, y+1, 0) & y=m \\ (x+1, y+1, z-1) & \text{иначе} \end{cases};$$

$$(x, y+1, z) \xrightarrow[r_o q_{12}]{} (x', y', z') =$$
$$= \begin{cases} (0, y-1, 0) & x=0 \text{ либо } x=1 \\ (x-1, y-1, z+1) & \text{иначе} \end{cases};$$

$$(x, y+1, z) \xrightarrow[p_1 q_{12}]{} (x', y', z') =$$
$$= \begin{cases} (x, y-1, z) & x, y, z>0 \\ (x, y-1, -2) & z=0, x \neq 0; \\ (0, y+2, 0) & y=m \end{cases}$$

$$(x, y+1, z) \xrightarrow[p_2 q_{12}]{} (x-1, y+2, z);$$

$$(x, y+1, z) \xrightarrow[p_{12} q_1]{} (x', y', z') =$$
$$= \begin{cases} (0, y+1, 0) & x=0 \\ (x', y, z-1) & x+y+z=m \end{cases};$$

$$(x, y+1, z) \xrightarrow[p_{12} q_2]{} (x', y', z') =$$
$$= \begin{cases} (0, y+2, 0) & y=m \\ (x+1, y+2, z-1) & \text{иначе} \end{cases};$$

$$(x, y+1, z) \xrightarrow[p_{12} q_{12}]{} (x', y', z') =$$
$$= \begin{cases} (0, y+2, z-1) & x, y, z>0 \\ (x, y+2, z) & \text{иначе} \end{cases};$$

3) переходы из состояний (x', y, z) ($x' = x - 1$) повторяют переходы из соответствующих состояний, принадлежащих пирамиде:

$$(x', y, z) \leftrightarrow \begin{cases} (0, y, 0) & x'=0 \\ (x', y, z) & \text{иначе} \end{cases}.$$

Переходы из состояний *второго* фиктивного экрана повторяют переходы из соответствующих состояний, принадлежащих пирамиде:

$$(x, y, z) \leftrightarrow (x-1, y, z+2) \text{ для } z=-2;$$

$$(x, y', z) \leftrightarrow \begin{cases} (x, 0, 0) & y = 0 \\ (x-1, 0, 0) & y = 1 \\ (0, y-1, 0) & y + z = m \\ (x, y, z+1) & \text{иначе} \end{cases}$$

для $x + y' + z = m + 2$.

Случайное блуждание рассматриваем в виде регулярной конечной цепи Маркова с переходной матрицей **P**. Зададим нумерацию так, как показано на рис. 2: сначала нумеруем состояния, принадлежащие основанию пирамиды ($z = 0$), далее — поперечным сечениям ($z = 1, 2, \ldots, m - 1$) и высоте пирамиды ($z = m$), затем — состояния первого фиктивного экрана, последними нумеруем состояния второго экрана.

Нумерацию начинаем с нуля, общее количество состояний в цепи будет

$$n = \sum_{i=0}^{m} \frac{(i+1)(i+2)}{2} + \frac{5m^2 + 7m + 4}{2}.$$

Далее, согласно введенной нумерации состояний и вышеуказанной схеме переходов, составляется матрица переходных вероятностей **P**. В данном исследовании был установлен вид матрицы **P** для произвольных значений параметра m. При составлении матрицы для каждого конкретного состояния определяются те состояния, в которые процесс переходит при выполнении допустимых операций, и вычисляются соответствующие вероятности переходов. Данный процесс был автоматизирован с помощью программы на языке C.

Следующим шагом является решение уравнения $\boldsymbol{\alpha}\mathbf{P} = \boldsymbol{\alpha}$, где $\boldsymbol{\alpha}$ — предельный вектор для полученной марковской цепи. Для этого применялась система Intel Math Kernel Library PARDISO. Данный инструмент используется для прямого решения систем линейных алгебраических уравнений с разряженной матрицей.

Элемент вектора α_i — это средняя доля времени, которое процесс проводит в состоянии i [18]. Для вычисления времени, проводимого процессом в состояниях, где происходят потери элементов очередей, нужно просуммировать элементы вектора $\boldsymbol{\alpha}$, соответствующие состояниям на экранах. При введенной нумерации это будут последние $\frac{5m^2 + 7m + 4}{2}$ элементов вектора.

Оценка сложности алгоритма:

$$O\left(\left(\sum_{i=0}^{m} \frac{(i+1)(i+2)}{2} + \frac{5m^2 + 7m + 4}{2}\right)^3\right).$$

Предложенная математическая модель анализируемого в статье нового способа работы с FIFO-очередями в общей памяти требует проверки ее адекватности. Для этого была разработана имитационная модель процесса, краткое изложение которой представлено в следующем разделе.

Имитационная модель

Обозначим через x текущую длину первой очереди, через y — второй, через z — расстояние между очередями. Общий объем памяти равен m. Тогда в качестве имитационной модели будем рассматривать случайное блуждание по кругу по часовой стрелке (см. рис. 1), при этом на нечетных шагах:

1) с вероятностью p_1 увеличивается длина первой очереди, соответственно, уменьшается расстояние z между очередями;

2) p_2 — увеличивается длина второй очереди;

3) p_{12} — одновременно увеличивается длина первой и второй очередей, расстояние z уменьшается;

4) r_o — процесс остается на месте.

На четных шагах:

1) с вероятностью q_1 уменьшается длина первой очереди;

2) q_2 — уменьшается длина второй очереди, соответственно, увеличивается расстояние z между очередями;

3) q_{12} — длина первой и второй очередей уменьшается одновременно, z увеличивается;

4) r_e — процесс остается на месте.

Для того чтобы в имитационной модели определять, какое действие будет выполняться, необходимо отрезок от 0 до 1 разделить между вероятностями так, чтобы их сумма была равна 1.

Далее с помощью датчика случайных чисел (использовался датчик, встроенный в транслятор gcc) генерируется последовательность, с помощью которой будет происходить блуждание по кругу, и подсчитывается количество нахождений очередей в переполнении, т. е. в случаях, когда $z = 0$ либо $x + y + z = m$.

Некоторые примеры численного анализа

Для определения случаев, когда целесообразно применять анализируемый в статье метод работы с очередями «Друг за другом», был проведен ряд экспериментов с различным набором исходных данных. Аналогичные вычисления были проведены для последовательного циклического способа организации очередей в общей памяти при оптимальном разбиении памяти [15]. Там же был рассмотрен случай, когда память заранее делится пополам при последовательном способе организации очередей. Это может быть полезно, если заранее не известны вероятностные

■ Сравнение потерь

Входные данные	Величина потерь при переполнении ($m = 10$)		
	Оптимальное разбиение	Разбиение пополам ($s = 5$)	Друг за другом
$p_1 = 0{,}25$, $p_2 = 0{,}25$, $p_{12} = 0{,}25$, $r_1 = 0{,}25$ $q_1 = q_2 = q_{12} = r_2 = 0{,}25$	0,089 ($s = 5$)	0,089	0,100
$p_1 = 0{,}30$, $p_2 = 0{,}20$, $p_{12} = 0{,}10$, $r_1 = 0{,}40$ $q_1 = q_2 = q_{12} = r_2 = 0{,}25$	0,012 ($s = 6$)	0,014	0,019
$p_1 = 0{,}35$, $p_2 = 0{,}15$, $p_{12} = 0{,}10$, $r_1 = 0{,}40$ $q_1 = q_2 = q_{12} = r_2 = 0{,}25$	0,019 ($s = 7$)	0,025	0,024
$p_1 = 0{,}40$, $p_2 = 0{,}10$, $p_{12} = 0{,}10$, $r_1 = 0{,}40$ $q_1 = q_2 = q_{12} = r_2 = 0{,}25$	0,035 ($s = 7$)	0,046	0,034
$p_1 = 0{,}45$, $p_2 = 0{,}05$, $p_{12} = 0{,}10$, $r_1 = 0{,}40$ $q_1 = q_2 = q_{12} = r_2 = 0{,}25$	0,064 ($s = 8$)	0,075	0,059

характеристики операций, производимых с очередями.

Некоторые результаты вычислений (все указанные результаты были подтверждены имитационными экспериментами) представлены в таблице. Так как аналитическое решение не было получено, нам нужно проводить вычисления для конкретных значений m. Величину $m = 10$ мы используем в качестве примера, взятые здесь вероятности являются теоретическими — для большей наглядности результатов. На практике же эти вероятности должны быть получены в результате предварительных статистических исследований.

Анализируя результаты, можно сказать, что для случая равных вероятностей метод «Друг за другом» дает наибольшие потери. Но с увеличением вероятности включения в одну очередь (в данном случае — в первую) и уменьшением вероятности включения в другую этот метод дает наименьшие потери.

Так, при вероятностях включения 0,4 в одну очередь и 0,1 в другую разница потерь между «Друг за другом» и оптимальным разбиением — 0,001. То есть из тысячи элементов (пакетов, в случае сетевых приложений FIFO-очередей) мы теряем в среднем на один элемент меньше, пуская очереди друг за другом, по кругу. А при вероятностях включения 0,45 и 0,05 соответственно — уже на пять элементов меньше.

С помощью разработанного программного комплекса можно находить оптимальный способ представления очередей для их конкретных вероятностных характеристик.

Заключение

В данной статье мы предлагаем и анализируем математическую модель, которая описывает новый способ управления работой двумя параллельными FIFO-очередями в общей памяти. Такая модель может быть полезна при разработке различных технических систем, в архитектуре которых необходим именно такой способ управления.

И хотя в настоящее время память становится все больше и дешевле, есть устройства, где ее размер ограничен архитектурой, и имеющийся ресурс необходимо использовать эффективно. Это могут быть различные сетевые устройства (например, маршрутизаторы). В них нельзя беспредельно увеличивать размер памяти, так как при больших размерах очередей значительно увеличивается время обработки элементов. Сетевые протоколы устроены так, что при большом времени обработки пакеты считаются потерянными, и их посылают снова, тем самым увеличивая загрузку сети (congestion collapse) [19].

Другими такими устройствами могут быть микросхемы и микропроцессоры, которые используются в системах, где потеря элементов очередей является допустимой, но нежелательной ситуацией (например, цифровая обработка сигналов или работа с мультимедийными приложениями [4]). В таких процессорах каждая очередь может содержать всего несколько десятков элементов. Предложенный в статье способ может являться одним из методов повышения надежности (как снижения доли потерянных элементов) работы таких устройств.

Остается открытым вопрос, какой из методов работы (разделение на четные-нечетные шаги или произвольное выполнение операций) будет оптимальным для тех или иных аппаратных или программных решений. Отметим, что в реальной телекоммуникационной системе вероятностные характеристики очередей будут меняться со временем. При этом разумно предположить, что вероятности появления новых элементов в очередях будут периодически повторяться. Тогда интервал времени такого периода (например, сутки) можно разбить на участки, где эти вероятности мало меняются, и на каждом промежутке времени выбирать тот или иной метод представления очередей, например, друг за другом по кругу в общей памяти или когда каждая очередь представлена циклически отдельно в своем участке.

Отдельной задачей в случае метода движения друг за другом по кругу может быть задача нахождения оптимальной точки начала движения второй очереди или оптимального сдвига очереди в случае если одна очередь догнала другую.

Оценить масштабируемость предлагаемого решения достаточно сложно. Даже для трех очередей возможны три варианта представления. Три очереди двигаются друг за другом в общей памяти; две очереди двигаются друг за другом в общей памяти, а третья представлена циклически в оставшейся памяти; или все три очереди представлены циклически, каждая в своем участке памяти. В случае n очередей число вариантов представления резко возрастает. Хотя, в принципе, можно предложить такой способ масштабирования решения: n очередей разбиваются на пары, и внутри каждой пары периодически производится смена метода представления, а в случае циклического раздельного представления очередей — оптимальное перераспределение памяти. Это делать проще, чем производить полное перераспределение памяти, но в таком случае достигается не глобальный, а некий локальный минимум средней доли потерянных при переполнении элементов очередей. Такой подход применяется в многоядерных процессорах.

Работа выполнена при финансовой поддержке РФФИ, грант 15-01-03404-а.

Литература

1. **Sedgewick R.** Algorithms in C++. Parts 1–4. — Addison-Wesley Professional, 1998. — 752 p.
2. **Bollapragada V., Murphy C., White R.** Inside Cisco IOS Software Architecture. — Cisco Press, 2000. — 240 p.
3. **Knuth D.** The Art of Computer Programming. Vol. 1. — Addison-Wesley Professional, 1997. — 672 p.
4. **Калачев А. В.** Многоядерные процессоры. — М.: БИНОМ, 2014. — 247 с.
5. **Аксенова Е. А., Соколов А. В., Драц А. В.** Оптимальное управление *n* FIFO-очередями на бесконечном времени // Информационно-управляющие системы. 2009. № 6. С. 46–54.
6. **Aksenova E. A., Sokolov A. V.** The Optimal Implementation of Two FIFO-Queues in Single-Level Memory // Applied Mathematics. 2011. Vol. 2. P. 1297–1302.
7. **Sokolov A. V., Drac A. V.** The Linked List Representation of n LIFO-Stacks and/or FIFO-Queues in the Single-Level Memory // Information Processing Letters. 2013. Vol. 13. P. 832–835.
8. **Соколов А. В., Драц А. В.** Моделирование некоторых методов представления n FIFO-очередей в памяти одного уровня // Эвристические алгоритмы и распределенные вычисления. 2014. Т. 1. № 1. С. 40–52.
9. **Соколов А. В.** О распределении памяти для двух стеков // Автоматизация эксперимента и обработки данных. 1980. С. 65–71.
10. **Yao A. C.** An Analysis of a Memory Allocation Scheme for Implementing Stacks // SIAM Journal on Computing. 1981. Vol. 10. P. 398–403.
11. **Flajolet P.** The Evolution of Two Stacks in Bounded Space and Random Walks in a Triangle // Lecture Notes in Computer Science. 1986. Vol. 223. P. 325–340.
12. **Louchard G., Schott R.** Probabilistic Analysis of Some Distributed Algorithms // Lecture Notes in Computer Science. 1990. Vol. 431. P. 239–253.
13. **Louchard G., Schott R., Tolley M., Zimmermann P.** Random Walks, Heat Equation and Distributed Algorithms // Journal of Computational and Applied Mathematics. 1994. N 53. P. 243–274.
14. **Maier R. S.** Colliding Stacks: A Large Deviations Analysis // Random Structures and Algorithms. 1991. N 2. P. 379–421.
15. **Соколов А. В., Барковский Е. А.** Оптимальное управление двумя параллельными FIFO-очередями на бесконечном времени // Информационно-управляющие системы. 2015. № 5. С. 65–71. doi:10.15217/issn1684-8853.2015.5.65
16. **Соколов А. В.** Математические модели и алгоритмы оптимального управления динамическими структурами данных. — Петрозаводск: ПетрГУ, 2002. — 216 с.
17. **Sokolov A. V., Barkovsky E. A.** Some Problems of Optimal Control of Two Parallel FIFO-queues // Intern. Conf. on Numerical Analysis and Applied Mathematics: Proc. 12th Int. Conf., Rhodes, 18–22 Sept. 2014. AIP Publishing, 2015. Vol. 1648. P. 520003.
18. **Kemeny J. G., Snell J. L.** Finite Markov Chains. — Van Nostrand, 1969. — 210 p.
19. **Tanenbaum A. S., Wetherall D. J.** Computer Networks. — Pearson Education, 2011. — 933 p.

UDC 004.942
doi:10.15217/issn1684-8853.2016.1.65

Management Model for Two Parallel FIFO Queues Moving One after Another in Shared Memory

Barkovsky E. A.[a], Applicant, barkevgen@gmail.com
Sokolov A. V.[a, b], Dr. Sc., Phys.-Math., Professor, avs@krc.karelia.ru
[a]Petrozavodsk State University, 33, Lenin St., 185910, Petrozavodsk, Russian Federation
[b]Institute of Applied Mathematical Research of Karelian Research Centre Russian Academy of Sciences, 11, Pushkinskaya St., 185910, Petrozavodsk, Russian Federation

Introduction: FIFO queue is a popular data structure in hardware and software applications. Various network devices and embedded operating systems use several FIFO queues located in a shared memory space. There are also multi-core processor architectures where two FIFO queues are allocated for each core. Software and hardware applications used to solve the described problems should improve the device reliability, reducing the average portion of the queue elements lost by overflow. **Purpose:** The goal is to construct and analyze a mathematical model of operating two FIFO queues moving one after another in a circle in the shared memory. On an odd step, elements are inserted into one of the queues. On an even step, elements are deleted. Both serial and parallel execution of the operations are possible. **Results:** A mathematical and a simulation models of this process for two queues were constructed, and numerical experiments based on theoretical data were performed. The mathematical model is built as a random walk on an integer pyramid with reflecting screens. An algorithm was proposed for the numbering of states; the form of the matrix of transition states for the obtained regular Markov chain was established; and the corresponding propositions were proved. Another algorithm and software were developed for calculating the average portion of queue elements lost by overflow. The peculiarity of this research is a specific execution of operations with queues: insertion and deletion of elements occur depending on the step (some amendments were made to maintain the homogeneity and regularity of the chain). The operations can be executed in parallel. **Practical relevance:** The model, algorithm and software complex proposed for the analysis of "One after another" queue movement method can be used in the design of network devices, such as routers, chips with multiple FIFO queues and other software or hardware devices where element loss is permitted but unwanted. Using this model, with given probabilities of the queues, you can choose the best queue representation method, for example, between the classical serial cyclic method and "One after another" method.

Keywords — Data Structures, FIFO queues, Random Walks, Regular Markov Chains.

References

1. Sedgewick R. *Algorithms in C++. Parts 1–4*. Addison-Wesley Professional, 1998. 752 p.
2. Bollapragada V., Murphy C., White R. *Inside Cisco IOS Software Architecture*. Cisco Press, 2000. 240 p.
3. Knuth D. *The Art of Computer Programming. Vol. 1*. Addison-Wesley Professional, 1997. 672 p.
4. Kalachev A. V. *Mnogoiadernye protsessory* [Multicore Architectures]. Moscow, BINOM Publ., 2014. 247 p. (In Russian).
5. Aksenova E. A., Sokolov A. V., Drac A. V. Optimal Control of the n FIFO-queues for Infinity Time. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2009, no. 6, pp. 46–54 (In Russian).
6. Aksenova E. A., Sokolov A. V. The Optimal Implementation of Two FIFO-Queues in Single-Level Memory. *Applied Mathematics*, 2011, vol. 2, pp. 1297–1302.
7. Sokolov A. V., Drac A. V. The Linked List Representation of n LIFO-Stacks and/or FIFO-Queues in the Single-Level Memory. *Information Processing Letters*, 2013, vol. 13, pp. 832–835.
8. Sokolov A. V., Drac A. V. Modeling of Several Methods of Representation of n FIFO-Queues in the Single-Level Memory. *Jevristicheskie algoritmy i raspredelennye vychislenija*, 2014, vol. 1, no. 1, pp. 40–52 (In Russian).
9. Sokolov A. V. About Memory Distribution for Two Stacks. *Avtomatizatsiia eksperimenta i obrabotki dannykh*, 1980, pp. 65–71 (In Russian).
10. Yao A. C. An Analysis of a Memory Allocation Scheme for Implementing Stacks. *SIAM Journal on Computing*, 1981, vol. 10, pp. 398–403.
11. Flajolet P. The Evolution of Two Stacks in Bounded Space and Random Walks in a Triangle. *Lecture Notes in Computer Science*, 1986, vol. 223, pp. 325–340.
12. Louchard G., Schott R. Probabilistic Analysis of Some Distributed Algorithms. *Lecture Notes in Computer Science*, 1990, vol. 431, pp. 239–253.
13. Louchard G., Schott R., Tolley M., Zimmermann P. Random Walks, Heat Equation and Distributed Algorithms. *Journal of Computational and Applied Mathematics*, 1994, no. 53, pp. 243–274.
14. Maier R. S. Colliding Stacks: A Large Deviations Analysis. *Random Structures and Algorithms*, 1991, no. 2, pp. 379–421.
15. Barkovsky E. A., Sokolov A. V. Optimal Control of Two Parallel FIFO Queues on an Infinite Time. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2015, no. 5, pp. 65–71 (In Russian). doi:10.15217/issn1684-8853.2015.5.65
16. Sokolov A. V. *Matematicheskie modeli i algoritmy optimal'nogo upravlenija dinamicheskimi strukturami dannyh* [Mathematical Models and Algorithms of Optimal Control of Dynamic Data Structures]. Petrozavodsk, PetrSU Publ., 2002. 216 p. (In Russian).
17. Sokolov A. V., Barkovsky E. A. Some Problems of Optimal Control of Two Parallel FIFO-queues. *Proc. 12th Int. Conf. "International Conference on Numerical Analysis and Applied Mathematics"*, AIP Publishing, 2015, vol. 1648, pp. 520003.
18. Kemeny J. G., Snell J. L. *Finite Markov Chains*. Van Nostrand, 1969. 210 p.
19. Tanenbaum A. S., Wetherall D. J. *Computer Networks*. Pearson Education, 2011. 933 p.

УДК 621.396.9
doi:10.15217/issn1684-8853.2016.1.74

ОБЕСПЕЧЕНИЕ СЕЛЕКТИВНОГО ДОСТУПА ПРИ ШИРОКОВЕЩАТЕЛЬНОЙ ПЕРЕДАЧЕ ИНФОРМАЦИИ

С. В. Штанько[a]*, канд. техн. наук, доцент*
Д. А. Лесняк[a]*, канд. техн. наук, старший преподаватель*
[a]*Военно-космическая академия им. А. Ф. Можайского, Санкт-Петербург, РФ*

Постановка проблемы: *одним из основных условий эффективного функционирования любой системы передачи информации является ее информационная безопасность. Объективным свойством защищенности информационных систем является ее постепенное снижение при неизменном составе средств защиты или при фиксированных их параметрах. Для поддержания информационной безопасности и предотвращения несанкционированного доступа осуществляется периодическая замена устаревших ключей в криптосистеме. В широковещательных системах передачи информации с селективным доступом имеются особенности обеспечения информационной безопасности, связанные с отсутствием обратного канала передачи информации. Это приводит к необходимости разработки специальных методов криптографической защиты для широковещательных систем информации.* ***Цель:*** *усовершенствование существующей криптографической системы защиты данных для повышения защищенности каналов широковещательных систем передачи информации с селективным доступом от несанкционированного доступа.* ***Результаты:*** *предложена модификация методов защиты информации в целях предотвращения возможности несанкционированного доступа к системе в конфликтных условиях взаимодействия с нарушителем, основанная на асимметричных криптографических системах с использованием математического аппарата эллиптических кривых. На базе предложенных криптографических методов защиты построена криптографическая система защиты данных, передаваемых по каналам широковещательной системы передачи информации с селективным доступом. Представлен пример генерации общего сеансового ключа при передаче информации в закрытом режиме работы.* ***Практическая значимость:*** *предложенные пути повышения защищенности каналов широковещательных систем передачи информации с селективным доступом от несанкционированного доступа позволяют повысить информационную безопасность системы на основе использования комбинированных криптосистем и математического аппарата эллиптических кривых.*

Ключевые слова *— широковещательные системы передачи информации с селективным доступом, несанкционированный доступ, криптоалгоритм, протокол управления ключами.*

Введение

Одним из элементов современной глобальной информационной инфраструктуры являются широковещательные системы передачи информации, которые активно внедряются и используются в области телекоммуникаций. Широковещательные системы передачи информации отличают такие особенности, как сплошное покрытие обслуживаемых территорий, а также ограниченное число передающих и значительно превосходящее их число приемных пунктов (абонентов) при отсутствии обратного канала передачи информации. Если в такой системе предусматривается селективный доступ, то в ней необходимо использовать криптографические методы обеспечения информационной безопасности. При этом обеспечение информационной безопасности в широковещательных системах передачи информации с селективным доступом (ШСПИ с СД) связано с отсутствием обратного канала передачи информации.

Постепенное снижение защищенности информационных систем при неизменном составе средств защиты или при их фиксированных параметрах, а также пространственная электромагнитная доступность, являющаяся свойством любых радиоканалов, создает условия для реализации угроз несанкционированного доступа (НСД) к радиоканалам передачи информации [1].

Для предотвращения возможности несанкционированного использования ШСПИ с СД в условиях конфликтного взаимодействия с нарушителем необходимо применять криптографические методы защиты информации. Для этого надо построить полноценную криптографическую систему защиты данных, передаваемых по радиоканалу ШСПИ с СД. Кроме того, в некоторых ШСПИ, помимо требований по криптостойкости (информационной скрытности), предъявляются требования по имитостойкости, и при осуществлении передачи конфиденциальной информации по каналам ШСПИ с СД принципиально необходимо выполнять процедуру аутентификации — подтверждения подлинности абонента [2, 3]. Дополнительная сложность реализации полноценной криптографической системы защиты данных в ШСПИ с СД заключается в отсутствии обратного канала, в то время как для построения необходимо реализовать эффективные процедуры смены ключевой информации.

Перечисленные проблемы защиты информации в ШСПИ с СД успешно решаются с использованием асимметричных криптосистем.

Построение системы защиты широковещательной системы связи на основе комбинированных криптосистем

В простейшем случае в ШСПИ с СД используется единый несменяемый ключ симметричного алгоритма, что делает такую систему уязвимой. Посредством этого ключа авторизованным пользователем расшифровывается получаемая им информация. Для повышения защищенности системы необходимо предусмотреть возможность смены ключей как в передающих (ПРД), так и в приемных (ПРМ) пунктах [4].

Для разработки полноценной криптосистемы следует обеспечить процедуру смены единого ключа для системы, а в лучшем случае — процедуру формирования рабочих ключей, защищаемых долговременным мастер-ключом, а также процедуру смены мастер-ключа [5].

Предпочтительным представляется применение комбинированных криптосистем, использующих преимущество симметричных по скорости и асимметричных по широким возможностям построения схем аутентификации и доставки ключей. В качестве алгоритмов асимметричной части комбинированной схемы рекомендуется прибегнуть к математическому аппарату эллиптических кривых как обладающий наилучшими криптографическими и скоростными характеристиками по сравнению с другими типами асимметричных алгоритмов [6]. Непосредственно шифрование защищаемых данных в этом случае осуществляется симметричным способом методом гаммирования, а для генерации сеансовых ключей и проведения процедуры аутентификации применяются асимметричные методы [7, 8].

Алгоритм шифрования и генерация ключа с использованием асимметричного алгоритма можно представить в следующем формализованном виде:

$$c = E_{k_о}(m); \quad (1)$$

$$m = D_{k_з}(c), \quad (2)$$

где E — шифрующее преобразование асимметричного алгоритма; D — расшифровывающее преобразование асимметричного алгоритма; m — исходный информационный массив; c — зашифрованный массив на ключе $k_з$; $k_з$ — закрытый ключ, сгенерированный случайным образом; $k_о$ — открытый ключ, полученный путем одностороннего преобразования [9]:

$$k_о = f(k_з). \quad (3)$$

Сеансовые ключи симметричной системы для шифрования и асимметричной для аутентификации (в случае необходимости) генерируются на основе алгоритма Диффи — Хеллмана с использованием долговременных ключей асимметричной системы и случайного числа либо меток времени (либо того и другого) для того, чтобы сеансовые ключи отличались друг от друга [3]. В случае однонаправленных каналов ШСПИ с СД алгоритм Диффи — Хеллмана является упрощенным в том смысле, что у всех абонентов долговременные ключи одинаковы и алгоритм не зависит от того, какому абоненту передается информация, и не требует передачи информации от наземного абонента [10].

Эллиптическую кривую над конечным полем Галуа GF_p можно представить в виде

$$E_p(a, b):\ y^2 = x^3 + ax + b(\mathrm{mod}\ p). \quad (4)$$

Здесь $E_p(a, b)$ — эллиптическая группа по модулю p, элементами которой (x, y) являются пары неотрицательных чисел, меньших p и удовлетворяющих уравнению кривой, а также точка в бесконечности O.

Групповой закон сложения точек $P_1 \oplus P_2$ имеет вид $P_1 \oplus P_2 = (x_3, y_3)$, где

$$x_3 = \left(\frac{y_2 - y_1}{x_2 - x_1}\right)^2 - x_2 - x_1; \quad (5)$$

$$y_3 = \left(\frac{y_2 - y_1}{x_2 - x_1}\right)(x_1 - x_3) - y_1. \quad (6)$$

При $P_1 = P_2 = (x_1, y_1)$ получаем $2P_1 = (x_2, y_2)$:

$$x_2 = \frac{(3x_1^2 + a)^2}{4y_1^2} - 2x_1; \quad (7)$$

$$y_2 = \frac{(3x_1^2 + a)}{2y_1}(x_1 - x_2) - y_1. \quad (8)$$

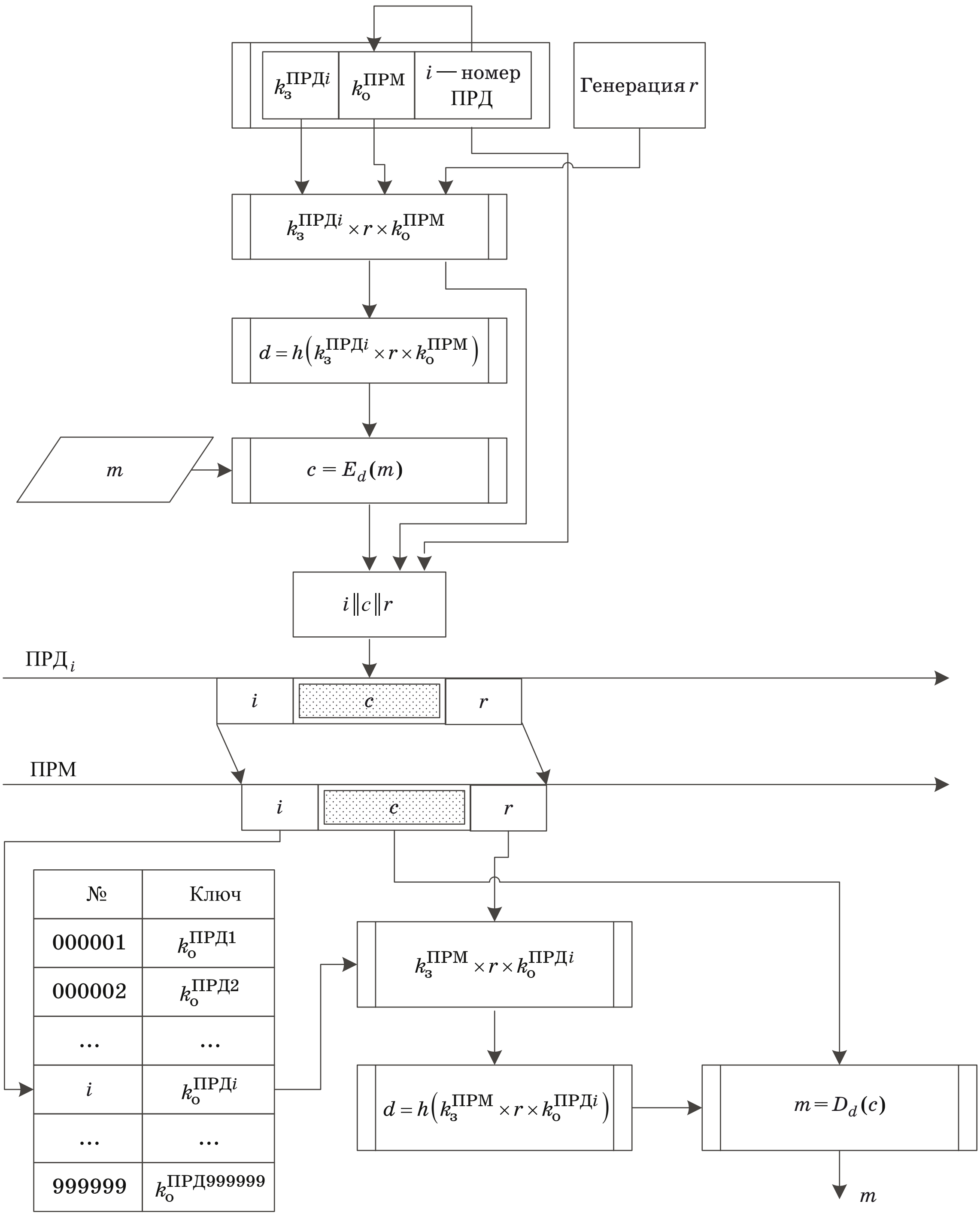
Для организации обмена ключевой информацией на основе асимметричных криптосистем необходима односторонняя функция. В качестве такой функции в математическом аппарате эллиптических кривых выступает умножение точки на число. Пусть p — простое число, G — генератор аддитивной циклической подгруппы группы точек эллиптических кривых, а P — произвольная точка, принадлежащая данной кривой.

Тогда любую точку P кривой $E(GF_p)$ можно представить как кратную генератору подгруппы в виде

$$P = n * G = \underbrace{G \oplus G \oplus \ldots \oplus G}_{n\ \text{раз}}, \quad (9)$$

где n — кратность данной точки генератору подгруппы; «$\oplus$» — знак групповой операции в группе точек кривой.

Алгоритм передачи информации по радиоканалу ШСПИ с СД с осуществлением криптографической защиты представлен на рисунке.



■ Алгоритм передачи информации в широковещательных системах с селективным доступом по радиоканалу ПРД — ПРМ с осуществлением криптографической защиты

Параметры алгоритма:

E — шифрующее преобразование симметричного алгоритма;

D — расшифровывающее преобразование симметричного алгоритма;

GF_p — конечное поле простой характеристики p;

$E_p(a, b)$ — эллиптическая кривая над полем GF_p в форме Вейрштрасса (кривая может быть задана на простом GF_p или расширенном GF_{qn} поле в аффинных либо проективных координатах);

$\#E(GF_p)$ — порядок группы точек кривой;

G — базовая точка — генератор подгруппы;

q — порядок циклической подгруппы;

h — односторонняя хеш-функция, принимающая значения в множестве двоичных векторов длины 256, определенная стандартом Р 34.11-2012 [11];

i — номер передающего пункта;

$k_з^{ПРДi}$ — закрытый долговременный ключ передающего пункта (число);

$k_о^{ПРДi}$ — открытый долговременный ключ передающего пункта (точка эллиптической кривой), где $k_о^{ПРДi} = k_з^{ПРДi} \times G$;

$k_з^{ПРМ}$ — закрытый долговременный ключ абонентов приемного пункта (число);

$k_о^{ПРМ}$ — открытый долговременный ключ абонентов приемного пункта (точка эллиптической кривой), где $k_о^{ПРМ} = k_з^{ПРМ} \times G$;

d — секретный сеансовый ключ шифрования данных (число);

r — случайное число.

Алгоритм передачи информации по каналу связи осуществляется следующим образом.

1. Передающий пункт формирует информационный массив m.

2. Передающий пункт генерирует случайное число r.

3. Передающий пункт вычисляет сеансовый ключ:

$$d = h\left(k_з^{ПРДi} \times r \times k_о^{ПРМ}\right) = h\left(k_з^{ПРДi} \times r \times k_з^{ПРМ} \times G\right).$$

4. Передающий пункт шифрует кадр m на ключе d:

$$c = E_d(m).$$

5. Передающий пункт передает комбинацию $i\|c\|r$.

6. Приемный пункт, получив комбинацию $i\|c\|r$, в матрице доступности находит открытый ключ передающего пункта: $i — k_о^{ПРДi}$.

7. Приемный пункт вычисляет сеансовый ключ d:

$$d = h\left(k_з^{ПРМ} \times r \times k_о^{ПРДi}\right) = h\left(k_з^{ПРМ} \times r \times k_з^{ПРДi} \times G\right).$$

8. Приемный пункт расшифровывает полученный кадр:

$$m = D_d(c).$$

В качестве r может быть использовано как собственно случайное число, так и временная метка τ и их совокупность $R = r \times \tau$.

Пример генерации общего сеансового ключа при передаче информации по каналу с осуществлением криптографической защиты (без использования хеш-функции)

Начальные условия:

Параметры эллиптической кривой:

модуль $p = 56eefdd4da5da8de9c6e7d42cd2406c58d71cb82ee7af4bc4121d57_{16}$;

коэффициенты:

$a = 420507356e57e70f37fb2108f63bb3b32815e46d237e396ec3dbf74_{16}$;

$b = 5f5f3cca03cf28a09cac8f15bb72cccf2dccb4a8ac45ca0bdc9d45_{16}$;

порядок циклической подгруппы

$q = 27bdb5dda5a728a04782f76efeb04676dc5c338eb26c48d9b7819d_{16}$;

базовая точка G:

$x = 9ff29581d0e93626157f777c31c6c2654661b7c4109a94c86d3b7d_{16}$;

$y = 4c5f6c882bce511fedf257cbf11d06a43dcf8d2754ec020d7c3190d_{16}$.

Ключи передающего пункта:

$k_з^{ПРДi} = 1759783648d84246da7a202c170e6bbff320924a65543153c10bb_{16}$;

$k_о^{ПРМi}$:

$x = 4b805969bc68df959ffd0534ad27468d87af03cc613661591dad9eb_{16}$;

$y = 45397a89d17ffe0913d73fff194c55f98f8fc3ddaf8e2265b69582d_{16}$.

Ключи приемного пункта:

$k_з^{ПРМi} = 2249d446910301796b96e840d82b7460f010b2be78be8755cc9d52_{16}$;

$k_о^{ПРМi}$:

$x = 12fea6bc3f5047b4c098806c386dbacb12a767dcde8592911175b81_{16}$;

$y = 45e07d359a5b34904193580de8519b4226679c7bbf07dbbddcce86f_{16}$.

1. Передающий пункт генерирует случайное число:

$r = cc7d574f7c697502baf9cbfb83d04e29fbd3efd5b018bc347f_{16}$.

2. Передающий пункт вычисляет сеансовый ключ: $d = k_з^{ПРДi} \times r \times k_о^{ПРМ} = 1759783648d84246da7a202c170e6bbff320924a65543153c10bb_{16} \times cc7d574f7c697502baf9cbfb83d04e29fbd3efd5b018bc347f_{16} \times (12fea6bc3f5047b4c098806c386dbacb12a767dcde8592911175b81_{16},\ 45e07d359a5b34904193580de8519b4226679c7bbf07dbbddcce86f_{16}) = 15e3c4b397db60794e42a6d84e8c590629c8d6d4d928676b9f18493255b792a2fbfff731beb290a13de2cec1a48fc31d74f39c27c3f302_{16}$.

3. Приемный пункт, получив r, вычисляет сеансовый ключ $d = k_з^{ПРМ} \times r \times k_о^{ПРДi} = 2249d446910301796b96e840d82b7460f010b2be78be8755cc9d52_{16} \times cc7d574f7c697502baf9cbfb83d04e29fbd3efd5b018bc347f_{16} \times (4b805969bc68df959ffd0534ad27468d87af03cc613661591dad9eb_{16},\ 45397a89d17ffe0913d73fff194c55f98f8fc3ddaf8e2265b69582d_{16}) = 15e3c4b397db60794e42a6d84e8c590629c8d6d4d928676b9f18493255b792a2fbfff731beb290a13de2cec1a48fc31d74f39c27c3f302_{16}$.

Таким образом, оба абонента обладают одним и тем же секретным ключом d. Очевидно, что если в алгоритм добавить хеш-функцию, как описано выше, результат вычислений также у обоих абонентов будет одинаков, так как аргументы хеш-функции у абонентов равны. В ходе дальнейшей передачи по каналу связи информация может быть зашифрована на полученном ключе.

В случае необходимости абонент на передающем пункте может сформировать аутентифицирующие блоки (аналог электронной цифровой подписи), используя свой закрытый ключ, а на приемных пунктах абоненты осуществляют верификацию аутентифицирующих блоков с использованием открытого ключа передающего пункта [12].

Долговременные ключи передающего и приемного пунктов ($k_з^{ПРМ}$ и $k_з^{ПРМi}$) также требуют замены. При этом закрытые ключи передающего пункта могут генерировать самостоятельно, поскольку каждый передающий пункт обладает личным закрытым ключом, а также вычислять соответствующие открытые ключи с последующей передачей их на приемный пункт. Замена же пары долговременных ключей приемным пунктом представляет большую проблему, так как приемный пункт не осуществляет передачу.

Одним из таких способов является периодическая (раз в месяц) замена на основе проведения циклической операции со старыми ключами. Смена может производиться как в жесткой привязке ко времени, так и по команде центра [13]. Для реализации циклической смены ключей необходим криптографически стойкий генератор псевдослучайной последовательности. В качестве аппарата, осуществляющего циклическую операцию, может быть также использован математический аппарат эллиптических кривых. Примером может служить генератор, основанный на умножении точки на число в группе точек эллиптической кривой.

Кроме того, для замены ключевой информации в приемном пункте можно хранить не один мастер-ключ, а множество, и централизованно переходить с одного на другой в случае необходимости. Чтобы исключить несанкционированный доступ к сигналу, доступ ко всем ключам в приемном пункте должен осуществляться по предъявлению пароля либо с использованием средств типа «электронных паролей», т. е. посредством ввода парольной информации с электронных носителей (специальный чип и т. д.).

Заключение

Таким образом, для разработки полноценной криптосистемы необходимо предусмотреть процедуры формирования рабочих ключей, защищаемых долговременным мастер-ключом, а также процедуру смены мастер-ключа. Предпочтительными являются комбинированные криптосистемы, использующие преимущество первых по скорости и вторых по широким возможностям построения схем аутентификации и доставки ключей. Окончательный вариант архитектуры зависит от требований, предъявляемых к криптографической системе защиты информации, наличия связных каналов, способов доставки ключей до приемных пунктов и т. д. В качестве алгоритмов асимметричной части комбинированной схемы рекомендуется применить математический аппарат эллиптических кривых как обладающий наилучшими криптографическими и скоростными характеристиками по сравнению с другими типами асимметричных алгоритмов.

Литература

1. **Малюк А. А.** Информационная безопасность: концептуальные и методологические основы защиты информации. — М.: Горячая линия–Телеком, 2004. — 280 с.
2. **Stallings W.** Cryptography and Network Security Principles and Practices. — New Jersey: Prentice Hall, 2000. — 592 p.
3. **Schneier B.** Applied Cryptography. Protocols, Algorthms and Source Code in C. — N. Y.: John Wiley, 1996. — 1027 p.
4. **Ростовцев А. Г., Маховенко Е. Б.** Введение в криптографию с открытым ключом. — СПб.: Мир и Семья, 2001. — 336 с.
5. **Иванов М. А.** Криптографические методы защиты информации в компьютерных системах и сетях. — М.: Кудиц-Образ, 2001. — 368 с.
6. **Корниенко А. А., Еремеев М. А., Ададуров С. Е.** Средства защиты информации на железнодорожном транспорте: Криптографические методы и средства. — М.: Маршрут, 2006. — 256 с.
7. **Мальцев Г. Н., Штанько С. В.** Протоколы аутентификации абонентов и защиты информации на основе асимметричных криптоалгоритмов // Проблемы информационной безопасности. Компьютерные системы. 2003. № 1. С. 51–56.
8. **Корниенко А. А., Штанько С. В.** Криптографический протокол защиты информации в радиоканалах сетевых спутниковых систем с использованием асимметричных алгоритмов // Информационно-управляющие системы. 2006. № 5(24). С. 21–26.
9. **Харин Ю. С., Берник В. И., Матвеев Г. В., Агиевич С. В.** Математические и компьютерные основы криптологии. — Мн.: Новое знание, 2003. — 382 с.
10. **Штанько С. В., Жукова Н. А.** Схемы аутентификации данных и пользователей в распределенных информационных системах // Известия СПбГЭТУ «ЛЭТИ». 2012. № 8. С. 46–51.
11. ГОСТ Р 34.11-2012. Информационная технология. Криптографическая защита информации. Функция хэширования. — М.: Стандартинформ, 2012. — 29 с.

12. ГОСТ Р 34.10-2012. Информационная технология. Криптографическая защита информации. Процессы формирования и проверки электронной цифровой подписи. — М.: Стандартинформ, 2013. — 18 с.
13. **Мальцев Г. Н., Панкратов А. В., Лесняк Д. А.** Исследование вероятностных характеристик изменения защищенности информационной системы от несанкционированного доступа нарушителей // Информационно-управляющие системы. 2015. № 1(74). С. 50–59. doi:10.15217/issn1684-8853.2015.1.50

UDC 621.396.9
doi:10.15217/issn1684-8853.2016.1.74

Providing Selective Access for Broadcasting Information Transfer

Shtanko S. V.[a], PhD., Tech., Associate Professor, craft2001@mail.ru
Lesniak D. A.[a], PhD., Tech., Lecturer, denislesnyk@mail.ru
[a]A. F. Mozhaiskiy Military Space Academy, 13, Zhdanovskaia St., 197082, Saint-Petersburg, Russian Federation.

Introduction: One of the main conditions for effective functioning of any information transfer system is its information security. An objective property of information system security is its gradual decrease at an invariable structure of the protection means, or when their parameters are fixed. To maintain the information security and prevent unauthorized access, the outdated keys in the cryptosystem are periodically replaced. In broadcasting systems of information transfer with selective access, there are features of ensuring information security related to the lack of a return channel of information transfer. Therefore we need to develop special methods of cryptographic protection for broadcasting information systems. **Purpose:** The goal is to improve the existing cryptographic system of data protection in order to increase the security of information transfer broadcasting channels with selective access. **Results:** A modification of data protection methods is proposed which should prevent unauthorized access to the system under conflict conditions of interaction with the violator. The modification is based on asymmetrical cryptographic systems with the mathematics of elliptic curves. On the base of the proposed methods, a cryptographic system is built to protect the data transferred via the channels of a broadcasting information system with selective access. An example of generating a common session key is given for the closed operation mode. **Practical relevance:** The proposed ways of improving the information transfer protection from unauthorized access can increase the information security of the system on the basis of using combined cryptosystems and mathematics of elliptic curves.

Keywords — Broadcasting Systems of Information Transfer with Selective Access, Unauthorized Access, Cryptoalgorithm, Key Management Protocol.

References

1. Maliuk A. A. *Informatsionnaia bezopasnost': kontseptual'nye i metodologicheskie osnovy zashchity informatsii* [Information Security: Conceptual and Methodological Bases of Information Security]. Moscow, Goriachaia liniia–Telecom Publ., 2004. 280 p. (In Russian).
2. Stallings W. *Cryptography and Network Security Principles and Practices*. New Jersey, Prentice Hall, 2000. 592 p.
3. Schneier B. Applied *Cryptography. Protocols, Algorthms and Source Code in C*. New York, John Wiley, 1996. 1027 p.
4. Rostovtsev A. G., Makhovenko E. B. *Vvedenie v kriptografiiu s otkrytym kliuchom* [Introduction to Cryptography With an Open Key]. Saint-Petersburg, Mir i Sem'ia Publ., 2001. 336 p. (In Russian).
5. Ivanov M. A. *Kriptograficheskie metody zashchity informatsii v komp'iuternykh sistemakh i setiakh* [Cryptographic Methods of Information Security in Computer Systems and Networks]. Moscow, Kudits-Obraz Publ., 2001. 368 p. (In Russian).
6. Korniyenko A. A., Yeremeyev M. A., Adadurov S. E. *Sredstva zashchity infor-matsii na zheleznodorozhnom transporte: Kriptograficheskie metody i sredstva* [Information Means of Protection on Railway Transport: Cryptographic Methods and Means]. Moscow, Marshrut Publ., 2006. 256 p. (In Russian).
7. Maltsev G. N., Shtanko S. V. Protocols of Authentication of Subscribers and Information Security on the Basis of Asymmetric Cryptoalgorithms. *Problemy informatsionnoi bezopasnosti. Komp'iuternye sistemy,* 2003, no. 1, pp. 51–56 (In Russian).
8. Kornienko A. A., Shtanko S. V. Information Security Protocol for Network Satellite Systems Using Asymmetric Algorithms. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2006, no. 5(24), pp. 21–26 (In Russian).
9. Kharin Iu. S., Bernik V. I., Matveev G. V., Agievich S. V. *Matematicheskie i komp'iuternye osnovy kriptologii* [Mathematical and Computer Fundamentals of Cryptology]. Minsk, Novoe znanie Publ., 2003. 382 p. (In Russian).
10. Shtanko S. V., Zhukova N. A. Schemes of Authentication of Data and Users in the Distributed Information Systems. *Izvestiia SPbGETU «LETI»*, 2012, no. 8, pp. 46–51 (In Russian).
11. State Standard R 34.11-2012. Information Technology. Cryptographic Information Security. Hashing Function. Moscow, Standartinform Publ., 2012. 29 p. (In Russian).
12. State Standard R 34.10-2012. Information Technology. Cryptographic Information Security. Processes of Formation and Check of a Digital Signature. Moscow, Standartinform Publ., 2013. 18 p. (In Russian).
13. Maltsev G. N., Pankratov A. V., Lesniak D. A. A Probabilistic Characteristics of Information System Security Changes under Unauthorized Access. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2015, no. 1(74), pp. 50–59 (In Russian). doi:10.15217/issn1684-8853.2015.1.50

УДК 621.391.251
doi:10.15217/issn1684-8853.2016.1.80

ТОЧНАЯ КОРРЕКТИРУЮЩАЯ СПОСОБНОСТЬ КОДОВ ГИЛБЕРТА ПРИ ИСПРАВЛЕНИИ ПАКЕТОВ ОШИБОК

Е. А. Крук[a]*, доктор техн. наук, профессор*
А. А. Овчинников[a]*, канд. техн. наук, доцент*
[a]*Санкт-Петербургский государственный университет аэрокосмического приборостроения, Санкт-Петербург, РФ*

Постановка проблемы: *для решения задачи повышения скорости обмена информацией в системах передачи и хранения данных требуется разрабатывать эффективные методы борьбы с помехами, возникающими при передаче, хранении и обработке информации, при использовании как можно меньшей избыточности. Для этого необходимо строить схемы кодирования, ориентированные на ошибки, характерные для конкретного канала связи. Большинство реальных каналов связи являются каналами с группированием ошибок, и типичная помеха в таком канале может описываться как пакет ошибок. Коды Гилберта, обладающие простыми процедурами кодирования и декодирования, хотя и имеют малое минимальное расстояние и неэффективны для исправления независимых ошибок, тем не менее могут быть использованы для исправления пакетов ошибок. Однако корректирующая способность этих кодов при исправлении пакетов ошибок оценивается лишь с помощью не всегда точных границ.* ***Цель:*** *получение точных значений максимальных длин исправляемых пакетов для кодов Гилберта в зависимости от параметров их конструкции.* ***Результаты:*** *разработана процедура, позволяющая для заданных параметров конструкции кодов Гилберта вычислять точное значение максимальной длины исправляемого пакета. В основе построения этой процедуры лежит анализ структуры проверочной матрицы кодов Гилберта, а также структуры пакетов, которые не могут быть исправлены при декодировании с помощью проверочной матрицы. Новизна результата заключается в том, что при любых параметрах конструкции процедура позволяет вычислять точное значение корректирующей способности соответствующего кода Гилберта.* ***Практическая значимость:*** *полученные точные значения корректирующей способности кодов Гилберта могут применяться для аналитических оценок вероятностей ошибок в каналах с памятью при использовании этих кодов, а также могут учитываться при выборе более эффективных схем кодирования в системах передачи и хранения данных.*

Ключевые слова *— каналы с группированием ошибок, коды Гилберта, коды с малой плотностью проверок на четность, исправление пакетов ошибок.*

Введение

Проектирование современных систем связи в подавляющем большинстве практических случаев основывается на моделях каналов связи (двоичного симметричного канала, канала с аддитивным белым гауссовым шумом), которые зачастую не являются адекватными реальным каналам связи, так как в этих моделях искажения отдельных передаваемых символов рассматриваются как статистически независимые события (ошибки). В то же время в реальных каналах связи присутствует эффект «памяти» (например, вследствие явлений замирания и рассеяния [1]), который может быть описан с помощью статистически зависимых ошибок. Для борьбы с этими явлениями в практических системах обычно используется процедура перемежения (интерливинга) [1].

Применение интерливинга обосновывается тем, что в рамках классической теории кодирования обычно строятся кодовые конструкции, ориентированные на исправление независимых ошибок, так как такие конструкции проще анализировать. Однако использование интерливера ведет к искажению реального влияния канала на передаваемые данные, канал трансформируется в канал без памяти, что уменьшает потенциально достижимые скорости передачи, а также увеличивает сложность и задержку как на передающей, так и на принимающей стороне [2, 3]. Таким образом, важной задачей является построение схем кодирования, ориентированных на типичные ошибки канала связи, в частности, на исправление пакетов ошибок (т. е. комбинаций ошибок, при которых первый и последний ошибочный символ некоторой принятой последовательности отстоят друг от друга не более чем на некоторое количество символов, называемое длиной пакета). Отметим, что эффект пакетирования ошибок не только часто встречается в современных каналах связи, но типичен также для систем хранения данных.

Из теории кодирования известны коды, исправляющие пакеты ошибок. Например, такими кодами являются коды Файра, к ним также можно отнести коды Рида — Соломона [4]. В последние десятилетия значительное внимание уделялось кодам с малой плотностью проверок на четность, в частности, их блочно-перестановочным конструкциям [5]. Простейшим частным случаем такой конструкции являются коды Гилберта [6], изначально предлагавшиеся именно для исправления пакетов ошибок. Однако точная корректирующая способность этих кодов при исправлении пакетов ошибок (т. е. максимальная длина исправляемого пакета) была неизвестна.

Процедура для вычисления корректирующей способности кодов Гилберта

Коды с малой плотностью проверок на четность (или LDPC-коды, low-density parity-check) были предложены Р. Галлагером [6, 7] и впоследствии исследовались во многих работах [8–11]. Обладая сравнительно плохим минимальным расстоянием, эти коды, тем не менее, обеспечивают крайне низкую вероятность ошибки при малой сложности декодирования. Было показано, что LDPC-коды могут превосходить турбокоды и приближаться к пропускной способности канала [12]. Кроме того, некоторые конструкции LDPC-кодов, особенно основанные на блочно-перестановочной конструкции, задают также циклические или квазициклические коды, что позволяет еще более эффективную реализацию кодеров и декодеров [3, 5, 13].

Простейшим частным случаем таких кодов являются коды Гилберта [14], которые могут быть описаны своей проверочной матрицей $\mathbf{H}_l$:

$$\mathbf{H}_l=\begin{bmatrix}\mathbf{I}_m & \mathbf{I}_m & \mathbf{I}_m & \dots & \mathbf{I}_m\\ \mathbf{I}_m & \mathbf{C} & \mathbf{C}^2 & \dots & \mathbf{C}^{l-1}\end{bmatrix}, \qquad (1)$$

где $\mathbf{I}_m$ — единичная ($m\times m$)-матрица; $\mathbf{C}$ — ($m\times m$)-матрица циклической перестановки:

$$\mathbf{C}=\begin{bmatrix}0 & 0 & 0 & \dots & 0 & 1\\ 1 & 0 & 0 & \dots & 0 & 0\\ 0 & 1 & 0 & \dots & 0 & 0\\ \dots & \dots & \dots & \dots & \dots & \dots\\ 0 & 0 & 0 & \dots & 1 & 0\end{bmatrix},$$

и $l\le m$.

Оценке корректирующей способности кодов Гилберта (а также их модификаций) при исправлении пакетов ошибок был посвящен ряд работ [15–18]. В статье [19] была получена оценка максимальной длины b исправляемого пакета для кодов, задаваемых матрицей (1):

$$b\le\min_{\gamma\in\{0,\ell-2\}}\max\{\gamma-1,m-\gamma-1\}. \qquad (2)$$

Однако оценка (2) является лишь верхней границей, не всегда выполняющейся с равенством. Точность этой границы ухудшается с ростом l.

Основной результат данной статьи заключается в следующей теореме.

Теорема. Коды с проверочной матрицей $\mathbf{H}_l$, задаваемые (1), могут исправлять одиночные пакеты ошибок максимальной длины b_l, где b_l вычисляется по первому из выполнившихся условий.

1. $b_3=m-1$, где m четно.
2. Если $l>\lceil m/2\rceil+1$, то

$$\begin{cases}b_l=m-\lceil m/2\rceil+1, & m \text{ нечетно;}\\ b_l=m/2-1, & m \text{ четно.}\end{cases}$$

3. Если $l\le\lceil m/2\rceil+1$, то
$b_l=m-l+1$, если $m\,\vdots\,(l-1)$;
$b_l=m-l+1$, если $\exists k>0$: $(m-l+3-k(l-1))\,\vdots\,(l-2)$;
$b_l=m-l+2$, если $\exists k>0$: $(m-k(l-3))\,\vdots\,(l-2)$;
$b_l=m-l+2$, если $\exists k>0$: $(m-k(l-2))\,\vdots\,(l-1)$;
$b_l=m-l+2$, если $\exists k>0$: $(m-k(l-1))\,\vdots\,(l-2)$.
4. Если ни одно из условий не выполнено, то $b_l=b_{l-1}$.

Для доказательства теоремы введем некоторые обозначения и докажем несколько лемм.

Представим матрицу (1) как $\mathbf{H}_l=[\mathbf{h}_0, \mathbf{h}_1, \dots, \mathbf{h}_{l-1}]$, где $\mathbf{h}_i$ — ($2m\times m$)-блоковый столбец, который мы будем называть блоком.

Код может исправлять одиночные пакеты ошибок длиной b, если и только если все пакеты длиной b находятся в разных смежных классах, т. е. не существует двух векторов ошибок $\mathbf{e}_1$ и $\mathbf{e}_2$ (образующих пакеты длиной, не превышающей b), для которых

$$\mathbf{e}_1\mathbf{H}_l^{\mathrm{T}}=\mathbf{e}_2\mathbf{H}_l^{\mathrm{T}}. \qquad (3)$$

Так как $\mathbf{e}_1$ и $\mathbf{e}_2$ — пакеты ошибок, анализ (3) может быть сведен к рассмотрению подматриц $\mathbf{H}_l$, состоящих из $2m$ строк и не более чем b столбцов. Очевидно, код с проверочной матрицей (1) не может исправлять пакеты ошибок длиной m, так как сумма всех столбцов любого из блоков матрицы $\mathbf{H}_l$ дает единичный столбец.

Каждый пакет $\mathbf{B}$ длиной $b<m$ затрагивает не более чем два соседних блока из $\mathbf{H}_l$, скажем, $\mathbf{h}_\gamma$ и $\mathbf{h}_{\gamma+1}$. Заменим в этих блоках все столбцы, не вошедшие в $\mathbf{B}$, нулями и сложим получившиеся матрицы $\mathbf{h}'_\gamma$ и $\mathbf{h}'_{\gamma+1}$, получив матрицу

$$Q(\mathbf{B})=(\mathbf{h}'_\gamma+\mathbf{h}'_{\gamma+1})^{\mathrm{T}}. \qquad (4)$$

Эта матрица имеет вид

$$Q^{\mathrm{T}}(\mathbf{B})=\begin{array}{cc} & \begin{array}{cccc:c:ccc} & & & & i & i+1 & & \end{array}\\ \begin{array}{c} \\ \\ \\ \\ i \\ \\ \\ \\ \\ \\ \\ \end{array} & \left[\begin{array}{cccc:c:ccc} 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0\\ 0 & 1 & \dots & 0 & 0 & 0 & \dots & 0\\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots\\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & 0\\ \hdashline 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0\\ \hdashline 0 & 0 & \dots & 0 & 0 & 1 & \dots & 0\\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots\\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 1\\ \hline & & & & 0 & & & \\ \multicolumn{4}{c:}{\tilde{\mathbf{C}}^{\gamma+1}} & \dots & \multicolumn{3}{c}{\tilde{\mathbf{C}}^{\gamma}}\\ & & & & 0 & & & \end{array}\right]\end{array},$$

где $\tilde{\mathbf{C}}^\gamma$ и $\tilde{\mathbf{C}}^{\gamma+1}$ — части матриц $\mathbf{C}^\gamma$ и $\mathbf{C}^{\gamma+1}$, а $i+1$ — начальная позиция (по модулю m) пакета $\mathbf{B}$, которую мы будем называть относительным началом $\mathbf{B}$.

Из (3) и (4) следует, что для того чтобы код $\mathbf{H}_l$ не мог исправлять пакеты ошибок длиной $b<m$, необходимо и достаточно, чтобы хотя бы для

одной пары пакетов $\mathbf{B}_1$ и $\mathbf{B}_2$ длиной b существовала пара векторов $\mathbf{x}_1$ и $\mathbf{x}_2$ длиной m таких, что

$$\mathbf{x}_1 Q(\mathbf{B}_1) = \mathbf{x}_2 Q(\mathbf{B}_2). \quad (5)$$

Для любого пакета $\mathbf{B}$ матрица $Q(\mathbf{B})$ может быть представлена как $Q(\mathbf{B}) = [Q_1(\mathbf{B}), Q_2(\mathbf{B})]$, где $Q_1(\mathbf{B})$ — единичная ($m \times m$)-матрица, у которой i-я строка заменена на нулевую. Тогда условие (5) может быть записано как

$$\mathbf{x}_1 Q_1(\mathbf{B}_1) = \mathbf{x}_2 Q_1(\mathbf{B}_2); \quad (6)$$

$$\mathbf{x}_1 Q_2(\mathbf{B}_1) = \mathbf{x}_2 Q_2(\mathbf{B}_2). \quad (7)$$

Удалив нулевую строку в $Q_1(\mathbf{B})$, мы получим ($(m-1)\times 2m$)-матрицу $Q'(\mathbf{B}) = [Q_1'(\mathbf{B}), Q_2'(\mathbf{B})]$.

Обозначим через $(\mathbf{y}\backslash s)$ вектор длиной $m-1$, полученный из $\mathbf{y}$ (длиной m) удалением s-й позиции. Тогда можно сформулировать следующую лемму.

Лемма 1. Условие (5) выполняется, если и только если существует вектор $\mathbf{y}$ длиной m такой, что

$$\begin{cases}(\mathbf{y} \setminus i)Q_2'(\mathbf{B}_1) = (\mathbf{y} \setminus j)Q_2'(\mathbf{B}_2);\\ y_i = y_j = 0,\end{cases}$$

где $i+1$, $j+1$ — относительные начала пакетов $\mathbf{B}_1$ и $\mathbf{B}_2$ соответственно.

Рассмотрим выражение (6). Матрицы $Q_1(\mathbf{B}_1)$ и $Q_1(\mathbf{B}_2)$ являются единичными ($m \times m$)-матрицами с нулевой i-й и j-й строкой соответственно. Это означает, что $\mathbf{x}_1 Q_1(\mathbf{B}_1)$ дает вектор $\mathbf{x}_1$ с нулевой i-й позицией, т. е. $x_i = 0$. Тогда из (6) следует, что векторы $\mathbf{x}_1$ и $\mathbf{x}_2$ совпадают и содержат нули на позициях i и j. То есть всегда может быть определен вектор $\mathbf{y} = \mathbf{x}_1 = \mathbf{x}_2$, у которого $y_i = y_j = 0$.

Перепишем (6) и (7) как

$$\begin{cases}(\mathbf{y} \setminus i)Q_1'(\mathbf{B}_1) = (\mathbf{y} \setminus j)Q_1'(\mathbf{B}_2);\\ (\mathbf{y} \setminus i)Q_2'(\mathbf{B}_1) = (\mathbf{y} \setminus j)Q_2'(\mathbf{B}_2).\end{cases}$$

Очевидно, первое уравнение выполняется для любого $\mathbf{y}$, у которого $y_i = y_j = 0$, что дает утверждение леммы. Лемма доказана.

Лемма 2. Для любого пакета $\mathbf{B}$

$$Q_2'(\mathbf{B}) = [\mathbf{I}_{m-1}, \mathbf{0}]\mathbf{C}^{\gamma},$$

где $\mathbf{0}$ — нулевой вектор-столбец длиной $m-1$, а γ — целое число от 0 до $l-1$.

Для доказательства леммы достаточно показать, что для любого пакета $\mathbf{B}$ соответствующая матрица $Q_2'(\mathbf{B})$ является циркулянтной ($(m-1) \times m$)-матрицей.

Пусть $\mathbf{B}$ — пакет длиной $m-1$, у которого первые $s \geq 0$ позиций приходятся на блок $\mathbf{h}_i$, а следующие $m-s-1$ — на блок $\mathbf{h}_{i+1}$. Тогда $Q_2'(\mathbf{B}) = [q_1(\mathbf{B}), q_2(\mathbf{B})]^{\mathrm{T}}$, где $q_1(\mathbf{B})$ — матрица из первых $m-s-1$ столбцов $\mathbf{C}^i$, а $q_2(\mathbf{B})$ — матрица из последних s столбцов $\mathbf{C}^{i-1}$. Матрицы $q_1(\mathbf{B})$ и $q_2(\mathbf{B})$ циркулянтны по построению, тогда осталось показать, что первый столбец $q_1(\mathbf{B})$ является циклическим сдвигом последнего столбца $q_2(\mathbf{B})$. Это следует из того факта, что $(m-s)$-й столбец $\mathbf{C}^{i-1}$ равен $(m-s-1)$-му столбцу $\mathbf{C}^i$. Лемма доказана.

Из лемм 1, 2 и условия (5) следует, что два пакета длиной $b < m$ имеют одинаковый синдром, если и только если существует вектор $\mathbf{y}$ длиной m и некоторое γ, для которых

$$\begin{cases}(\mathbf{y} \setminus i, 0) = (\mathbf{y} \setminus j, 0)\mathbf{C}^{\gamma};\\ y_i = y_j = 0,\end{cases} \quad (8)$$

где $(\mathbf{y}\backslash i,\ 0)$ — вектор длиной m, полученный из $(\mathbf{y}\backslash i)$ дописыванием нулевого элемента.

Поставим в соответствие вектору $\mathbf{y}$ многочлен $y(x) = \sum_{k=0}^{m-1} y_k x^k$. Тогда можно записать (8) в полиномиальном представлении

$$\begin{cases}(y(x) \setminus i) = (y(x) \setminus j)x^{\gamma} \bmod x^m - 1;\\ y_i = y_j = 0,\end{cases} \quad (9)$$

(коэффициенты $(y(x)\backslash i)$ определяются вектором $(\mathbf{y}\backslash i,0)$).

Лемма 3. Если $y(x)$ — многочлен, удовлетворяющий (9), то для любого ненулевого y_k выполняется одно из условий

$$y_k = y_{(k-\gamma) \bmod m}, \quad y_k = y_{(k-\gamma+1) \bmod m},$$

$$y_k = y_{(k-\gamma-1) \bmod m}.$$

Для доказательства отметим, что многочлен $(y(x)\backslash i)$ может быть представлен как

$$(y(x) \setminus i) = \sum_{k=0}^{i-1} y_k x^k + \sum_{k=i}^{m-2} y_{k+1} x^k = A^i(x) + B^i(x). \quad (10)$$

Тогда $((y(x)\backslash j)x^{\gamma}) \bmod x^m - 1$ можно записать как

$$\begin{aligned}&\left((y(x) \setminus j)x^{\gamma}\right) \bmod x^m - 1 =\\ &= \left(\sum_{k=0}^{j-1} y_k x^{k+\gamma} + \sum_{k=j}^{m-2} y_{k+1} x^{k+\gamma}\right) \bmod x^m - 1 =\\ &= \left(A^{\gamma}(x) + B^{\gamma}(x)\right) \bmod x^m - 1.\end{aligned} \quad (11)$$

Из (9) следует равенство коэффициентов при одинаковых степенях у многочленов (10) и (11).

Рассмотрим коэффициенты при степенях $k = \overline{0, i-1}$ (т. е. принадлежащие $A^i(x)$). Они соответствуют коэффициентам при тех же степенях либо из $A^{\gamma}(x)$, либо из $B^{\gamma}(x)$. Однако из (11) следует, что любая степень k из $((y(x)\backslash j)x^{\gamma}) \bmod x^m - 1$ может быть представлена как $(s+\gamma) \bmod m$. Тогда

$$y_k \in A^i(x) = y_k \in A^{\gamma}(x) = y_{(k-\gamma) \bmod m} \text{ или}$$

$$y_k \in A^i(x) = y_k \in B^{\gamma}(x) = y_{(k-\gamma+1) \bmod m}. \quad (12)$$

Аналогично для степеней $k=\overline{i, m-2}$

$$y_k \in B^i(x)=y_{k-1} \in A^\gamma(x)=y_{(k-\gamma) \bmod m} \text{ или}$$

$$y_k \in B^i(x)=y_{k-1} \in B^\gamma(x)=y_{(k-\gamma-1) \bmod m}. \qquad (13)$$

Из (12) и (13) следует утверждение леммы.

Следствие 1. Для любого $y(x)$, удовлетворяющего (9), существуют неотрицательные a_1, a_2, a_3 такие, что

$$a_1(\gamma+1)+a_2\gamma+a_3(\gamma-1)=m. \qquad (14)$$

Действительно, пусть y_k — ненулевой элемент вектора **y**. Тогда, в соответствии с леммой 2, один из элементов $y_{(k-\gamma)\bmod m}$, $y_{(k-\gamma+1)\bmod m}$ или $y_{(k-\gamma-1)\bmod m}$ также является ненулевым. Продолжая таким образом, мы циклически «вернемся» в y_k, и может быть найден ненулевой y_s, для которого выполняется одно из следующих условий:

$$y_s=y_{(s-\gamma) \bmod m}=y_k;$$

$$y_s=y_{(s-\gamma+1) \bmod m}=y_k;$$

$$y_s=y_{(s-\gamma-1) \bmod m}=y_k. \qquad (15)$$

Таким образом, существуют такие неотрицательные a_1, a_2, a_3, для которых

$$a_1(\gamma+1)+a_2\gamma+a_3(\gamma-1)=0 \bmod m. \qquad (16)$$

Для доказательства следствия 1 достаточно показать, что возвращение в y_k происходит не более чем за один проход по вектору **y**.

Если это не так, то либо **y** состоит из всех единиц и представляет собой неисправляемый пакет длиной m, либо в нем есть нулевые элементы, и тогда найдется γ', возможно, меньшее, чем γ из (16), для которого также выполняется условие (9), и до момента выполнения (15) совершается лишь один проход по вектору **y**, что завершает доказательство.

Следствие 1 также означает, что из всех γ, для которых выполняется (9), необходимо выбирать наименьшее ненулевое значение.

Назовем секцией вектора **y** последовательность, состоящую из единицы и последующих нулей. Например, **y** = 100010000 содержит две секции длиной 4 и 5.

Из следствия 1 следует, что **y** является конкатенацией секции длиной γ, $\gamma+1$ и $\gamma-1$. Следующие лемма и следствие уточняют положение этих секций.

Лемма 4. Для вектора, удовлетворяющего (9), справедливо следующее неравенство

$$j+\gamma-1 \le m-1. \qquad (17)$$

В самом деле, пусть y_{j+1} — последний ненулевой элемент **y** [в противном случае (17) выполняется очевидным образом]. Вспомним, что элемент y_{j+1} равен единице, так как j — относительное начало пакета ошибок. Предположим, что коэффициент y_0 в $y(x)$ всегда равен единице. Если это не так, соответствующий вектор **y** может быть циклически сдвинут с сохранением всех результатов (9), леммы 3 и следствия 1.

Тогда из леммы 3 и следствия 1 возможны следующие случаи:

$$y_0=y_{m \bmod m}=y_{(m-\gamma-1) \bmod m}=y_{j+1};$$

$$y_0=y_{m \bmod m}=y_{(m-\gamma+1) \bmod m}=y_{j+1};$$

$$y_0=y_{m \bmod m}=y_{(m-\gamma) \bmod m}=y_{j+1},$$

откуда мы получаем $j+\gamma-1 \le m-1$. Лемма доказана.

Следствие 2. Если $i \le j+\gamma$, то среди секций, формирующих **y**, секция длиной $\gamma-1$ либо отсутствует, либо встречается ровно один раз и расположена последней.

Из леммы 4 и неравенства (17) многочлен (11) может быть представлен как

$$((y(x)\setminus j)\cdot x^\gamma) \bmod x^m-1=\sum_{k=0}^{j-1} y_k x^{k+\gamma}+$$

$$+\sum_{k=j}^{m-\gamma-2} y_{k+1}x^{k+\gamma}+\sum_{k=m-\gamma-1}^{m-2} y_{k+1}x^{(k+\gamma) \bmod m}=$$

$$=\sum_{k=0}^{\gamma-2} y_{(k-\gamma+1) \bmod m}x^k+y_j x^{\gamma-1}+\sum_{k=j+\gamma}^{m-2} y_{k-\gamma+1}x^k=$$

$$=A^m(x)+y_j x^{\gamma-1}+B^m(x)+C^m(x). \qquad (18)$$

Как мы предположили ранее, пусть $y_0=1$. Вначале рассмотрим случай, когда относительное начало обоих пакетов не совпадает с началом какого-либо блока $\mathbf{h}_\gamma$ матрицы (1). Тогда $i+1 \ge \gamma-1$ или $i \ge \gamma-2$. Рассмотрим случаи $i=\gamma-2$ и $i=\gamma-1$. В первом случае из (10) и (18) имеем $y_{i+1}=y_{-1 \bmod m}=y_{m-1}$, что возможно, только если **y** содержит только 1. Если нет других решений (9), то максимальная длина исправляемого пакета составляет $m-1$. Мы исключим такие векторы из рассмотрения.

Во втором случае для $i=\gamma-1$ имеем $y_{i+1}=y_j$, что невозможно, так как элемент y_{i+1} всегда ненулевой, тогда как y_j всегда равен нулю. Отсюда $i \ge \gamma$, т. е. степени от 0 до $\gamma-1$ полностью принадлежат $A^i(x)$, а y_0 является единственным ненулевым коэффициентом из $A^m(x)$ (это следует из того, что минимальное количество позиций между ненулевыми элементами составляет $\gamma-1$, и коэффициент при $x^{\gamma-1}$ равен нулю). Степени, большие $\gamma-1$, принадлежат $B^m(x)$ и $C^m(x)$.

Пусть $i \le j+\gamma$. Тогда из (10) и (18) имем $C^m(x) \subset B^i(x)$, и выполняется одно из следующих равенств:

$$y_k=y_{(k-\gamma+1) \bmod m}, \ 0 \leqslant k<\gamma;$$

$$y_k=y_{k-\gamma}, \ \gamma \leqslant k<i;$$

$$y_k = y_{k-\gamma-1},\ \ i \leqslant k < (j+\gamma);$$

$$y_k = y_{k-\gamma},\ \ (j+\gamma) \leqslant k \leqslant m-2.$$

Таким образом, условие $y_k = y_{(k-\gamma+1) \bmod m}$ может выполняться только для одного ненулевого элемента **y**, а именно y_0. Из этого также следует, что для $i < j + \gamma$ справедливо $a_3 = 1$, и единственная секция длиной $\gamma - 1$ в **y** является последней.

Рассмотрим частный случай $i = j + \gamma$. Тогда $C^m(x) = B^i(x)$, и вектор **y** не содержит секции длиной $\gamma + 1$, он содержит только (последнюю) секцию длиной $\gamma - 1$, а все остальные секции имеют длину γ (другими словами, $a_1 = 0$, $a_3 = 1$). Следствие доказано.

Проанализируем результаты леммы 3 и ее следствий. Пусть у нас есть многочлен $y(x)$, удовлетворяющий (9), и соответствующий вектор **y**. Если мы обозначим через i^0 и j^0 количество нулевых позиций в **y** перед $(i + 1)$-й и $(j + 1)$-й позициями, тогда длина неисправляемого пакета оценивается как $b = \max(\mathrm{m} - i^0, m - j^0)$.

Однако неизвестно, как решать (9), имея только m и l. С другой стороны, известны свойства многочлена $y(x)$, описанные в лемме 3 и следствиях. Очевидно, что, определив коэффициенты a_1, a_2 и a_3 из (14), мы определим многочлен $y(x)$, удовлетворяющий (9). В то же время для некоторого m может существовать несколько решений (14), в этом случае мы должны выбрать худший случай, т. е. многочлены, задающие минимальную длину неисправляемого пакета. Далее мы покажем, что коэффициенты a_1, a_2 и a_3, так же как и минимальная длина неисправляемого пакета, зависят от взаимного расположения относительных начал пакетов i и j.

Рассмотрим случай, когда начала одного или обоих пакетов совпадают с началом соответствующего блока матрицы $\mathbf{H}_l$. Тогда значения $i + 1$ или (и) $j + 1$ должны быть равны нулю, и отсюда $i = -1 \bmod m$ или (и) $j = -1 \bmod m$ (так как из $i = -1 \bmod m$ следует $(\mathbf{y} \backslash i, 0) = \mathbf{y}$).

Следующая лемма связывает параметры m, γ, коэффициенты a_1, a_2 и a_3.

Лемма 5. Для вектора **y** длиной m и параметров i, j, γ, удовлетворяющих (8), значения коэффициентов a_1, a_2, a_3 из (14) определяются в зависимости от значений i и j в соответствии с таблицей.

Рассмотрим последовательно все возможные расположения относительных начал i и j пакетов ошибок, используя полиномиальные представления (10) и (18).

1. Пусть $i \leq j$ и $i = j = -1$ (этот случай соответствует первой строке таблицы). Это означает, что выполняется уравнение $y(x) = (y(x) \backslash i) = (y(x) \backslash j)$, т. е.

$$(y(x) \setminus i) = \sum_{k=0}^{m-1} y_k x^k;$$

$$((y(x) \setminus j) x^{\gamma}) \bmod x^m - 1 = \sum_{k=0}^{m-1} y_k x^{(k+\gamma) \bmod m},$$

отсюда $y_k = y_{(k-\gamma) \bmod m}$ для всех k. С учетом (14) получаем $a_1 = 0$, $a_3 = 0$. Длина неисправляемого пакета в этом случае $b = m - \gamma - 1$.

2. Теперь пусть $i \leq j$, $i \neq -1$, $j \neq -1$. При доказательстве следствия 2 было показано, что если $i \neq -1$, тогда $i \geq \gamma$. Рассмотрим различные возможные значения i и j.

Во-первых, пусть $i = \gamma$. В этом случае из (10) и (18) получаем $B^m(x) \subset B^i(x)$, $C^m(x) \subset B^i(x)$. В многочлене (10) коэффициент y_{j+1} всегда соответствует степени x^j, а степень x^j в многочлене (18) всегда принадлежит $B^m(x)$, поэтому $y_{i+1} = y_0$, $y_{j+1} = y_{j-\gamma}$. То есть перед $(i + 1)$-й и $(j + 1)$-й позициями **y** нет секции длиной $\gamma + 1$, тогда длина неисправляемого пакета составляет $m - \gamma$.

Во-вторых, пусть $(\gamma + 1) \leq i < (j + \gamma)$. В этом случае степень x^i принадлежит $B^m(x)$, $y_{i+1} = y_{i-\gamma}$, $y_{j+1} = y_{j-\gamma}$, что является обобщением предыдущего случая, и длина неисправляемого пакета также составляет $m - \gamma$.

Рассмотренные случаи соответствуют второй строке таблицы.

3. Пусть $i \geq j + \gamma$. Как было показано при доказательстве следствия 2, если $i = j + \gamma$, тогда **y** содержит одну секцию длиной $\gamma - 1$, а другие секции имеют длину γ. Однако, так как ни i, ни j в рассматриваемом случае не равны -1, перед относительными началами пакетов находятся секции длиной γ, поэтому $b = m - \gamma + 1$. Если $i > j + \gamma$, степень x^i принадлежит $C^m(x)$ и $B^m(x) \subset A^i(x)$, поэтому $y_{i+1} = y_{i-\gamma+1}$, $y_{j+1} = y_{j-\gamma+1}$. В этом случае в векторе **y** перед относительными началами пакетов также находятся секции длиной γ, поэтому снова $b = m - \gamma + 1$.

■ Относительные начала пакетов i, j, коэффициенты (14) и минимальные длины b неисправляемых пакетов

№	i, j	b	Коэффициенты	γ
1	$i = -1, j = -1$	$m - \gamma - 1$	$a_1 = 0, a_3 = 0$	$l - 1$
2	$-1 < i < (j+\gamma), j \neq -1$	$m - \gamma$	$a_1 \neq 0, a_3 = 1$	$l - 2$
3	$i \geq (j+\gamma), j \neq -1$	$m - \gamma + 1$	$a_1 = 0, a_3 \neq 0$	$l - 2$
4	$i = -1, j \neq -1$	$m - \gamma + 2$	$a_1 = 0, a_3 \neq 0$	$l - 1$
5	$i \neq -1, j = -1$	$m - \gamma + 1$	$a_1 \neq 0, a_3 = 0$	$l - 2$

Теперь необходимо рассмотреть случаи, когда i или j равны -1.

4. Пусть $i = -1 = m-1$, $j \neq -1$. Тогда (10) и (18) могут быть записаны как

$$(y(x)\setminus i) = \sum_{k=0}^{m-1} y_k x^k = A^i(x);$$

$$\left((y(x)\setminus j)x^{\gamma}\right) \bmod x^m - 1 =$$

$$= A^m(x) + y_j x^{\gamma-1} + B^m(x) + C^m(x).$$

Отсюда при любом k справедливо одно из равенств: $y_k = y_{(k-\gamma) \bmod m}$, $y_k = y_{(k-\gamma+1) \bmod m}$. Другими словами, коэффициент a_1 в (14) равен нулю. Это совпадает с предыдущим случаем, когда $B^m(x) \subset A^i(x)$ и перед позицией $j + 1$ находится $\gamma-1$ нулей. Однако перед позицией $i + 1$ находится $\gamma-2$ нулей, поэтому необходимо выбрать $b = \max(m-\gamma + 1, m-\gamma + 2) = m-\gamma + 2$. Таким образом, в этом случае $b = m-\gamma + 2$, $a_3 \neq 0$, $a_1 = 0$.

5. Пусть $i \neq -1$, $j = -1 = m-1$. Тогда

$$(y(x)\setminus i) = \sum_{k=0}^{i-1} y_k x^k + \sum_{k=i}^{m-2} y_{k+1} x^k = A^i(x) + B^i(x);$$

$$((y(x)\setminus j)x^{\gamma}) \bmod x^m - 1 = \sum_{k=0}^{m-1} y_{(k-\gamma) \bmod m} x^k.$$

Имеем $y_{i+1} = y_{i-\gamma}$, $y_{j+1} = y_0 = y_{(-\gamma) \bmod m}$. Это дает $b = m-\gamma + 1$, $a_1 \neq 0$, $a_3 = 0$.

Рассмотренные случаи завершают доказательство леммы 5.

Если известны значения коэффициентов из (14), можно вычислить минимальную длину неисправляемого пакета. Однако в третьей и четвертой строках таблицы мы имеем одинаковые значения коэффициентов, но при этом различные значения длин пакетов. Тем не менее с учетом значений γ можно видеть, что на самом деле значения длин пакетов для этих случаев совпадают. Проанализируем связь между значениями γ (неизвестного при построении кода) и l, являющимся параметром кода.

Вначале рассмотрим $i \neq -1$, $j = -1 = m-1$ (пятая строка таблицы). Как уже было показано, вектор **y** для таких значений i и j состоит из секций длиной γ и $\gamma + 1$, причем последняя секция имеет длину γ. Так как $i \neq -1$ и $y_0 = 1$, первый пакет ошибок затрагивает блоки $\mathbf{h}_0$ и $\mathbf{h}_1$, а во втором пакете ошибок позиция, соответствующая столбцу в $Q_2(\mathbf{B}_2)$ с единицей в первой строке (считая от нуля) или, эквивалентно, в $(m + 1)$-й строке блока $\mathbf{h}_\gamma$, должна содержать ненулевое значение. Отсюда следует, что номер этой позиции $(0-\gamma) \bmod m = m-\gamma$. Но это возможно, только если второй пакет соответствует блоку $\mathbf{h}_{\gamma+1}$. Тогда верхняя диагональ $Q_2(\mathbf{B}_2)$ содержит $\gamma + 1$ единиц, а номер столбца, содержащего второй ненулевой элемент этой диагонали (т. е. позиция единицы в $(m + 1)$-й строке блока $\mathbf{h}_{\gamma+1}$), равен $m-\gamma$. Таким образом, так как пакет $\mathbf{B}_2$ занимает один блок, номер этого блока – $\gamma + 1$, а не γ, поэтому значение γ не может превышать $l-2$.

Рассмотрим более детально связь между значениями γ и l. Очевидно, что $\gamma \leq l-1$, так как пакет начинается в блоке $\mathbf{h}_\gamma$. Однако пакет может занимать два блока, и необходимо проверить, затрагивается ли пакетом блок с номером, превышающим $l-1$. Пусть для некоторого l код может исправлять пакеты ошибок длиной, не превышающей b_l. Тогда найдутся два неисправимых пакета длиной b_{l+1}, занимающих первый и последний блоки $\mathbf{H}_l$ (предполагается, что второй пакет заканчивается в последнем блоке, но может начинаться в предпоследнем). Если второй пакет не затрагивает последний блок, это означает, что эти пакеты не исправляются кодом с количеством блоков, меньшим l, тогда мы можем рассмотреть этот более короткий код с двумя пакетами в первом и последнем блоках.

Далее, каждый вектор **y**, определяемый неисправляемыми пакетами [см. (3) и (8)] и являющийся решением (8) и (14), имеет соответствующее значение γ, равное либо $l - 1$, либо $l-2$. Рассмотрим код с количеством блоков, увеличенным на единицу. Длина неисправляемого пакета может либо сократиться, либо остаться прежней. В первом случае имеем $\gamma' = \gamma + 1$, во втором $\gamma' = \gamma$. Если для $\gamma' = \gamma + 1$ нет решений уравнения (14) относительно a_1, a_2, a_3, тогда $b_{l+1} = b_l$ (где b_l обозначает максимальную длину исправляемого пакета для кода с параметром l).

Продолжая рассматривать значения γ аналогичным образом, получим значения, приведенные в последнем столбце таблицы.

Подводя итоги проведенного анализа, можно сформулировать следующие условия для вычисления длины пакета ошибок.

1. Если выполняется условие в первой строке таблицы, это значит, что $m \vdots \gamma$ или $m \vdots (l-1)$ для заданных m и l, и код не может исправлять пакеты ошибок длиной $b = m-l + 2$.

2. $\exists k > 0$: $(m - (\gamma-1)-k(\gamma + 1)) \vdots \gamma$ или $\exists k > 0$: $(m - l + 3-k(l-1)) \vdots (l-2)$, тогда $b = m-l + 2$.

3. $\exists k > 0$: $(m - k(\gamma-1)) \vdots \gamma$ или $\exists k > 0$: $(m - k(l-3)) \vdots (l-2)$, тогда $b = m-l + 3$.

4. $\exists k > 0$: $(m - k(\gamma-1)) \vdots \gamma$ или $\exists k > 0$: $(m - k(l-2)) \vdots (l-1)$, тогда $b = m-l + 3$.

5. $\exists k > 0$: $(m - k(\gamma + 1)) \vdots \gamma$ или $\exists k > 0$: $(m - k(l-1)) \vdots (l-2)$, тогда $b = m-l + 3$.

Имея параметры кода m и l, необходимо вначале рассмотреть условия, дающие меньшую длину неисправляемого пакета (т. е. условия 1 и 2). Затем рассматриваются условия 3, 4 и 5. Если нет выполнившихся условий, то неисправляемый пакет не находится в последнем блоке $\mathbf{H}_l$, тогда необходимо уменьшить l и повторить процедуру.

Следующие две леммы завершают доказательство теоремы.

Лемма 6. Если $l > \lceil m/2 \rceil + 1$, то код, определяемый (1), может исправлять одиночные пакеты ошибок длиной $b \le b_l$, где

$$\begin{cases} b_l = m - \lceil m/2 \rceil + 1, & m \text{ нечетно;} \\ b_l = m/2 - 1, & m \text{ четно.} \end{cases} \quad (19)$$

В соответствии с леммой 5 и ее следствиями для каждого ненулевого элемента **y** найдутся два других ненулевых элемента слева и справа от данного, отстоящих на γ, $\gamma + 1$ или $\gamma - 1$ позиций (т. е. ненулевые элементы соседних секций). Если $l > \lceil m/2 \rceil + 1$, то максимальное значение γ может превысить $m/2$, это означает, что количество секций в **y** минимально и равно двум, т. е. пакет ошибок содержит лишь две единицы (начало и конец пакета), переходящие друг в друга при сдвиге на γ позиций.

Дальнейшее увеличение l (превышающее $\lceil m/2 \rceil + 1$) не приведет к уменьшению длины неисправляемого пакета. Очевидно, что в этом случае максимальная длина исправляемого пакета определяется (19).

Лемма 7. Для $l = 3$ и нечетного m код с проверочной матрицей (1) может исправлять одиночные пакеты ошибок длиной $b = m - 1$.

Ясно, что для $l = 3$ пакеты ошибок могут занимать лишь первый и последний блоки $\mathbf{H}_3$. В противном случае γ равнялся бы единице и **y** состоял бы только из единиц. Тогда $\gamma = 2$, что возможно только при $i = -1$, $j = -1$, и в соответствии с таблицей длина пакета составляет $m - \gamma + 1$, что равно $m - 1$ при $\gamma = 2$.

Легко проверить, что для четных m и $l = 3$ всегда можно определить вектор **y**, удовлетворяющий (9) и содержащий единицы на четных позициях и нули на нечетных.

Из результатов, приведенных в таблице, леммах 6 и 7, следует утверждение теоремы.

Заключение

Рассмотрена корректирующая способность кодов Гилберта при исправлении одиночных пакетов ошибок. Сформулирована процедура, позволяющая вычислить точное значение максимальной длины исправляемого пакета в зависимости от параметров кода. Можно отметить, что обобщение кодов Гилберта путем добавления дополнительных строк-блоков к проверочной матрице может улучшить корректирующую способность кода, однако для любого блочно-перестановочного LDPC-кода эта величина будет меньше размера блока.

Литература

1. **Proakis J., Salehi M.** Digital Communications. — McGraw-Hill, 2007. — 1150 p.
2. **Krouk E. A., Ovchinnikov A. A.** Metrics for Distributed Systems // Problems of Redundancy in Information and Control Systems (REDUNDANCY): XIV Intern. Symp., 1–5 June 2014. P. 66–70. doi:10.1109/RED.2014.7016708
3. **Krouk E., Semenov S.**, et al. Modulation and Coding Techniques in Wireless Communications/Edited by E. Krouk, S. Semenov. — John Wiley & Sons, 2011. — 680 p. doi:10.1002/9780470976777
4. **MacWilliams F., Sloane N.** The Theory of Error-Correcting Codes. — North-Holland publishing company, 1983. — 782 p.
5. **Lin S., Ryan W.** Channel Codes: Classical and Modern. — Cambridge university press, 2009. — 710 p.
6. **Gallager R. G.** Low-Density Parity-Check Codes // IRE Transactions on Information Theory. Jan. 1962. Vol. 8. N 1. P. 21–28. doi:10.1109/TIT.1962.1057683
7. **Gallager R. G.** Low Density Parity Check Codes. — Cambridge, MA: MIT Press, 1963. — 90 p.
8. **MacKay D. J. C.** Good Error-Correcting Codes Based on Very Sparse Matrices // IEEE Transactions on Information Theory. Mar. 1999. Vol. 45. N 2. P. 399–431. doi:10.1109/18.748992
9. **MacKay D. J. C., Neal R. M.** Near Shannon Limit Performance of Low Density Parity Check Codes // Electronics Letters. 13 Mar. 1997. Vol. 33. N 6. P. 457–458. doi:10.1049/el:19970362
10. **Richardson T. J., Urbanke R. L.** The Capacity of Low-Density Parity-Check Codes Under Message-Passing Decoding // IEEE Transactions on Information Theory. Feb. 2001. Vol. 47. N 2. P. 599–618. doi:10.1109/18.910577
11. **Zyablov V. V., Pinsker M.S.** Estimation of the Error-Correction Complexity for Gallager Low-Density Codes // Problems of Information Transmission. 1975. Vol. 11. N 1. P. 23–36.
12. **Chung S.-Y., Forney G. D., Richardson T. J., Urbanke R.** On the Design of Low-Density Parity-Check Codes within 0.0045 dB of the Shannon Limit // Communications Letters. IEEE. Feb. 2001. Vol. 5. N 2. P. 58–60. doi:10.1109/4234.905935
13. **Козлов А. В., Крук Е. А., Овчинников А. А.** Подход к построению блочно-перестановочных кодов с малой плотностью проверок на четность // Изв. вузов. Приборостроение. 2013. Т. 56. № 8. С. 9–14.
14. **Gilbert E. N.** A Problem in Binary Encoding // Proc. of the Symp. in Applied Mathematics. 1960. Vol. 10. P. 291–297.
15. **Arazi B.** The Optimal Burst Error-Correcting Capability of the Codes Generated by f(x)=(x^p+1)(x^q+1)/

(x+1) // Information and Control. 1978. Vol. 39. N 3. P. 303–314.
16. **Bahl L., Chien R.** On Gilbert Burst-Error-Correcting Codes // IEEE Transactions on Information Theory. May 1969. Vol. 15. N 3. P. 431–433. doi:10.1109/TIT.1969.1054310
17. **Neumann P. G.** A Note on Gilbert Burst-Correcting Codes // IEEE Transactions on Information Theory. Jul. 1965. Vol. 11. N 3. P. 377–384. doi:10.1109/TIT.1965.1053780
18. **Zhang W., Wolf J. K.** A Class of Binary Burst Error-Correcting Quasi-Cyclic Codes // IEEE Transactions on Information Theory. May 1988. Vol. 34. N 3. P. 463–479. doi:10.1109/18.6026
19. **Krouk E. A., Semenov S. V.** Low-Density Parity-Check Burst Error-Correcting Codes // Algebraic and Combinatorial Coding Theory: International Workshop. Leningrad, 1990. P. 121–124.

UDC 621.391.251
doi:10.15217/issn1684-8853.2016.1.80

Exact Burst-Correction Capability of Gilbert Codes

Krouk E. A.[a], Dr. Sc., Tech., Professor, ekrouk@vu.spb.ru
Ovchinnikov A. A.[a], PhD, Tech., Associate Professor, mldoc@vu.spb.ru
[a]Saint-Petersburg State University of Aerospace Instrumentation, 67, B. Morskaia St., 190000, Saint-Petersburg, Russian Federation

Introduction: In order to increase the data exchange rates in data transmission and storage systems, you have to efficiently reduce the noise which comes out during the transmission, storing and processing of the data, using as small redundancy as possible. The coding schemes should consider typical errors, most common in the given communication channel. Most real communications channels are channels with memory, and a typical noise can be described as an error burst. Gilbert codes with simple encoding and decoding procedures have a poor minimal distance, being ineffective for correcting independent errors; nevertheless, they can be used for bursts correction. However, the burst-correcting capability of these codes can be estimated only by bounds which are not always accurate. **Purpose:** The goal is to obtain exact values for the maximal lengths of error bursts correctable by Gilbert codes depending on their construction parameters. **Results:** A procedure is developed which allows you to calculate the exact value for the maximal length of a single burst correctable by Gilbert code, depending on the code's construction parameters. To build this procedure, we analyzed the structure of the parity-check matrix of Gilbert codes and the structure of the bursts which cannot be corrected by parity-check matrix decoding. The novelty of this result is that the procedure allows you to calculate the exact burst-correction capability for any parameters of the Gilbert code. **Practical relevance:** The obtained accurate values for Gilbert code burst-correction capability can be used for analytical estimation of error probabilities in channels with memory when using Gilbert codes. They also can be taken into account to choose more effective coding schemes in communication and storage systems.

Keywords — Channels with Memory, Gilbert Codes, Low-Density Parity-Check Codes, Burst Error Correction.

References

1. Proakis J., Salehi M. *Digital Communications*. McGraw-Hill, 2007. 1150 p.
2. Krouk E. A., Ovchinnikov A. A. Metrics for Distributed Systems. *XIV Intern. Symp. "Problems of Redundancy in Information and Control Systems" (REDUNDANCY)*, 1–5 June 2014, pp. 66–70. doi:10.1109/RED.2014.7016708
3. Krouk E., Semenov S., et al. *Modulation and Coding Techniques in Wireless Communications*. Edited by E. Krouk, S. Semenov. John Wiley & Sons, 2011. 680 p. doi:10.1002/9780470976777
4. MacWilliams F., Sloane N. *The Theory of Error-Correcting Codes*. North-Holland publishing company, 1983. 782 p.
5. Lin S., Ryan W. *Channel Codes: Classical and Modern*. Cambridge university press, 2009. 710 p.
6. Gallager R. G. Low-Density Parity-Check Codes. *IRE Transactions on Information Theory*, January 1962, vol. 8, no. 1, pp. 21–28. doi:10.1109/TIT.1962.1057683
7. Gallager R. G. *Low Density Parity Check Codes*. Cambridge, MA, MIT Press, 1963. 90 p.
8. MacKay D. J. C. Good Error-Correcting Codes Based on Very Sparse Matrices. *IEEE Transactions on Information Theory*, Mar. 1999, vol. 45, no. 2, pp. 399–431. doi:10.1109/18.748992
9. MacKay D. J. C., Neal R. M. Near Shannon Limit Performance of Low Density Parity Check Codes. *Electronics Letters*, 13 Mar. 1997, vol. 33, no. 6, pp. 457–458. doi:10.1049/el:19970362
10. Richardson T. J., Urbanke R. L. The Capacity of Low-Density Parity-Check Codes under Message-Passing Decoding. *IEEE Transactions on Information Theory*, Feb. 2001, vol. 47, no. 2, pp. 599–618. doi:10.1109/18.910577
11. Zyablov V. V., Pinsker M. S. Estimation of the Error-Correction Complexity for Gallager Low-Density Codes. *Problems of Information Transmission*, 1975, vol. 11, no. 1, pp. 23–36.
12. Chung S.-Y., Forney G. D., Richardson T. J., Urbanke R. On the Design of Low-Density Parity-Check Codes within 0.0045 dB of the Shannon Limit. *Communications Letters*, IEEE, Feb. 2001, vol. 5, no. 2, pp. 58–60. doi:10.1109/4234.905935
13. Kozlov A., Krouk E., Ovchinnikov A. An Approach to Development of Block-Commutative Codes with Low Density of Parity Check. *Izvestiia vuzov. Priborostroenie*, 2013, vol. 8, pp. 9–14 (In Russian).
14. Gilbert E. N. A Problem in Binary Encoding. *Proc. of the Symp. in Applied Mathematics*, 1960, vol. 10, pp. 291–297.
15. Arazi B. The Optimal Burst Error-Correcting Capability of the Codes Generated by f(x)=(x^p+1)(x^q+1)/(x+1). *Information and Control*, 1978, vol. 39, no. 3, pp. 303–314.
16. Bahl L., Chien R. On Gilbert Burst-Error-Correcting Codes. *IEEE Transactions on Information Theory*, May 1969, vol. 15, no. 3, pp. 431–433. doi:10.1109/TIT.1969.1054310
17. Neumann P. G. A Note on Gilbert Burst-Correcting Codes. *IEEE Transactions on Information Theory*, Jul. 1965, vol. 11, no. 3, pp. 377–384. doi:10.1109/TIT.1965.1053780
18. Zhang W., Wolf J. K. A Class of Binary Burst Error-Correcting Quasi-Cyclic Codes. *IEEE Transactions on Information Theory*, May 1988, vol. 34, no. 3, pp. 463–479. doi:10.1109/18.6026
19. Krouk E. A., Semenov S. V. Low-Density Parity-Check Burst Error-Correcting Codes. *International Workshop Algebraic and Combinatorial Coding Theory*, Leningrad, 1990, pp. 121–124.

УДК 621.396.96
doi:10.15217/issn1684-8853.2016.1.88

СВЕРХРАЗРЕШЕНИЕ РАДИОЛОКАЦИОННЫХ ЦЕЛЕЙ ПРИ ВОЗДЕЙСТВИИ АКТИВНЫХ ШУМОВЫХ ПОМЕХ ПО ОСНОВНОМУ И БЛИЖНИМ БОКОВЫМ ЛЕПЕСТКАМ ДИАГРАММЫ НАПРАВЛЕННОСТИ АНТЕННЫ РЛС

А. А. Чижов[a], *доктор техн. наук, доцент*
[a]*Военная академия войсковой противовоздушной обороны Вооруженных Сил Российской Федерации им. Маршала Советского Союза А. М. Василевского, Смоленск, РФ*

Постановка проблемы: *актуальными остаются проблемы пространственной компенсации активных шумовых помех, действующих по основному и ближним боковым лепесткам диаграммы направленности антенны РЛС. Рядом исследователей отмечается эффект «искажения» результирующей диаграммы направленности антенны РЛС или «смещения» ее максимума при пространственной компенсации помех, действующих по основному и ближним боковым лепесткам. При этом возможности и особенности сверхразрешения радиолокационных целей при воздействии активных шумовых помех исследованы недостаточно.* ***Цель:*** *исследовать возможности и особенности сверхразрешения радиолокационных целей при воздействии активных шумовых помех с близких к пеленгам целей угловых направлений, а также обосновать способ пространственной компенсации таких помех, обеспечивающий формирование при этом несмещенных оценок пеленгов одиночных и групповых сосредоточенных целей.* ***Результаты:*** *вскрыта причина смещенности оценок пеленгов целей при согласованной обработке сигналов в условиях воздействия активных шумовых помех, а именно не совсем корректное использование при формировании оценок всех составляющих функции правдоподобия. Предложен основанный на корректном использовании метода максимального правдоподобия способ обработки эхосигналов одиночных и групповых сосредоточенных целей при воздействии активных шумовых помех с близких к пеленгам целей угловых направлений, обеспечивающий формирование несмещенных оценок пеленгов, а также других параметров целей. Предложенный способ сверхразрешения радиолокационных целей эффективен при воздействии активных шумовых помех с произвольной пространственной корреляцией.* ***Практическая значимость:*** *предложенную обработку возможно использовать как составную часть процедуры проекционного сверхразрешения, а именно как процедуру оценки конфигурации портрета радиолокационной цели при воздействии активных шумовых помех.*

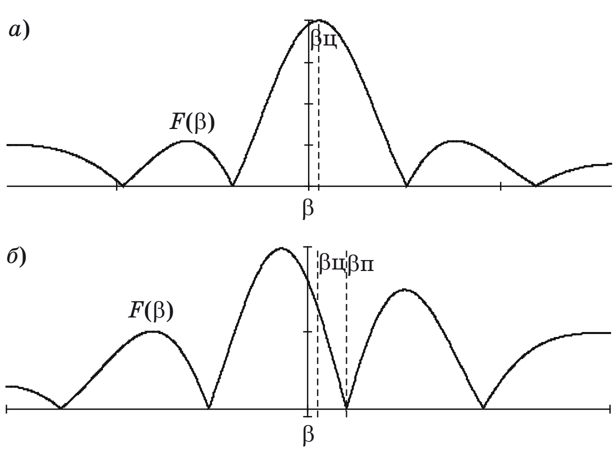
Ключевые слова *— активная шумовая помеха, сверхразрешение, защита от активных шумовых помех, компенсация активных шумовых помех.*

Введение

Проблемы пространственной компенсации активных шумовых помех (АШП), особенно действующих по основному и ближним боковым лепесткам диаграммы направленности антенны (ДНА) РЛС, сами по себе являются не до конца исследованными и на современном этапе. Классические подходы к решению указанных проблем изложены, например, в работе [1].

Следует подчеркнуть, что в работе [1] и в других работах [2, 3] отмечается эффект «искажения» результирующей ДНА РЛС или «смещения» ее максимума при пространственной компенсации АШП, действующих по основному и ближним боковым лепесткам ДНА (рис. 1).

Результаты аналитического расчета ДНА РЛС в отсутствие АШП представлены на рис. 1, *а*, а результаты расчета «искаженной» результирующей ДНА РЛС при классической пространственной компенсации АШП, действующей в основном лепестке ДНА, — на рис. 1, *б*. Рисунки наглядно демонстрируют, что, наряду с формированием «нуля» в ДНА РЛС в направлении АШП, применение классической пространственной компенсации АШП приводит к «смещению» максимума ДНА и, как следствие, к смещенности в оценках пеленга цели (даже одиночной). Это одна из проблем, решению которых посвящена статья.



■ *Рис. 1.* ДНА РЛС в отсутствие АШП (*а*) и результирующая ДНА РЛС при пространственной компенсации АШП, действующей в основном лепестке ДНА (*б*) ($\beta_ц$ и $\beta_п$ — пеленги цели и АШП)

Вместе с тем в настоящее время много работ посвящено различным методам и способам так называемого сверхразрешения радиолокационных целей, достаточно полный обзор и сравнительная оценка эффективности которых приведены в работах [4–6]. Из представленных в [5, 6] и других работах оценок следует, что максимальной энергетической эффективностью, приближающейся к потенциально возможной (так называемому хелстромовскому пределу), обладает проекционный метод сверхразрешения. При этом если вопросы радиолокационного сверхразрешения в отсутствие АШП исследованы достаточно глубоко, то возможности и особенности сверхразрешения при воздействии АШП, тем более действующих по основному и ближним боковым лепесткам ДНА РЛС, исследованы недостаточно.

В настоящей статье исследованы возможности и особенности сверхразрешения радиолокационных целей при воздействии АШП с близких к пеленгам целей угловых направлений, а также обоснован способ пространственной компенсации АШП, обеспечивающий формирование при этом несмещенных оценок пеленгов одиночных и групповых сосредоточенных целей.

Способ сверхразрешения радиолокационных целей при воздействии АШП

Для того чтобы уяснить причины «смещенности» результирующей ДНА РЛС при компенсации АШП с близких к пеленгам целей угловых направлений, проанализируем составляющие функции правдоподобия в условиях наличия АШП.

Пусть с m-канальной антенной решетки (АР) РЛС снимается эхосигнал цели (в общем случае групповой), принимаемый на фоне АШП:

$$\mathbf{y}(t) = \mathbf{s}(\boldsymbol{\alpha}, t)\mathbf{E} + \mathbf{n}(t), \tag{1}$$

где $\mathbf{y}(t)$ — комплексная огибающая сигнала на выходе АР (вектор-функция размера $m\times 1$); $\mathbf{s}(\boldsymbol{\alpha}) = [\mathbf{s}(\alpha_1,\ t)\ \mathbf{s}(\alpha_2,\ t)\ \ldots\ \mathbf{s}(\alpha_n,\ t)]$ — матрица-функция эхосигналов отдельных целей; $\mathbf{E} = [E_1\ E_2\ \ldots\ E_n]^{\mathrm{T}}$ — вектор комплексных амплитуд эхосигналов отдельных целей; $\boldsymbol{\alpha} = [\alpha_1\ \alpha_2\ \ldots\ \alpha_n]^{\mathrm{T}}$ — вектор информативных параметров эхосигналов отдельных целей (например, пеленгов целей); $n(t)$ — шум наблюдения (собственные шумы АР и АШП, характеризуемые корреляционной матрицей $\theta(t,\ t') = \Theta\delta(t - t')$. Следует отметить, что для сокращения выкладок в статье рассмотрен частный, но достаточно часто распространенный случай, когда собственные шумы АР и АШП задаются гауссовыми и некоррелированными по времени (белыми), но все последующие рассуждения можно повторить и для более общего случая. Также необходимо указать, что вопросы оценки θ в статье не рассматриваются, считается, что она оценивается с достаточной точностью в тех каналах дальности и (или) скорости, где заведомо отсутствуют радиолокационные цели.

Применительно к (1) функция правдоподобия (ФП) описывается следующим нормальным распределением (индекс Н обозначает операцию эрмитова сопряжения — одновременного транспонирования и комплексного сопряжения):

$$p(\mathbf{y}/\boldsymbol{\alpha}, \mathbf{E}) = \mathrm{const}\exp\Big(-\int\big(\mathbf{y}(t) - \mathbf{s}(\boldsymbol{\alpha},\ t)\mathbf{E}\big)^{\mathrm{H}} \times \times \boldsymbol{\Theta}^{-1}\big(\mathbf{y}(t) - \mathbf{s}(\boldsymbol{\alpha}, t)\mathbf{E}\big)\mathrm{d}t\Big). \tag{2}$$

Исследование на экстремум ФП $p(\mathbf{y}/\boldsymbol{\alpha},\ \mathbf{E})$ классически заменяется исследованием на экстремум ее натурального логарифма, который после раскрытия скобок в (2) задается выражением

$$\ln p(\mathbf{y}/\boldsymbol{\alpha}, \mathbf{E}) = 2\mathrm{Re}\big(\mathbf{E}^{\mathrm{H}} Z(\boldsymbol{\alpha})\big) - \mathbf{E}^{\mathrm{H}}\mathbf{Q}(\boldsymbol{\alpha})\mathbf{E} + \mathrm{const}, \tag{3}$$

где $\mathbf{Z}(\boldsymbol{\alpha}) = \int \mathbf{s}^{\mathrm{H}}(\boldsymbol{\alpha},\ t)\boldsymbol{\Theta}^{-1}\mathbf{y}(t)\mathrm{d}t$ — векторный корреляционный интеграл; $\mathbf{Q}(\boldsymbol{\alpha}) = \int \mathbf{s}^{\mathrm{H}}(\boldsymbol{\alpha},\ t)\boldsymbol{\Theta}^{-1}\mathbf{s}(\boldsymbol{\alpha},\ t)\mathrm{d}t$ — матрица Грама эхосигналов отдельных целей.

Следует подчеркнуть, что классическая (при локации одиночной цели и $n = 1$) согласованная обработка с пространственной компенсацией АШП подразумевает поиск максимума лишь одной составляющей ФП, а именно поиск максимума модуля корреляционного интеграла:

$$\hat{\boldsymbol{\alpha}}_{\text{согл}} = \arg\max_{\boldsymbol{\alpha}} \left|\mathbf{Z}(\boldsymbol{\alpha})\right|. \tag{4}$$

Одним из множителей в подынтегральном выражении для корреляционного интеграла является матрица, обратная матрице $\boldsymbol{\Theta}$, что и приводит к формированию «нуля» в ДНА РЛС в направлении АШП (см. рис. 1, *б*), т. е. обеспечивает компенсацию АШП. Вместе с тем, как указывалось выше (см. рис. 1), применение классической пространственной компенсации АШП приводит к «смещению» максимума ДНА и, как следствие, к смещенности в оценках пеленга цели (даже одиночной). Последнее противоречит общеизвестному факту о несмещенности оценок, получаемых по методу максимального правдоподобия.

Следует несколько более подробно остановиться на указанном противоречии, причина которого состоит в том, что оценки (4) не являются в данном случае оценками максимального правдоподобия, и зачастую воспринимаемая как постулат логика, что максимум модуля корреляционного интеграла соответствует максимуму ФП, в рассматриваемом случае (а также в более общем случае локации групповой цели) неправомерна.

Необходимо подчеркнуть, что максимум модуля корреляционного интеграла соответствует максимуму ФП только в случае локации одиночной цели (при $n = 1$) в отсутствие АШП (и, если совсем строго, при некоррелированных собственных шумах АР). При этом составляющая ФП $\mathbf{E}^{\mathrm{H}}\mathbf{Q}(\boldsymbol{\alpha})\mathbf{E}$, имеющая смысл энергии ожидаемого эхосигнала ($\boldsymbol{\Theta}$ в этом случае — диагональная матрица), не зависит от $\boldsymbol{\alpha}$, а максимум модуля корреляционного интеграла соответствует максимуму ФП.

Зависимость $\mathbf{Q}$ от $\boldsymbol{\alpha}$ при локации одиночной цели и наличии АШП, а также зависимость натурального логарифма ФП (3) от $\boldsymbol{\alpha}$ для того же случая представлены на рис. 2.

На рисунке продемонстрирован тот факт, что если при формировании оценок пеленгов целей учитывать все составляющие ФП, то оценки максимального правдоподобия действительно являются несмещенными.

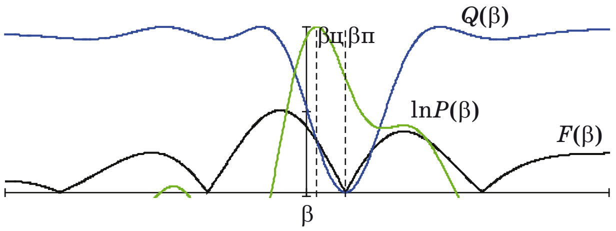
Необходимо оговориться, что графические зависимости на рис. 2 построены для нехарактерной для практики простейшей ситуации, когда комплексная амплитуда эхосигнала одиночной цели известна. Основанный на методе максимального правдоподобия способ формирования оценок вектора информативных параметров $\boldsymbol{\alpha}$ в общем случае (в том числе и при локации групповой сосредоточенной цели) должен быть основан на следующем выражении:

$$\hat{\boldsymbol{\alpha}} = \arg\max_{\boldsymbol{\alpha}} \mathbf{Z}^{\mathrm{H}}(\boldsymbol{\alpha})\mathbf{Q}^{-1}(\boldsymbol{\alpha})\mathbf{Z}(\boldsymbol{\alpha}). \quad (5)$$

Выражение (5) получено максимизацией (3) с учетом квадратичной зависимости натурального логарифма ФП (3) от $\mathbf{E}$ и возможности аналитического исключения $\mathbf{E}$ из процесса поиска экстремума.

Действительно, квадратичная форма $(\mathbf{Z}(\boldsymbol{\alpha}) - \mathbf{Q}(\boldsymbol{\alpha})\mathbf{E})^{\mathrm{H}}\mathbf{Q}(\boldsymbol{\alpha})^{-1}(\mathbf{Z}(\boldsymbol{\alpha}) - \mathbf{Q}(\boldsymbol{\alpha})\mathbf{E})$ при любом фиксированном $\boldsymbol{\alpha}$ имеет глобальный минимум по $\mathbf{E}$ при $\mathbf{E} = \mathbf{Q}^{-1}(\boldsymbol{\alpha})\mathbf{Z}(\boldsymbol{\alpha})$. Вместе с тем эту же форму возможно записать в виде $\mathbf{Z}^{\mathrm{H}}(\boldsymbol{\alpha})\mathbf{Q}^{-1}(\boldsymbol{\alpha})\mathbf{Z}(\boldsymbol{\alpha}) - \mathbf{E}^{\mathrm{H}}\mathbf{Z}(\boldsymbol{\alpha}) - \mathbf{Z}^{\mathrm{H}}(\boldsymbol{\alpha})\mathbf{E} + \mathbf{E}^{\mathrm{H}}\mathbf{Q}(\boldsymbol{\alpha})\mathbf{E} = \mathbf{Z}^{\mathrm{H}}(\boldsymbol{\alpha})\mathbf{Q}^{-1}(\boldsymbol{\alpha})\mathbf{Z}(\boldsymbol{\alpha}) - \ln p(\boldsymbol{\alpha}, \mathbf{E})$, откуда непосредственно следует, что $\hat{\mathbf{E}} = \arg\max_{\mathbf{E}} \ln p(\mathbf{y}/\boldsymbol{\alpha}, \mathbf{E}) = \mathbf{Q}^{-1}(\boldsymbol{\alpha})\mathbf{Z}(\boldsymbol{\alpha})$ при том, что

$$\arg\max_{\boldsymbol{\alpha}} \mathbf{Z}^{\mathrm{H}}(\boldsymbol{\alpha})\mathbf{Q}^{-1}(\boldsymbol{\alpha})\mathbf{Z}(\boldsymbol{\alpha}) = \arg\max_{\boldsymbol{\alpha}} \ln p\left(\mathbf{y}/\boldsymbol{\alpha}, \hat{\mathbf{E}}\right).$$



■ *Рис. 2.* Зависимость $\mathbf{Q}$ от $\boldsymbol{\alpha}$, а также зависимость натурального логарифма ФП от $\boldsymbol{\alpha}$ при локации одиночной цели и наличии АШП

Последнее свидетельствует о справедливости выражения (5).

Приведенное доказательство не использует операцию дифференцирования $\ln p(\boldsymbol{\alpha}, \mathbf{E})$ по $\mathbf{E}$, которая вследствие квадратичной зависимости $\ln p(\boldsymbol{\alpha}, \mathbf{E})$ от комплексного вектора $\mathbf{E}$ невозможна. Для полной математической строгости приведенного доказательства следует также добавить, что матрица Грама эхосигналов отдельных рассеивателей $\mathbf{Q}$ является эрмитово сопряженной ($\mathbf{Q} = \mathbf{Q}^{\mathrm{H}}$) и, в типовых ситуациях, невырожденной.

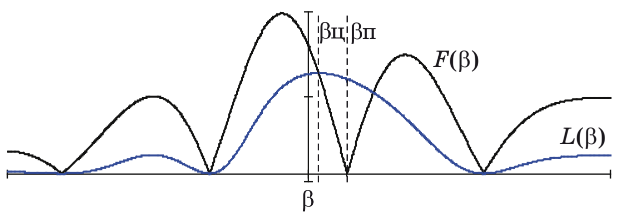
Оценка эффективности предложенного способа сверхразрешения радиолокационных целей при воздействии АШП

Аналитически рассчитанная зависимость $L(\beta) = \mathbf{Z}^{\mathrm{H}}(\beta)\mathbf{Q}^{-1}(\beta)\mathbf{Z}(\beta)$ для случая локации одиночной цели при наличии АШП показана на рис. 3. Из анализа рисунка видно, что формируемые согласно (5) оценки пеленга одиночной цели являются несмещенными.

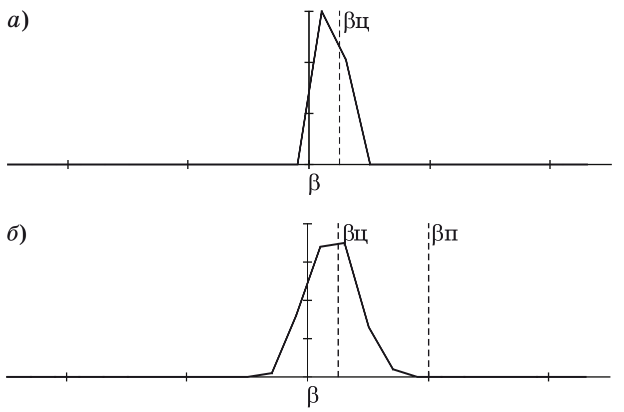
Результаты имитационного математического моделирования процесса пеленгации одиночной цели в отсутствие и при наличии АШП представлены на рис. 4, *а* и *б*. При этом в ходе моделирования обработки эхосигналов согласно (5) использовалась оценка корреляционной матрицы шумовых помех, полученная по 16 реализациям шума.

Анализ рис. 4 позволяет сделать выводы, во-первых, о работоспособности исследуемого способа, а во-вторых, о несмещенности формируемых оценок пеленга одиночной цели при наличии АШП в основном лепестке ДНА РЛС.

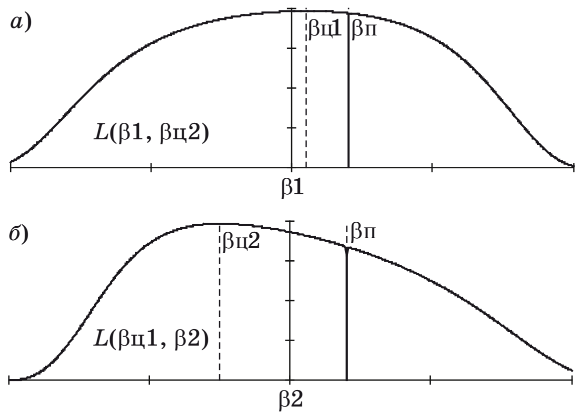
Анализ рис. 5, *а* и *б*, где приведены аналитически рассчитанные сечения зависимости $L(\beta_1, \beta_2) = \mathbf{Z}^{\mathrm{H}}(\beta_1, \beta_2)\mathbf{Q}^{-1}(\beta_1, \beta_2)\mathbf{Z}(\beta_1, \beta_2)$ для случая локации парной сосредоточенной цели при наличии АШП, показывает, что формируемые согласно (5) оценки пеленгов отдельных целей из состава парной также являются несмещенными.



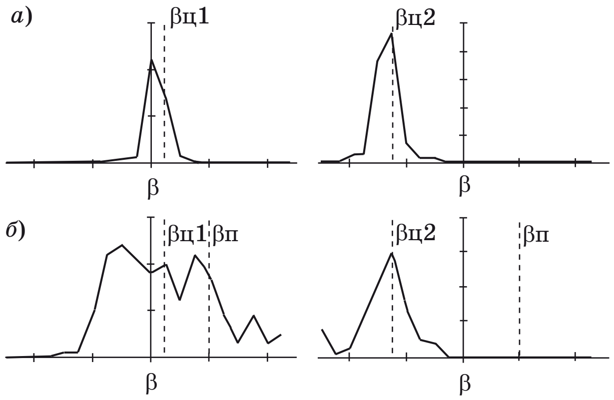
■ *Рис. 3.* Зависимость $L(\beta)$=$\mathbf{Z}^{\mathrm{H}}(\beta)\mathbf{Q}^{-1}(\beta)\mathbf{Z}(\beta)$ для случая локации одиночной цели при наличии АШП (для сравнения также приведена результирующая ДНА РЛС при пространственной компенсации АШП, действующей в основном лепестке ДНА РЛС)



■ *Рис. 4.* Распределение оценок пеленгов одиночной цели в отсутствие (*а*) и при наличии (*б*) АШП (отношение сигнал/шум 26 дБ, отношение помеха/шум 60 дБ, удаление пеленга АШП от пеленга цели 0,25 ширины ДНА РЛС)



■ *Рис. 5.* Сечения зависимости $L(\beta_1, \beta_2)$ для случая локации парной цели при наличии АШП: *а* — вертикальное сечение, параллельное оси β_1, проходящее через точку $\beta_{ц2}$ на оси β_2; *б* — вертикальное сечение, параллельное оси β_2, проходящее через точку $\beta_{ц1}$ на оси β_1



■ *Рис. 6.* Распределение оценок пеленгов парной цели в отсутствие (*а*) и при наличии (*б*) АШП (отношение сигнал/шум по каждой цели 26 дБ, отношение помеха/шум 60 дБ, взаимное удаление пеленгов целей 0,5 ширины ДНА РЛС, удаление пеленга АШП от пеленга ближайшей цели 0,25 ширины ДНА РЛС

Результаты имитационного математического моделирования процесса пеленгации парной сосредоточенной цели в отсутствие и при наличии АШП представлены на рис. 6, *а*, *б*. При этом, как и ранее, в ходе моделирования обработки эхосигналов согласно (5) использовалась оценка корреляционной матрицы шумовых помех, полученная по 16 реализациям шума.

Результаты моделирования свидетельствуют о работоспособности исследуемого способа при типовых отношениях сигнал/шум, в том числе в ситуациях, когда в пределах основного лепестка ДНА РЛС присутствует несколько целей и АШП. Естественно, что точность пеленгации отдельных целей из состава групповой сосредоточенной и особенно цели, пеленг которой находится вблизи пеленга источника АШП (см. рис. 6), несколько ниже, чем точность пеленгации одиночной цели (см. рис. 4), однако при достаточном для соответствующих условий локации отношении сигнал/шум точностные характеристики обработки сопоставимы с аналогичными, полученными в отсутствие АШП.

Заключение

Таким образом, основанный на методе максимального правдоподобия предложенный способ обработки эхосигналов одиночных и групповых сосредоточенных целей при воздействии АШП с близких к пеленгам целей угловых направлений обеспечивает формирование несмещенных оценок их пеленгов, а также других параметров, т. е. является способом сверхразрешения радиолокационных целей при воздействии АШП с произвольной пространственной корреляцией. Предложенную обработку возможно использовать как составную часть процедуры проекционного сверхразрешения [5–6], а именно как процедуру оценки конфигурации портрета радиолокационной цели при воздействии АШП.

Литература

1. **Ширман Я. Д., Манжос В. Н.** Теория и техника обработки радиолокационной информации на фоне помех. — М.: Радио и связь, 1981. — 416 с.
2. **Григорян Д. С., Торбин С. А., Герасимов В. В.** Защита моноимпульсного радиопеленгатора от активной шумовой помехи, действующей по основным лепесткам диаграмм направленности //

Вестник Концерна ПВО «Алмаз-Антей». 2014. № 2. С. 103–112.
3. **Pan D., Cheng N., Bian Y.** Method Eases Monopulse System Jamming// Microwaves & RF. Aug. 2014. P. 82–86.
4. **Чижов А. А.** Сверхрэлеевское разрешение. Т. 1: Классический взгляд на проблему. — М.: Красанд, 2010. — 96 с.
5. **Чижов А. А.** Сверхрэлеевское разрешение. Т. 2: Преодоление фактора некорректности обратной задачи рассеяния и проекционная радиолокация. — М.: Красанд, 2010. — 104 с.
6. **Чижов А. А.** Метод разрешения групповых сосредоточенных целей // Радиотехника. 2009. № 10. С. 4–12.

UDC 621.396.96
doi:10.15217/issn1684-8853.2016.1.88

Super-Resolution of Radar Targets under Jamming on Primary and Middle Lateral Lobes of Radar Antenna Directivity Diagram

Chizhov A. A.[a], Dr. Sc., Tech., Associate Professor, rtshouse@mail.ru
[a]The Marshal of the Soviet Union A. M. Vasilevsky Army Air Defense Military Academy of the Armed Forces of the Russian Federation, 2, Kotovskogo St., 214027, Smolensk, Russian Federation

Introduction: Spatial compensation of jamming on the primary and middle lateral lobes of a radar antenna directivity diagram is still a serious problem. Some researchers noted the effects of "distortion" of the resulting antenna pattern, or an "offset" of its maximum when the noise on the primary and middle lateral lobes is compensated spatially. The features and capabilities of super-resolution radar targets under the influence of jamming is studied insufficiently. **Purpose:** The goal of this paper is to explore the capabilities and features of super-resolution radar targets under the impact of jamming with angular direction targets close to bearing, and to substantiate the method of spatial compensation of such noise, providing the formation of unbiased estimates of bearing for single and group targets. **Results:** It has been found out why the estimates of target bearing were shifted when the signals were processed coherently under active noise. This shift took place because all the likelihood function components were used in a not completely correct way. An alternative way is proposed, based on a correct usage of maximum likelihood method. This way ensures that the bearing estimates do not shift. The proposed method of radar target super-resolution is effective under the impact of jamming with arbitrary spatial correlation. **Practical relevance:** The proposed treatment can be used as an integral part of the projective superresolution procedure, more precisely as a procedure for evaluating the configuration of a radar target portrait under the impact of jamming.

Keywords — Active Noise Jamming, Super-resolution, Active Noise Jamming Protection, Active Noise Jamming Compensation.

References

1. Shirman Ia. D., Manzhos V. N. *Teoriia i tekhnika obrabotki radiolokatsionnoi informatsii na fone pomekh* [Theory and Technique of Processing Radar Information on the Interferences Background]. Moscow. Radio i sviaz' Publ., 1981. 416 p. (In Russian).
2. Grigoryan D. S., Torbin S. A., Gerasimov V. V. Monopulse Radar Protection from the Active Noise Interference Acting on the Main Lobe of Antenna Diagram. *Vestnik Kontserna PVO "Almaz-Antei"*, 2014, no. 2, pp. 103–112 (In Russian).
3. Pan D., Cheng N., Bian Y. Method Eases Monopulse System Jamming. *Microwaves & RF*, August 2014, pp. 82–86.
4. Chizhov A. A. *Sverkhreleevskoe razreshenie*. Tom 1. *Klassicheskii vzgliad na problemu* [Super Resolution. Vol. 1. Classic View on the Problem]. Moscow, Krasand Publ., 2010. 96 p. (In Russian).
5. Chizhov A. A. *Sverkhreleevskoe razreshenie*. Tom 2. *Preodolenie faktora nekorrektnosti obratnoi zadachi rasseianiia i proektsionnaia radiolokatsiia* [Super Resolution. Vol. 2. Overcoming the Incorrectness Factor of the Inverse Dispersion Problem and Projection Radiolocation]. Moscow. Krasand Publ., 2010. 104 p. (In Russian).
6. Chizhov A. A. The Method of Resolution of Multiple Lumped Targets. *Radiotekhnika*, 2009, no. 10, pp. 4–12 (In Russian).

УДК 621.38
doi:10.15217/issn1684-8853.2016.1.93

АНАЛИЗ ПРОЦЕССА ДИАГНОСТИРОВАНИЯ ДИНАМИЧЕСКИХ СИСТЕМ ПО OSTN-ДИАГРАММАМ IDEF3-ТЕХНОЛОГИИ

Г. С. Бритов[a]*, канд. техн. наук, доцент*
[a]*Санкт-Петербургский государственный университет аэрокосмического приборостроения, Санкт-Петербург, РФ*

Постановка проблемы: *большинство известных работ по диагностированию динамических систем посвящены описанию методов диагностирования, таких как параметрическая идентификация, использование наблюдателей состояния, аналитической избыточности и др. Однако анализ технологического процесса диагностирования не проводится.* ***Цель:*** *проанализировать технологический процесс диагностирования динамических систем, получив вероятности состояний процесса на основе OSTN-диаграмм IDEF3-технологии; развить подход, позволяющий формализовать анализ IDEF3-моделей по OSTN-диаграммам. В соответствии с ним должен осуществляться расчет вероятностей состояний процессов диагностирования.* ***Методы:*** *построены рекуррентные уравнения вероятностей состояний, возникающих при диагностировании динамических систем, на основе теории марковских процессов.* ***Результаты:*** *разработан метод построения стохастической матрицы рекуррентных уравнений, для которых известны начальные условия. Суть метода в том, что с помощью разработанных PFDD-диаграмм диагностирования динамических систем строятся OSTN-диаграммы IDEF3-технологии. Благодаря этим диаграммам осуществляется не только полное описание технологического процесса диагностирования, но и определение его состояний.* ***Выводы:*** *полученные результаты позволяют оценить качество технологического процесса диагностирования динамических систем. Проведена оценка систем тестового и функционального диагностирования.*

Ключевые слова *— диаграммы IDEF3-технологии, состояния системы, переходные интенсивности, стохастическая матрица, стационарные вероятности состояний.*

Введение

Технологические процессы диагностирования реализуются в системах технического диагностирования. Они могут быть тестовыми и функциональными. Тестовые системы проверяют исправность объекта диагностирования в специальном режиме, а функциональные системы в рабочем режиме контролируют правильность функционирования объекта.

Методы диагностирования зависят от вида исследуемого объекта. В статье рассматриваются системы диагностирования динамических систем (ДС). Тестирование линейных ДС подробно описано в целом ряде работ [1–10]. Обычно применяемые методы основаны на формировании диагностических признаков, по которым принимается решение об исправности системы.

Диагностические признаки, основанные на таких важных характеристиках линейных ДС, как нули и операторные нормы, приведены в работах [8, 9], а использование математических моделей рассмотрено в работах [2–4]. Организация тестирования линейных ДС на основе их передаточных функций представлена в работах [1, 10].

Следует отметить методы тестирования на основе специальных входных сигналов системы. Так, в статье [6] рассчитывается комплементарный сигнал, обеспечивающий переход системы за заданное время из нулевых начальных условий опять в нулевые, конечные условия. А в статье [7] предлагается подавать на вход системы аннулирующий сигнал, на который отсутствует выходная реакция системы. Классическим примером такого сигнала является гармонический сигнал, частота которого равна передаточному нулю системы.

Системы функционального диагностирования ДС рассмотрены в работах [11–14]. Проверка правильности функционирования линейных систем основана на разработке устройства функционального диагностирования [11, 13]. С его помощью формируется диагностический признак, который должен оставаться в допуске, пока система работает без ошибок. Появление ошибок вызывается сбоями и отказами в системе. На сбои диагностический признак реагирует как на импульсное входное воздействие. При отказах реакция диагностического признака соответствует постоянному входному воздействию.

Функциональное диагностирование нелинейных ДС описано в работах [12, 14]. Для формирования диагностических признаков используется предложенная авторами алгебра функций.

Организация диагностирования ДС требует разработки соответствующих технологических процессов. Для этого можно использовать стандартную IDEF3-технологию. Известно, что PFDD-диаграммы этой технологии позволяют строить модели технологических процессов [15], а OSTN-диаграммы — выполнять их анализ [16]. Применение IDEF3-технологии к системам технического диагностирования выполнено в работе [17].

Целью статьи является представление PFDD- и OSTN-диаграмм для систем тестового и функционального диагностирования линейных ДС и вероятностный анализ состояний процессов, проходящих в этих системах.

IDEF3-модели систем диагностирования

Будем рассматривать систему технического диагностирования ДС с точки зрения задач и технологических процессов, в ней происходящих. Поэтому можно использовать стандартную IDEF3-технологию, PFDD-диаграммы которой позволяют построить модели технологических процессов, а OSTN-диаграммы — выполнить их анализ.

Контекстная PFDD-диаграмма процесса тестового диагностирования ДС показана на рис. 1.

Оператор должен выполнить диагностирование на основании результатов тестирования ДС с помощью диагностических признаков.

PFDD-диаграмма второго уровня декомпозиции процесса тестового диагностирования ДС представлена на рис. 2. Показано семь единиц работы, или операций, которые обеспечивают тестирование. X-разветвление J2 позволяет получить результаты либо от исправной диагностической системы (ИДС), либо от неисправной диагностической системы (НИДС). Сформированный диагностический признак (ДП) дает возможность принять решение либо об исправности, либо неисправности ДС (X-разветвление J4). Неисправную систему следует восстановить и вновь протестировать.

По PFDD-диаграмме, представленной на рис. 2, можно построить диаграммы состояний процесса диагностирования, отличающиеся числом состояний, а следовательно, глубиной анализа. На рис. 3 показаны соответствующие OSTN-диаграммы процесса тестового диагностирования ДС.

На схеме рис. 3, *а* три состояния:

1) C_1 — тестовое диагностирование исправной ДС;

2) C_2 — тестовое диагностирование неисправной ДС;

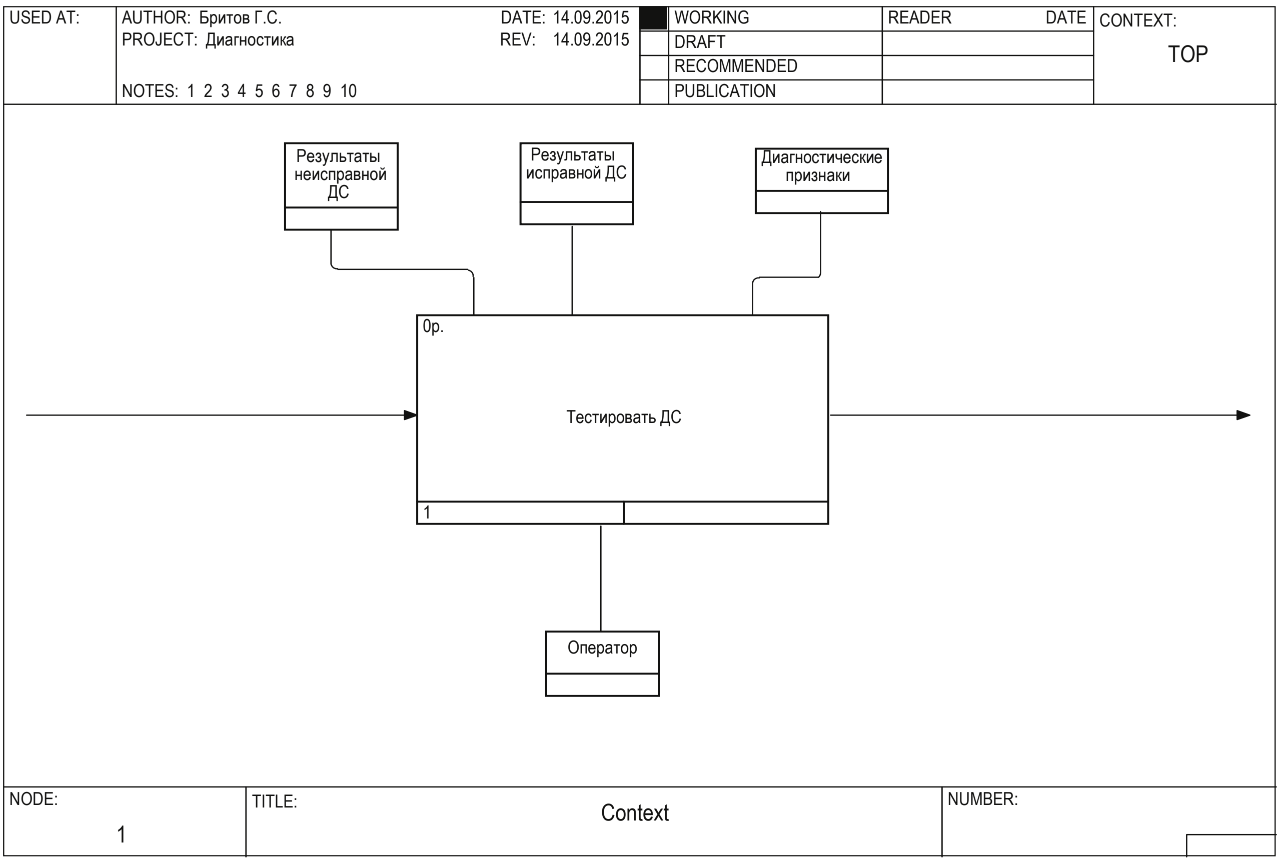
3) C_3 — восстановление системы.

Коэффициенты передачи между состояниями определяют интенсивность:

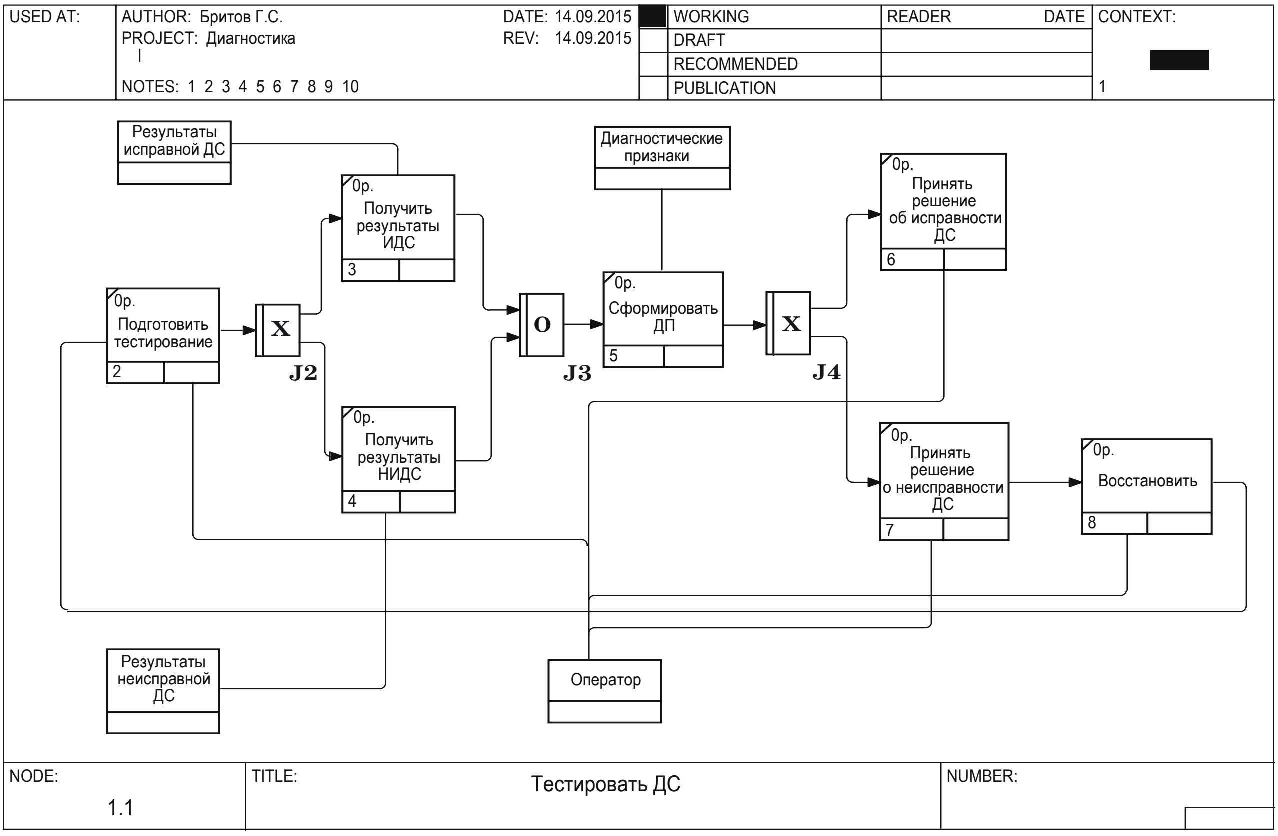
a_1 — появления дефектов;

a_2 — обнаружения дефектов;

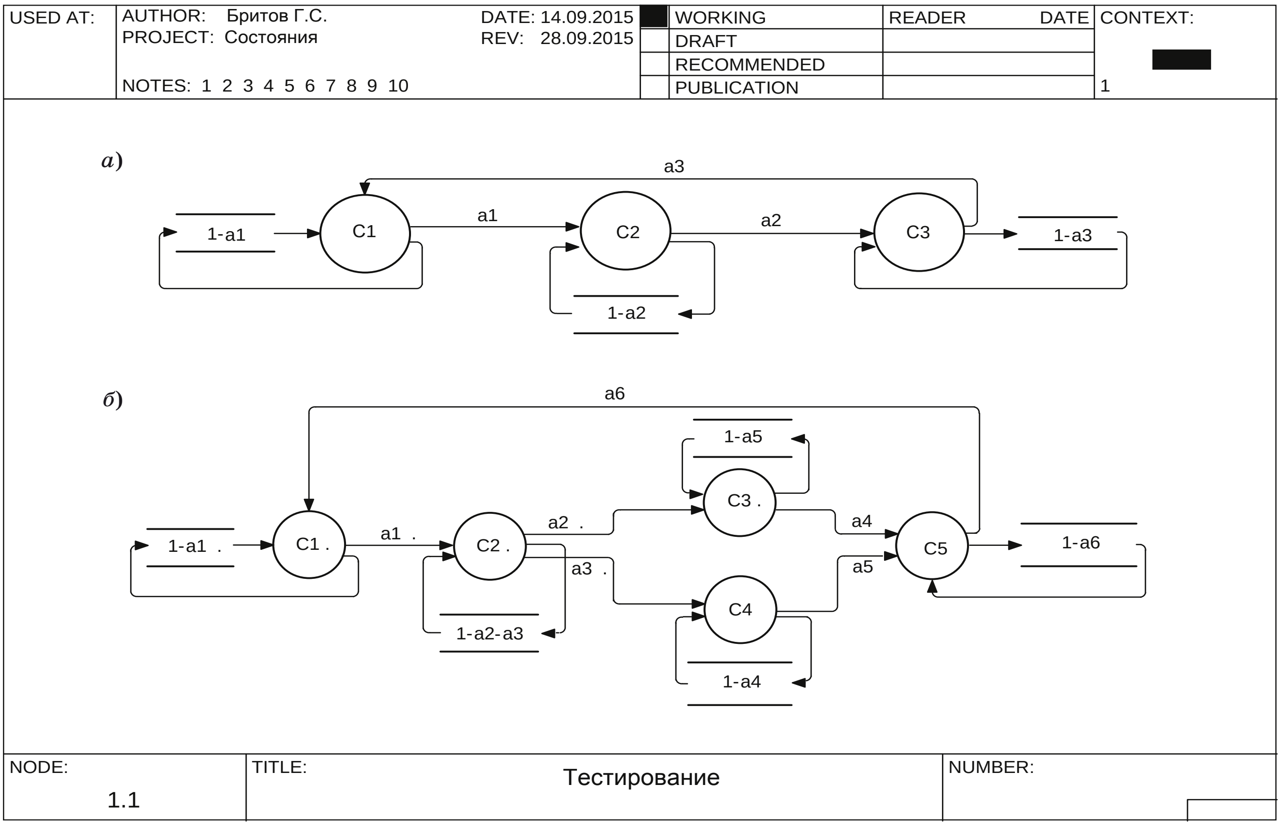
a_3 — исправления дефектов.



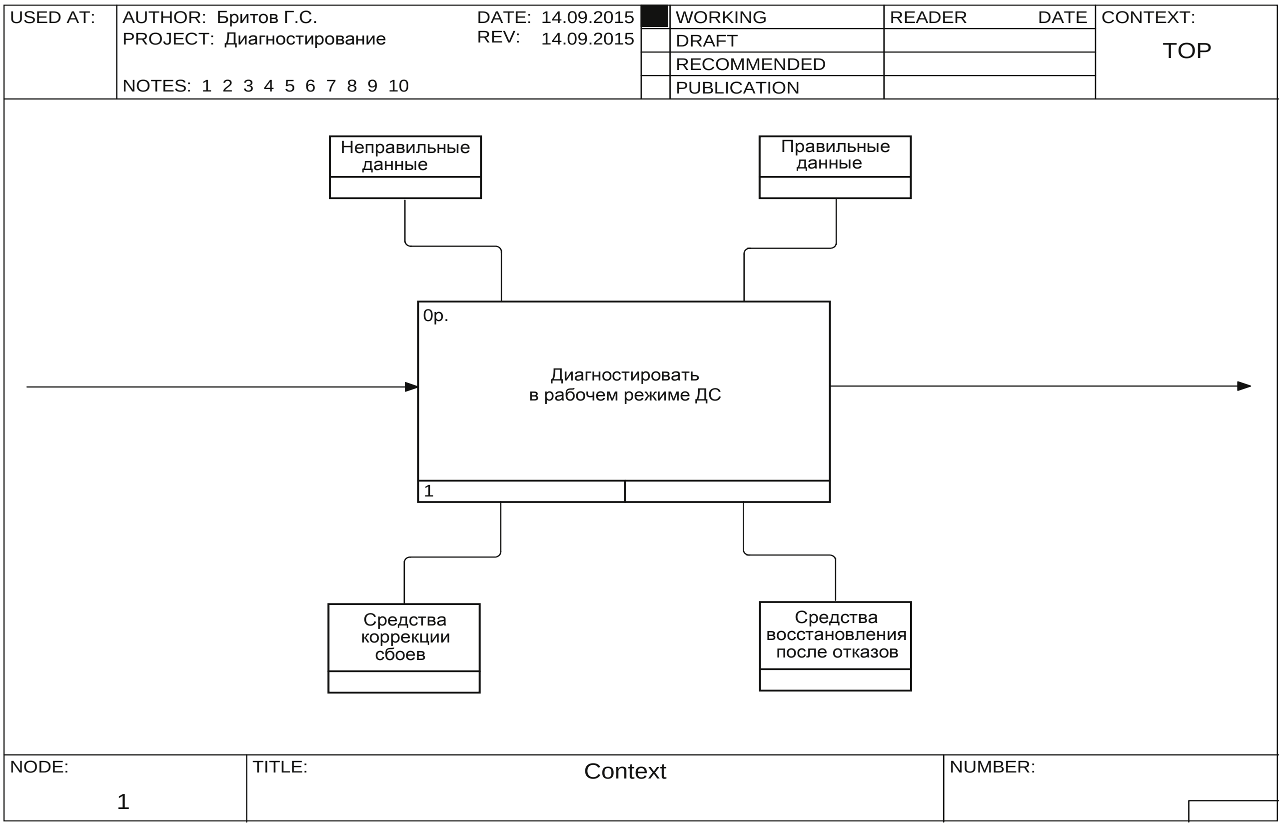
■ *Рис. 1.* Контекстная PFDD-диаграмма процесса тестового диагностирования ДС



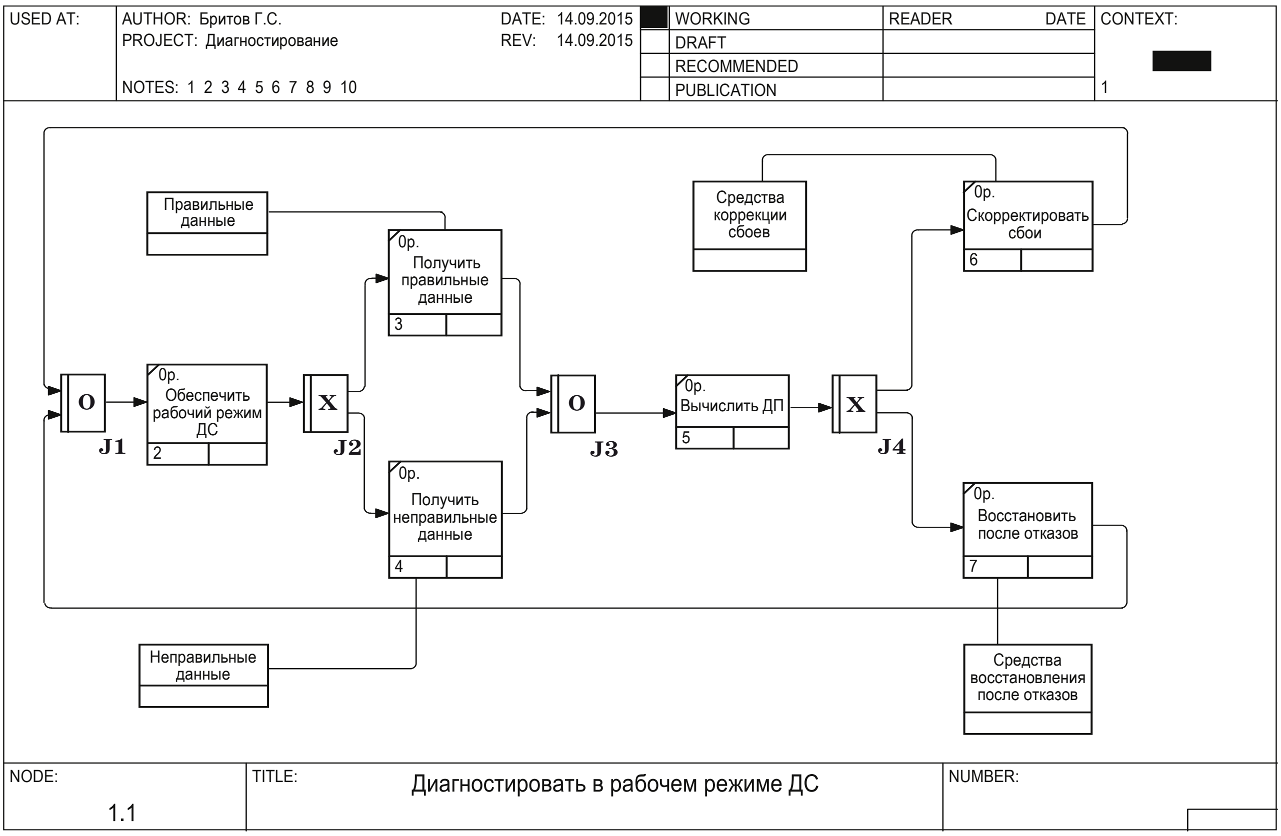
■ *Рис. 2.* PFDD-диаграмма второго уровня декомпозиции процесса тестового диагностирования ДС



■ *Рис. 3.* OSTN-диаграммы процесса тестового диагностирования ДС с тремя (*а*) и пятью (*б*) состояниями



■ *Рис. 4.* Контекстная PFDD-диаграмма процесса функционального диагностирования ДС



■ *Рис. 5.* PFDD-диаграмма второго уровня декомпозиции процесса функционального диагностирования ДС

Пять состояний на рис. 3, *б*:

1) C_1 — тестовое диагностирование исправной ДС;

2) C_2 — тестовое диагностирование неисправной ДС;

3) C_3 — обнаружение дефектов;

4) C_4 — локализация дефектов;

5) C_5 — восстановление системы.

Коэффициенты передачи между состояниями определяют интенсивность:

a_1 — появления дефектов;

a_2 — перехода к обнаружению дефектов;

a_3 — перехода к исправлению дефектов;

a_4 — решения об обнаружении дефектов;

a_5 — решения об исправлении дефектов;

a_6 — восстановления системы.

Процесс функционального диагностирования ДС, контекстная PFDD-диаграмма которого представлена на рис. 4, осуществляется в рабочем режиме ДС на основе получаемых данных о входных и выходных сигналах системы. После обнаружения отказов или сбоев в работе системы используются средства восстановления и коррекции.

На PFDD-диаграмме второго уровня декомпозиции процесса функционального диагностирования ДС (рис. 5) показано шесть единиц работы, или операций, которые обеспечивают непрерывный контроль за работой системы. X-разветвление J2 позволяет получить либо правильные данные от исправной системы, либо неправильные данные от неисправной системы. Вычисленный ДП дает возможность принять решение либо о появлении сбоев, либо о появлении отказов в ДС (X-разветвление J4). Неисправную систему следует откорректировать после сбоев или восстановить после отказов и вновь обратиться к диагностированию.

По PFDD-диаграмме, представленной на рис. 5, можно построить диаграмму состояний процесса диагностирования, содержащую четыре состояния. На рис. 6 показана OSTN-диаграмма процесса функционального диагностирования ДС.

Имеется четыре состояния на рис. 6:

1) C_1 — поступление правильных данных;

2) C_2 — поступление неправильных данных;

3) C_3 — коррекция сбоев;

4) C_4 — восстановление после отказов.

Коэффициенты передачи между состояниями определяют интенсивность:

a_1 — появления дефектов;

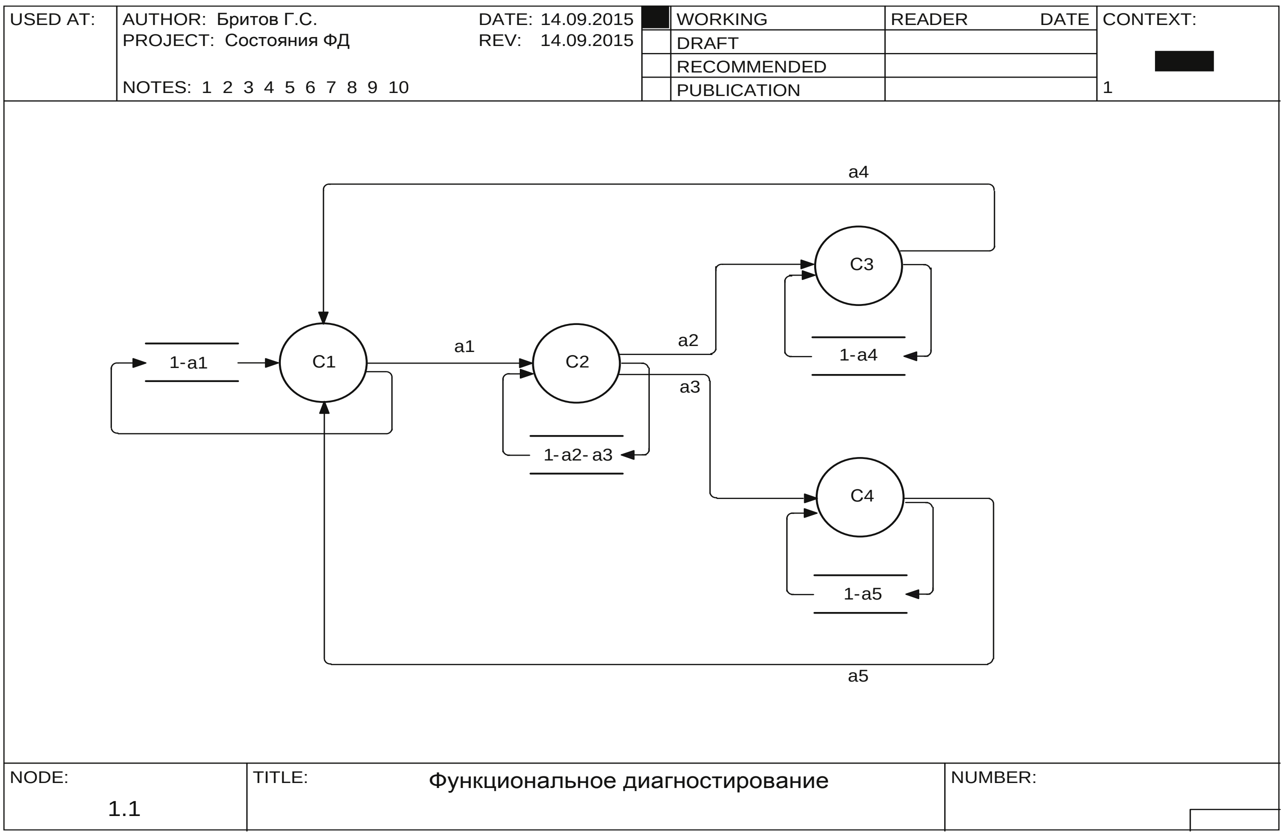
a_2 — обнаружения сбоев;

a_3 — обнаружения отказов;

a_4 — коррекции сбоев;

a_5 — восстановления после отказов.

Полученные диаграммы позволяют выполнить анализ процессов диагностирования, используя уравнения вероятностей построенных состояний.



■ ***Рис. 6.*** OSTN-диаграмма процесса функционального диагностирования ДС

Расчет вероятностей состояний процессов диагностирования

Для расчета вероятностей состояний процессов диагностирования, рассмотренных в предыдущем разделе, примем следующие допущения.

1. Переходы между состояниями в OSTN-диаграммах являются случайными событиями.

2. Вероятности этих событий являются переходными вероятностями и распределены по закону Пуассона.

3. Вероятность появления i-го события за время Δt есть $a_i \Delta t$.

4. Время $\Delta t = 1$.

5. Единица времени выбирается так, чтобы в течение времени исследования интенсивности можно было бы считать постоянными и меньшими 1.

Рекуррентные уравнения получающегося при принятых допущениях марковского процесса с тремя состояниями C_1, C_2, C_3 из OSTN-диаграммы, показанной на рис. 3, *а*, имеют вид

$$p_1(t+1) = (1-a_1)\cdot p_1(t) + a_3 \cdot p_3(t),\ p_1(0) = 1;$$
$$p_2(t+1) = (1-a_2)\cdot p_2(t) + a_1 \cdot p_1(t),\ p_2(0) = 0;$$
$$p_3(t+1) = (1-a_3)\cdot p_3(t) + a_2 \cdot p_2(t),\ p_3(0) = 0.$$

Матрица этих уравнений

$$\mathbf{A} = \begin{bmatrix} 1-a_1 & 0 & a_3 \\ a_1 & 1-a_2 & 0 \\ 0 & a_2 & 1-a_3 \end{bmatrix}$$

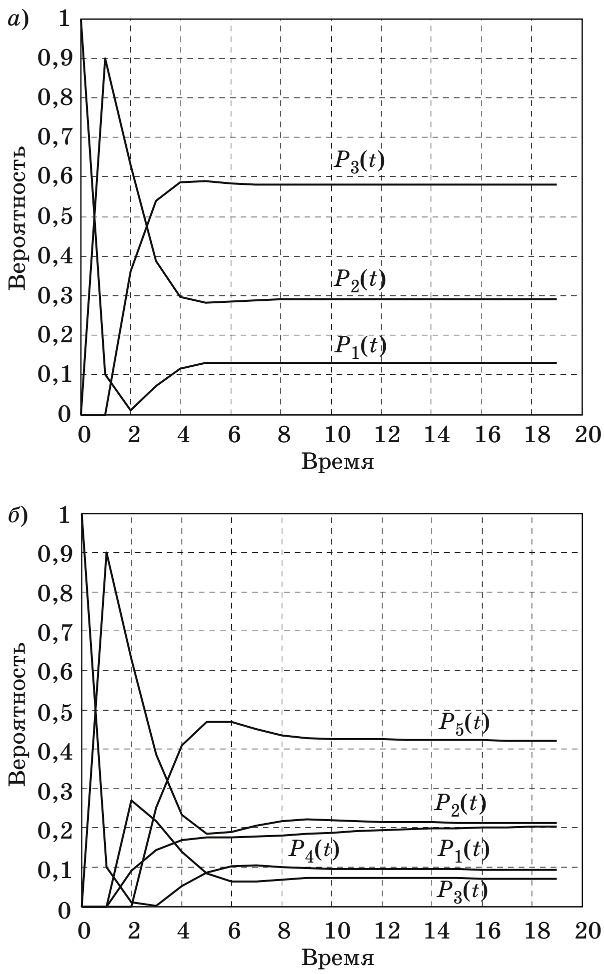
является стохастической матрицей, и ее спектральный радиус равен 1. Поэтому существует стационарный режим соответствующего марковского процесса. Расчет стационарных вероятностей состояний выполним, используя инструменты математического пакета MatLab:

```
>> syms a1 a2 a3 A L V P
>> A=[1-a1 0 a3; a1 1-a2 0; 0 a2 1-a3]
A =
[ 1 - a1,      0,     a3]
[     a1, 1 - a2,      0]
[      0,     a2, 1 - a3]
 >> [L,V]=eig(A);
>> V(1,1)=1
>> P=factor(L(:,1)/sum(L(:,1)))
P =
 (a3*a2)/(a1*a2 + a1*a3 + a2*a3)
 (a3*a1)/(a1*a2 + a1*a3 + a2*a3)
 (a1*a2)/(a1*a2 + a1*a3 + a2*a3)
```

Полученные формулы позволяют достаточно быстро рассчитать стационарные вероятности P и принять решение о качестве процесса диагностирования. Желательно, чтобы состояние диагностирования исправной системы C_1 происходило чаще, а состояния неисправной системы C_2 и ее восстановления C_3 — как можно реже. Следовательно, необходимо обеспечить такие интенсивности, чтобы стационарная вероятность P_1 была близка к 1, а стационарные вероятности P_2 и P_3 были, соответственно, малы. Тогда для оценки качества диагностирования можно использовать коэффициент готовности системы, который совпадает с вероятностью P_1.

Если задать конкретные значения интенсивностей, то можно получить графики марковского процесса. Например, графики вероятностей трех состояний при $a_1 = 0{,}9$, $a_2 = 0{,}4$, $a_3 = 0{,}2$ показаны на рис. 7, *а*. Из него следует, что при такой большой интенсивности появления дефектов вероятность состояния диагностирования исправной ДС будет около 0,1, что говорит о том, что коэффициент готовности системы равен 0,1. Значит, получаем либо плохое качество процесса диагностирования, либо плохое качество ДС.

За плохое качество процесса диагностирования отвечает оператор. Он должен обеспечивать правильный процесс диагностирования. Плохое



■ ***Рис. 7.*** Вероятности трех (*а*) и пяти (*б*) состояний марковского процесса системы тестового диагностирования качество диагностируемой ДС будет усложнять работу оператора.

Рекуррентные уравнения марковского процесса с пятью состояниями C_1, C_2, C_3, C_4, C_5 из OSTN-диаграммы, показанной на рис. 3, *б*, имеют следующий вид:

$$p_1(t+1)=(1-a_1)\cdot p_1(t)+a_6\cdot p_5(t),\ p_1(0)=1;$$
$$p_2(t+1)=(1-a_2-a_3)\cdot p_2(t)+a_1\cdot p_1(t),\ p_2(0)=0;$$
$$p_3(t+1)=(1-a_4)\cdot p_3(t)+a_2\cdot p_2(t),\ p_3(0)=0;$$
$$p_4(t+1)=(1-a_5)\cdot p_4(t)+a_3\cdot p_2(t),\ p_3(0)=0;$$
$$p_5(t+1)=(1-a_6)\cdot p_5(t)+a_4\cdot p_3(t)+a_5\cdot p_4(t),\ p_5(0)=0.$$

Матрица этих уравнений

$$\mathbf{A}=\begin{bmatrix}1-a_1 & 0 & 0 & 0 & a_6\\ a_1 & 1-a_2-a_3 & 0 & 0 & 0\\ 0 & a_2 & 1-a_4 & 0 & 0\\ 0 & a_3 & 0 & 1-a_5 & 0\\ 0 & 0 & a_4 & a_5 & 1-a_6\end{bmatrix}$$

вновь стохастическая, поэтому можно выполнить расчет стационарных вероятностей:

```
>> syms a6
>> A=[1-a1 0 0 0 a6; a1 1-a2-a3 0 0 0; 0 a2 1-a4 0 0; 0 a3 0 1-a5  0;0 0 a4 a5 1-a6]
A =
[ 1 - a1,           0,      0,      0,     a6]
[     a1, 1 - a3 - a2,      0,      0,      0]
[      0,          a2, 1 - a4,      0,      0]
[      0,          a3,      0, 1 - a5,      0]
[      0,           0,     a4,     a5, 1 - a6]
>> [L,V]=eig(A);
>> V(1,1)
ans =
1
 >> P=factor(L(:,1)/sum(L(:,1)))
P =
 (a6*a4*a5*(a2 + a3))/(a1*a2*a4*a5 + a1*a3*a4*a5 + a1*a2*a5*a6 + a1*a3*a4*a6 + a1*a4*a5*a6 + a2*a4*a5*a6 + a3*a4*a5*a6)
(a6*a1*a4*a5)/(a1*a2*a4*a5 + a1*a3*a4*a5 + a1*a2*a5*a6 + a1*a3*a4*a6 + a1*a4*a5*a6 + a2*a4*a5*a6 + a3*a4*a5*a6)
(a2*a6*a1*a5)/(a1*a2*a4*a5 + a1*a3*a4*a5 + a1*a2*a5*a6 + a1*a3*a4*a6 + a1*a4*a5*a6 + a2*a4*a5*a6 + a3*a4*a5*a6)
(a3*a6*a1*a4)/(a1*a2*a4*a5 + a1*a3*a4*a5 + a1*a2*a5*a6 + a1*a3*a4*a6 + a1*a4*a5*a6 + a2*a4*a5*a6 + a3*a4*a5*a6)
(a1*a4*a5*(a2 + a3))/(a1*a2*a4*a5 + a1*a3*a4*a5 + a1*a2*a5*a6 + a1*a3*a4*a6 + a1*a4*a5*a6 + a2*a4*a5*a6 + a3*a4*a5*a6)
>>
```

Полученные формулы существенно сложнее предыдущих. Поэтому оценку стационарных вероятностей удобнее выполнять по графикам марковского процесса, построенным при конкретных значениях интенсивностей.

Графики вероятностей пяти состояний при значениях интенсивностей a_1 = 0,9, a_2 = 0,3, a_3 = 0,1, a_4 = 0,9, a_5 = 0,1, a_6 = 0,2 показаны на рис. 7, *б*. Вновь, как и раньше, большая интенсивность появления дефектов приводит к малой стационарной вероятности состояния C_1. Значит, коэффициент готовности системы равен 0,1. Соответственно, получаем плохое качество системы.

Рекуррентные уравнения марковского процесса с четырьмя состояниями C_1, C_2, C_3, C_4 из OSTN-диаграммы, показанной на рис. 6, имеют вид

$$p_1(t+1)=(1-a_1)\cdot p_1(t)+a_4\cdot p_3(t)+a_5\cdot p_4(t),\ p_1(0)=1;$$
$$p_2(t+1)=(1-a_2-a_3)\cdot p_2(t)+a_1\cdot p_1(t),\ p_2(0)=0;$$
$$p_3(t+1)=(1-a_4)\cdot p_3(t)+a_2\cdot p_2(t),\ p_3(0)=0;$$
$$p_4(t+1)=(1-a_5)\cdot p_4(t)+a_3\cdot p_2(t),\ p_4(0)=0.$$

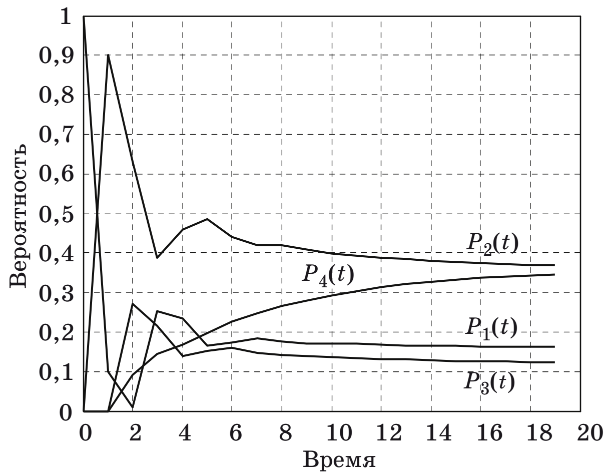
Матрица этих уравнений

$$\mathbf{A}=\begin{bmatrix}1-a_1 & 0 & a_4 & a_5\\ a_1 & 1-a_2-a_3 & 0 & 0\\ 0 & a_2 & 1-a_4 & 0\\ 0 & a_3 & 0 & 1-a_5\end{bmatrix}.$$

Для стохастической матрицы выполним расчет стационарных вероятностей:

```
>> syms a4 a5
>> A=[1-a1 0 a4 a5; a1 1-a2-a3 0 0; 0 a2 1-a4 0; 0 a3 0 1-a5]
A =
[ 1 - a1,           0,     a4,     a5]
[     a1, 1 - a3 - a2,      0,      0]
[      0,          a2, 1 - a4,      0]
[      0,          a3,      0, 1 - a5]
>> [L,V]=eig(A);
>> V(1,1)=1
>> P=factor(L(:,1)/sum(L(:,1)))
P =
(a5*(a2 + a3)*a4)/(a1*a2*a5 + a1*a3*a4 + a1*a4*a5 + a2*a4*a5 + a3*a4*a5)
(a5*a1*a4)/(a1*a2*a5 + a1*a3*a4 + a1*a4*a5 + a2*a4*a5 + a3*a4*a5)
(a2*a5*a1)/(a1*a2*a5 + a1*a3*a4 + a1*a4*a5 + a2*a4*a5 + a3*a4*a5)
(a1*a3*a4)/(a1*a2*a5 + a1*a3*a4 + a1*a4*a5 + a2*a4*a5 + a3*a4*a5)
```

Графики вероятностей четырех состояний при интенсивностях a_1 = 0,9, a_2 = 0,3, a_3 = 0,1, a_4 = 0,9, a_5 = 0,1 представлены на рис. 8.



■ *Рис. 8.* Вероятности четырех состояний марковского процесса системы функционального диагностирования

Относительно стационарной вероятности состояния C_1 можно сделать вывод, аналогичный предыдущим выводам для процессов с тремя и пятью состояниями.

Приведенные графики вероятностей состояний процессов диагностирования динамической системы и формулы стационарных режимов являются результатами анализа организации тестового и функционального диагностирования. Этот анализ позволяет оценить качество предлагаемой организации и определяет проектирование соответствующих систем технического диагностирования.

Заключение

Изложен подход к анализу организации тестового и функционального контроля ДС, основанный на использовании PFDD- и OSTN-диаграмм технологической модели системы. Показано, что для рассматриваемых диаграмм ДС могут быть построены рекуррентные уравнения марковских процессов с различным числом состояний. Предлагаемая процедура вероятностного анализа тестового и функционального диагностирования приводит к временным графикам марковских процессов и расчетным формулам, позволяющим получить стационарные вероятности состояний исследуемых процессов в организации диагностирования ДС. Результаты компьютерных расчетов показывают эффективность такого анализа и возможность использования его при проектировании систем технического диагностирования.

Работа поддержана грантами РФФИ 14-08-00327 и 14-08-00399.

Литература

1. **Бритов Г. С.** Метод тестового диагностирования линейных динамических систем // Информационно-управляющие системы. 2015. № 1. С. 77–85. doi:10.15217/issn1684-8853.2015.1.77
2. **Бритов Г. С.** Верификация, валидация и тестирование компьютерных моделей линейных динамических систем // Информационно-управляющие системы. 2013. № 2. С. 75–83.
3. **Воронин В. В., Шалобанов С. С.** Диагностирование непрерывных динамических систем методом пробных отклонений параметров модели // Информатика и системы управления. 2010. № 1(23). С. 121–127.
4. **Бритов Г. С., Мироновский Л. А.** Расчет тестового режима линейных систем управления // Приборы и системы. Управление, контроль, диагностика. 2006. № 11. С. 44–49.
5. **Мироновский Л. А.** Тестовый контроль линейных систем управления // Информационно-управляющие системы на железнодорожном транспорте. 2005. № 5. С. 3–8.
6. **Мироновский Л. А.** Диагностирование линейных систем методом комплементарного сигнала // Приборы и системы. Управление, контроль, диагностика. 2002. № 5. С. 52–57.
7. **Мироновский Л. А.** Диагностирование систем управления методом аннулирующего сигнала // Информационно-управляющие системы на железнодорожном транспорте. 2001. № 5. С. 3–7.
8. **Егоров А. Н., Мироновский Л. А.** Использование нулей динамических систем в задачах технической диагностики // Электронное моделирование. 1996. № 6. С. 34–42.
9. **Архангельский О. И., Мироновский Л. А.** Диагностирование динамических систем с помощью операторных норм // Электронное моделирование. 1995. № 5. С. 40–49.
10. **Мироновский Л. А.** Тестовый контроль передаточных функций стационарных объектов // Изв. вузов. Приборостроение. 1984. № 10. С. 22–26.
11. **Бритов Г. С., Мироновский Л. А.** Автоматизированное проектирование устройств функционального диагностирования // Информационно-управляющие системы. 2010. № 2. С. 55–61.
12. **Shumsky A. Y., Zhirabok A. N.** Nonlinear Diagnostic Filter Design: Algebraic and Geometric Points of View // Int. J. Appl. Math. Comput. Sci. 2006. Vol. 16. N 1. P. 115–127.
13. **Мироновский Л. А.** Функциональное диагностирование динамических систем. — М.: Изд-во МГУ, 1998. — 340 с.
14. **Жирабок А. Н.** Функциональное диагностирование на основе соотношений паритета // Автоматика и Телемеханика. 1998. № 2. С. 133–143.

15. **Бритов Г. С.** Метод формального описания PFDD-диаграмм IDEF3-технологии // Информационно-управляющие системы. 2014. № 2. С. 25–32.
16. **Бритов Г. С.** Методы построения OSTN-диаграмм IDEF3-технологии // Системный анализ и логистика. 2014. Вып. 11. С. 3–12.
17. **Бритов Г. С.** IDEF3-модели систем технического диагностирования // Научная сессия ГУАП: сб. докл. в 3 ч. Ч. 1: Технические науки. СПб.: ГУАП, 2014. С. 8–16.

UDC 621.38
doi:10.15217/issn1684-8853.2016.1.93

Analyzing the Process of Dynamic System Diagnostics by OSTN Diagrams of IDEF3 Technology

Britov G. S.[a], PhD, Tech., Associate Professor, britovgs@gmail.com
[a]Saint-Petersburg State University of Aerospace Instrumentation, 67, B. Morskaia St., 190000, Saint-Petersburg, Russian Federation

Introduction: Most available works on dynamic system diagnostics simply discuss the diagnostic methods like parametric identification, use of state observers, analytical redundancy, etc. However, they do not analyze the technological process of the diagnostics. **Purpose:** The goal is to analyze the process of dynamic system diagnostics, obtaining the probabilities of the process states on the base of OSTN diagrams of IDEF3 technology. We also should find a way to formalize the analysis of IDEF3 models by OSTN diagrams. According to this analysis, the probabilities of the states of the diagnostics process should be calculated. **Methods:** Using the theory of Markov processes, we have built recurrence equations for probabilities of states encountered during the diagnostics of dynamic systems. **Results:** A method is developed to build a stochastic matrix of recurrence equations with known initial conditions. Using PFDD diagrams developed for dynamic systems of diagnostics, OSTN diagrams of IDEF3 technology are built. These diagrams allow you to fully describe the technological process of the diagnostics and to define its states. **Conclusion:** The obtained results help to evaluate the quality of the technological process of diagnosing dynamic systems. In the article, test and functional diagnostic systems have been evaluated.

Keywords — IDEF3 Technology Diagrams, System States, Transition Intensities, Stochastic Matrix, Stationary Probabilities of States.

References

1. Britov G. S. A Method for Testing Linear Dynamic Systems. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2015, no. 1, pp. 77–85 (In Russian). doi:10.15217/issn1684-8853.2015.1.77
2. Britov G. S. Verification, Validation and Testing of Computer Models of Linear Dynamic Systems. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2013, no. 2, pp. 75–83 (In Russian).
3. Voronin V. V., Shalobanov S. S. Diagnosing Continuous Dynamic Systems by the Method of Test Deviations of Parameters of the Model. *Informatika i sistemy upravleniia*, 2010, no. 1(23), pp. 121–127 (In Russian).
4. Britov G. S., Mironovskii L. A. Estimation Testing Procedure of Linear Control Systems. *Pribory i sistemy. Upravlenie, kontrol', diagnostika*, 2006, no. 11, pp. 44–49 (In Russian).
5. Mironovskii L. A. Test Checking of Linear Control Systems. *Informatsionno-upravliaiushchie sistemy na zheleznodorozhnom transporte*, 2005, no. 5, pp. 3–8 (In Russian).
6. Mironovskii L. A. Diagnosis of Linear Systems by Complementary Signal Method. *Pribory i sistemy. Upravlenie, kontrol', diagnostika*, 2002, no. 5, pp. 52–57 (In Russian).
7. Mironovskii L. A. Diagnosis Control Systems by Invalidating Signal Method. *Informatsionno-upravliaiushchie sistemy na zheleznodorozhnom transporte*, 2001, no. 5, pp. 3–7 (In Russian).
8. Egorov A. N., Mironovskii L. A. Using Nulls of Dynamic Systems in Technical Diagnostic Tasks. *Elektronnoe modelirovanie*, 1996, no. 6, pp. 34–42 (In Russian).
9. Arkhangel'skii O. I., Mironovskii L. A. Diagnosis Dynamic Systems by Operator Norms. *Elektronnoe modelirovanie*, 1995, no. 5, pp. 40–49 (In Russian).
10. Mironovskii L. A. Test Checking of Transfer Functions of Stationary Objects. *Izvestiia vuzov. Priborostroenie*, 1989, no. 10, pp. 22–26 (In Russian).
11. Britov G. S., Mironovsky L. A. Automated Design of Functional Diagnosing Devices. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2010, no. 2, pp. 55–61 (In Russian).
12. Shumsky A. Y., Zhirabok A. N. Nonlinear Diagnostic Filter Design: Algebraic and Geometric Points of View. *Int. J. Appl. Math. Comput. Sci.*, 2006, vol. 16, no. 1, pp. 115–127.
13. Mironovskii L. A. *Funkcional'noe diagnostirovanie dinamicheskih sistem* [Functional Diagnostics of Linear Dynamic Systems]. Moscow, MGU Publ., 1998. 256 p. (In Russian).
14. Zhirabok A. N. Functional Diagnosis on the Basis of the Ratios of Parity. *Avtomatika i Telemekhanika*, 1998, no. 2, pp. 133–143 (In Russian).
15. Britov G. S. A Method for Formal Definition of PFDD-Diagrams of IDEF3-Technology. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2014, no. 2, pp. 25–32 (In Russian).
16. Britov G. S. Methods of Construction of OSTN-diagrams of IDEF3-technology. *Sistemnyi analiz i logistika*, 2014, no. 11, pp. 3–12 (In Russian).
17. Britov G. S. IDEF3-model of the Systems of Technical Diagnostics. *Nauchnaia sessiia GUAP*. Chast' 1. *Tekhnicheskie nauki* [Scientific Session SUAE. Vol. 1. Technical Science], 2014, pp. 8–16 (In Russian).

УДК 519.2
doi:10.15217/issn1684-8853.2016.1.102

АНАЛИЗ ДОСТОВЕРНОСТИ ТЕСТОВОЙ ФОРМЫ КОНТРОЛЯ ЗНАНИЙ

Р. С. Кударов[a]**,** *канд. техн. наук, доцент*
Ю. Сутугинене[a]**,** *PhD, старший преподаватель*
[a]*Петербургский государственный университет путей сообщения Императора Александра I, Санкт-Петербург, РФ*

Цель: *разработать методику статистического анализа достоверности тестовой формы контроля знаний, отличающуюся от известных подходов наличием правила вычисления критериальных баллов теста по эмпирическим данным экзаменационных оценок.* ***Методы:*** *предложен вероятностно-статистический подход к проверке достоверности теста в виде методики последовательного выборочного анализа, позволяющей выносить суждения на заданном уровне значимости.* ***Результаты:*** *разработана методика оценки достоверности теста, включающая в себя корреляционный анализ результатов тестирования и экзаменационных оценок, проверку гипотезы о виде закона распределения результатов тестирования и построение правила вычисления критериальных баллов на основе эмпирических частот экзаменационных оценок.* ***Практическая значимость:*** *разработанная методика способствует научно обоснованному внедрению тестов в учебный процесс, показано ее успешное применение для оценки достоверности тестов по дисциплине «Методика преподавания русского языка как иностранного (РКИ)».*

Ключевые слова *— достоверность теста, метод выборочного исследования, корреляционный анализ, критерий хи-квадрат.*

Введение

В настоящий момент в образовании сложилась ситуация, обеспечивающая сосуществование многообразных форм и методов контроля. Это устные опросы, письменные контрольные и самостоятельные работы, рефераты, курсовые, лабораторные и проектные работы, дневниковые записи, журналы наблюдений, коллоквиумы, семинары и индивидуальные собеседования, зачеты и экзамены. Каждая из форм имеет свои особенности, достоинства и недостатки, в заметной степени влияющие на результаты контроля. С начала XX в. к этим формам добавились тестовые методы, которые в разных странах то применялись, то отвергались, в зависимости от общей ситуации и числа сторонников этого метода. Дополнительно отметим, что широкое распространение компьютерных технологий привело к формированию методов оценки знаний, используемых в различных дистанционных (электронных) системах образования, например в системе Moodle.

В системе высшей школы на сегодняшний день все более востребованными и актуальными средствами контроля знаний учащихся выступают педагогические тесты. В отличие от традиционных средств контроля (устных и письменных экзаменов и зачетов по билетам, контрольных работ и пр.), именно педагогические тесты имеют научно обоснованные критерии качества [1, 2]. Кроме того, тесты обладают очевидными преимуществами перед другими методами педагогического контроля. В первую очередь, это эффективность и мобильность контроля при помощи емких тестовых заданий, что выражается в минимальных затратах времени, усилий и средств, при этом индивидуализация контроля позволяет справедливо оценить каждого отдельного испытуемого по единым критериям [3]. Объективность и качественность оценивания достигаются за счет стандартизации процедуры тестирования, а также стандартизации и проверки показателей качеств заданий и тестов в целом. Отметим, что использование тестовых методик исключает возможность возникновения конфликтов из-за несогласия с экзаменационной оценкой, способствуя тем самым установлению доброжелательных отношений как между студентами, так и между преподавателем и студентами.

Однако внедрение тестов вместо контроля по экзаменационным билетам приобретает закономерный характер лишь при установленной высокой связи результатов теста с экзаменационными оценками. Следовательно, вопрос проверки достоверности такой замены становится особенно актуальным.

В настоящем исследовании сопоставление результатов двух контрольных процедур осуществлено на примере итогового экзамена по дисциплине «Методика преподавания русского языка как иностранного (РКИ)», проводимого на кафедрах русского языка в вузах Литвы (Литовского эдукологического университета, Вильнюсского университета), выпускающих бакалавров, будущих учителей РКИ. Всего в эксперименте приняли участие 100 студентов.

В преддверии данного экзамена, предполагающего ответы испытуемых на экзаменационные вопросы, студенты прошли процедуру контроля по апробированным тестам. Таким образом был достигнут одинаковый уровень подготовки учащихся и к экзамену, и к тестированию, что в итоге позволило получить более точные результаты для дальнейшего их анализа.

Вероятностно-статистический подход к проверке достоверности теста

С математической точки зрения выстроим понятие достоверности разработанного теста в виде двух взаимодополняющих утверждений:

1. Результаты тестирования и экзаменационные оценки одной и той же группы студентов связаны прямой линейной связью.
2. Правило преобразования результатов тестирования в пятибалльную (или десятибалльную) систему учитывает реальное распределение экзаменационных (зачетных) оценок.

Первое утверждение позволяет избежать абсурдной ситуации, когда успешно выполнившие тест студенты получают низкие оценки на экзамене/зачете или наоборот. Следует добиться такой структуры и содержания теста, чтобы отклонение результатов тестирования от экзаменационных оценок в этом смысле стало незначимым.

Второе утверждение дает возможность избежать неопределенности и искажения информации при переходе от шкалы результатов тестирования к десятибалльной системе оценок. Правило преобразования баллов в оценки, сконструированное с учетом экспериментальных данных, максимально приближает идею тестирования к классическим процедурам проведения экзамена.

Если относительно разработанного теста имеют место оба утверждения, то этот тест будем считать достоверным: количественная констатация уровня знаний испытуемых (ранжирование), проводимая с помощью тестирования, аналогична работе экзаменатора. Конечно же, в этом случае полностью исключаются влияния случайности выбора билета с вопросами и «человеческого фактора» как в отношении студента, так и самого преподавателя.

Далее в работе будем полагать, что результат тестирования студента есть случайная величина X, а экзаменационная оценка того же студента по той же дисциплине есть случайная величина Y. Такое допущение оправдано тем, что на результаты тестирования и на экзаменационные оценки студентов оказывает влияние множество факторов, и ни при каких условиях точно предугадать результаты заранее нельзя.

На традиционном экзамене фактические знания оцениваются по умению их преподнести. Здесь сказывается уровень развития устной речи, активный словарный запас, кругозор, коммуникабельность, стрессоустойчивость и многие другие личностные качества экзаменуемого. Кроме того, влияние оказывает сложившееся впечатление экзаменатора о работоспособности обучающегося в течение учебного семестра. Не следует исключать и случайность выбора теоретических вопросов и выдачи практических задач на экзамене.

К факторам, влияющим на результаты тестирования, относятся следующие умения экзаменуемого: способность четко выполнять задания, понимать их формулировку, строго и грамотно оформлять свою работу; сосредоточенность и внимание (невозможность исправлять ошибки). Только при хорошей самоорганизации обучающегося возможна длительная работа над тестом, содержащим определенное количество заданий по разным темам.

Вместе с тем, независимо от формы контроля знаний, результат аттестационной процедуры всегда зависит от ряда обстоятельств, не относящихся к учебному процессу, но влияющих на психофизиологическое состояние студента (ситуация в личной жизни).

Случайный характер результатов теста и экзаменационных оценок дает возможность применять классические методы теории вероятностей и математической статистики.

Методы теории вероятностей и математической статистики, применяемые для обоснования достоверности разработанного теста

При исследовании связи между результатами тестирования и экзаменационными оценками целесообразно применять следующие элементы корреляционного анализа: вычисление выборочного коэффициента корреляции [4], оценку значимости выборочного коэффициента корреляции [4], вычисление коэффициента детерминации [5].

Вид закона распределения результатов тестирования рекомендуется устанавливать по критерию хи-квадрат Пирсона [6–9].

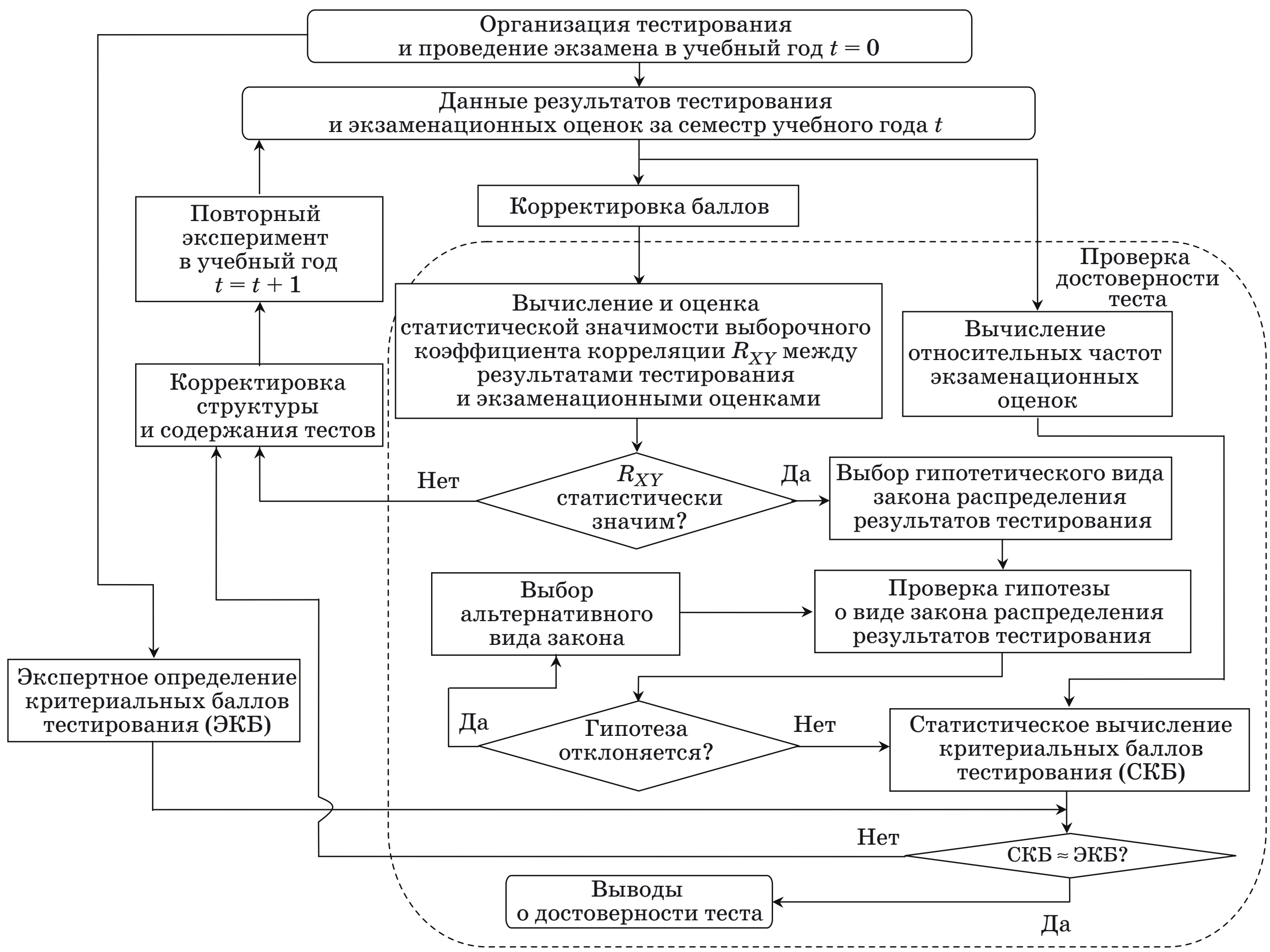
Удовлетворительные результаты использования описанных выше статистических методов позволяют построить правило вычисления критериальных баллов на основе сопоставления экспериментальных частот экзаменационных оценок с установленным теоретическим законом результатов теста.

Частоты экзаменационных оценок n_i представляют собой количество студентов, сдавших экзамен на i баллов. Следовательно, относительная частота n_i/n (где n — количество студентов) есть доля сдавших экзамен на i баллов студентов от общего их числа.

■ ***Таблица 1.*** Эмпирические критериальные баллы

ЭО* i	Частота ЭО	Относительная частота ЭО	Накопленная относительная частота ЭО	Квантиль закона	Критериальный балл
Пятибалльная экзаменационная оценка					
2	n_2	n_2/n	$k_2 = n_2/n$	x_{k2}	—
3	n_3	n_3/n	$k_3 = n_2/n + n_3/n$	x_{k3}	$[x_{k3}]$
4	n_4	n_4/n	$k_4 = n_2/n + n_3/n + n_4/n$	x_{k4}	$[x_{k4}]$
5	n_5	n_5/n	$k_5 = n_2/n + n_3/n + n_4/n + n_5/n$	x_{k5}	$[x_{k5}]$
Десятибалльная экзаменационная оценка					
4	n_4	n_4/n	$k_4 = n_4/n$	x_{k4}	—
5	n_5	n_5/n	$k_5 = n_4/n + n_5/n$	x_{k5}	$[x_{k5}]$
...	...	...	...	...	...
9	n_9	n_9/n	$k_9 = n_4/n + n_5/n + \ldots + n_9/n$	x_9	$[x_{k9}]$
10	n_{10}	n_{10}/n	1	—	$[x_{k10}]$

* ЭО — экзаменационная оценка.



■ ***Рис. 1.*** Блок-схема статистической проверки достоверности тестовой формы контроля знаний

Используя накопленные эмпирические относительные частоты $k_i = \sum_{r=4}^{i} n_r / n$, эту закономерность распределения экзаменационных оценок перенесем на теоретический закон результатов тестирования с помощью понятия квантиля x_{ki} [10]:

$$P(X < x_{ki}) = k_i,$$

где X — случайная величина результата теста (в баллах); x_{ki} — квантиль порядка k_i.

Тогда округленный до целого значения квантиль $[x_{ki}]$ будет представлять собой нижний критериальный балл для оценки i. То есть предлагается правило, которое работает следующим образом: вероятность того, что случайно выбранный результат теста будет меньше критериального балла $[x_{ki}]$, равна накопленной частоте k_i. Тем самым ранжирование испытуемых по знаниям, произведенное с помощью тестирования, оказывается близким к аналогичному ранжированию на экзамене.

Вычисление критериальных баллов на основе эмпирических данных экзаменационных оценок с округлением до целого числа по двум экзаменационным оценкам представляется в виде табл. 1.

В настоящем подходе рекомендуется согласовывать статистически полученные критериальные баллы $[x_{ki}]$ с экспертными критериальными баллами (ЭКБ$_i$) согласно следующему соотношению:

$$[x_{ki}] \in (\text{ЭКБ}_i - 10\% \cdot x_{\max}; \text{ЭКБ}_i + 10\% \cdot x_{\max}),$$

где $[x_{ki}]$ — округленный до целого значения нижний критериальный балл для оценки i; (ЭКБ$_i$) — нижний экспертный критериальный балл для оценки i; $x_{\max}$ — максимальный скорректированный результат теста.

Представим вероятностно-статистический подход к проверке достоверности теста с помощью описанных выше статистических процедур в виде блок-схемы (рис. 1).

Реализация разработанной методики к тестам по дисциплине «Методика преподавания РКИ»

Исходные данные скорректированных баллов тестирования и экзаменационных оценок сведем в табл. 2.

Вычислив средние выборочные величины $\overline{x} = 18{,}7$ и $\overline{y} = 6{,}63$, получим значение выборочного коэффициента корреляции: $R_{XY} = 0{,}87$.

Найденное значение $R_{XY} = 0{,}87$ близко к единице и свидетельствует о высокой степени прямой линейной связи между скорректированными баллами и экзаменационными оценками студентов, прошедших контрольные испытания.

Поскольку выборочное значение $R_{XY} = 0{,}87$ больше критического $R_{0{,}05} = 0{,}80$, то принимается предположение о статистической значимости выборочного коэффициента корреляции: можно утверждать, что связь между скорректированными баллами и экзаменационной оценкой близка к прямой линейной.

Учитывая очевидную тематическую связь между процедурой тестирования и экзаменом, а также их непосредственную последовательность во времени, имеет смысл проведенный анализ дополнить вычислением коэффициента детерминации: $R_{XY}^2 = (R_{XY})^2 = (0{,}87)^2 = 0{,}76$.

Коэффициент детерминации показывает, что вариация экзаменационных оценок на 76 % обусловливается результатами тестирования, что может свидетельствовать о высокой степени аналогичности двух аттестационных процедур.

При использовании *критерия* χ^2 выдвинем нулевую гипотезу H_0, состоящую в том, что статистические данные соответствуют нормальному закону распределения. В качестве альтернативной гипотезы H_1 примем предположение о том, что статистические данные не соответствуют нормальному закону распределения.

Произведем группировку вариационного ряда таким образом, чтобы в каждой группе оказались примерно равные частоты. Результаты равномерной группировки скорректированных баллов приведем в табл. 3.

По таблице интегральной функции $\Phi_0(x)$ Лапласа восстановим теоретические значения вероятностей $p_j = \Phi_0\left(\frac{x_{j,\max} - \overline{x}}{s}\right) - \frac{1}{2}$ нормального распределения $N(18{,}70;159{,}71)$. Обозначив кратко $b_j = \frac{x_{j,\max} - \overline{x}}{s}$, сведем результаты расчетов в табл. 4.

Вычисленные частоты n_j и приведенные вероятности np_j нормального закона запишем в табл. 5.

По данным табл. 5 вычислим расчетную величину хи-квадрат:

$$\chi^2_{\text{расчет}} = \sum_{j=1}^{8} \frac{(n_j - np_j)^2}{np_j} =$$

$$= \frac{(14 - 13{,}392)^2}{13{,}392} + \ldots + \frac{(12 - 6{,}339)^2}{6{,}339} = 9{,}72.$$

На уровне статистической значимости $\alpha = 0{,}05$, при количестве групп $k = 8$ и числе параметров гипотетического закона $l = 2$ по таблице [9] квантилей распределения хи-квадрат найдем критическую величину $\chi^2_{1-\alpha}(k - l - 1)$:

$$\chi^2_{1-0{,}05}(8 - 2 - 1) = \chi^2_{0{,}95}(5) = 11{,}07.$$

Поскольку расчетное $\chi^2_{\text{расчет}} = 9{,}72$ не превосходит критическое $\chi^2_{0{,}95}(5) = 11{,}07$, нет основа-

■ ***Таблица 2.*** Исходные данные скорректированных баллов (СБ) и экзаменационных оценок (ЭО)

№ п/п i	СБ x_i	ЭО y_i	№ п/п i	СБ x_i	ЭО y_i	№ п/п i	СБ x_i	ЭО y_i	№ п/п i	СБ x_i	ЭО y_i
1	–8,75	4	26	8,8	6	51	18	5	76	26,5	8
2	–8	4	27	11,3	4	52	18	7	77	27	8
3	–8	4	28	11,7	6	53	18,2	7	78	28,5	7
4	–5,5	4	29	11,4	6	54	18,3	5	79	28,8	8
5	–4	4	30	11,5	5	55	18,3	7	80	30,8	8
6	–3	4	31	11,5	6	56	18,8	7	81	31,8	8
7	–2,7	5	32	11,7	6	57	19,7	7	82	31,8	7
8	3,7	4	33	11,5	6	58	19,7	7	83	32,6	8
9	4,1	5	34	11,8	6	59	20	6	84	32,7	8
10	4,3	6	35	12,3	6	60	20,6	7	85	32,8	8
11	4,5	4	36	12,5	7	61	21,6	7	86	32,9	9
12	4,5	5	37	13,3	6	62	21,7	7	87	37,55	8
13	4,7	5	38	14	6	63	22,3	7	88	37,8	9
14	4,7	6	39	14,3	6	64	22,3	6	89	38,2	6
15	4,8	5	40	14,5	7	65	22,8	7	90	38,7	8
16	5	5	41	14,7	6	66	23	7	91	39,2	9
17	5,3	4	42	15	6	67	23,6	7	92	39,3	9
18	5,5	5	43	15	9	68	23,9	7	93	39,5	10
19	5,6	5	44	16,2	6	69	24,7	7	94	40,3	9
20	6,2	5	45	16,2	6	70	25	5	95	40,5	9
21	6,5	5	46	16,8	8	71	25,3	8	96	41	9
22	6,7	5	47	17,2	7	72	25,5	8	97	41,8	10
23	6,7	8	48	17,2	7	73	25,6	8	98	50,5	9
24	7,7	5	49	17,5	7	74	25,7	8	99	50,7	10
25	7,8	6	50	17,8	7	75	26,5	7	100	54	10

■ ***Таблица 3.*** Частоты равномерно сгруппированных скорректированных баллов

Показатель	j							
	1	2	3	4	5	6	7	8
$x_{j,\min}$, нижняя граница интервала	$-\infty$	4,70	9,50	14,50	18,20	23,00	28,50	38,00
$x_{j,\max}$, верхняя граница интервала	4,70	9,50	14,50	18,20	23,00	28,50	38,00	$+\infty$
n_j, частота скорректированного балла	14	13	13	13	13	12	10	12

■ ***Таблица 4.*** Вероятности нормального закона распределения

Показатель	j							
	1	2	3	4	5	6	7	8
b_j	–1,108	–0,728	–0,340	–0,056	0,300	0,617	1,527	$+\infty$
$\Phi_0(b_j)$	–0,366	–0,267	–0,133	–0,022	0,118	0,231	0,437	0,5
$p_j = \Phi_0(b_{j-1}) - \Phi_0(b_j)$	0,134	0,099	0,134	0,111	0,140	0,113	0,205	0,063

■ ***Таблица 5.*** Приведенные вероятности кривой нормального закона распределения

Показатель	j							
	1	2	3	4	5	6	7	8
n_j	14	13	13	13	13	12	10	12
np_j	13,392	9,931	13,350	11,107	14,026	11,329	20,524	6,339

■ ***Таблица 6.*** Частоты и относительные частоты вариантов скорректированных баллов

Показатель	j						
	4	5	6	7	8	9	10
n_j, частота экзаменационной оценки	10	16	21	24	16	9	4
n_j / n, относительная частота экзаменационной оценки	0,1	0,16	0,21	0,24	0,16	0,09	0,04

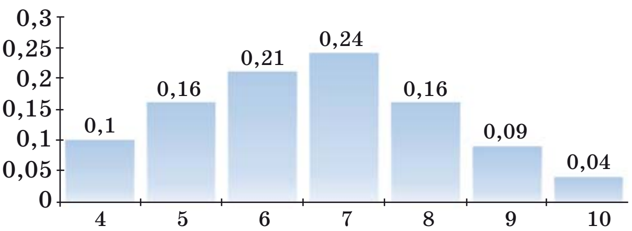
■ ***Таблица 7.*** Критериальные баллы

ЭО	Частота ЭО	Относительная частота ЭО	Накопленная относительная частота ЭО	Квантиль нормального закона N(18,90;159,71)	Критериальный балл
4	10	0,1	0,1	2,51	—
5	16	0,16	0,26	10,57	3
6	21	0,21	0,47	17,75	11
7	24	0,24	0,71	25,70	18
8	16	0,16	0,87	32,94	26
9	9	0,09	0,96	40,83	33
10	4	0,04	1	—	41

ния отклонять гипотезу о соответствии распределения скорректированных баллов нормальному закону N(18,70;159,71).

По данным экзаменационных оценок, содержащихся в табл. 2, построим гистограмму относительных частот. Для этого вычислим частоту n_j каждой экзаменационной оценки (j принимает целые значения от 4 до 10) и вычислим их относительные частоты n_j/n (табл. 6).

Полученные относительные частоты n_j/n экзаменационных оценок изобразим на рис. 2 в виде гистограммы.



■ ***Рис. 2.*** Относительные частоты экзаменационных оценок

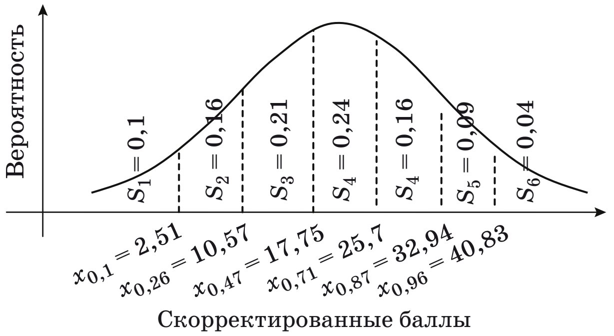
Зная относительные частоты экзаменационных оценок, получим критериальные баллы для результатов тестирования. Критериальные баллы вычислим как квантили закона распределения N(18,70;159,71) скорректированных баллов с округлением до целого числа. Графически такие квантили разбивают фигуру, заключенную под кривой нормального закона скорректированных баллов, на части, площади которых совпадают с соответствующими относительными частотами экзаменационных оценок (рис. 3).

Вычисление критериальных баллов на основе эмпирических данных экзаменационных оценок с округлением до целого числа представим в табл. 7.

Нижний критериальный балл для оценки 5, посчитанный статистически, равен 3 (т. е. $[x_{k5}]$), а экспертный критериальный балл для этой оценки равен 7,82. Максимальный результат теста $x_{\max} = 54$:

$$[x_{k5}] \in (7{,}82 - 10\%\cdot 54;\ 7{,}82 + 10\%\cdot 54),$$

$$[x_{k5}] \in (2{,}42;\ 13{,}22).$$



■ ***Рис. 3.*** Квантили нормального закона $N(18{,}70;159{,}71)$

■ ***Рис. 4.*** Частоты экзаменационных оценок и результатов теста

Нижний критериальный балл для оценки 10, посчитанный статистически, равен 41 (т. е. $[x_{k10}]$), а экспертный критериальный балл для этой оценки равен 40,74:

$$[x_{k10}] \in (40{,}74 - 10\%\cdot 54;\ 40{,}74 + 10\%\cdot 54),$$

$$[x_{k10}] \in (35{,}34;\ 46{,}14).$$

Следовательно, можно сделать вывод, что критериальные баллы, полученные статистически, согласуются с экспертными критериальными баллами ЭКБ$_i$.

Воспользуемся полученными критериальными баллами для преобразования скорректированных баллов тестирования в десятибалльную шкалу и сравним полученные результаты графически с гистограммой относительных частот экзаменационных оценок на рис. 4.

Построенное авторами правило обработки результатов тестирования демонстрирует вполне приемлемые показатели — очевидна схожесть гистограммы преобразованных баллов тестирования с гистограммой эмпирических частот экзаменационных оценок.

Таким образом, тесты по дисциплине «Методика преподавания РКИ» рекомендуются к внедрению в учебный процесс в аттестационных целях вместо экзамена по билетам, а выстроенный подход к обоснованию адекватности теста, наглядно представленный выше в виде схемы с указанием последовательных действий, успешно можно применять для разработки новых тестовых батарей.

Заключение

Разработанная методика статистического анализа достоверности тестовой формы контроля знаний значительно упрощает работу по внедрению тестов в учебный процесс. Отметим, что построенный подход согласуется с известной теорией IRT и является относительно этой теории более простым алгоритмом, позволяющим с заданной точностью и минимальными затратами добиться поставленной цели. В качестве преимуществ предлагаемого подхода относительно IRT можно отметить сравнительно низкие затраты на его введение в учебный процесс: формирование команды специалистов по внедрению подхода, ожидание вычислительных процедур обработки результатов, наглядность результатов внедрения подхода и др.

Сама форма представления методики допускает автоматизацию. В перспективе предполагается создание соответствующего компьютерного программного обеспечения. Отличие такого программного обеспечения от существующих прикладных статистических пакетов (SPSS, Statistica др.) заключается в легкости его освоения, а также в минимальных требованиях к пользователям и технике. Умение по входным данным (результатам тестирования и экзаменационным оценкам) оперативно получать информацию о достоверности теста, вычислять критериальные баллы теста, полностью согласующиеся с границами экзаменационных оценок, представляет неоценимую пользу при внедрении любого банка тестов в различные учебные заведения.

Литература

1. **Аванесов В. С.** Основы научной организации педагогического контроля в высшей школе. — М.: НИТУ «МИСиС», 1989. — 167 с.
2. **Анастази А., Урбина С.** Психологическое тестирование. — СПб.: Питер, 2006. — 688 с.
3. **Ким В. С.** Тестирование учебных достижений: монография. — Уссурийск: Изд-во УГПИ, 2007. — 214 с.
4. **Кобзарь А. И.** Прикладная математическая статистика. Для инженеров и научных работников. — М.: Физматлит, 2006. — 816 с.
5. **Айвазян С. А., Енюков И. С., Мешалкин Л. Д.** Прикладная статистика: Исследование зависимос-

тей: справ. изд. / под ред. С. А. Айвазяна. — М.: Финансы и статистика, 1985. — 487 с.
6. **Вадзинский Р. Н.** Справочник по вероятностным распределениям. — СПб.: Наука, 2001. — 295 с.
7. Handbook of Applicable Mathematics. Vol. VI: Statistics. Part A. /Ch. Ed. Walter Ledermann, Ed. Emlyn Lloyd. — John Wiley & Sons, 1984. — 366 p.
8. Handbook of Applicable Mathematics. Vol. VI: Statistics. Part B. /Ch. Ed. Walter Ledermann, Ed. Emlyn Lloyd. — John Wiley & Sons, 1984. — 522 p.
9. **Большев Л. Н., Смирнов Н. В.** Таблицы математической статистики. — М.: Наука. Гл. ред. физ.-мат. лит., 1983. — 416 с.
10. Вероятность и математическая статистика: энциклопедия / под ред. Ю. В. Прохорова. — М.: Большая Российская энциклопедия, 2003. Репр. изд. — 912 с.

UDC 519.2
doi:10.15217/issn1684-8853.2016.1.102

Analysis of Validity of Test Forms of Knowledge Control

Kudarov R. S.[a], PhD, Tech., Associate Professor, krs.83@mail.ru
Sutuginiene J.[a], PhD, Senior Lecturer, sutuginiene@mail.ru
[a]Petersburg State Transport University, 9, Moskovsky Pr., 190031, Saint-Petersburg, Russian Federation

Purpose: The goal is to develop a technique for statistical analysis of validity of the test form of knowledge control which would differ from the available approaches by having a rule for cut-off score calculation according to the empirical data of the exam scores. **Methods:** A statistical-probabilistic approach was proposed for test validity control as a method of selective analysis which allows you to make judgements according to the specified level of significance. **Results:** A technique has been developed for test validity evaluation. It includes correlation analysis of the test results and exam scores, verification of the hypothesis of the test results distribution law and the construction of the rule of calculating cut-off scores based on the empirical frequency of the exam scores. **Practical relevance:** The developed technique promotes scientifically grounded introduction of test forms into the educational process. The article demonstrates the successful usage of this technique for the evaluation of test validity in the course ‘Didactics and Teaching Russian as a Foreign Language’.

Keywords — Test Validity, Selective Research, Correlation Analysis, Chi-Square Criterion.

References

1. Avanesov V. S. *Osnovy nauchnoi organizatsii pedagogicheskogo kontrolia v vysshei shkole* [The Basis of Scientific Organization of Pedagogical Control in Higher School]. Moscow, Natsional’nyi issledovatel’skii tekhnologicheskii universitet «MISiS» Publ., 1989. 167 p. (In Russian).
2. Anastazi A., Urbina S. *Psikhologicheskoe testirovanie* [Psychological Testing]. Saint-Petersburg, Piter Publ., 2006. 688 p. (In Russian).
3. Kim V. S. *Testirovanie uchebnykh dostizhenii* [Testing of the Education Achievements]. Ussuriysk, UGPI Publ., 2007. 214 p. (In Russian).
4. Kobzar A. I. *Prikladnaia matematicheskaia statistika* [Applied Mathematical Statistics]. Moscow, Fizmatlit Publ., 2006. 816 p. (In Russian).
5. Aivazian S. A., Eniukov I. S., Meshalkin L. D. *Prikladnaia statistika: Issledovanie zavisimostei* [Applied Statistics: Dependence Research]. Moscow, Finansy i statistika Publ., 1985. 487 p. (In Russian).
6. Vadzinskii R. N. *Spravochnik po veroiatnostnym raspredeleniiam* [Reference Book on Probabilistic Distribution]. Saint-Petersburg, Nauka Publ., 2001. 295 p. (In Russian).
7. *Handbook of Applicable Mathematics. Vol. VI. Statistics. Part A.* Ch. Ed. Walter Ledermann, Ed. Emlyn Lloyd. John Wiley & Sons, 1984. 366 p.
8. *Handbook of Applicable Mathematics. Vol. VI. Statistics. Part B.* Ch. Ed. Walter Ledermann, Ed. Emlyn Lloyd. John Wiley & Sons, 1984. 522 p.
9. Bol’shev L. N., Smirnov N. V. *Tablitsy matematicheskoi statistiki* [Tables of Mathematical Statistics]. Moscow, Nauka Publ., 1983. 416 p. (In Russian).
10. *Veroiatnost’ i matematicheskaia statistika: Entsiklopediia* [Probability and Mathematical Statistics: Encyclopaedia]. Ed. by Iu. V. Prokhorov. Moscow, Bol’shaia Rossiiskaia entsiklopediia Publ., 2003. 912 p. (In Russian).

УДК 004.45
doi:10.15217/issn1684-8853.2016.1.110

ВОПРОСЫ ТЕРМИНОЛОГИИ. КАК ПРАВИЛЬНО: ПРАВА И/ИЛИ РАЗРЕШЕНИЯ?

А. В. Гордеев[a], *доктор техн. наук, профессор*
[a]*Санкт-Петербургский государственный университет аэрокосмического приборостроения, Санкт-Петербург, РФ*

Введение: *в абсолютном большинстве публикаций, в том числе в учебной литературе, термины «права» и «разрешения», которые используются при описании методов управления доступом и конкретных правил доступа к информационным ресурсам в информационных системах, отождествляются, т. е. авторы не видят между ними различий, что приводит к тому, что достаточно часто существенно изменяется и искажается смысл.* ***Цель:*** *анализ понятий «права» и «разрешения» с определением областей их корректного применения.* ***Результаты:*** *анализ происхождения этих терминов, их истинного значения и механизмов реализации в информационных системах показал, что термин «права» непосредственно связан с понятием авторизации, и его следует использовать для обозначения полномочий пользователя, а термин «разрешения» — для описания тех действий, которые можно или нельзя применять по отношению к конкретным экземплярам информационных объектов.* ***Практическая значимость:*** *корректное использование терминов «права» и «разрешения» не только повысит культурный уровень специалистов в области информационных систем, но и приведет к более полному взаимопониманию при общении и точности изложения особенностей информационных систем.*

Ключевые слова *— права, разрешения, терминология, управление доступом.*

Введение

В литературе по POSIX-системам [1–3] при описании условий доступа тому или иному пользователю или процессу к некоторому файлу достаточно часто встречается словосочетание «File Rights» — права на файл. Англоязычный термин «Rights» переводится на русский язык как «права». В то же самое время при работе с современными операционными системами компании Microsoft, которые в настоящем в основном используют файловую систему NTFS, и при чтении соответствующей документации на английском языке мы сталкиваемся с использованием термина «Permissions» — «разрешения». В большинстве версий этих операционных систем, локализованных для Российской Федерации (например, в распространенных Windows 8.1 и в серверной версии Windows Server 2008), на вкладке Безопасность (Security) свойств файла можно увидеть слова «разрешения» [4–7]. Однако в технической литературе наши авторы очень часто употребляют термин «права на файлы» при описании не только POSIX-систем, но и Windows-систем. Эти же термины употребляют и при описании условий доступа к другим информационным ресурсам. Поэтому следует, разобрав эти термины, понять, что и когда лучше использовать.

Права на файлы в POSIX-системах

Как известно, к POSIX-системам относятся все те операционные системы, которые удовлетворяют спецификациям POSIX [1]. Они описаны в серии стандартов, разработанных Комитетом 1003 при американском Институте инженеров по электротехнике и электронике (IEEE), и своей целью, прежде всего, имели унификацию и стандартизацию UNIX-систем. Широко известные и все более популярные системы GNU/Linux являются POSIX-системами. К POSIX-системам относятся различные BSD-системы, семейство MAC OS, семейство систем реального времени QNX и многие другие. Во всех этих системах при описании доступа к файлам принято использовать термин «права на файлы». И речь там идет о праве на чтение некоторого файла, праве на его исполнение и праве на запись в этот файл. Эти права определяются для владельца файла, для группы этого владельца и для всех остальных пользователей. Это классика, и ее знают все, кто изучал POSIX-системы.

Разрешения на доступ к файлам в Windows-системах

После завершения эпохи операционных систем Windows 9x, которые использовали файловые системы FAT32, мы в большинстве случаев сталкиваемся с тем, что нужные нам файлы располагаются на томах с NTFS. Эта файловая система появилась в 1993 г. и изначально была ориентирована на использование в корпоративном секторе, поэтому она получила развитые механизмы разграничения доступа. Система NTFS имеет несколько версий, в настоящее время мы преимущественно сталкиваемся с NTFS v.5.x. Для тех, кто изучил реальные возможности этой файловой системы и часто ее использует, должны быть очевидны такие понятия, как «атомарные разрешения», «стандартные разрешения», «специальные (или особые) разрешения» [6], хотя в современных интерфейсах работы со свойствами безопасности теперь таких словосочетаний уже не встретишь. Термин «разрешения» в компании Microsoft используют как по отношению к локальным файлам и каталогам, так и по отношению к сетевому доступу к папкам с файлами, хотя это совершенно разные разрешения.

Этот же термин — «разрешения» — используют при описании дискреционного метода управления доступом. Напомним, что это метод, при котором информационные ресурсы имеют списки управления доступом (Access Control List — ACL), хотя в теории вместо списков могут применяться матрицы. Каждый элемент списка указывает на разрешение или запрет для некоторого принципала безопасности на конкретные операции, которые потенциально можно выполнить над информационным ресурсом. Сами принципалы безопасности указываются специальными идентификаторами. Количество разрешений и их конкретный состав зависят от класса объекта (класса информационного ресурса).

Однако термин «разрешения» почти не используется ни в обиходе специалистов по информационным технологиям, ни в нашей русскоязычной специальной литературе. Даже когда речь идет о системах и продуктах компании Microsoft [7]. Вместо этого почему-то практически повсеместно используют термин «права». И даже в диссертациях авторы часто используют термин «права», в то время как по смыслу речь идет о «разрешениях». Возможно, что это происходит либо из-за неправильного (некорректного) перевода англоязычной технической литературы, либо «по инерции», поскольку UNIX-системы появились намного (на 20 лет!) раньше, и уже сложились некоторые терминологические традиции.

Понятие «права» в Windows-системах и различия между правами и разрешениями

Что касается термина «права», то в контексте систем компании Microsoft в англоязычной литературе синонимом для слова «Rights» может выступить термин «Privileges» — привилегии, льготы. О правах здесь говорят по отношению к пользователям (или их процессам). Эти права (или привилегии) определяются для локальных встроенных групп. Например, во всех современных системах Windows есть такие группы, как *Администраторы*, *Опытные пользователи*, *Операторы настройки сети*, *Пользователи*, *Пользователи удаленного рабочего стола*, *Операторы архива*, *Репликаторы*, *Читатели журнала событий*, *Гости* и некоторые другие. Любой пользователь может входить в несколько групп, и его права при этом суммируются. Максимально возможными правами обладают члены группы *Администраторы*, а минимально возможными правами обладают *Гости*. Если пользователя не включить ни в одну из вышеупомянутых встроенных локальных групп, то его права не определены, и в первых операционных системах класса NT от компании Microsoft такой пользователь даже не мог войти в систему. Для исправления этого казуса правом работы в современных системах стали обладать члены специальной группы *Прошедшие проверку* (Authenticated Users). А для тех случаев, когда в систему можно зайти без аутентификации (т. е. без ввода пароля), что предпочитает делать большинство пользователей, правом работать на компьютере стали все интерактивные пользователи.

Итак, если учетную запись некоторого пользователя поместить в ту или иную встроенную локальную группу, то он начинает обладать соответствующими *правами*. Заметим, что эти права вступают в силу для пользователя после перерегистрации его в системе, а не с момента включения в группу. Дело в том, что после прохождения аутентификации (процедуры проверки подлинности) пользователь и все его вычислительные процессы получают так называемый маркер доступа. Этот маркер включает в себя идентификатор учетной записи пользователя и идентификаторы всех тех групп, в которые входит его учетная запись. Все обращения (запросы) к операционной системе на выполнение тех или иных операций или сервисов всегда сопровождаются маркером доступа.

Чтобы понять механизм действия прав и разрешений и осознать различия между ними, рассмотрим чуть более подробно работу подсистемы безопасности с вышеупомянутыми ACL. Подсистема безопасности NT берет поочередно ACE-элементы из списка и проверяет их на совпадение с одним из принципалов безопасности, которые составляют маркер доступа. При совпадении сигнатур подсистема безопасности проверяет возможность выполнения запрошенной операции, анализируя ситуацию: на данное разрешение стоит запрет, разрешение, или это действие/разрешение в текущем ACE вообще не определено. Если в ACE на данную операцию для совпадающего принципала из маркера доступа встречается запрет, то далее не проводится никаких проверок и в запрашиваемой операции отказывается. Если запрета нет, то осуществляется дальнейший перебор элементов из маркера доступа, а затем берется следующий элемент.

В конечном итоге, мы либо получаем явный запрет на выполнение операции, либо получаем явное разрешение, либо не получаем разрешение, но при отсутствии явного запрета. Последнее означает, что можно не разрешать то или иное действие без явного запрета. Именно на этом механизме и реализуются как разрешения, так и права пользователей. С той только разницей, что, определяя разрешения на доступ к объекту, можно в явном виде запрещать выполнение той или иной операции для конкретного пользователя, в то время как при определении прав никогда не пользуются запретом, благодаря чему можно обеспечить суммирование прав пользователей.

Теперь собственно о различии разрешений и прав. Если список управления доступом принадлежит обычным файлам, а не системным утилитам или каким-либо иным средствам управления системой, то следует говорить о разрешениях. Если же список управления доступом относится к системной программе (например, оснастке управления дисками или апплету настройки сетевого интерфейса), то речь идет о правах. Другими словами, права определяются для класса объектов, а разрешения — для отдельных экземпляров или подмножеств экземпляров этого класса.

Важно отметить, что права имеют приоритет перед разрешениями. Например, есть право владения объектом (информационным ресурсом). Если пользователь является владельцем ресурса, то он может изменять ACL этого объекта, даже если в ACL для этого пользователя стоят явные запреты на редактирование списка.

Определение (назначение) прав для встроенных локальных групп осуществляется за счет указания в ACL для той или иной системной программы (утилиты) соответствующего ACE. При этом указание идентификаторов безопасности возможно именно для встроенных групп, так как эти идентификаторы известны заранее. Так, во всех ACL для системных программ указан идентификатор безопасности группы *Администраторы*. Именно поэтому они могут делать в системе все. На некоторые системные программы помимо группы *Администраторы* могут быть указаны и некоторые другие группы. Редактирование подобных ACL приводит к переопределению прав пользователей либо к явному определению прав для вновь созданных групп безопасности.

Заключение

Итак, при рассмотрении вопросов доступа к информационным ресурсам в POSIX-системах можно и нужно использовать термин «права» на доступ, хотя не будет ошибкой сказать «разрешения» на доступ. Но если мы имеем дело не с UNIX-подобными системами, то при рассмотрении вопросов доступа к тем или иным ресурсам (например, к тем же файлам) следует употреблять термин «разрешения». Именно так в основном и поступают IT-специалисты, которые говорят на английском языке (они используют термин Permissions, а не Rights).

Литература

1. **Робачевский А. М.** Операционная система UNIX. — СПб.: БХВ-Петербург, 2002. — 528 с.
2. **Немет Э., Снайдер Г., Сибасс С., Хейн Т.** UNIX: руководство системного администратора. Для профессионалов: пер. с англ. — СПб.: Питер, Издательская группа BHV, 2002. — 928 с.
3. **Федорчук А. В.** Доступный UNIX: Linux, FreeBSD, DragonFlyBSD, NetBSD, OpenBSD. — СПб.: БХВ-Петербург, 2006. — 672 с.
4. **Холме Д.** Эффективное администрирование. Ресурсы Windows Server 2008, Windows Vista, Windows XP, Windows Server 2003: пер. с англ. — М.: Русская редакция; СПб.: БХВ-Петербург, 2009. — 768 с.
5. **Моримото Р., Ноэл М., Драуби О., Мистри М.** и др. Microsoft Windows Server 2008 R2. Полное руководство: пер. с англ. — М.: Вильямс, 2012. — 1456 с.
6. **Кастер Х.** Основы Windows NT и NTFS: пер. с англ. — М.: Изд. отдел «Русская редакция «ТОО Channel Trading Ltd», 1996. — 440 с.
7. Управление правами на доступ к данным в Microsoft Office 2010. https://support.office.com/ru-ru/article/%D0%A3%D0%BF%D1%80%D0%B0%D0%B2%D0%BB%D0%B5%D0%BD%D0%B8%D0%B5-%D0%BF%D1%80%D0%B0%D0%B2%D0%B0%D0%BC%D0%B8-%D0%BD%D0%B0-%D0%B4%D0%BE%D1%81%D1%82%D1%83%D0%BF-%D0%BA-%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D0%BC-%D0%B2-Microsoft-Office-2010-c7a70797-6b1e-493f-acf7-92a39b85e30c?ui=ru-RU&rs=ru-RU&ad=RU (дата обращения: 20.05.2015).

UDC 004.45
doi:10.15217/issn1684-8853.2016.1.110

Terminology Issues: What Should We Say, Rights and/or Permissions?

A. V. Gordeyev[a], Dr. Sc., Tech., Professor, avg@aanet.ru
[a]Saint-Petersburg State University of Aerospace Instrumentation, 67, B. Morskaia St., 190000, Saint-Petersburg, Russian Federation

Introduction: In most literature, including textbooks, the terms “rights” and “permissions” are usually identified. Both these terms are used for describing access control methods and specific rules about the access to data resources in informational systems. Authors often do not see any difference between “rights” and “permissions”. As a result, the meaning can be seriously distorted. **Purpose:** In this paper, we analize the ideas behind "rights" and "permissions", defining the areas of correct usage of these terms. **Results:** The analysis of etymology of these terms, their true meaning and the ways of their realization in information systems showed that the term "rights" is directly associated with the idea of autorization and should be used to designate the user's authority, while "permissions" should be used to describe the actions which can or cannot be performed over specific information objects. **Practical relevance:** The correct usage of the terms “rights” and “permissions” will increase the cultural level of information system specialists. Moreover, it will lead to clearer mutual understanding during the communication on informational system issues.

Keywords — Rights, Permissions, Terminology, Access Control.

References

1. Robachevsky A. M. *Operatsionnaia sistema UNIX* [Operating System UNIX]. Saint-Petersburg, BKhV-Peterburg Publ., 2002. 528 p. (In Russian).
2. Nemeth E., Snyder G., Hein T. R. *UNIX and Linux System Administration Handbook*. 4th Edition. Prentice Hall, 2010. 1344 p.
3. Fedorchuk A. V. *Dostupnyi UNIX: Linux, FreeBSD, DragonFlyBSD, NetBSD, OpenBSD* [Available UNIX: Linux, FreeBSD, DragonFlyBSD, NetBSD, OpenBSD]. Saint-Petersburg, BKhV-Peterburg Publ., 2006. 672 p. (In Russian).
4. Holme Dan. *Windows Administration Resource Kit: Productivity Solutions for IT Professionals Published by Microsoft Press*. Redmond, Washington. 700 p.
5. Morimoto Rand, Noel Michael, Droubi Omar, Mistry Ross. *Microsoft Windows Server 2008 R2 Unleashed*. Sams., 2010. 1680 p.
6. Custer Helen. *Inside the Windows NT File System*. Redmond, Microsoft Press, 1994. 91 p.
7. *Upravlenie pravami na dostup k dannym v Microsoft Office 2010* [Data Access Control in Microsoft Office 2010]. Available at: https://support.office.com/ru-ru/article/%D0%A3%D0%BF%D1%80%D0%B0%D0%B2%D0%BB%D0%B5%D0%BD%D0%B8%D0%B5-%D0%BF%D1%80%D0%B0%D0%B2%D0%B0%D0%BC%D0%B8-%D0%BD%D0%B0-%D0%B4%D0%BE%D1%81%D1%82%D1%83%D0%BF-%D0%BA-%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D0%BC-%D0%B2-Microsoft-Office-2010-c7a70797-6b1e-493f-acf7-92a39b85e30c?ui=ru-RU&rs=ru-RU&ad=RU (accessed 20 May 2015).

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. Рукописи не возвращаются.

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

СВЕДЕНИЯ ОБ АВТОРАХ

**БАЛОНИН
Николай
Алексеевич**

Профессор кафедры вычислительных систем и сетей Санкт-Петербургского государственного университета аэрокосмического приборостроения.
В 1982 году окончил Ленинградский электротехнический институт им. В. И. Ульянова (Ленина) по специальности «Автоматика и телемеханика».
В 2008 году защитил диссертацию на соискание ученой степени доктора технических наук.
Является автором более 90 научных публикаций, в том числе трех монографий.
Область научных интересов — теория динамических систем, теория идентификации, теория операторов, теория матриц, вычислительные методы, интернет-робототехника, интернет-книги с исполняемыми алгоритмами, научные социальные сети.
Эл. адрес: korbendfs@mail.ru

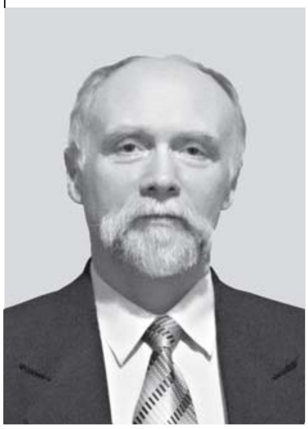
**БАРКОВСКИЙ
Евгений
Александрович**

Аспирант лаборатории информационных компьютерных технологий Института прикладных математических исследований Карельского научного центра РАН, Петрозаводск.
В 2012 году окончил Петрозаводский государственный университет по специализации «Информационные технологии в агропродовольственной сфере».
Является автором семи научных публикаций и одного учебного пособия.
Область научных интересов — задачи оптимального управления параллельными динамическими структурами данных, FIFO-очереди, деки с ограниченным входом.
Эл. адрес:
barkevgen@gmail.com

**БРИТОВ
Георгий
Семенович**

Доцент кафедры информационных систем Санкт-Петербургского государственного университета аэрокосмического приборостроения.
В 1963 году окончил Ленинградский институт авиационного приборостроения по специальности «Авиационное приборостроение».
В 1968 году защитил диссертацию на соискание ученой степени кандидата технических наук.
Является автором более 100 научных публикаций и 13 патентов на изобретения.
Область научных интересов — теория надежности и техническая диагностика информационных систем.
Эл. адрес:
bgs@ibi.metrocom.ru

**ГОРДЕЕВ
Александр
Владимирович**



Профессор кафедры вычислительных систем и сетей Санкт-Петербургского государственного университета аэрокосмического приборостроения.
В 1975 году окончил Ленинградский институт авиационного приборостроения по специальности «Электронные вычислительные машины».
В 1998 году защитил диссертацию на соискание ученой степени доктора технических наук.
Является автором более 90 научных публикаций.
Область научных интересов — моделирование параллельных вычислительных систем, распределенные вычисления, вычислительные сети.
Эл. адрес: avg@aanet.ru

**ЕГОРОВ
Игорь
Валерьевич**

Аспирант кафедры компьютерных систем и программных технологий Института компьютерных наук и технологий Санкт-Петербургского государственного политехнического университета Петра Великого.
В 2012 году окончил с отличием факультет технической кибернетики Санкт-Петербургского государственного политехнического университета по специальности «Информатика и вычислительная техника».
Является автором двух научных публикаций.
Область научных интересов — надежность вычислительных систем, программная инженерия, проектирование вычислительных систем.
Эл. адрес:
iegorov@kspt.icc.spbstu.ru

**ЖАРИНОВ
Игорь
Олегович**

Заведующий кафедрой машинного проектирования бортовой электронно-вычислительной аппаратуры Санкт-Петербургского национального исследовательского университета информационных технологий, механики и оптики, руководитель учебно-научного центра АО «ОКБ «Электроавтоматика», Санкт-Петербург.
В 2000 году окончил Санкт-Петербургский государственный университет аэрокосмического приборостроения по специальности «Радиоэлектронные системы».
В 2011 году защитил диссертацию на соискание ученой степени доктора технических наук.
Является автором более 100 научных публикаций.
Область научных интересов — проектирование бортовой вычислительной техники.
Эл. адрес: igor_rabota@pisem.net

КОСТИШИН
Максим
Олегович

Аспирант, ассистент кафедры машинного проектирования бортовой электронно-вычислительной аппаратуры Санкт-Петербургского национального исследовательского университета информационных технологий, механики и оптики, инженер АО «ОКБ «Электроавтоматика», Санкт-Петербург.
В 2012 году окончил Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики по специальности «Управление и информатика в технических системах».
Является автором 12 научных публикаций.
Область научных интересов — геоинформационные системы, системы индикации авионики.
Эл. адрес:
maksim@kostishin.com

КРУК
Евгений
Абрамович

Профессор, проректор по науке, директор Института информационных систем и защиты информации, заведующий кафедрой безопасности информационных систем Санкт-Петербургского государственного университета аэрокосмического приборостроения, заслуженный деятель науки РФ, лауреат премии Правительства Санкт-Петербурга.
В 1974 году окончил Ленинградский институт авиационного приборостроения по специальности «Автоматические системы управления».
В 1999 году защитил диссертацию на соискание ученой степени доктора технических наук.
Является автором более 120 научных публикаций.
Область научных интересов — теория кодирования, криптография.
Эл. адрес: ekrouk@vu.spb.ru

КУДАРОВ
Руслан
Серикович

Доцент кафедры математики и моделирования Петербургского государственного университета путей сообщения Императора Александра I.
В 2005 году окончил Петербургский государственный университет путей сообщения по специальности «Прикладная математика и информатика».
В 2009 году защитил диссертацию на соискание ученой степени кандидата технических наук.
Является автором 23 научных и учебно-методических публикаций.
Область научных интересов — теория вероятностей, прикладная статистика, эконометрика.
Эл. адрес: krs.83@mail.ru

ЛАХИН
Олег
Иванович

Руководитель направления, ООО «НПК «Разумные решения», Самара.
В 1998 году окончил Поволжский институт информатики, радиотехники и связи по специальности «Программное обеспечение вычислительной техники и автоматизированных систем», в 2006 году — Самарский государственный экономический университет по Федеральной программе подготовки управленческих кадров.
Является автором 31 научной публикации.
Область научных интересов — мультиагентные технологии для создания интеллектуальных систем управления ресурсами в реальном времени, использующих принципы самоорганизации и эволюции.
Эл. адрес: lakhin@yandex.ru

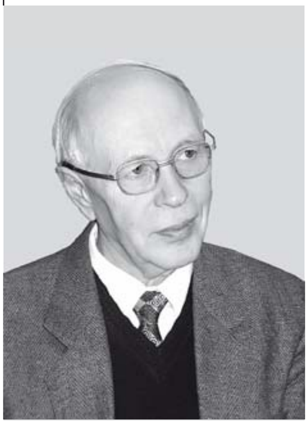
ЛЕСНЯК
Денис
Александрович

Старший преподаватель кафедры космических радиотехнических систем Военно-космической академии им. А. Ф. Можайского, Санкт-Петербург.
В 2001 году окончил Военный инженерно-космический университет им. А. Ф. Можайского по специальности «Радиоэлектронные системы».
Является автором 21 научной публикации.
Область научных интересов — защита информации, информационная безопасность.
Эл. адрес: denislesnyk@mail.ru

МАРКЕЛОВ
Владимир
Владимирович

Ведущий инженер-программист АО «ОКБ «Электроавтоматика», Санкт-Петербург.
В 1996 году окончил Балтийский государственный технический университет им. Д. Ф. Устинова по специальности «Космические аппараты и разгонные блоки».
Является автором одной научной публикации.
Область научных интересов — бортовые комплексы летательных аппаратов, навигация и управление движением летательных аппаратов, алгоритмическое обеспечение бортовых вычислительных систем.
Эл. адрес:
vvmarkel@gmail.com

МЕЛЕХИН
Виктор
Федорович



Профессор кафедры компьютерных систем и программных технологий Санкт-Петербургского политехнического университета Петра Великого, заслуженный работник высшей школы РФ.
В 1960 году окончил Ленинградский политехнический институт имени М. И. Калинина по специальности «Автоматика и телемеханика».
В 1984 году защитил диссертацию на соискание ученой степени доктора технических наук.
Является автором 230 научных публикаций, 52 изобретений, 23 патентов, четырех монографий, шести учебников и 14 учебных пособий.
Область научных интересов — теория и технология проектирования вычислительных систем и устройств.
Эл. адрес:
melekhin@kspt.icc.spbstu.ru

МИШУРОВА
Наталья
Валерьевна

Инженер 1-й категории ОАО «Ракетно-космическая корпорация «Энергия» им. С. П. Королёва», Королёв.
В 2008 году окончила Московский государственный технический университет им. Н. Э. Баумана по специальности «Подъемно-транспортные машины».
Является автором 21 научной публикации.
Область научных интересов — космонавтика, средства контроля и управления космическими летательными аппаратами.
Эл. адрес: post@rsce.ru

НЕЧАЕВ
Владимир
Анатольевич

Доцент кафедры машинного проектирования бортовой электронно-вычислительной аппаратуры Санкт-Петербургского национального исследовательского университета информационных технологий, механики и оптики, главный конструктор АО «ОКБ «Электроавтоматика», Санкт-Петербург.
В 2001 году окончил Санкт-Петербургский государственный университет аэрокосмического приборостроения по специальности «Приборостроение».
Является автором 13 научных публикаций.
Область научных интересов — геоинформационные системы, системы индикации авионики, навигационные системы.
Эл. адрес: nil-12@mail.ru

НОВИКОВ
Павел
Андреевич

Аспирант кафедры информационных и вычислительных систем факультета автоматизации и интеллектуальных технологий Петербургского государственного университета путей сообщения Императора Александра I.
В 2013 году окончил Петербургский государственный университет путей сообщения Императора Александра I по специальности «Программное обеспечение вычислительной техники и автоматизированных систем».
Область научных интересов — искусственные нейронные сети, машинное обучение.
Эл. адрес: imortan@gmail.com

ОВЧИННИКОВ
Андрей
Анатольевич

Доцент кафедры безопасности информационных систем Санкт-Петербургского государственного университета аэрокосмического приборостроения.
В 2000 году окончил Санкт-Петербургский государственный университет аэрокосмического приборостроения по специальности «Информационные системы в экономике».
В 2004 году защитил диссертацию на соискание ученой степени кандидата технических наук.
Является автором 20 научных публикаций и пяти патентов на изобретения.
Область научных интересов — теория помехоустойчивого кодирования, криптография.
Эл. адрес: mldoc@ieee.org

ПЯТАЕВА
Анна
Владимировна

Аспирант Сибирского государственного аэрокосмического университета им. академика М. Ф. Решетнёва, Красноярск.
В 2003 году окончила магистратуру Сибирского государственного аэрокосмического университета им. академика М. Ф. Решетнёва по специальности «Информатика и вычислительная техника».
Область научных интересов — анализ изображений, интеллектуальные информационные системы.
Эл. адрес: anna4u@list.ru

**СЕРГЕЕВ
Михаил
Борисович**

Профессор, заведующий кафедрой вычислительных систем и сетей Санкт-Петербургского государственного университета аэрокосмического приборостроения, директор НИИ информационно-управляющих систем Санкт-Петербургского национального исследовательского университета информационных технологий, механики и оптики.
В 1980 году окончил «ЛЭТИ».
В 2001 году защитил диссертацию на соискание ученой степени доктора технических наук.
Является автором более 100 научных публикаций и 14 патентов на изобретения.
Область научных интересов — теория разрядных вычислений, методы проектирования спецпроцессоров для систем контроля и управления и др.
Эл. адрес: mbse@mail.ru

**СОКОЛОВ
Андрей
Владимирович**

Ведущий научный сотрудник Института прикладных математических исследований Карельского научного центра РАН, профессор кафедры информатики и математического обеспечения Петрозаводского государственного университета.
В 1974 году окончил Ленинградский государственный университет по специальности «Прикладная математика».
В 2006 году защитил диссертацию на соискание ученой степени доктора физико-математических наук.
Является автором более 100 научных публикаций.
Область научных интересов — задачи оптимального управления динамическими структурами данных, динамическое распределение памяти, FIFO-очереди.
Эл. адрес: avs@krc.karelia.ru

**СОЛЕНЫЙ
Сергей
Валентинович**

Доцент кафедры технической физики, электромеханики и робототехники Санкт-Петербургского государственного университета аэрокосмического приборостроения.
В 2005 году окончил Донецкий национальный технический университет по специальности «Электротехнические системы электропотребления», в 2006 году — по специальности «Организационно-экономическое и правовое обеспечение деятельности предприятий».
В 2011 году защитил диссертацию на соискание ученой степени кандидата технических наук.
Является автором 67 научных публикаций и одного патента на промышленный образец.
Область научных интересов — электробезопасность, пожаробезопасность и взрывобезопасность, развитие систем «Умный дом» и Smart Grid и др.
Эл. адрес: ssv555ssv@yandex.ru

**СКОРЮПИНА
Екатерина
Григорьевна**

Аналитик ООО «НПК «Разумные решения», Самара.
В 2015 году окончила Самарский государственный аэрокосмический университет им. академика С. П. Королева по специальности «Информатика и вычислительная техника».
Является автором пяти научных публикаций.
Область научных интересов — использование онтологий и мультиагентных технологий для создания интеллектуальных систем поддержки принятия решений при управлении сложными техническими объектами.
Эл. адрес:
kat-skorupina@yandex.ru

**СОЛЁНАЯ
Оксана
Ярославовна**

Доцент кафедры технической физики, электромеханики и робототехники Санкт-Петербургского государственного университета аэрокосмического приборостроения.
В 2007 году окончила Донецкий национальный технический университет по специальности «Электротехнические системы электропотребления», в 2009 году — по специальности «Организационно-экономическое и правовое обеспечение деятельности предприятий».
В 2014 году защитила диссертацию на соискание ученой степени кандидата технических наук.
Является автором 35 научных публикаций и одного патента на изобретение.
Область научных интересов — энергосбережение и повышение энергетической эффективности развития методов оценки безопасности технологических объектов.
Эл. адрес: osolenaya@list.ru

**СУТУГИНЕНЕ
Юлия**

Старший преподаватель кафедры русского языка Петербургского государственного университета путей сообщения Императора Александра I.
В 2008 году окончила магистратуру Санкт-Петербургского государственного университета по специальности «Филология».
В 2015 году в Литовском эдукологическом университете (г. Вильнюс) защитила диссертацию на соискание ученой степени PhD.
Является автором 17 научных публикаций.
Область научных интересов — методика преподавания русского языка как иностранного, вопросы теории и практики языкового и педагогического тестирования, дистанционное обучение.
Эл. адрес: sutuginiene@mail.ru

ФАВОРСКАЯ
Маргарита
Николаевна

Профессор, заведующая кафедрой информатики и вычислительной техники Сибирского государственного аэрокосмического университета им. академика М. Ф. Решетнёва, Красноярск. Почетный работник высшего профессионального образования РФ.
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.
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

ЧИЖОВ
Анатолий
Анатольевич

Доцент, заместитель начальника кафедры радиолокационного вооружения Военной академии войсковой противовоздушной обороны Вооруженных сил РФ им. Маршала Советского Союза А. М. Василевского, Смоленск.
В 1998 году окончил Военную академию ПВО Сухопутных войск РФ по специальности «Специальные радиотехнические системы», в 2007 году — по специальности «Управление воинскими частями и соединениями».
В 2011 году защитил диссертацию на соискание ученой степени доктора технических наук.
Является автором более 150 научных публикаций и шести патентов на изобретения.
Область научных интересов — сверхразрешение, оптимизация стохастических систем.
Эл. адрес: rtshouse@mail.ru

ШТАНЬКО
Сергей
Владимирович

Заместитель начальника кафедры космических радиотехнических систем Военно-космической академии им. А. Ф. Можайского, Санкт-Петербург.
В 1999 году окончил Военный инженерно-космический университет им. А. Ф. Можайского по специальности «Радиоэлектронные системы».
В 2004 году защитил диссертацию на соискание ученой степени кандидата технических наук.
Является автором 46 научных публикаций и восьми учебных изданий.
Область научных интересов — защита информации в радиотехнических комплексах управления, сбора и передачи информации.
Эл. адрес: craft2001@mail.ru

ЮРЫГИНА
Юлия
Сергеевна

Руководитель проекта, ООО «НПК «Разумные решения», Самара.
В 2009 году окончила Ульяновский государственный технический университет по специальности «Программное обеспечение вычислительной техники и автоматизированных систем», 2010 году — по специальности «Бухгалтерский учет, анализ и аудит».
Является автором семи научных публикаций.
Область научных интересов — использование мультиагентных технологий и онтологий для создания интеллектуальных систем анализа и диагностики работы сложных технических изделий, поддержки принятия решений в случае возникновения нештатных и аварийных ситуаций.
Эл. адрес:
Yurygina@smartsolutions-123.ru

ЯКОВЛЕВ
Евгений
Леонидович

Начальник учебной лаборатории кафедры математического и программного обеспечения Военно-космической академии им. А. Ф. Можайского, Санкт-Петербург.
В 2005 году окончил Военно-космическую академию им. А. Ф. Можайского по специальности «Системы управления летательными аппаратами».
Является автором одной научной публикации.
Область научных интересов — численная теория массового обслуживания, нейронные сети.
Эл. адрес: evgen-1932@yandex.ru

Уважаемые авторы!

При подготовке рукописей статей необходимо руководствоваться следующими рекомендациями.

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

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, электронные адреса авторов, которые по требованию ВАК должны быть опубликованы на страницах журнала. При написании аннотации не используйте аббревиатур и не делайте ссылок на источники в списке литературы.

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 — эта информация будет опубликована в ссылке на первой странице.

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»), т. к. 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 в текст не заверстываются и предоставляются отдельными исходными файлами, поддающимися редактированию:

— рисунки, графики, диаграммы, блок-схемы предоставляйте в виде отдельных исходных файлов, поддающихся редактированию, используя векторные программы: Visio 4, 5, 2002—2003 (*.vsd); Coreldraw (*.cdr); Excel (*.xls); Word (*.doc); AdobeIllustrator (*.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 подрисуночных подписей 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;

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

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

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

Список литературы предоставляйте в двух вариантах: первый на языках оригиналов и второй — перевод (не траслитерация, а перевод) списка на английский язык.

Более подробно правила подготовки текста с образцами изложены на нашем сайте в разделе «Оформление статей».

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

ИНФОРМАЦИОННО-УПРАВЛЯЮЩИЕ СИСТЕМЫ

ISSN 1684-8853

9 771684 885009